Atatürk Üniversitesi Ziraat Fakültesi Ders Yayınları No: 185



2024

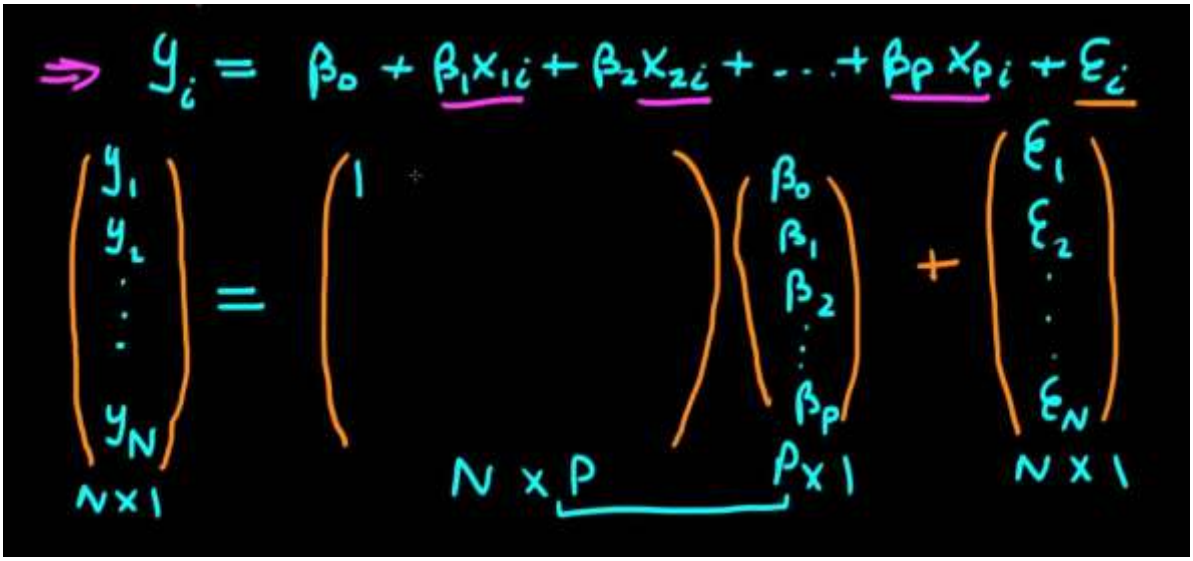
Ekonometri

Teori ve Uygulama

D e r s N o t l a r ı

Prof. Dr. Fahri YAVUZ

T a r ı m E k o n o m i s i B ö l ü m ü



Erzurum - 2001

ÖN SÖZ

Ekonometri, ekonomik faaliyetin öznesi olan birimlerin davranışlarını rakamsal olarak tanımlayabilmek ve tahmin edilen parametreleri kullanarak ekonomik analizler yapabilmek için çok sayıda metotlar sunan bir bilim dalıdır. Temel olarak regresyon modelleriyle yapılan ekonometri çalışmaları, ekonomik bir olguyu veya ekonomik birimlerin davranışlarını tanımlayan unsurların ne ölçüde açıklayıcı olduğunu ilgili modeli tahmin ederek ortaya koymaktadır. Tahmin edilen ekonometri modelleri, ekonomi problemlerinin çözümünde, bazı politik senaryoların analizinde ve ileriye yönelik kestirmelerin yapılmasında kullanılmaktadır. Tüm bunlar, ekonomistlerin ve özellikle tarım ekonomistlerinin ekonometri yöntemlerini kullanabilme yetkinliğine sahip olmalarının gerekliliğine işaret etmektedir.

Bu ders notları, büyük ölçüde Maddala'nın "Introduction to Econometrics" adlı kitabından ve diğer kaynaklardan yararlanılarak, hem lisans hem de lisansüstü öğrenciler için hazırlanmıştır. Birinci ve ikinci bölümde ekonometri, ekonomi teorisi ilişkisi ve istatistik ile ilgili temel kavramlar ve bilgiler verilmiştir. Daha çok lisans öğrencilerine yönelik olan üçüncü ve dördüncü bölümlerde basit regresyon, beşinci ve altıncı bölümlerde ise çoklu regresyon konuları çeşitli yönleriyle ele alınmıştır. Birinci ve ikinci bölümlerde kısmen, yedinci, sekizinci ve dokuzuncu bölümlerde ise çoğunlukla lisansüstü öğrencilere yönelik olarak sırasıyla ekonomi teorisi ekonometri ilişkileri, istatistik teorisi, ekonometri problemleri, kukla değişkenler, sınırlı bağımlı değişken modelleri ve simultane denklem modelleri ele alınmıştır. Konular işlenirken, regresyon modellerinin nasıl tahmin edileceği hususu kısaca anlatıldıktan sonra, ağırlıklı olarak modellerin nasıl yorumlanacağı üzerinde durulmuştur. Her konuyla ilgili uygulamalı örnekler sunulmuş, sınıfta yapılabilecek örnek uygulamalar verilmiş ve bu örnekler ekonometri yazılımı çıktıları ile desteklenmiştir. Benzer mantıkla çalışan farklı yazılımlar rahatça kullanılabilir. Böylece öğrencilerin konuları; model oluşturarak, bu modelleri tahmin ederek ve çıkarım yaparak öğrenmesi amaçlanmıştır.

Ders notlarında (*) ile işaretlenen başlıklar, ileri seviyede konular olup, lisansüstü öğrenciler için hazırlandığından lisans öğrencileri bu kısımları çalışmayabilirler. Öğrenciler çalışırken, teori kısmını anlamaya çalışmalı ve uygulamalarla bu öğrendiklerini pekiştirmelidir. Lisans ve lisansüstü öğrencilere yönelik olarak hazırlanan bu ders notlarının, ekonometri analizi yöntemlerini kullanan ekonomist ve tarım ekonomistlerine, açıklamalı yazılım çıktılarıyla yararlı olacağı kanaatindeyim. İlk defa 2001 yılında basılan bu ders notları çok defa güncellenmesine rağmen hala eksiklikler içermesi muhtemeldir. Ders notlarının iyileşmesine katkı sağlayacak her türlü eleştiri için şimdiden teşekkür ederim. Bu vesile ile ders notlarının öğrencilere ve yararlanacak tüm okurlara faydalı olmasını dilerim.

Fahri Yavuz
Ocak 2024, Erzurum

İÇİNDEKİLER

ÖN SÖZ i

İÇİNDEKİLER iii

I. EKONOMETRİYE GİRİŞ

1.1. Ekonometri Nedir? 1

1.2. Ekonomi ve Ekonometri Modelleri 2

1.3. Ekonometrinin Amaçları ve Çalışma Süreci 4

*1.4. Ekonometri ve Ekonomi Teorisi 7

*1.5. Üretim Fonksiyonlarının Tahmini 7

*1.6. Tüketici Talep Fonksiyonlarının Tahmini 12

II. TEMEL İSTATİSTİK BİLGİLER

2.1. Giriş 15

2.2. Toplam ve Çarpım Sembolleri ve Bazı Kullanımları 16

*2.3. İhtimal 17

*2.4. Şansa Bağlı Değişkenler ve İhtimal Dağılımı 18

*2.5. Normal Dağılım ve İlgili İhtimal Dağılımları 19

*2.6. Klasik İstatistik Çıkarım (Inference) 20

2.7. Nokta Tahmini 21

2.8. Aralık Tahmini 22

*2.9. Tahmin Edicilerin Özellikleri 23

2.10. Ana Kitleden Elde Edilen Örnekler İçin Örnekleme Dağılımları 24

2.11. Hipotez Testi 24

2.12. Güven Aralığı ve Hipotez Testleri Arasındaki İlişki 28

YAZILIM UYGULAMASI 29

III. BASİT REGRESYON

3.1. Giriş 31

3.2. İlişkilerin Belirlenmesi 32

3.3. Momentler Metodu 36
3.4. En Küçük Kareler Yöntemi 38
3.5. Basit Regresyon Modelinde İstatistiksel Çıkarım 43
YAZILIM UYGULAMASI 49

IV. BASİT REGRESYON ANALİZLERİ

4.1. Kesişme Katsayısı Olmaksızın Regresyon 53
4.2. Ters Regresyon 53
4.3. Basit Regresyon Modeli İçin Varyans Analizi 56
4.4. Basit Regresyon Modeli İle Kestirme 57
4.5. Anormal Gözlemler 60
4.6. Regresyon Denklemleri İçin Alternatif Fonksiyonel Formlar 65
YAZILIM UYGULAMASI 67

V. ÇOKLU REGRESYON

5.1. Giriş 71
5.2. İki Açıklayıcı Değişkene Sahip Bir Model 72
5.3. Çoklu Regresyonda İstatistiksel Yorum 76
BİLGİSAYAR UYGULAMASI 84

VI. ÇOKLU REGRESYON ANALİZLERİ

6.1. Regresyon Katsayılarının Yorumu 89
6.2. Kısmi Korelasyon ve Çoklu Korelasyon 89
6.3. Basit, Kısmi ve Çoklu Korelasyon Arasındaki İlişkiler 91
6.4. Çoklu Regresyon Modelinde Kestirme 94
6.5. Varyans Analizi ve Hipotez Testleri 96
6.6. İlgili Değişkenlerin İhmali Ve İlgisiz Değişkenlerin Mevcudiyeti 97
6.7. Serbestlik Derecesi ve R^2 100
6.8. İstikrarlılık (Stability) Testleri 101
YAZILIM UYGULAMASI 103

VII. EKONOMETRİ PROBLEMLERİ

*7.1. Çok Varyanslılık (Heteroskedasticity) …..... 107
*7.2. Otokorelasyon (Autocorrelation) 111
*7.3. Çoklu Eş Doğrusallık (Multicollinearity) 114
YAZILIM UYGULAMASI …….... 119

VIII. KUKLA VE SINIRLI BAĞIMLI DEĞİŞKEN MODELLERİ

8.1. Kesişme Katsayısındaki Değişmeler İçin Kukla (Dummy) Değişkeni 131
*8.2. Eğim Katsayısındaki Değişmeler İçin Kukla Değişkeni 135
*8.3. Çapraz Denklem Kısıtları İçin Kukla Değişkeni ……..…… 136
*8.4. Sınırlı Bağımlı Değişken Modelleri 138
*8.5. Doğrusal İhtimal Modeli 139
*8.6. Probit ve Logit Modeller 139
YAZILIM UYGULAMASI ……..... 144

IX. SİMULTANE DENKLEM MODELLERİ

*9.1. Giriş 153
*9.2. Dahili (Endojen) ve Harici (Eksojen) Değişkenler 154
*9.3. Tanımlama Problemi: İndirgenmiş Form ile Tanımlama 155
*9.4. Tanımlama İçin Gerek ve Yeter Şartlar 159
*9.5. Araç Değişkenler Yöntemi 163
*9.6. İki Aşamalı En Küçük Kareler Yöntemi 167
*9.7. Sınırlı Bilgiyle Maksimum Olabilirlik Metodu (LIML) 173
*9.8. Simulasyon Denklem Modellerinde OLS'nin Kullanılması 174
*9.9. Üç Aşamalı En Küçük Kareler Yöntemi 178
YAZILIM UYGULAMASI ….... 179

X. ZAMAN SERİSİ ANALİZLERİNE GİRİŞ

*10.1. Giriş ……..…… 189
*10.2. Zaman Serisi Analizi Nedir? ……..…….. 189
*10.3. Zaman Serisi Analizleri Neden Kullanılıyor? ..…… 190

*10.4. Zaman Serisi Analizi Türleri ……..... 191
*10.5. Zaman Serisi Analiz Modelleri ve Teknikleri …….. 192
*10.6. Zaman Serisi Analizi Örnek Kitapları …….. 193
KAYNAKLAR 194
EK 1. TABLOLAR 195

Not: Başlıklardaki * işareti lisansüstü eğitimde kullanılan konulara işaret eder.

I. EKONOMETRİYE GİRİŞ

1.1. Ekonometri Nedir?

Ekonometrinin kelime olarak tam karşılığı "ekonomi biliminde ölçüm" dür. Ekonomi biliminin çoğu konuları ölçüm ile ilgili olduğundan, bu tanımlama çok geniş anlamlı olur. Örneğin GSYH, istihdam, para arzı, ihracat, ithalat ve fiyat endeksleri gibi makroekonomi göstergeleri ölçülmektedir. O halde "Ekonometri" nasıl tanımlanabilir? Ekonometri; ekonomi teorilerine ampirik yani gözleme dayalı mana vermek ve ekonomi teorilerini doğrulama veya aksini ifade etmek amacıyla hazır veya üretilmesi gereken ekonomi verilerini kullanarak istatistik ve matematik metotları uygulamaya koyan bir bilim dalıdır. Ekonometri biliminin diğer bir tanımı ise, "istatistik yöntemlerin ekonomik olguların analizinde kullanılmasıyla ilgili çalışma alanı" şeklinde yapılabilir.

Ekonometri çalışan kişi, en başta bir ekonomisttir ve açıklamaya çalıştığı problemlerin ampirik analizlerini geliştirmede ekonomi teorisini kullanabilme kabiliyetine sahiptir. Bazen, ekonomi teorisinin uygun bir şekilde formüle edilmesi için matematik bilmesi gerekir. Ekonomik verilerin bulunması, toplanması ve ekonomik teori kavramlarını gözlenebilen değişkenler haline getirmede ise veri işlemeyi bilmelidir. Çoğu zaman, ekonomik ilişkileri ifade eden katsayıların tahmini ve ekonomik olayların kestirilmesi için bilgisayar başında saatler harcayan bir uygulamalı istatistikçidir. Aynı zamanda, ekonomi biliminin tabiatı gereği ortaya çıkan ampirik problemlere uygun İstatistik teknikleri geliştirmek için becerisini kullanan bir istatistik teorisyeni olabilir.

İstatistik daha çok denemeye dayalı ve kontrol edilebilen değişkenlerin kullanıldığı araştırmaların analizi için geliştirilmiş bir bilim dalıdır. Diğer tarafta, ekonomik olayların analizinde kullanılan değişkenlerin kontrol edilebilir olmaması ve gözleme dayalı verilerin kullanılması, ekonometriyi istatistikten ayırır. Bu durum, birtakım ekonometri problemlerini ortaya çıkarmakta ve bu problemlerin çözümü için yeni tekniklerin geliştirilmesini gerekli kılmaktadır. İşte bu noktada Ekonometrinin önemi ortaya çıkmaktadır. Ekonometri aynı zamanda Ekonomi Matematiğinden ayrılır, çünkü Ekonomi Matematiğinde sadece Matematiğin uygulaması vardır ve gerçek hayata ait veriler kullanılarak ampirik bir mananın belirlenmesi söz konusu değildir.

İstatistik araçlarının ekonomik verilere uygulanması çok uzun bir geçmişe sahiptir. İlk deneysel talep doğrusu 1699 yılında Charles Davenant tarafından yayınlandı ve ilk istatistiksel çalışmalar İtalyan istatistikçi Rudolfo Enini tarafından 1907 yılında yapılmıştır. Fakat ekonometrinin gelişmesi 1930 yılında Ekonometri Derneğinin kurulması ve 1933 yılında "Econometrica" dergisinin yayınlanması ile hızlanmıştır.

1950 ve 1960'lı yıllarda ekonometri tahminleri ve sonuçların istatistik yorumu kısmı, araştırıcıların harcadığı zaman açısından çok fazla yer işgal etmiştir. Ekonometri ile uğraşanlar, zamanlarının çoğunu belirlenen ekonometri modellerinin istatistik yöntemler kullanılarak tahminine harcamışlardır. Model tanımlama için çok az zaman

ayrılmış ve bu konuya olan ilgi daha az olmuştur. 1940'larda bu durum değiştirilmek istenmiş fakat istatistiksel analizlerde çok önemli problemler ortaya çıkmıştır. Bu yüzden 1950 ve 1960'lı yıllar alternatif tahmin modellerinin ve alternatif bilgisayar programlarının yenilenmesine harcanmıştır. Modellerin belirlenmesinde ortaya çıkan hatalara veya gözlem hatalarına fazla önem verilememiştir. Hem hız hem de hafıza açısından yüksek kapasiteye sahip bilgisayarların üretilmesiyle bütün bunlar değişmiştir. Tahmin problemleri artık ortadan kalkmış ve ekonometri ile uğraşanlar dikkatlerini ekonometri analizlerinin diğer yönlerine vermişlerdir.

Herhangi bir ekonomik verinin istatistiksel analizi yapılmadan önce, ilgili ekonomik teorinin matematiksel formülünün açıkça yazılması gerekir. Çok basit bir örnek olması bakımından talep eğrisi aşağı doğru meyillidir demek yeterli değildir. Bu ifadeyi matematiksel bir formda yazmak gereklidir. Bu da değişik şekillerde olabilir. Örneğin **q**'yu talep edilen miktar **p**'yi fiyat olarak alırsak, aşağıdaki şekillerde yazabiliriz.

$$\mathbf{q} = \alpha + \beta\mathbf{p} \quad \beta < 0 \qquad \text{veya} \qquad \mathbf{q} = \alpha\, \mathbf{p}^{\beta} \quad \beta < 0$$

İleride de göreceğimiz gibi karşılaştığımız en önemli problemlerden biri, ekonomik teorinin fonksiyonel formlar hakkında çok az bilgi verici olmasıdır. Bu yüzden fonksiyonel formların belirlenmesi, istatistik ve matematik yöntemlerin de kullanılmasını gerekli kılmaktadır.

1.2. Ekonomi ve Ekonometri Modelleri

Ekonometri ile uğraşan kişinin karşılaşacağı ilk iş bir ekonometri modelinin formüle edilmesidir. O halde model nedir? Bir model gerçek dünyada olup bitenlerin basitleştirilmiş şeklidir. Örneğin talep edilen portakal miktarı portakalın fiyatına bağlıdır demek basit bir temsildir. Çünkü portakalın talep miktarını belirleyen diğer birçok faktörün de olduğu düşünülebilir. Mesela tüketici geliri, portakal suyunun kahveden daha çok sağlığa faydalı olduğu düşüncesi, elmanın fiyatındaki yükseliş ve düşüşler bu faktörlerden sayılabilir. Fakat bu şekilde saymaya devam edersek bunun sonu gelmez. Benzin fiyatı bile portakal talebini etkileyen bir faktör olabilir.

Çoğu bilim adamı modellerin basitliği lehinde tartışmalar yapmışlardır. Çünkü basit modeller, verileri anlamayı, içeriğini sezmeyi ve test etmeyi kolaylaştırmaktadır. Çok karmaşık olan gerçek olay ve olguları açıklamak için basit model seçimi,

1. modelin gereğinden çok fazla basitleştirilmiş olması,
2. varsayımların gerçekçi olmaması gibi iki temel eleştiriye yol açmaktadır.

Portakal talebi örneğinde olduğu gibi, portakalın talep miktarının sadece portakalın fiyatına bağlı olduğunu söylemek çok basitleştirme ve aynı zamanda gerçekçi olmayan bir varsayımdır. Bir modelin çok basit bir şekilde oluşturulmasına yönelik eleştiri karşısında, basit bir modelle başlamanın ve takiben daha kapsamlı modeller oluşturmanın daha iyi olacağı düşünülebilir. Diğer taraftan bazı bilim insanları ise, çok

genel bir modelle başlayıp, eldeki verileri temel alarak modeli basitleştirmek yoluna gidilebileceğini ifade etmektedirler. Meşhur istatistikçilerden biri ”bir model fil kadar büyük olmalıdır” diyor. Bu alternatif yaklaşıma rağmen, burada basit modellerle başlanılacak ve geliştirilerek daha kapsamlı modeller inşa edilecektir.

Diğer bir eleştiri ise daha önce bahsedildiği gibi gerçekçi olmayan varsayımlardan hareketle modelin oluşturulmasıdır. Bu eleştiriye karşı diğer bir istatistikçi Friedman “bir teorinin varsayımları hiç bir zaman açıklayıcı olarak gerçekçi değildir” şeklinde bir ifade kullandıktan sonra özetle şöyle diyor. “Bir teorinin varsayımları hakkında sorulan ilgili soru, bu varsayımların gerçekçi olup olmadıkları değil -ki hiçbir zaman değiller- fakat eldeki amaç için yeterince uygun yaklaşım olup olmadığıdır. Bu soru, teorinin oluşturulan modelle çalışıp çalışmadığına bakılarak cevaplandırılabilir ki bunun manası da yeterince doğru tahminleri modelin üretip üretmediğidir”.

Portakalın talebi ile ilgili örneğimize dönersek, talebin sadece portakalın fiyatına bağlı olduğunu söylememiz tanımlama yönünden gerçek dışı bir varsayım olur. Bununla beraber diğer değişkenlerin modele alınması, örneğin gelir ve elmanın fiyatı gibi, modeli daha gerçekçi kılmaz. Bu modelin bile gerçekçi olmayan varsayımlar üzerine kurulduğu kabul edilir. Çünkü bu model, sağlık konusundaki düşünceler gibi diğer birçok değişkeni model dışına iter. Fakat önemli olan, hangi modelin portakal talebini tahmin etmede daha kullanışlı olduğudur. Bu hususta karar, elimizde mevcut olan ve elde edebileceğimiz veriler dikkate alınarak verilir.

Uygulamada, amacımızla ilgili olarak kabul ettiğimiz bütün değişkenleri modele dâhil ederiz ve kalan değişkenleri şansa bağlı değişken olan hata terimi içine attığımızı varsayarız. Bu da bize ekonomi ve ekonometri modeli arasındaki farkı verir. Bir ekonomi modeli, yaklaşık olarak ekonomideki bir birimin davranışını tanımlayan varsayımların tümünü içine alan bir modeldir. Bir ekonometri modeli ise aşağıdaki unsurları içerir.

- Bir ekonomik modelinden bir veya birkaç davranış denklemi çıkarılabilir. Bu denklemler, ilgili ve gözlenebilen bazı değişkenler ile modelin amacıyla ilgili olmadığı düşünülen ve aynı zamanda gözlenemeyen olayları temsil ettiği varsayılan değişkenleri içine alan hata terimlerinden oluşur.
- Gözlem yoluyla elde edilen değişkenlerde gözleme hatalarının olup olmadığını belirten bir ifadeyi içerir.
- Hata teriminin ve ölçüm hatalarının ihtimal dağılımının bir tanımını ifade eden bir unsuru içerir.

Bu belirlemelerden sonra, ekonomik bir modelin ampirik tahmini yanında geçerliliğini test etmeye, geleceğe yönelik kestirmeler yapmaya veya politik analizler de kullanmaya devam edilebilir.

Ekonomi teorisinde bir talep modelinin en basit örneği ele alınırsa, bir ekonometri modeli aşağıdaki unsurları içerir.

- Davranış denklemi, $\mathbf{q = \alpha + \beta p + u}$ şeklinde ifade edilirse, burada **q** talep edilen miktar, p ise fiyattır. Burada **p** ve **q** gözlenen değişkenler **u** ise hata terimidir.
- Burada $\mathbf{E(u/p) = 0}$ olur. Yani hata terimi ile **p** arasında bir ilişki yoktur ve değişik gözlemler için **u**'nun değerleri bağımsızdır. Yine bu hata terimi, sıfır ortalama ve σ^2 varyans ile normal bir dağılıma sahiptir.

Bu belirlemelerden sonra, ampirik olarak talep kanunu veya $\beta < 0$ hipotezi test edilebilir. Aynı zamanda geleceğe yönelik kestirmeler ve politik amaçlar için tahmin edilen talep fonksiyonu kullanılabilir.

1.3. Ekonometrinin Çalışma Alanı ve Süreci

Ekonometrinin amaçlarını aşağıdaki gibi sıralayabiliriz.

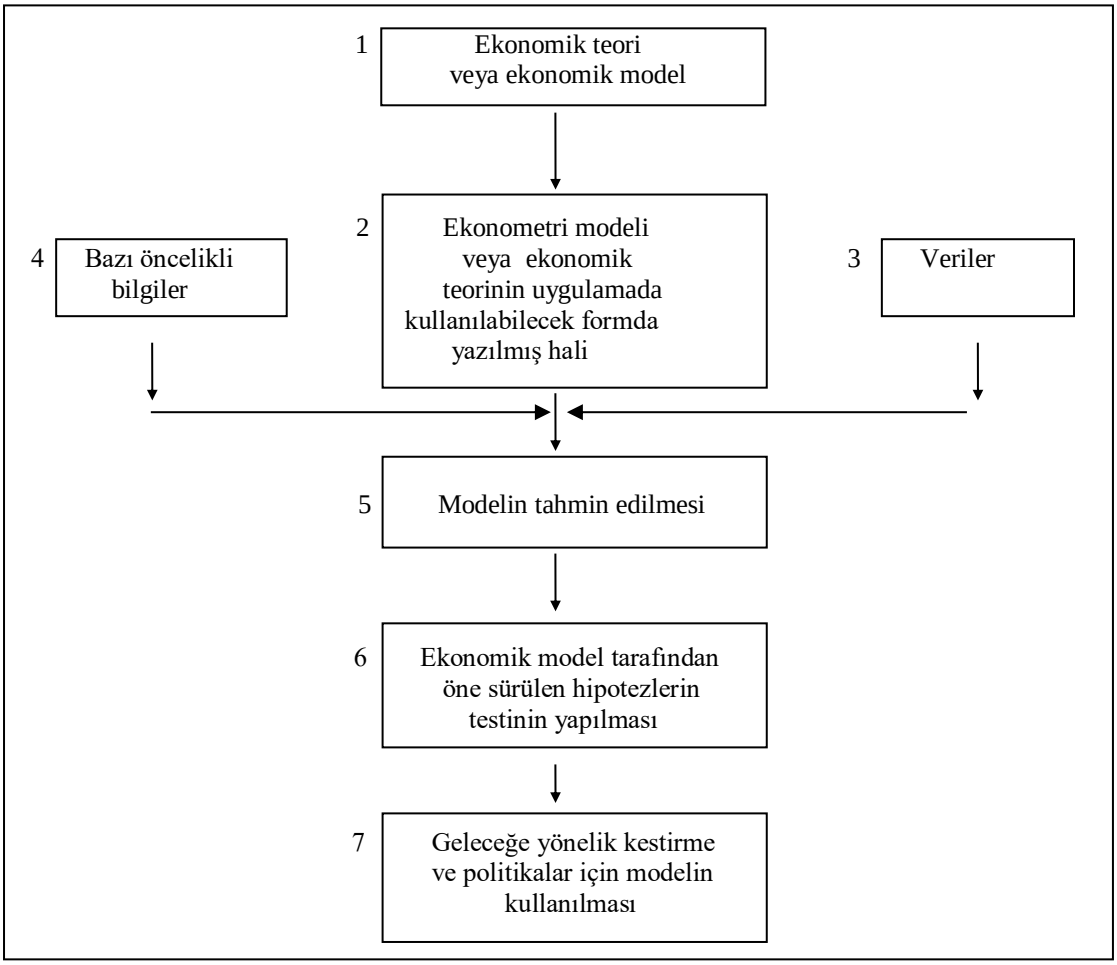
- Ekonometri modellerinin formüle edilmesi, yani ekonomi modellerinin ampirik olarak test edilebilmesi için bir denklem haline getirilmesi gerekir. Genellikle, fonksiyonel formun seçilmesi ve değişkenlerin stokastik şeklinin belirlenmesi gibi zorunluluklar olduğundan bir ekonomi modelinden ekonometri modelinin çıkarılmasının değişik yolları vardır. Bir ekonometri çalışmasında bu aşama, modelin belirlenmesi ile ilgili kısımdır.
- Gözlenerek temin edilen veriler dikkate alınarak fonksiyonel formu belirlenen modelin tahmin ve test edilmesi: Bu aşama, bir ekonometri çalışmasının katsayıların tahmini ve istatistik yorumu kısmını oluşturur.
- Tahmin edilen modellerin gelecekle ilgili kestirmelerde ve politika analizi amaçlı kullanılması son aşamayı oluşturur.

Sistematik olarak ekonometri analizinin içerdiği safhalar, Şekil 1.1'deki gibi gösterilebilir. Kutular içindeki ifadeler açık olduğundan dolayı onlar hakkında açıklama yapmaya gerek yoktur. Açıklama isteyen kutu, sadece öncelikli bilgileri içeren dördüncü kutudur. Bu bilgiler, ekonomik teoriden gelebileceği gibi önceki ampirik çalışmalardan da elde edilebilecek ve modeldeki bilinmeyen katsayılar hakkında sahip olunan güvenilir herhangi bir bilgidir.

Ancak Şekil 1.1'deki sıralamayla ilgili olarak dikkate alınması gerekecek derecede önemli bir eksiklik söz konusudur. Daha önceleri de bu eksiklik fark edilmiş olsa da ilk defa 1970'li yıllarda bu tek yönlü süreç hakkında tartışmalar ortaya çıkmıştır. Bu tartışmaların üçünü aşağıdaki gibi sıralayabiliriz.

- Şekil 1.1'de ekonomi teorilerinin ekonometri modeli üzerinde yapılan testinden ekonomi teorilerinin formüle edilmesi ters yönde yani 6. kutudan 1. kutuya bir bilgi akımı yoktur. Ekonometri çalışanların sadece ekonomi teorileri doğrultusunda

değerlendirme yapmadığı tartışıla gelmiştir. Ekonometri analizleri yapanların, ekonomi teorilerini olduğu gibi aldıkları ve test ettikleri ve testlerden hiç bir şey öğrenmedikleri doğru değildir. Bu nedenle 6. kutudan 1. kutuya doğru bir akışın olması gerekmektedir.



Şekil 1.1. Ekonometri çalışmasında basamakların sistematik olarak sıralanışı

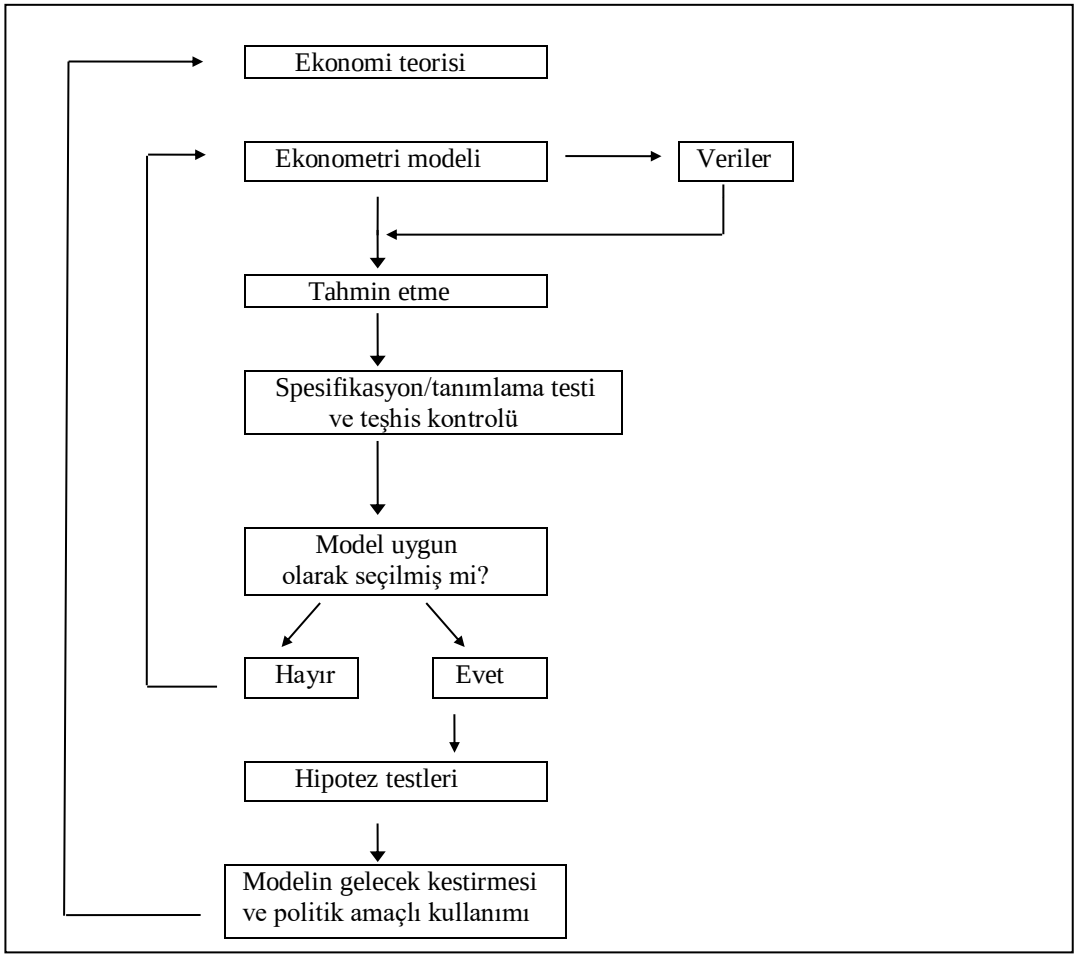
- Aynı şey veri toplama işlemi için de doğrudur. Ekonometri analizi yapanlar sadece toplayabildikleri veya var olan verileri kullanmazlar. Eğer veriler yeterli değil ise yeni verilere ihtiyaç duyarlar. Yani 2. ve 5. kutulardan 3. kutuya doğru bir geri dönüş olabilmesi gerekmektedir.
- Hipotez testinin sadece ekonomik teori tarafından öngörülen hipotezler üzerinde olması gerektiği eğilimi de ayrıca tartışılmaktadır. Bu da ikinci kutuda belirlenen modelin doğru olup olmadığı varsayımına bağlıdır. Fakat orijinal modelin doğruluğunun da test edilmesi gerekebilir. Bu yüzden modelin belirlenmesi ve teşhisi için yeni kutulara ihtiyaç olabilir. Yani, 6. kutudan 2. Kutuya da bir dönüş olması gerekmektedir. Bu da modelin tam tanımlama testi, ekonomi teorisi için yeni bir model tespitini gerekli kılabilir anlamına gelmektedir.

Yukarıda ileri sürülen değerlendirmeler doğrultusunda oluşturulan ekonometri çalışması sürecini gösteren sistematik Şekil 1.2'de verilmiştir. Orijinal kutuların bazıları ya silinmiş ya da yoğunlaştırılmıştır. Şekil 1.2'deki sistematik tanımlama sadece şekil olarak verilmiş, fakat metin halinde açıklanmamıştır.
Burada dikkate alınacak önemli husus geriye bilgi akımı olmasıdır. Bu akımlar;

• Ekonometri analizi sonuçlarından ekonomi teorisine,

• Tanımlama testi ve teşhis kontrollerinden ekonomi modelinin yenilenmiş şekline,

• Ekonometri modelinden verilere gitmektedir.



Şekil 1.2. Bir Ekonometri çalışmasında aşamaların yeni sistematiğe göre sıralanışı

Şimdiye kadar yalnız bir teoriden bahsedildi, hâlbuki birbiriyle yarışan birçok teori vardır. Ekonometrinin temel amaçlarından biri de birbirleriyle yarışan teorilerden birini seçmede yardımcı olmaktır. İşte bu model seçimi problemidir. Bu problem ilerideki konularda tartışılacaktır.

*1.4. Ekonometri ve Ekonomi Teorisi

Ekonometrinin amaçlarından biri de ekonomi teorilerinin testidir. Bu durumda önemli bir soru ortaya çıkıyor. Böyle bir testi neler oluşturuyor? Ekonomi teorisinin başarılı bir testinin göstergesi olarak bir ekonometri modelindeki tahmin edilen katsayıların işaretlerinin doğruluğunu rapor etmek alışıla gelen bir yöntemdir. Bu yaklaşım ekonomik teorinin doğrulanması yaklaşımı olarak da isimlendirilebilir. Bu yaklaşımdaki problem, bir bilim adamının deyimiyle "Ekonomi biliminin birçok sahasında değişik ekonometri çalışmaları birbiriyle çatışan sonuçlar veriyor. Eldeki mevcut verilerle çoğu zaman hangi sonucun doğru olduğuna karar vermenin etkin bir yolu yoktur. Bu neticeden olarak karşı hipotez uzun yıllardan beri ortada olmaya devam ediyor."

Eğer bir model, önceki alternatiflere oranla daha iyi tahminleri veriyorsa bu model ekonomi teorisinde daha gerçekçi ve geçerlidir. Bu yüzden eldeki modelin önceki modellerle mukayesesinin yapılması gerekli görülmektedir. Alternatif teorilerin değerlendirilmesindeki bu yaklaşım son yıllarda büyük bir ilgi çekmektedir.

Politik problemlere dönük ekonomi analizlerinin çoğunda, gerek üretim gerekse tüketim ile ilgili ekonomideki birimlerin davranışlarını belirleyen bazı parametreleri bilmek gerekmektedir. Örneğin; Etin talep elastikiyeti nedir? Et endüstrisinde verimler, sabit, artan veya azalan oranlarda mı değişmektedir? Tarımsal üretimde sermaye ve işgücü arasındaki ikamenin derecesi nedir? gibi sorulara verilecek cevaplar, politik analizler için önemli imalara sahiptir. Bu soruların cevapları ekonomideki birimin davranışını tanımlayan veriler analiz edilerek bulunabilir. Böyle bir inceleme genellikle regresyon analizi gibi istatistiksel araçların kullanımını gerektirir. Ekonomiyle ilgili araştırmalarda bu gibi istatistik yöntemlerinin nasıl kullanılacağını gösteren bilim dalına ekonometri denilmektedir.

Günümüzde istatistik hesaplarının nasıl yapılacağından daha çok nelerin ve niçin hesap edileceği, sonuçların nasıl yorumlanacağı, modelde nelerin yanlış olabileceği ve modelin nasıl geliştirilebileceğinin cevabını bilmek önem kazanmıştır. Belki 1960'lı yıllarda nasıl hesap edileceği sorusu yeterli kapasitede bilgisayar teknolojisi henüz gelişmediğinden dolayı önem arz etmekteydi. Fakat bugün gelişmiş bilgisayarlarla hesaplama işi kolayca yapılabilmektedir. Bu yüzden ekonomi teorisinin ekonometri bilimine olan katkısı günümüzde daha fazla önem kazanmıştır.

*1.5. Üretim Fonksiyonlarının Tahmini

Varsayalım ki, herhangi bir tarım alt sektörü için üretim fonksiyonunun tahminini yapmakla ilgileniyoruz ve elimizde çok sayıda çiftçiye ait girdi ve çıktıları gösteren yatay-kesit verileri var. Bu durumda bu verilere ait bir üretim ilişkisini nasıl tahmin edeceğiz. İlk karşılaşacağımız sorun, bu üretim işlemine ait bir fonksiyonel formun belirlenmesi olacaktır. Yani üretimin girdilerle nasıl bir ilişkiye sahip olduğunun saptanması gerekmektedir. Diyelim ki biz, ekonomi teorisine uygun sınırlamaları içeren,

tahmin etme problemleri fazla olmayan esnek ve basit bir fonksiyonel form arzuluyoruz. İleride de göreceğimiz gibi bu özellikleri bir arada fonksiyonel bir formda bulmak çok kolay değildir.

Şu an için bir fonksiyonel form üzerinde durmaya çalışalım. Genel Cobb-Douglas formunun üretim fonksiyonunu ele alarak, i tarım işletmesinin üretimini x_1, x_2 ve x_3 girdilerini kullanarak $\mathbf{y = a\ x_1^b\ x_2^c\ x_3^d}$ fonksiyonu üzerinde üretim ile üretim faktörleri arasındaki ilişkiyi göstermeye çalışalım. Burada ***a, b, c,*** **ve** ***d*** tahmin edilecek parametrelerdir.

Üretim fonksiyonuna ait girdiler nelerdir? Bunlar klasik ekonomideki üretim faktörleri olan işgücü (**I**), sermaye (**S**), ve arazi (**T**)'dir. Hammaddeler genellikle arazi veya doğal kaynaklar adı altında ele alınır. Bu sebeple bizim kullanacağımız fonksiyonel form, i tarım işletmesine ait üretimin ($\mathbf{Y_i}$), o işletmede kullanılan sermaye ($\mathbf{S_i}$), işgücü ($\mathbf{I_i}$) ve arazinin ($\mathbf{T_i}$) bir fonksiyonu olduğunu kabul edelim. Bu fonksiyon denklem 1'deki gibi eşitliğin iki tarafının doğal logaritması alınarak daha iyi ifade edilebilir.

$$\mathbf{Ln\ Y_i = a + b\ ln\ S_i + c\ ln\ I_i + d\ ln\ T_i} \qquad (1.1)$$

Böyle bir doğrusal logaritmik form, en küçük kareler tekniği kullanılarak tahmin edilmeğe çok elverişlidir. Eğer sadece gerekli verileri doğru bir şekilde elde edebilirsek, tahmin yapmaya hazırız demektir. Burada bazı sorunlar ortaya çıkmaktadır. Üç üretim faktöründen her biri sadece miktar değil, aynı zamanda miktarlar toplamıdırlar. Sermaye, kullanılan birçok makinenin ağırlıklı ortalaması olabilir. İşgücü, yine aynı şekilde değişik işgücü tiplerinin toplamı olabilir.

Her bir üretim faktöründe ortaya çıkan hesaplama probleminin, birim sermayenin belirlenmesinde daha önemli bir sorun olarak karşımıza çıktığını görmekteyiz. Üretim, birim zamana isabet eden mal miktarı olarak ölçüldüğünden, sermaye de makine saati olarak ölçülmelidir. Bunun iki avantajı vardır. Birincisi, aynı sayıdaki makinenin az veya çok yoğun olarak kullanılması ihtimalini göz önünde bulundurmasıdır. İkincisi ise, teknolojik değişikliklerden dolayı makinenin yıllara göre değişik seviyede sermaye hizmeti sağlamasını hesaba katmasıdır.

Şimdilik yukarıda saydığımız sorunları çözdüğümüzü varsayarak tarım işletmelerine ait kabul edilebilir yatay-kesit girdi ve çıktı verilerine sahip olduğumuzu düşünelim. Üretim işlemine ait parametrelerin tahminini nasıl yapacağımız üzerinde durmaya çalışalım. Önceden de işaret edildiği gibi ilk akla gelecek yöntem, doğal logaritması alınmış çıktı/bağımlı değişken ve yine doğal logaritması alınmış girdilerin/bağımsız değişken regresyon analizini yapmak olacaktır. Acaba bu yöntem bize kabul edilebilir parametre tahminlerini sağlayabilecek midir?

Bu sorunun cevabı, geniş ölçüde, bu parametre tahminlerinin ne amaçla kullanılacağına bağlıdır. Örneğin, kullanılan üretim girdisi seviyeleri verilen bir tarım işletmesinin geleceğe ait üretim seviyesini kestirmek istiyorsak, yukarıda adı geçen teknikle elde

edilen tahminler belki kabul edilebilir. Fakat eğer, değişik faktörlerin marjinal ürününü -daha fazla işgücü ilavesinde ne kadar daha fazla ürün üreteceğimizi- kestirmeğe çalışıyorsak, yukarıda izah edilen tahminler belki de yetersiz olacaktır. Bunun niye böyle olduğunu görmek için modelin tahmini sırasında ortaya çıkacak problemi daha açık bir şekilde ortaya koymaya çalışalım.

Büyük harflerle ifade edilen değerlerin doğal logaritması alınmış şeklini küçük harflerle ifade edip modelimizi denklem 2'deki gibi yazabiliriz. Burada $\mathbf{u_i}$, ortalaması 0, varyansı σ^2 olan şansa bağlı bir değişkendir.

$$\mathbf{y} = \boldsymbol{a} + \boldsymbol{b}\,\mathbf{s_i} + \boldsymbol{c}\,\mathbf{i_i} + \boldsymbol{d}\,\mathbf{t_i} + \mathbf{u_i} \qquad (1.2)$$

Bu stokastik özelliği şu şekilde izah edebiliriz. Eğer biz bu tarım alt sektöründeki bir işletmeyi şansa bağlı olarak seçmek ve o işletmeye ait girdileri gözlemek durumundaysak, bu işletmeye ait çıktı seviyesinin beklenen değeri $\boldsymbol{a} + \boldsymbol{b}\mathbf{s_i} + \boldsymbol{c}\mathbf{i_i} + \boldsymbol{d}\mathbf{t_i}$ olacaktır. Biz burada ***a***, ***b***, ***c*** ve ***d*** parametrelerinin tahmin edilmesi ile ilgileniyoruz.

Sapmasız tahminciler içinde en düşük varyansa sahip tahmincilerin en iyi doğrusal tahminciler olduğunu belirten Gauss-Markov teorisine göre, bu parametrelerin iyi tahminleri, eşitliğin sağ tarafındaki değişkenlerle hata terimi ($\mathbf{u_i}$) arasında bir korelasyon yoksa sağlanabilir. Uygulamalarda bu zorunluluğun geçerliliği nedir? Bu soruyu cevaplandırmak için, hata teriminin tabiatını dikkate almamız gerekmektedir. Ne gibi etkiler hata terimi tarafından temsil edilmektedir? Bunun genel açıklaması, model dışında kalan bütün değişkenlerin toplam etkisini temsil ettiği şeklindedir. Tahmin edilen katsayıların kabul edilebilirliği hakkındaki soruyu cevaplandırmak için elimizdeki modelin dışında kalan değişkenlerin ne tip değişkenler olduğunu kendi kendimize sormamız gerekmektedir.

Genel olarak, model dışında iki çeşit değişken vardır. Birinci tip değişken, ne araştırmacı olarak bizim tarafımızdan elde edilebilen ne de çiftçi tarafından gözlenebilen değişkendir. Verilerdeki hatalar, girdi seviyelerindeki bilinmeyen sapmalar, hava durumu gibi tahmin edilemeyen faktörler ve benzeri bilgiler bu tip değişkenlere örnektirler. Çiftçiler bu değişkenleri göremediklerinden, onların seçtikleri girdi seviyeleri ile bu bilinmeyen etkiler arasında bir korelasyon olmaması doğaldır. Model dışında kalan ikinci tip değişken daha önemli problemlere sahiptir. Bu, çiftçi tarafından gözlenen fakat bizim tarafımızdan elde edilemeyen değişkendir. Girdi kalitesindeki ve işletmelerin kullandığı teknolojiler arasındaki sapmalar gibi çiftçiler tarafından bilinen model dışındaki faktörler bu değişkenlere örnektir. Model dışındaki değişkenlerle ortaya çıkan sorun, bu değişkenlerle modelde kullanılan değişkenler arasında bir korelasyonun kaçınılmaz olmasıdır. Eğer tarım işletmesinde karar verenler bu etkileri gözlerlerse, optimum girdi seviyesini belirlerken bu bilgileri mutlaka dikkate alacaklardır. Bu yüzden eşitliğin sağ tarafındaki değişkenler hata teriminin bir kısmından bağımsız olmayacaktır ve bu da sapmalı (bias) tahminlere sebep olacaktır.

Bu hususu daha iyi açıklamak için bir tarımsal üretim fonksiyonunu nasıl tahmin edebileceğimizi genişletilmiş bir örnekle inceleyelim. Varsayalım ki bir i tarım işletmesinde mısır üretimi ($\mathbf{M_i}$), ekilen mısır tohumu miktarına ($\mathbf{K_i}$) ve büyüme sezonundaki güneşli günlerin sayısına ($\mathbf{G_i}$) bağlıdır. Şimdilik varsayalım ki, bunlar mısır üretimini etkileyen iki değişkendir ve üretim ilişkisi açık olarak $\mathbf{M_i = K_i^{\alpha}\, G_i^{1-\alpha}}$ şeklinde verilmektedir. Bunun logaritmasını alırsak, bu fonksiyonu denklem 3'deki gibi yazabiliriz.

$$\ln M_i = \alpha \ln K_i + (1-\alpha) \ln G_i \qquad (1.3)$$

Şimdi varsayalım ki sadece $\mathbf{M_i}$ ve $\mathbf{K_i}$'ye ait verilere sahibiz, yani hava durumu ile ilgili bilgilere sahip değiliz. Bu durumda $\ln M_i = \alpha \ln K_i + u_i$ fonksiyonunu tahmin etmek için en küçük kareler yöntemini uygulamamızın α'nın iyi bir tahminini vereceğini düşünmemiz doğru olur mu? Burada $\ln K_i$ ile $u_i = (1-\alpha) \ln G_i$ arasında bir ilişki olmadığını varsayabilir miyiz? Eğer çiftçiler $\mathbf{K_i}$'yi seçmeden önce $\mathbf{G_i}$'yi gözlemleyemiyorlarsa, $\mathbf{K_i}$'nin seçimi $\mathbf{G_i}$ tarafından etkilenemez. Dolayısıyla en küçük kareler yöntemi α'nın iyi bir tahminini vermelidir. Şimdi yine varsayalım ki, üretim ilişkisi her bir tarım işletmesinde mevcut olan arazinin kalitesine ($\mathbf{R_i}$) de bağlıdır. Bu durumda üretim fonksiyonunu denklem 4'deki gibi yazabiliriz.

$$M_i = R_i K_i^{a} G^{1-a} \quad \text{veya} \quad \ln M_i = \ln R_i + a \ln K_i + (1 - \alpha) \ln G_i \qquad (1.4)$$

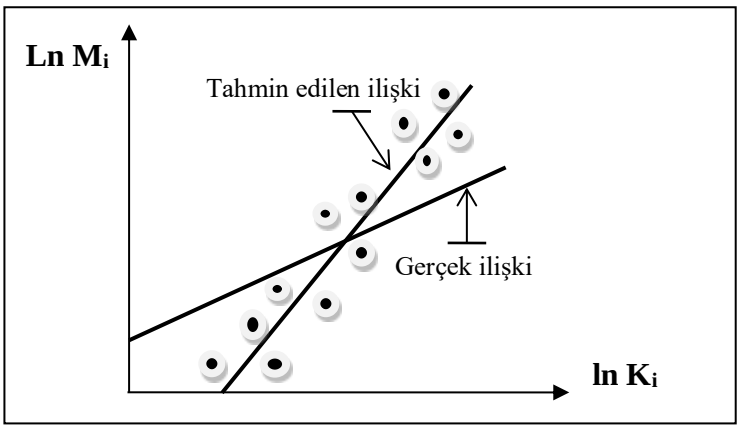
Daha önce belirtildiği gibi, $\mathbf{G_i}$'nin ne çiftçiler ne de araştırmacı olarak bizim tarafımızdan gözlenemediğini varsayalım. Buna ilave olarak, $\mathbf{R_i}$'nin çiftçiler tarafından gözlendiğini fakat bizim elimizde bu verilerin olmadığını varsayalım. Bizim açımızdan $\mathbf{G_i}$ ve $\mathbf{R_i}$'nin ikisi de şansa bağlı değişkenlerdir. Bu durumda $\ln M_i = \alpha \ln K_i + u_i$ regresyonu α'nın iyi bir tahminini bize verecek mi? Bu sorunun cevabı hayır, çünkü çiftçi i $\mathbf{R_i}$'yi gözlediği için $\mathbf{K_i}$ çiftçinin sahip olduğu bu bilgiden etkilenmektedir. $\mathbf{K_i}$ ile hata terimi arasında korelasyon olacağından α'nın sapmalı tahmini sonucunu verecektir. Aslında biz çiftçilerin kendi arazilerinin kalitesi hakkındaki bilgilerini $\mathbf{K_i}$'nin seçiminde ne şekilde kullanacaklarını, çiftçinin davranış biçimini irdeleyerek ortaya koyabiliriz. Çiftçi, her müteşebbis gibi kârını en yüksek yapmayı amaçlar ve bu doğrultuda hareket eder.

Çiftçi için kârın en yüksek yapılması problemi $\max p^*(R_i K_i^{\alpha} G_i^{1-\alpha}) - q^*(K_i)$ şeklinde yazılabilir. Burada **p** mısırın fiyatı ve **q** tohumun fiyatıdır. Basitleştirmek için, çiftçinin $\mathbf{G_i}$'yi kendi ortalama değerinde sabit olduğunu varsayalım. Bu durumda, kârı maksimum yapan $\mathbf{K_i}$ girdi seviyesi, fonksiyonun K_i'ye göre birinci dereceden kısmi türevi alınarak elde edilir ki bu da $p\,\alpha\, R_i K_i^{\alpha-1} G^{1-\alpha} - q = 0$'dır. Bu eşitliği $\mathbf{K_i}$'yi verecek şekilde çözersek denklem 5'i elde ederiz

$$K_i = (\alpha\, p\, R_i / q)^{1/1-a}\, G_i \qquad (1.5)$$

Bu eşitlikten açıkça anlaşılıyor ki çiftçinin $\mathbf{R_i}$ üzerine olan gözlemi ne kadar mısır ekeceği kararına ve dolayısıyla ne kadar üreteceğine etki etmektedir. Varsayalım ki **ln** $\mathbf{K_i}$ ve **ln** $\mathbf{M_i}$ üzerine olan gözlemlerimiz şekil 1.3'deki gibi bir görünüme sahiptir. Ortalama seviyesinde sabitleştirilmiş $\mathbf{R_i}$ için **ln** $\mathbf{K_i}$ ve **ln** $\mathbf{M_i}$ **arasındaki** gerçek ilişki de yine aynı şekilde çizilmiştir.

Çok fazla $\mathbf{K_i}$ girdisine sahip bir tarım işletmesi büyük bir ihtimalle daha geniş ölçüde $\mathbf{R_i}$'ye de sahiptir. Çünkü bu tarım işletmesine ait üretim normal seviyede arazi kalitesine sahip bir işletmenin üretim seviyesinden daha fazla olacaktır. Bu yüzden geniş $\mathbf{K_i}$ ile ilgili veri noktaları ortalama kalitede araziye sahip çiftliklere ait gerçek ilişkinin üzerinde olacaktır. Aynı şekilde düşük $\mathbf{K_i}$ girdilerine sahip çiftliklerin üretimi ortalama arazi kalitesine sahip çiftliklerin üretiminden daha az olacaktır.



Şekil 1.3. Sapmalı tahmin örneği

Net sonuç, böyle veri noktalarından tahmin edilen doğrunun ***a***'nın gerçek değerinden daha büyük bir ***a*** tahminini vermesidir. Bu tamamen daha önce bahsedilen sapmalı tahmin tipidir. Buradaki sorun, geniş ölçüdeki üretim verilerinin tamamen çok miktarda kullanılan mısır tohumu girdisine bağlı olmamasıdır. Halbuki, hem üretim seviyesini ve hem de kullanılan girdi miktarını belirleyen üçüncü bir değişken olan arazi kalitesi burada ihmal edilmektedir.

Bu tip sapmalı tahmine ekonometri çalışmalarında çok rastlanmaktadır. Regresyon teorisi, kontrol edilebilen denemeleri analiz etmek için geliştirilmiştir. Bu denemelerde hatalar mevcut olmasına rağmen, bunların etkileri açıklayıcı değişkenlerin etkilerinden ayıklanabilmektedir. Ekonomide, açıklayıcı değişkenler genelde verilmez, fakat bunun yerine ekonomik birimler tarafından seçilir. Bu seçim genelde ekonomik birimin temin edebildiği bilgilerin ışığı altında yapılır. Dikkate alınan bütün bu hususlar tarım işletmesinde karar verme sürecinin bütün ayrıntılarıyla incelenmesini yani üretime etki eden faktörlerin belirlenmesinde büyük bir titizlik gösterilmesini gerekli kılmaktadır. Sadece böyle bir yolla pratiğe uygulanan bir model ekonomi teorisine uygun olabilir ve geliştirilebilir.

Burada verilen örneklerde sadece iki faktörlü üretim fonksiyonlarıyla ilgilenilmesine rağmen gerçek dünyada uygulamaların çok girdili üretim işlemleri üzerine olduğu açıktır. Birisi gerçekten bir üretim işleminin parametrelerini tahmin etmeye uğraşıyorsa, onun çok sayıda girdiği bulundurabilecek bir teknoloji tanımlamasına ihtiyacı vardır. Örneğin, Cobb- Douglas formu *n* girdi durumuna kolayca ayarlanabilen bir teknoloji şeklidir.

Bu form, parametrelerin tahmininde doğrusal olduğundan dolayı kullanışlı olmaktadır. Fakat, Cobb-Douglas formu ciddi ve belki de gerçekçi olmayan sınırlamaları üretim işlemine dahil eder. Örneğin, her iki üretim faktörü arasındaki ikame elastikiyeti, Cobb-Douglas teknolojisi için belirlenmiş olan 1'e eşit olması gerekmektedir. İlk etapta bu gibi sınırlamaların modele konulması için özel bir sebep görünmemektedir. Bu yüzden teknolojinin yapısından dolayı daha az sınırlamalar içeren bir üretim fonksiyonu formuna sahip olmak istenen durumdur. Bu durumda böyle bir form olan Sabit İkame Elastikiyetli (Constant Elasticity of Substitution-CES) üretim fonksiyonu kullanılabilir.

Adından da anlaşılabileceği gibi, üretim fonksiyonu iki faktör arasındaki ikame elastikiyetinin birim elastikiyetten farklı olmasına izin verir. Fakat, CES fonksiyonu üretim faktörlerinin her biri arasındaki ikame elastikiyetini aynı olmaya zorlar. CES fonksiyonu doğru yöndeki bir adımdır, fakat hala çok fazla sınırlayıcı olduğu söylenebilir. Kolayca tahmin yapmak için yeterince basit ve ekonomik parametreler üzerine gereğinden çok sınırlamalar koymayacak üretim fonksiyonu formunu bulmanın zor olmayacağı düşünülebilir. Fakat, aslında böyle bir formu bulmak oldukça zordur. Üretim fonksiyonunun teknolojik özelliklerini kullanan çalışmaların çoğu ya CES ya da Cobb-Douglas formunu kullandığını söylemek doğru bir ifade olur.

Bilindiği gibi, ekonomik davranma biçimi olan maliyetin minimum ve kârın maksimum yapılması, matematiksel olarak maliyet ve kâr fonksiyonları kullanılarak ifade edilebilmektedir. Üretim fonksiyonunun ekonometride kullanışlı olma özelliği de bu fonksiyonlar ile ilgili olmasındandır. Bu yüzden maliyet fonksiyonu için fonksiyonel bir form bulabiliriz. Örneğin, böyle bir form daha fazla esnekliğe izin verdiği gibi, tahmin edilmesi de istenen ölçüde kolay olabilir. Bu tip maliyet fonksiyonlarının esnek formlarının bulunmasında ekonomi teorisinde izah edilen iki taraflılıktan (duality) faydalanılabilir. Değişik üretim fonksiyonlarına ait parametreleri hesap etmek için uygun bir maliyet fonksiyonunu seçmek ve onun parametrelerini tahmin etmemiz gerekmektedir. Bu amaç için en yaygın olarak kullanılan maliyet fonksiyonlarından biri translog maliyet fonksiyonudur.

*1.6. Tüketici Talep Fonksiyonlarının Tahmini

Kısaca belirtmek gerekirse, talep fonksiyonlarının tahminindeki sorun, önceki kısımda tartışılan üretim fonksiyonlarının tahminindeki problemlerle benzerlik göstermektedir. Talep fonksiyonları üzerine yapılan çalışma onu zorlaştıran ilave üç soruna daha sahiptir.

- Üretim tahmininde sistemin çıktı seviyesini yani üretilen ürün miktarını gözleyebilmemize rağmen talep tahmininde sistemin çıktısı olan fayda seviyesini gözlememiz mümkün değildir.
- Üretim tahmininde bütün firmaların aynı teknoloji seviyesinde çalıştığını varsaymak mantıklı görünmektedir. Fakat talep sisteminin tahmininde fayda fonksiyonu ve gelir seviyesinin birimden birime farklılık gösterdiği açıktır.
- Üretim tahmininde her bir firmanın çıktısını doğrudan gözleriz. Talep tahmininde biz çoğunlukla sadece satın alınan miktarları gözleyebiliyoruz.

Yatay-kesit verileri mevcut olan tek bir mal için uygulamada en çok karşılaşılan talep tahmin problemini önce dikkate alalım. Böylece belli sayıda birime ait tüketilen mal miktarına ($\mathbf{Q_i}$) ve o birimlere ait ve model dışında belirlenen fiyat ($\mathbf{P_i}$) ile ilgili gözlemlere sahip olduğumuzu düşünelim. Bu durumda ilk yapılacak iş, talep denkleminin fonksiyonel formunu seçmek olacaktır. Sık olarak kullanılan formlardan biri doğrusal formdur ve denklem 6'daki gibi yazılabilir.

$$\mathbf{Q_i} = \boldsymbol{a} + \boldsymbol{b}\mathbf{Pi} + \mathbf{u_i} \qquad (1.6)$$

Daha kullanışlı olanı ise çarpanlı $\mathbf{Q_i} = \boldsymbol{A}\ \mathbf{P_i}^{\boldsymbol{b}}\ \mathbf{u_i}$ talep denklemidir. Bu denklemin her iki tarafının da logaritmasını alırsak yeni talep formu denklem 7'deki gibi yazılabilir.

$$\mathbf{ln}\ \mathbf{Q_i} = \mathbf{ln}\ \boldsymbol{A} + \boldsymbol{b}\ \mathbf{ln}\ \mathbf{P_i} + \mathbf{ln}\ \mathbf{u_i} \qquad (1.7)$$

Bu form, direk olarak talebin fiyat elastikiyetini (b) verdiğinden dolayı kullanışlı bir form olma özelliğine sahiptir.

Önceden olduğu gibi burada da, $\mathbf{u_i}$ şansa bağlı hata terimi olarak alınmaktadır. Gauss-Markow Teoremi, $\mathbf{P_i}$ ile $\mathbf{u_i}$ arasında bir korelasyon olmadığı zaman en küçük kareler yöntemi bize $\boldsymbol{b}$ parametresinin en iyi tahminini vereceğini ifade etmektedir. Bunun böyle olduğu gerçekten doğru mudur? Bu soruya cevap vermek için model dışındaki hangi tür etkilerin $\mathbf{u_i}$'de toplandığını kendimize sormamız gerekmektedir. Temel tüketici davranışı teorisinden bilindiği gibi talep fonksiyonları genelde fiyatların ve gelirlerin fonksiyonudurlar. Bu etkiler hata teriminin bir parçası olarak mutlaka dikkate alınmalıdır.

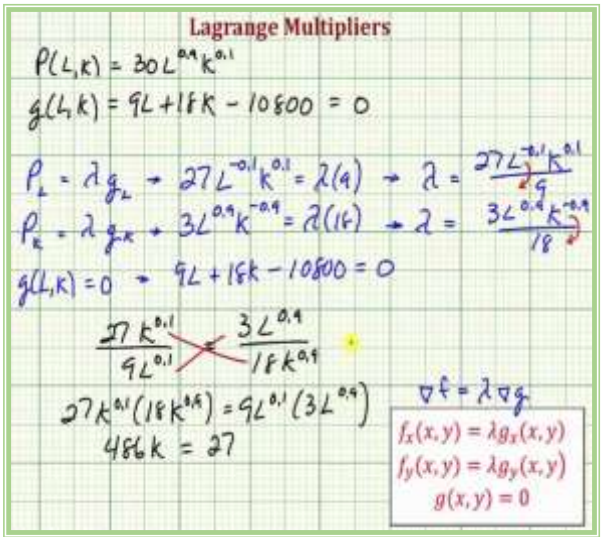
Şimdi gelir ve diğer fiyatlar ile $\mathbf{P_i}$ arasında bir korelasyon olmadığını düşünebiliriz. Varsayalım ki, kahvenin talebini fiyatın bir fonksiyonu olarak tahmin ediyoruz. Elimizde çayın fiyatına ait veriler yok ama tüketiciler bu gibi fiyatların farkındalar. Kahvenin talebi büyük bir ihtimalle çayın fiyatına bağlı olacaktır ve böylece çayın fiyatı hata teriminin bir parçası olacaktır. Eğer örnekteki çay ve kahve fiyatları arasında yüksek bir korelasyon varsa en küçük kareler tekniği, modeldeki parametrelerin sapmalı tahminlerini üretebilir.

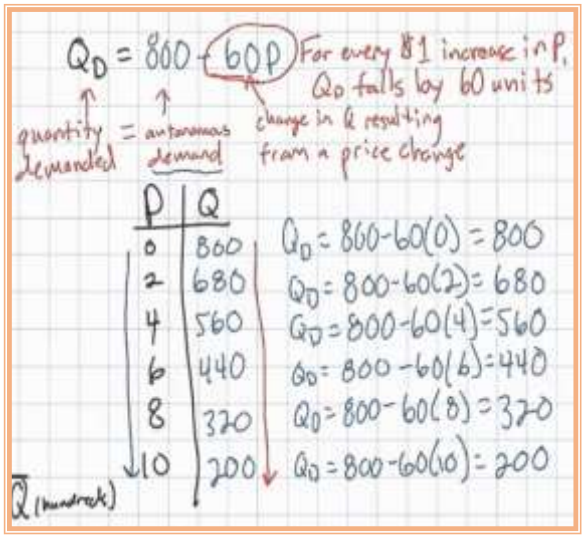
En basit talep tahmininde bile ilgili malın fiyatı ve gelir dışındaki parametreleri model içinde bulundurmak çoğu zaman gereklidir. Bütün bu nedenlerden dolayı talep

tahminlerinde sistem halindeki denklemler önemli rol oynamaktadır. Burada ilk akla gelen ve çok yaygın olarak kullanılan Doğrusal Harcama Sistemi (Linear Expenditure System-LES) ve İdeale Yakın Talep Sistemi (Almost Ideal Demand System-AIDS) örneklerini sıralayabiliriz.

Denklemler sistemi, belli standart teknikler kullanılarak tahmin edilebilir. Bu tahmin yöntemlerinde çok sayıda mal veya mal gruplarının tahminleri birlikte aynı anda yapıldığından mal veya mal grupları arasındaki etkileşim gerçek hayatta olduğu gibi dikkate alınmış olur. Talep sistemlerinde genellikle ilgili malın talebi, o malın fiyatına, diğer malların fiyatına ve tüketicinin gelirine bağlıdır. Üretim fonksiyonlarının tahmininde olduğu gibi, burada da talep sistemlerinin yeni fonksiyonel formlarını bulmak için de çift taraflılık (duality) teorisi kullanılmaktadır. Ekonomi teorisine göre fiyat ve gelirin fonksiyonu olan dolaylı fayda fonksiyonunun tanımlanması, tüketilen mal miktarının fonksiyonu olan dolaysız fayda fonksiyonunun tanımına denktir. Bu durumda hem dolaylı hem de dolaysız fayda fonksiyonlarının maksimum yapılmasından hareketle talep fonksiyonları elde edilir.

Burada bütün tüketici gelirlerini gözlemlediğimizi ve bütün tüketicilerin aynı tercihlere sahip olduklarını mümkün olmamasına rağmen özellikle varsaydık. Benzer fayda fonksiyonu varsayımı aslında o kadar kötü bir varsayım değildir. Beğenilerin dağılımı problemini gözden geçirmenin mantıklı bir yolu; tüketicilerin, verilen bir fayda fonksiyonunun etrafına dağılmış fayda fonksiyonlarına sahip olduğunu düşünmemizdir. Sonra eğer faydaları gözleyebilseydik, şansa bağlı olarak bir i biriminin seçimini ve onun dolaylı fayda fonksiyonunu $v_i(p,y_i) = v(p,y_i) + u_i$ şeklinde düşünebilirdik. Burada hata terimleri çok açık bir yoruma sahiptirler ve bunlar popülasyondaki farklı beğenilerin bir sonucu olarak vardır. Eğer tercihler belli bir ortalama fayda fonksiyonu etrafında yakın olarak dağılmış olduğu kabul edilebilirse, toplam talep davranışı, neoklasik talep eğrileri artı tesadüfî hata terimi tarafından temsil edilebilir. Hangi tip model tanımlamasının en kullanışlı olacağı kararı tamamen ampirik sonuçlara bağlıdır.





II. TEMEL İSTATİSTİK BİLGİLER

2.1. Giriş

Bu başlık altında ders kapsamı içerisinde kullanılacak bazı temel bilgilere yer verilmeye çalışılacak ve dolayısıyla daha önce öğrenilen temel istatistik bilgilerinin bir gözden geçirilmesi söz konusu olacaktır. Bu kısım daha çok özet olarak ve ayrıntıya inmeden verilecektir. Bu bölüm iki açıdan faydalı olacaktır. Birincisi istatistik derslerinde anlatılan konuların bir gözden geçirilmesi sağlanacaktır. İkincisi de ihtiyaç duyulduğunda bir referans olarak kullanılacaktır.

Özetlenmiş rakamların ifade ettiği gerçekler ile ilgili olarak istatistik, günlük hayatta da kullanılmasına rağmen bilimsel anlamda daha geniş bir mana ifade etmektedir. Genelde verilerin toplanması, analiz edilmesi ve yorumlanmasını içerir. İstatistik, verileri özetleme metotlarının belirlenmesi, bunların yorumlanması ve sunulmasına yardımcı olmak için geliştirilmiştir. Bu yöntemlerin uygulamasına deskriptif istatistik denilmektedir. İstatistik çalışmalarının anahtar iki unsuru, ana kitle ve örnektir. Deskriptif istatistik, tablo, grafik veya rakamları kullanarak bir ana kitle veya örneğe ait verilerin özetini sunar. İstatistiksel yorum ise, ana kitle üzerinde örnekleme yapılarak elde edilen özet bilgilerle, ana kitleye ait özellikler hakkında sonuç çıkarılmasıdır.

İstatistik analizlerde genellikle bir ana kitleye ait bilgiler, onu temsil eden örnekler kullanılarak temin edilir. Bunun nedeni, çoğu zaman ana kitlelerin çok büyük olması dolayısıyla bütün verilerin toplanmasının hem zaman hem de maddi açıdan büyük külfet getirmesi hatta bunu yapmanın bazen imkânsız oluşudur. Eğer uygun bir örnekleme yapılırsa, tüm bilgileri toplama çoğu zaman gereksizdir. Örnek, ana kitlenin bir alt grubudur. Gerekli bilgiler belirlenen örneklerden alınarak ana kitle hakkında sonuçlara varılır. Ana kitlenin özelliklerine göre örneklemenin yöntemi değişir. Örneğin, eğer ana kitle önemli özellikleri itibariyle homojen bir yapı arz ediyorsa tam şansa bağlı, yok eğer heterojen ise o zaman tabakalı örnekleme veya ana kitlenin özelliğine göre farklı örnekleme yöntemleri kullanılabilir.

İstatistik, diğer bir yönüyle de belirsizlik bilimidir. İstatistikte, nedir? sorusuyla değil, ne olabilirdi? ne olabilir? ve olma ihtimali nedir? gibi sorular ile uğraşılır. "Eğer bütçe açıkları kestirildiği gibi yüksek olursa, faiz oranları da senenin ikinci yarısında yüksek olarak kalabilir", "Tanınan bir firmanın hisse senedi fiyatları altı ay içinde şimdikinden daha yüksek olabilir" ve "Yüksek olarak tahmin edilen etin gelir elastikiyeti, tüketici gelirlerinde reel artış olduğunda et tüketiminin artacağını göstermektedir" gibi tam kesinlik arz etmeyen ifadelerle birtakım ekonomik yorumlar yapılabilir.

2.2. Toplam ve Çarpım Sembolleri ve Bazı Kullanımları

Örnek verileri kullanılırken genelde iki notasyon çok yaygın olarak kullanılmaktadır. Eğer n sayıda örnek söz konusu ise, toplam ($\sum$) ve çarpım işaretinin ($\prod$) kullanılması gerekmektedir. Bu notasyonlar, ekonometri modellerinin yazılmasında ileride

kullanılabileceğinden burada izah edilmesinde fayda vardır. Toplam işareti aşağıdaki gibi gösterilebilir.

$\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$ şeklinde ifade edilebilir.

İstatistikte ortalama bu notasyonla gösterilebilir. Ana kitle ortalaması, $\mu = \sum x_i / n$ olarak ifade edilebilir ve $\mu = (x_1+x_2+x_3+ ..... + x_n) / n$ şeklinde açılımı yapılabilir. Bu sembollerin bazı önemli özellikleri aşağıdaki gibidir:

1) $\sum_{i=1}^{n} c = nc$, eğer c bir sabit sayı ise

2) $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

3) $\sum_{i=1}^{n} (c + b x_i) = n c + b \sum_{i=1}^{n} x_i$

Bu sembollerin kullanılması durumunda kolaylık açısından sadece $\sum_i x_i$ veya $\sum x_i$ şeklinde yazılabilir. Bazen çift toplam işareti de kullanılabilir ki bu aşağıdaki gibidir.

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} = x_{11} + x_{12} + ...+ x_{1m}$$
$$+x_{21} + x_{22} + ...+x_{2m}$$
$$. \quad . \quad .$$
$$. \quad . \quad .$$
$$+ x_{n1} + x_{n2}+.....+x_{nm}$$

Yine burada da toplam işaretleri sadece $\sum_i$, $\sum_j$, $\sum_{ij}$ şeklinde yazılabilir. Ayrıca eğer denilse ki $\sum_i \sum_j x_{ij}$ $i \neq j$ ve i, 1'den 3'e j, 1'den 2 'ye gidiyorsa bu durumda açılım,

$\sum_i \sum_j x_{ij} = x_{11} + x_{12} + x_{21} + x_{22} + x_{31} + x_{32}$ yerine

$\sum\sum_{i \neq j} x_{ij} = x_{12}+x_{21}+x_{31}+x_{32}$ şeklinde yazılabilir.

Diğer bir örnek ele alınırsa,

$$(x_1 + x_2 + x_3)^2 = x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_2 x_3 + 2x_1x_3$$

bu da aşağıdaki gibi yazılabilir.

$$(\sum_i x_{ij})^2 = \sum_i \sum_j x_i x_j = \sum_{i \neq j} x_i^2 + \sum_i \sum_j x_i x_j$$

Çarpan işareti $\prod$ de aşağıdaki gibi tanımlanmaktadır.

$$\prod_{i=1}^{n} x_i = x_1 x_2 .... x_n$$

Burada da kolaylık açısından çarpan işareti yine sadece $\prod_i x_i$ veya $\prod x_i$ şeklinde yazılabilir. Toplam işaretinde olduğu gibi yine çift çarpan işareti kullanılabilir. Örneğin x_1, x_2, x_3 ve y_1, y_2 değişkenleri mevcutsa bu durum $\prod_i \prod_j x_i y_j = x_1 x_2 x_3 y_1 y_2$ şeklinde yazılabilir.

Bir ana kitlenin varyansı, her bir değerin ortalamadan sapmalarının karelerinin toplamının ana kitle sayısına bölümü ile bulunur. Bu da $\sigma^2 = \sum (x_i - \mu)^2 / N$ şeklinde ifade edilebilir. Ayrıca bu $(\sum x_i^2 - N\mu^2) / N$ veya $\sum (x_i^2 / N) - \mu^2$ şeklinde de yazılabilir. Ana kitlenin standart sapması, varyansın karekökü̈dür ve bu da σ'ya eşittir. Bu ders notlarında örnek varyansı, standart sapması ve ortalaması sırasıyla s^2, s ve x_{ort} ile gösterilecektir. Örnek varyansı hesap edilirken payda da N yerine n-1 kullanılır.

*2.3. İhtimal

İhtimal terimi, belli bir ifade ile alakalı belirsizliğe rakamsal bir değer vermektir. İhtimal ilk zamanlar kumar oyunlarında uygulanmak amacıyla ortaya çıkarılmıştır. Tahminen kumarcıların çoğunun bazı ihtimal hesapları yapmış olmalarına rağmen, ilk sistematik ihtimal hesaplarını yapanın bir İtalyan doktor olan Gerolama Cardaro (1501-1576) olduğu kabul edilmektedir. Cardaro'ya göre ihtimal aşağıdaki şekilde ifade edilebilir.

İhtimal = (istenen sonuç sayısı) / (toplam mümkün sonuç sayısı)

Bu ihtimalin klasik tanımıdır. Varsayalım hilesiz bir zar atılmaktadır. Bir altı elde etmenin ihtimali nedir? Mümkün olabilecek sonuç sayısı 6 (1,2,3,4,5,6) dır. Bu yüzden 6 elde etmenin ihtimali 1/6'dır.

İhtimalin diğer bir tanımı, çok sayıda tekrar edilmesi sonucu bir olayın ortaya çıkmasının nispi sıklığı şeklinde yapılabilir. Örneğin, 6 elde etme ihtimali, zarı çok sayıda atıp elde edilen 6 rakamlarının adet olarak toplamının toplam atış sayısına bölünmesi suretiyle hesap edilebilir. Bu gözlediğimiz oransal değer bize 6 elde etme ihtimalini verir. Hiç kimsenin bu yolla ihtimal hesap edecek kadar sabırlı olamayacağı düşünülebilir, fakat bunun örnekleri de vardır. Örneğin, Danimarka'da Kerrizk adında birisi demir bir parayı 10 bin defa atıyor ve tura gelme sıklığı çok değişme göstermesine rağmen sonuçta 0,5'e yakın bir değer olan 0,507 değerinde kalıyor.

İhtimalin sıklık derecesi fikri, ilk defa Fransız matematikçi Poisson tarafından 1837 yılında ortaya atılmıştır. Buna göre eğer n sayıda deneme varsa ve n(E)'de E olayının ortaya çıkış sayısı ise P(E) ile ifade edilen E olayının ihtimali,

$$P(E) = \lim_{n \to \infty} n(E) / n \quad \text{olarak ifade edilir.}$$

Klasik görüşe göre ihtimal, istenen durum sayısının toplam mümkün sayıya oranı olarak tanımlanan teoriksel bir sayıdır. Sıklık derecesi görüşüne göre ise, tekrarların sayısı sonsuza giderken ihtimal, gözlenen nispi sıklığın bir limitidir.

Üçüncü bir görüş olan subjektif görüşe göre ise, ihtimal kişisel inançlar temeline dayanır. Farz edelim ki gelecek hafta yapılacak bir futbol maçında Galatasaray'ın kazanma ihtimalinin ¾ olduğu birileri tarafından iddia edilmektedir. Galatasaray'ın kazanması durumunda 10 bin TL'yi bu kişilerin alacağı, fakat yenilmesi durumunda 30 bin TL'yi bu kişilerin vereceği bir bahse girebiliyorlarsa bu dürüst bir bahis olur. Yok, eğer böyle bir bahse girmek istenmiyorsa Galatasaray'ın yenmesi konusunda subjektif ihtimal hesabı 3/4 den daha azdır.

*2.4. Şansa Bağlı Değişkenler ve İhtimal Dağılımı

Eğer her bir gerçek sayı a için X'in a'dan küçük ve a'ya eşit değerleri aldığı bir ihtimal $P(X \leq a)$ mevcut ise X değişkeninin bir şansa bağlı olduğu söylenir. Bundan sonra şansa bağlı değişkenler büyük harflerle (X, Y, Z), bu değişkenlere ait her bir değer ise küçük harflerle (x, y, z) temsil edileceklerdir. Bu yüzden $P(X=x)$ bir ihtimaldir ve burada şansa bağlı değişken olan X, x değerini alır. $P(x_1 \leq X \leq x_2)$ X tesadüfi değişkeninin x_1 ve x_2 arasında aldığı değerlerin ihtimalidir ve her iki taraf da kapsam dâhilindedir.

Eğer şansa bağlı değişken olan X sadece belirli sayıda değerleri alabildiği varsayılıyorsa buna kesikli şansa bağlı bir değişken denilir. Eğer değişken, belli bir aralık içindeki değerlerin tümünü kapsıyorsa, bu şansa bağlı değişkene sürekli değişken denilir. Kesikli değişkene örnek olarak, belli bir iş yerine günün ilk saatinde gelen müşteri sayısı verilebilir. Sürekli şansa bağlı değişkene örnek ise Türkiye'deki insanların gelir seviyeleri verilebilir. Gerçek uygulamada devamlı şansa bağlı değişkenlerin kullanımı daha yaygındır, çünkü matematik teorisinde kullanımı daha basittir. Örneğin, gelirin devamlı şansa bağlı değişken olduğu ifade edilse bile aslında devamlı değişken olduğunda kesinlik yoktur. Fakat böyle bir yaklaşım kullanışlı olur.

Şansa bağlı bir X değişkeninin değişik değerleri için ihtimalleri veren bir formül, kesikli tesadüfi değişkenler durumunda bir ihtimal dağılımı, devamlı tesadüfi değişkenler durumunda ihtimal yoğunluk fonksiyonu olarak adlandırılırlar. Bu da genellikle f(x) tarafından temsil edilir.

Genel olarak, bir devamlı şansa bağlı değişken için X'in herhangi bir gerçek değerinin ortaya çıkma ihtimali sıfır olabildiğinden ihtimal, belli bir mesafe olarak tartışılır. Bu ihtimaller istenen bir aralık üzerine entegre edilerek temin edilirler. Örneğin, eğer $P(a \leq x \leq b)$ isteniyorsa bu aşağıdaki formülle verilir.

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$

Burada, c gibi bir sayıdan küçük veya eşit değerleri alan X şansa bağlı değişkeninin ihtimali çoğunlukla $F(c) = P(x \le c)$ şeklinde yazılabilir. F(x) fonksiyonu x'in değişik değerleri için, kümülatif ihtimalleri temsil eder ve bu yüzden kümülatif dağılım fonksiyonu olarak adlandırılır.

$$F(c) = P(X \le c) = \int_{-\infty}^c f(x)\,dx$$

*2.5. Normal Dağılım ve İlgili İhtimal Dağılımları

Eğer bir şansa bağlı değişken olan X'in ihtimal dağılımı verilirse, X'in belli bir aralık içerisindeki ihtimali belirlenebilir. İhtimalleri tablo haline getirilen ve çok geniş sayıdaki olaylar için uygun açıklamalar yaptığı kabul edilen bazı ihtimal dağılışları mevcuttur. Bunlar normal dağılım ve bu dağılıştan elde edilen t, χ^2, ve F dağılımlarıdır. Bunların dışında başka dağılımlar da vardır, fakat bu ders kapsamında daha çok bunlar üzerinde durulacak ve bu dağılımlar kullanılacaktır. Normal dağılımın ekonomik değişkenleri açıklamada uygun olup olmadığı konusunda ortada bir kuşku mevcuttur. Ancak, değişkenler normal dağılmamış olsalar da, bu değişkenlerin transformasyonu düşünülebilir ve böylece transformasyonu yapılmış değişkenler böylece normal dağılım gösterirler.

Normal Dağılım: Normal dağılım, çok değişik bilimsel sahalarda yapılan istatistik uygulamalarında en yoğun olarak kullanılan çan eğrisi şeklinde bir dağılıştır. Bu dağılımın ihtimal yoğunluk fonksiyonu aşağıdaki gibi yazılır.

$$f(x) = (1/\sigma\sqrt{2\pi})\, e^{[-1/2\sigma^2 (x-\mu)^2]} \qquad -\infty < x < +\infty$$

Ortalaması μ ve varyansı σ^2 dir. Eğer x, ortalaması μ, ve varyansı σ^2 olan normal dağılışa sahip olursa bu durum kısa ve özlü bir şekilde $x \sim N(\mu,\sigma^2)$ olarak ifade edilir. Normal dağılışın önemli bir özelliği, normal olarak dağılmış değişkenlerin herhangi bir doğrusal fonksiyonu da normal olarak dağılmıştır. Değişkenler bağımlı olsa da olmasa da bu özellik geçerlidir. Eğer $x_1 \sim N(\mu_1,\sigma_1^2)$ ve $x_2 \sim N(\mu_2,\sigma_2^2)$ ise ve x_1 x_2 arasındaki korelasyon ρ ise bu durumda

$$a_1x_1 + a_2x_2 \sim N\,(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2\rho a_1a_2\sigma_1\sigma_2),$$

$$x_1+x_2 \sim N\,(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2) \text{ ve}$$

$$x_1-x_2 \sim N\,(\mu_1 - \mu_2,\ \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2) \text{ olur.}$$

Normal dağılıma ilave olarak, sıklıkla kullanılacak olan diğer dağılımlar da vardır. Bunlar χ^2, t ve F dağılımlarıdır. Bunlar normal dağılımdan çıkarılmış olan ve aşağıda açıklandığı gibi tanımlanan dağılımlardır.

χ^2 Dağılımı: Eğer x_1, x_2, ... x_n ortalaması 0 ve varyansı 1 olan bağımsız normal değişkenler ise, yani $x_i \sim IN(0,1)$, i =1, 2, ..., n ise o zaman $Z = \sum x_i^2$ 'nin n serbestlik derecesiyle χ^2 dağılımına sahip olduğu söylenir ve $Z \sim \chi^2_n$ şeklinde yazılır. İndis n serbestlik derecesini gösterir. χ^2_n dağılımı, n sayıdaki bağımsız standart normal değişkenlerin karesinin toplamının dağılımıdır.

Eğer $x_i \sim IN(0,\sigma^2)$ ise o zaman, $Z= \sum_i (x_i^2 / \sigma^2)$ şeklinde tanımlanmalıdır. χ^2 dağılımı aynı zamanda, normal dağılımın özelliğinden farklı olmasına ve daha çok sınırlayıcı olmasına rağmen toplayıcı özelliğe sahiptir. Bu özellik, eğer $Z_1 \sim \chi^2_n$, $Z_2 \sim \chi^2_m$ ve Z_1 ve Z_2 bağımsızlarsa o zaman $Z_1+Z_z \sim \chi^2_{n+m}$ şeklinde ifade edilebilir. Burada not edilmesi gereken husus, herhangi bir genel doğrusal kombinasyon değil, bağımsızlığa duyulan ihtiyaç ve sadece basit toplamaların dikkate alınabilmesidir. Bu sınırlı özellik bile pratik uygulamalarda çok faydalı olmaktadır. Bu sınırlı özelliği bile taşımayan çok sayıda dağılımlar mevcuttur.

***t* Dağılımı:** Eğer $x \sim N(0,1)$ ve $y \sim \chi^2_n$ ve x ve y bağımsızlarsa, $Z = x / \sqrt{y/n}$, n serbestlik derecesi ile *t* dağılımına sahiptir. Bu $Z \sim t_n$ şeklinde yazılır ve indis n yine serbestlik derecesini temsil eder. Bu yüzden *t* dağılımı, bir standart normal değişkenin ortalaması alınmış bağımsız χ^2 değişkeninin (χ^2 değişkeni kendi serbestlik derecesine bölünmüştür) karekökünе bölünmüş şeklinin dağılımıdır. Burada *t* dağılımı, normal dağılım gibi simetrik bir ihtimal dağılımıdır, fakat normal dağılımdan daha yassı ve uzun uçlara sahiptir. Serbestlik derecesi sonsuza yaklaştıkça *t* dağılımı da normal dağılıma yaklaşır.

F Dağılımı: Eğer $y_1 \sim \chi^2_{n1}$ ve $y_2 \sim \chi^2_{n2}$ ve y_1 ve y_2 bağımsızlarsa $Z = (y_1/n_1) / (y_2/n_2)$, n_1 ve n_2 serbestlik dereceleri ile F-dağılımına sahiptir. Bunu $Z \sim F_{n1, n2}$ gibi ifade ederiz. Buradaki ilk indis n_1, paydaki serbestlik derecesini, ikinci indis n_2, paydadaki serbestlik derecesini ifade eder. Bu yüzden F-dağılımı, χ^2'ye ortalanmış bağımsız iki değişkenin oranının dağılımıdır.

*2.6. Klasik İstatistik Çıkarım (Inference)

İstatistik çıkarım, kullandığımız gözlenen verilerin kullanma yöntemini açıklayan sahadır ki bu saha, üretilen verilerin üretim işlemi hakkında veya verilerin geldiği ana kitle hakkında bir takım sonuçlar çıkarır. Buradaki varsayım eldeki verileri üreten bilinmeyen bir işlem vardır ki bu işlem bir ihtimal dağılımı ile açıklanır ve bilinmeyen bazı parametreler tarafından tanımlanır. Örneğin, bir normal dağılım için bilinmeyen parametreler μ ve σ^2'dir.

Geniş anlamda konuşulursa, istatistik çıkarım, Klasik ve Beyziyen olmak üzere iki başlık altında sınıflandırılır. Klasik çıkarım iki varsayım temeline oturur. Bunlar (1) örnek verileri sadece verilerle ilgili bilgileri içerir ve (2) çıkarım için kullanılan değişik yöntemlerin inşası ve esas olarak benzer şartlar altındaki uzun dönem davranışı değerlendirilmesi temeline dayanır.

Beyziyen çıkarımda örneğe ait bilgiler, daha önceden üretilen mevcut olan bilgilerle birleştirilir. Farz edilsin ki, bilinen ortalama (μ) ve standart sapmaya (σ^2) sahip normal bir ana kitleden n sayıda (y_1, y_2,...., y_n) bir şansa bağlı örnek alındı ve μ hakkında çıkarımlar yapılması isteniyor. Klasik çıkarımda, örnek ortalaması olan y_{ort} μ'nün tahmini olarak alınır ve bunun varyansı da σ^2/n olarak kabul edilir. Beyziyen çıkarımda μ ile ilgili öncelikli bilgi vardır ve öncelik dağılımı olarak bilinen bir ihtimal dağılımıyla ifade edilir. Öncelik dağılımı ortalaması μ_0 ve standart sapması σ^2_0 olan normal bir dağılım olduğu kabul edilirse, bu dağılım örnekten elde edilen bilgi ile μ'nün sonraki dağılımı olarak bilinen dağılımı sağlamak için birleştirilir. Bu dağılımın normal olduğu gösterilebilir çünkü sonraki ortalama, örnek ortalaması ve öncelikli bilgilerin ortalaması arasında uzanır. Sonraki varyans hem örnek hem de öncelik varyansından daha az olur.

Bu ders kapsamında arzu edilenden daha fazla detay olacağından Beyziyen çıkarım metodu tartışılmayacaktır. Klasik çıkarım biçimine geri dönülürse, klasik çıkarımı, nokta tahmini, aralık tahmini ve hipotez testleri olarak üç başlık altında tartışmak geleneksel bir yoldur.

2.7. Nokta Tahmini

İhtimal dağılımının, bir parametre olan θ'yı içinde bulundurduğu ve bu dağılımdan n sayıda örneğin elde mevcut olduğu yani bu ihtimal dağılımından y_1, y_2,..., y_n örnek verilerinin elde edildiği kabul edilsin. Nokta tahmininde, bu gözlemlerden bir $g(y_1, y_2,..., y_n)$ fonksiyonu oluşturulur ve g'nin θ'nın tahmincisi olduğu söylenir. Yaygın terminoloji, $g(y_1, y_2,..., y_n)$ fonksiyonunun bir tahmin edici olduğunu ve belli bir örnekteki θ'nın değerini tahmin etmek için kullanılır. Bu yüzden tahmin edici şansa bağlı bir değişken, tahmin ise bu şansa bağlı değişkenin belli bir değeridir. Örneğin θ bir ana kitlenin ortalaması ve $g(y_1, y_2,..., y_n) = 1/n \sum yi = y_{ort}$ örneğin ortalaması ise, y_{ort} da θ'nın tahmin edicisidir. Belli bir örnekte eğer $y_{ort} = 4$ ise burada 4, θ'nın bir tahmin edilen değeridir.

2.8. Aralık Tahmini

Aralık tahmininde, örnek gözlemlerinin $g_1(y_1, y_2,..., y_n)$ ve $g_2(y_1, y_2,..., y_n)$ şeklindeki iki fonksiyonunu oluşturulur ve denilir ki θ, belli bir ihtimalle g_1 ve g_2 arasında uzanır. Hipotez testinde, θ hakkında bir fikir ileri sürülür. Örneğin hipotez H_0: $\theta = 4$ şeklinde

ifade edilir ve hipotezi kabul veya reddetme temeline bağlı olarak bu hipotez doğrultusunda örnekteki olayın derecesi incelenir.

Aralık tahmininde örnek gözlemlerinin iki fonksiyonu olan $g_1(y_1,y_2,....,y_n)$ ve $g_2(y_1,y_2,...,y_n)$ oluşturulur. Bu durumda $P(g_1 < \theta < g_2) = \alpha$, verilen bir ihtimal olarak ifade edilir. Burada α, güven katsayısı, g_1 ve g_2 aralığı ise güven aralığı olarak adlandırılır. θ bir parametre veya bilinmeyen bir sabit olduğundan dolayı, yukarıdaki ihtimal ifadesi θ hakkında değil, g_1 ve g_2 hakkında bir ifadedir. Bunun ima ettiği şey, eğer $g_1(y_1,y_2,...y_n)$ ve $g_2(y_1,y_2,...y_n)$ formülleri, değişik örneklerle tekrar edilerek kullanırsa ve her seferinde güven aralığı oluşturulursa, o zaman bu durumların %100 α'sında verilen aralık, gerçek değeri içinde bulunduracaktır.

Güven aralığını oluşturmak için örnekleme dağılımlarının nasıl kullanılacağına bir örnek olarak, ortalaması μ ve varyansı σ^2 olan bir normal ana kitleden n sayıdaki bağımsız $y_1,y_2,..,y_n$ gözlemlerine sahip bir örnek dikkate alınsın. O zaman,

$$(n-1)\, s^2 / \sigma^2 \sim \chi^2_{n-1} \quad \text{ve} \quad \sqrt{n}(y_{ort} - \mu) / s \sim t_{n-1}$$

ki burada y_{ort} ve s^2 örnek ortalaması ve örnek varyansıdır. Eğer bir örnek büyüklüğü n = 20 ve serbestlik derecesi n-1= 19 ise χ^2 tablosuna 19 serbestlik derecesiyle bakıldığında,

$$P\,(\, 19\, s^2 / \sigma^2 > 32{,}9\,) = 0{,}025 \qquad (2.1)$$

$$P\,(\, 10{,}1 < 19\, s^2 / \sigma^2 < 30{,}1\,) = 0{,}90 \qquad (2.2)$$

Ayrıca, 19 serbestlik derecesiyle *t* tablosu kullanıldığında,

$$P\,(\, -2{,}093 < \sqrt{n}(y_{ort} - \mu) / s < 2{,}093\,) = 0{,}95 \qquad (2.3)$$

Bir önceki 2.2 eşitliğinden aşağıdaki eşitlik elde edilir.

$$P\,(\, 19\, s^2 / 30{,}1 < \sigma^2 < 19\, s^2 / 10{,}1\,) = 0{,}90 \qquad (2.4)$$

ve eğer $s^2 = 9{,}0$ ise bu durumda σ^2 için % 90 güven aralığı (5,7 - 16,9) olarak hesap edilir. Yine yukarıdaki 2.3 denkleminden,

$$P\,(\, y_{ort} - (2{,}093\, s / \sqrt{n}) < \mu < y_{ort} + (2{,}093\, s / \sqrt{n}\,)) = 0{,}95 \qquad (2.5)$$

elde edilir. Eğer $y_{ort} = 5$ ve $s = 3{,}0$ ise biz μ için %95 güven sınırlarını (3,6 – 6,4) olarak elde edilir.

Bu aralıklara iki taraflı aralık denir. Tek taraflı aralıklar da oluşturulabilirler. Örneğin, ilk baştaki 2.1 eşitliğinden,

$$P\,(\, \sigma^2 < 19\, s^2 / 32{,}9\,) = 0{,}025$$

$$P\,(\, \sigma^2 > 1\,9\, s^2 / 32{,}9\,) = 0{,}975$$

Eğer $s^2 = 9$ ise, σ^2 için % 97,5 sağ taraflı güven aralığını (5,205 - ∞) olarak elde edilir. μ için de tek taraflı benzer güven aralığı hesap edilebilir.

Pratikte bilinmesi gereken şey, nokta tahmin edicinin (g), aralık tahmin edicinin (g_1, g_2) ve hipotez testinin yönteminin nasıl oluşturulacağıdır. Klasik istatistik yorumda bütün bunlar örnekleme dağılımlarına dayanırlar. Örnekleme dağılımları, örnek gözlemlerinin fonksiyonlarının ihtimal dağılımlarıdırlar. Örneğin, örnek ortalaması y_{ort}, örnek gözlemlerinin bir fonksiyonudur ve onun ihtimal dağılımı y_{ort}'nın örnek dağılımı olarak adlandırılır. Klasik istatistik çıkarım metodunda, tahmin edicilerin özellikleri, onun örnekleme dağılımlarının özellikleri içerisinde tartışılırlar.

*2.9. Tahmin Edicilerin Özellikleri

Bu derste sık sık bahsedilecek ve dolayısıyla bilinmesi gereken tahmin edicilerin istenen bazı özellikleri vardır. Bunlar;

1. Sapmasızlık (unbiasedness)
2. Etkinlik (efficiency) ve
3. Tutarlılık (consistency) olarak sıralanabilir. Bunların ilk ikisi küçük örnek, üçüncüsü ise bir büyük örnek özelliğidir.

Bir tahmin edici g, eğer $E(g) = \theta$ yani g'nin örnekleme dağılımının ortalaması θ'ya eşit ise sapmasız olduğu söylenir. Bunun demek istediği şey, eğer biz her bir örnek için g'yi hesaplar ve bu işlemi sonsuz defa tekrar edersek bütün bu tahminlerin ortalaması θ'ya eşit olacaktır. Eğer $E(g) \neq \theta$ ise g'nin sapmalı (bias) olduğu söylenir ve $E(g) - \theta$'yı sapma olarak kabul ederiz. Sapmasızlık istenen bir özelliktir fakat maliyetinin yüksek olmaması gerekir. Farz edilsin ki g_1 ve g_2 diye iki tahmin edici mevcut ve g_1 θ'dan uzak değerleri almasına rağmen onun ortalaması hala θ'ya eşitken g_2 θ'ya yakın dizilir fakat onun ortalaması θ'dan biraz uzaktadır. Bu durumda sapmalı olmasına rağmen varyansı daha küçük olduğundan g_2 tercih edilebilir. Eğer tahmin edicinin varyansı geniş olursa, tahminin gerçek değerden uzak olduğu yerde bazı istenmeyen örneklere sahip olunabilir. Bu yüzden istenen ikinci önemli özellik tahmin edicinin küçük varyansa sahip olmasıdır.

Etkinlik özelliği tahmin edicinin varyansı ile ilgilidir ve doğal olarak nispi bir kavramdır. Eğer g sapmasız bir tahmin edici ise ve sapmasız tahmin ediciler sınıfında minimum varyansa sahip ise g'nin etkin bir tahmin edici olduğu söylenir. Bu durumda g minimum varyanslı tahmin edici olarak adlandırılır. Eğer doğrusal bir tahmin edici dikkate alınıyorsa, yani $g = c_1y_1 + c_2y_2 + \ldots + c_ny_n$ ve burada c'ler g'nin sapmasız ve minimum varyansa sahip olduğu seçilen sabitler ise g, sapmasız en iyi doğrusal tahminci (Best Linear Unbiased Estimator - BLUE) olarak adlandırılır.

Çoğu zaman sapmasızlık ve etkinlik gibi istenen küçük örnek özelliklerine sahip tahmin edicileri bulmak mümkün değildir. Bu gibi durumlarda istenen büyük örnek özelliğine bakılır. Bunlar, asimtotik özellikler olarak adlandırılır. Tutarlılık, asimtotik sapmasızlık ve asimtotik etkinlik gibi özellikler bunlardandır. Tutarlılık özelliği, genellikle büyük verilerde aranır. Varsayalım ki, θ_n, n örnek büyüklüğüne sahip θ'nın bir tahmin

edicisidir. Yani, örnek büyüklüğü artırılarak, θ_n tahmin edicisi, keyfi olarak yakın ihtimal ile gerçek değere yaklaşması sağlanabilir. Böylece, örnek büyüklüğü sonsuza yaklaştıkça θ'nın tahmin hatasının sıfıra yaklaştığı söylenebilir.

2.10. Normal Bir Ana Kitlenin Örnekleri için Örnekleme Dağılımları

En yaygın olarak kullanılan örnekleme dağılımları, normal ana kitleden elde edilen örnekler için kullanılan dağılımlardır. Örneğin, n sayıda bağımsız gözlemlerin bir örneği yani, ortalaması μ ve varyansı σ^2 olan normal bir ana kitleden $y_1, y_2,...., y_n$ şeklinde elde bir örnek var olsun. Bu durumda,

$y_{ort} = (1/n) \sum y_i$ örnek ortalaması

$s^2 = (1/ n-1) \sum (y - y_{ort})^2$ örnek varyansı olarak hesaplanır. Böylece,

- Örnek ortalaması olan y_{ort}'nın örnekleme dağılımı, μ ortalama ve σ^2/n varyans ile normaldir.
- $(n-1)\ s^2 / \sigma^2$, (n-1) serbestlik derecesi ile bir χ^2 dağılımına sahiptir. Ayrıca y_{ort}'nın ve s^2'nin dağılımları bağımsızdırlar.
- $\sqrt{n}(y_{ort}-\mu)/\sigma \sim N(0,1)$ ve $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$ olduğundan ve bu dağılımlar bağımsız olduklarından $\sqrt{n}(y_{ort}-\mu)/s$, (n-1) serbestlik derecesi ile bir *t*-dağılımına sahiptir.
- Ayrıca $E(y_{ort}) = \mu$ ve $E(s^2) = \sigma^2$ olduğundan y_{ort} ve s^2 sırasıyla μ ve σ^2 için sapmasız tahmin edicilerdir.

Bu örnekleme dağılımları, aralık tahminlerini elde etmek ve μ ve σ^2 hakkındaki hipotezleri test etmek için de kullanılacaktır.

2.11. Hipotez Testi

Hipotez testinin ne anlama geldiği, hipotez testlerinin çeşitleri, hipotez testlerinin nasıl test edildiği, test sonuçlarının ne anlama geldiği, test sonuçlarını nasıl algılamamız gerektiği, hipotez testlerinde ne gibi hataların yapılabileceği gibi konular, hipotez testinin önemli unsurlarıdır.

• Bir istatistiksel hipotez, örneğin çıkarıldığı ana kitledeki bazı parametre değerleri hakkındaki bir ifadedir.

• Bir parametrenin spesifik bir değere sahip olduğunu söyleyen hipoteze nokta hipotezi denir. Belli aralıkta uzandığını söyleyen hipoteze ise aralık hipotezi denir. Örneğin eğer μ, ana kitle ortalaması ise H: $\mu = 4$ bir nokta hipotezidir. H: $4 \leq \mu \leq 7$ ise bir aralık hipotezidir.

• Bir hipotez testi, örnek değeri ve hipotez olarak belirtilen ana kitle değeri arasında gözlenen farklılığın gerçek mi yoksa değişen bir varyasyondan mı kaynaklandığına

cevap verme metoduna denir. Örneğin, ana kitle ortalaması $\mu = 6$ ve örnek ortalaması $y_{ort} = 8$ olduğunda, buradaki farkın gerçek mi yoksa değişen varyasyondan mı kaynaklandığının bilinmek istenmesi bir hipotez testidir.

• Test ettiğimiz hipotezi, sıfır hipotezi olarak adlandırırız ve çoğu zaman H_0 ile gösteririz. Alternatif hipotez de H_1 ile gösterilir. Doğru olduğu halde H_0'ı reddetme ihtimali, gerçekte önem seviyesi olarak adlandırılır. Verilerden elde edilen değer ve H_0 hipotezi altında beklenen değer arasında gözlenen farkın gerçek veya varyasyon değişmesine bağlı olduğunu test etmek için test istatistiği kullanılır. Test istatistiği için arzu edilen kriter, örnekleme dağılımının kullanımının kolay, tercih edilir ve tablo haline getirilmiş olmasıdır. Bu tablolar normal, *t*, χ^2 ve *F* dağılımları için hazır olarak vardır ve böylece seçilen test istatistiği çoğu zaman bu dağılıma sahip olanlar için kullanılır.

• Gözlenen önem seviyesi veya *P*-değeri, test istatistiğinin gözlenen değerinden uzak veya daha uzak olan bir test istatistiği getirmenin ihtimalidir. Bu ihtimal sıfır hipotezinin doğru olması temeline dayanılarak hesap edilir. Örneğin ortalaması μ ve varyansı σ^2 olan bir normal dağılıştan n kadar bağımsız gözlemlerin örneği dikkate alınmış olsun. Burada aşağıdaki hipotezlerin test edilmesi isteniyor.

$$H_0: \mu = 7 \quad \text{karşı hipotez} \quad H_1: \mu \neq 7$$

Bu durumda kullanılacak test istatistiği, (n-1) serbestlik derecesi ile t dağılımına sahip olan $t = \sqrt{n}(y_{ort} - \mu) / s$ dir. Örneğin, $n = 25$, $y_{ort} = 10$, $s = 5$. H_0'ın doğru olması durumunda t'nin gözlenen değeri $t_0 = 3$ dür. Burada *t*'nin yüksek pozitif değerleri H_0 hipotezinin aleyhine bir durumdur. *P*-değeri, serbestlik derecesi (n-1) = 24 olduğundan $P=P(t_{24}>3)$ olur ki, bu gözlenen önem seviyesidir.

• Testin sonucunun önemli veya önemsiz olduğunun söylenmesi ve gerçek *P*-değerinin rapor edilmemesi yaygın bir uygulamadır. Bu iki terimin manası aşağıdaki gibidir. *İstatistiksel olarak önemli*: H_0 hipotezi ile örnek değerleri arasındaki farkı örnekleme varyasyonunun açıklayamaması halidir. *İstatistiksel olarak önemli değil*: H_0 hipotezi ile örnek değeri arasındaki farkı örnekleme varyasyonunun açıklaması durumudur.

Aynı zamanda, önemli ve yüksek derecede önemli terimleri geleneksel olarak 0.05 veya 0.01 seviyesinde sırasıyla önemli olduğunu ifade etmektedir. Fakat örneğin % 5 önem seviyesinde *P*-değerinin 0.055 ve 0.045 olduğu iki durumu dikkate alındığında, birinci durumda farkın önemli olmadığı fakat ikinci durumda da önemli olduğu, her iki durumda da örnek olayının marjinal olarak farklı olmasına rağmen söylenecektir. Buna benzer olarak 0.80 ve 0.055 *P*-değeri olan iki testin önemli olmadığı her iki durumda da sıfır hipotezi ile örnek değeri arasında çok büyük fark olmasına rağmen kabul edilebilecektir. Bu yüzden çoğu bilgisayar programı *P*-değerlerini de verir.

• İstatistiksel önem ve pratikteki önem aynı şeyler değillerdir. İstatistiksel olarak çok önemli olan bir sonuç pratik açıdan hiçbir öneme sahip olmayabilir. Örneğin, beklenen ortalama ağırlığı 450 gram olan yer fıstığı kutuları ile ilgili yapılan örnekleme ile

ağırlıkların gerçek örnek ortalaması 449,5 gram bulunmuştur. Bu farkın, yeterince geniş örneğe veya yeterince küçük örnekleme varyansına sahip olunduğunda istatistiksel olarak önemli fakat pratik olarak önemli olmadığı sonucuna varılabilir. Diğer tarafta çok ince ölçekli aletlerle ilgili olarak örneğin, eğer bir parçanın beklenen uzunluğu 10 cm ve örnek ortalaması 9,9 ise ve örnek büyüklüğü n küçük ve s değeri büyük ise farklılık istatistiksel olarak önemli çıkmayabilir fakat pratikte çok önemli olabileceği söylenebilir.

• Test istatistiği, seçilen önemlilik seviyesinde istatistiksel olarak önemli ise H_0 reddedilir ve yine seçilen belli bir önemlilik seviyesinde test istatistiği önemli değil ise H_0 hipotezi reddedilmez. Gerçekte H_0 doğru veya yanlış olabilir. Bu her iki gerçek durumunda ortaya çıkabilecek iki sonuç çizelge 2.1'deki 4 ihtimale neden olur.

Çizelge 2.1. Hipotez testlerinde ortaya çıkan hatalar

	Gerçek	
Testin Sonucu	H_0 doğru iken	H_0 yanlış iken
Önemli (H_0'ı red)	Birinci tip hata (α)	Doğru sonuç
Önemsiz (H_0'ı kabul)	Doğru sonuç	İkinci tip hata (β)

Yapabilecek iki tip hata vardır, (1) doğru olduğu halde H_0'ı reddetmek ki bu birinci tip hata veya α hatası olarak bilinir, (2) doğru olmadığı halde H_0'ı reddetmemek ki bu ikinci tip hata veya β hatası olarak bilinir. Bu yüzden,

$$\alpha = P\ (H_0\text{ 'ı reddetme} / H_0 \text{ doğru})$$

$$\beta = P\ (H_0\text{ 'reddetmeme} / H_0 \text{ doğru değil})$$

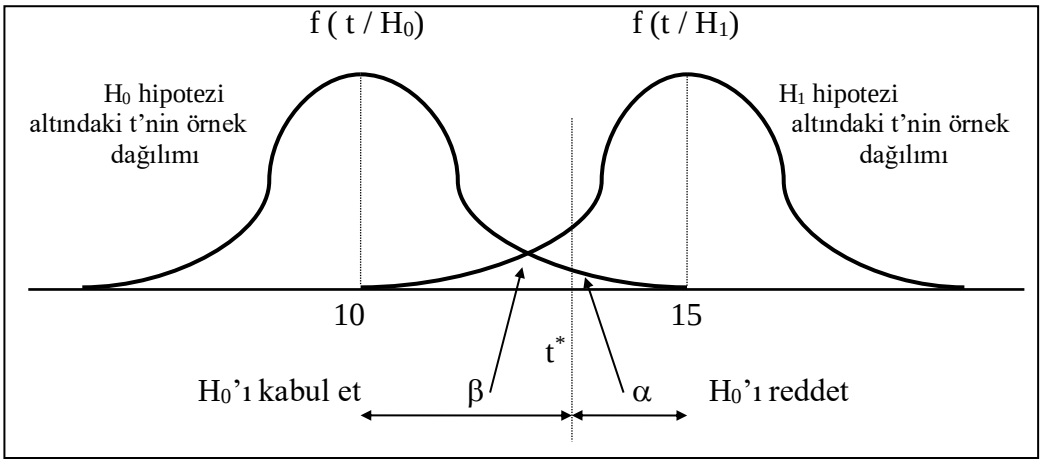
α önceden de tanımlandığı gibi sadece önemlilik seviyesidir ve $(1-\beta)$ ise testin gücü olarak bilinir. Testin gücü alternatif H_1 hipotezi belirlenmeden hesap edilemez. Yani H_0 doğru değil demek, H_1 doğru demektir. Örneğin, varyansı σ^2 ve ortalaması μ olan normal bir ana kitle için H_0: $\mu = 10$ karşılığı H_1: $\mu = 15$ şeklinde bir hipotez testi problemi dikkate alınmış olsun. Kullanılan test istatistiği $t = \sqrt{n}(x_{ort} - \mu) / S$ ve eldeki örnek verileri de n, x_{ort} ve S'ye olsun. Burada α'yı hesap etmek için $\mu = 10$ ve β'yı hesap etmek için de $\mu = 15$ kullanılır. Bu durumda iki hata aşağıdaki gibi ifade edilebilir.

$$\alpha = P(\ t > t^* \mid \mu = 10)$$

$$\beta = P(\ t < t^* \mid \mu = 15)$$

Burada t^*, H_0 hipotezini kabul edip etmemede kullandığımız t'nin kesim noktasıdır. H_0 ve H_1 hipotezleri altındaki t'nin dağılımı şekil 2.1'de gösterilmiştir. Örneğimizde α, H_0 hipotezi altında t-dağılımındaki sağ kuyruktaki alan, β da H_1 hipotezi altındaki t-dağılımının sol kuyruğundaki alandır.

Eğer alternatif hipotez H_1: $\mu < 0$ ise, o zaman H_1 altındaki *t*-dağılımı, H_0 hipotezinin altındaki *t*-dağılımının solunda olacaktır. Bu durumda α, H_0 hipotezinin altındaki *t*-dağılımının sol taraftaki kuyruk alanı ve β da H_1 hipotezi altındaki *t*-dağılımının sağ tarafındaki kuyruk alanı olacaktır.



Şekil 2.1. Hipotez testinde birinci ve ikinci tip hatalar

Tavsiye edilen alışılmış yöntem ki buna Neyman-Person yaklaşımı da denilir, belli bir seviyede α'yı sabit tutup β'yı minimum yapmaktır. Yani en fazla güce sahip test istatistiğini seçmektir. Uygulamada kullandığımız örneğin *t*, χ^2, ve F gibi testlerin en güçlü testler olduğu tespit edilmiştir.

• Neyman-Pearson teorisini kabul etmeyen bazı istatistikçiler de vardır. Örneğin, bazılarına göre, bir hipotezin kabul veya reddi için karar kuralı olarak bir önemlilik testini çok ciddiye almak çok büyük bir basitleştirmedir. Bu tip kararlar sadece deneysel bir olgu üzerine değil daha kompleks olgular üzerine yapılırlar. Bu yüzden önemlilik testinin amacı, verilerdeki olayın güçlülüğünü, 0 ile 1 arasında değişen bir ölçüyle ifade edilen bir hipoteze karşı sadece rakamsal olarak ifade etmekten ileri gitmemelidir. Yoksa bir sonucu kabul veya reddetme teklifinde bulunma değildir. Diğer tarafta bazı istatistikçiler var ki bunlar, önemlilik seviyesinin örnek büyüklüğüne bağlı olması gerektiğini iddia etmektedirler.

Önceden belirlenmiş önemlilik seviyesi, örnek büyüklüğü yeterince büyük olduğunda bütün sıfır hipotezlerinin reddedilebilmesinden dolayı problem olmaktadır. Bu, binlerce gözlemi olan geniş yatay kesit verilerini kullananların çoğu zaman karşılaştığı bir durumdur. Hemen hemen bütün katsayılar % 5 seviyesinde önemli bulunmaktadır. Bir istatistikçi, çok geniş örnek verileri için daha düşük önemlilik seviyesini, küçük örnek verileri için daha yüksek önemlilik seviyelerini kullanılması gerektiğini iddia etmiştir. Başka bir istatistikçi, regresyon modellerinde küçük örnekler için % 5'den çok daha

yukarı, büyük örnekler için % 5'den daha aşağı önem seviyelerinin kullanılması gerektiğini belirtmiştir.

Çoğu zaman testin amacı tahmin sürecini basitleştirmektir. Bu bir ön test problemidir ki burada test ilerdeki tahminler için ön bilgi sunar. Bu tip ön testler durumunda kullanılan önemlilik seviyesinin yaygın olan % 5'lik testten çok daha yüksek olmasının uygun olacağı tespit edilmiştir (bazen % 25'ten % 50'ye ve hatta % 99'a). Dikkate alınması gereken önemli husus önemlilik testinin birçok amacının olması ve sadece % 5 önem seviyesinin sürekli olarak kullanılmamasıdır.

2.12. Güven Aralığı ve Hipotez Testleri Arasındaki İlişki

Güven aralığı ve hipotez testleri arasında çok yakın bir ilişki vardır. Örneğin, bir hipotezin % 5 önemlilik seviyesinde testi isteniyor ise, dikkate alınan parametreler için bir % 95 güven aralığı tespit edilebilir ve hipotezi test edilen değerin bu aralık içinde olup olmadığı görülebilir. Eğer böyle ise, o zaman hipotez reddedilmez, yok eğer değilse o zaman hipotez reddedilir.

Örneğin, varyansı σ^2 ve ortalaması μ olan normal dağılımdan alınan 20 tane bağımsız gözlemin örnek ortalaması $y_{ort} = 5$, varyansı $s^2 = 9$'dur. Aralık tahmininindeki örnekte görüldüğü gibi güven aralığı 3,6 – 6,4'dür. Burada varsayalım ki H_0: $\mu = 7$ ve H_1: $\mu \neq 7$ şeklinde bir hipotez ve karşı hipotez kuruluyor. Eğer % 5 önemlilik seviyesini dikkate alınırsa, H_0 hipotezinin reddedilmesi gerekir. Çünkü

$$\left| \sqrt{n}\,(y_{ort} - 7) / S \right| > 2{,}093$$

2.093 noktası 19 serbestlik derecesindeki *t* tablosundan bir noktadır ki burada P (-2,093 < t < 2,093) = 0,95 veya P ($|t| > 2{,}093$) = 0,05. Bu örnekte $\sqrt{n}(y-7)/s = -3$ olduğundan H_0 reddedilir. Aslında her ne zaman H_0, μ değerini belirlenen güven aralığı (3.6 - 6.4) dışında bir değer olarak ifade ederse o zaman H_0 % 5 önemlilik seviyesinde reddedilecektir.

Tek kuyruklu test için bir taraflı güven aralığının dikkate alınır. Eğer H_0: $\mu = 7$ ve H_1: $\mu < 7$ ise o zaman $t = \sqrt{n}\,(y_{ort} - 7) / S$'nin yüksek değerleri için H_0 reddedilir. Ayrıca t tablosundan 19 serbestlik derecesi ile P($t < -1{,}73$) = 0,05 bulunur. Eğer $t < -1.73$'ü gözlemlersek H_0 reddedilir. Örnekte bu değer -3 olduğundan H_0 reddedilir. Bunun karşılığı olan % 95 tek taraflı güven aralığı aşağıdaki şekilde verilir.

$$P(-1{,}73 < (\sqrt{n}\,(y_{ort} - \mu) / s)) = 0{,}95$$

Burada n = 20, $y_{ort} = 5$, S = 3 yukarıda ki denklemde yerine konulduğunda ($-\infty$, 6,15) güven aralığını verir.

YAZILIM UYGULAMASI

```
*******************************************************************************
  SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
  ** Copyright (C)  1999  by K.J. White - All Rights Reserved **

  FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
  AT: ATATURK UNIVERSITESI - ERZURUM

  FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
*******************************************************************************
```

Uygulamalı Örnek 2.1

```
*******************************************************************************
|_*Burada ve bundan sonraki bölümlerdeki SHAZAM komutları dosyası,
|_*verilerin nasıl okunacağı ve bazı deskriptif istatistiklerin nasıl elde
|_*edileceğini vermektedir. Verilerin okunmasının burada iki metodu
|_*gösterilmektedir. Birincisinde veriler, doğrudan SHAZAM komut dosyasına
|_*yüklenmektedir. İkincisi ise veriler için ayrı bir dosya
|_*kullanılmaktadır. Başında yıldız (*) olan satırlar, yorum yapan
|_*satırlardır ve SHAZAM programı tarafından dikkate alınmaz SAMPLE komutu,
|_*dikkate alınacak gözlemler aralığını belirler. Koyu renkli sıralar
|_*SHAZAM komutlarını, diğerleri ise çıktıları gösteriyor.

|_SAMPLE 1 30

|_*READ komutu, verileri programa dahil eder ve değişkenleri isimlendirir.
|_*Değişken isimleri sekiz karakterden fazla olmamalıdır. Eğer fazla olursa
|_*SHAZAM bunu kısaltır ve uyarı mesajı verir. Veriler ayrı bir dosyada
|_*saklandığı zaman, verilerin bulunduğu dosya READ komutunu takip eder.

|_READ X Y
    2 VARIABLES AND       30 OBSERVATIONS STARTING AT OBS       1
|_*GENR komutu yeni değişkenlerin oluşturulmasını sağlar
|_*
|_*PRINT komutu belirlenen verileri SHAZAM output dosyasına yazar
|_*
|_PRINT X Y XMINY
       X              Y             XMINY
    42.00000       25.00000       17.00000
    65.00000       37.00000       28.00000
    76.00000       83.00000      -7.000000
    92.00000       36.00000       56.00000
    37.00000       73.00000      -36.00000
    47.00000       23.00000       24.00000
    27.00000       97.00000      -70.00000
    23.00000       36.00000      -13.00000
    63.00000       70.00000      -7.000000
    40.00000       51.00000      -11.00000
    70.00000       39.00000       31.00000
    82.00000       36.00000       46.00000
    90.00000       82.00000       8.000000
    68.00000       30.00000       38.00000
    82.00000       72.00000       10.00000
    28.00000       39.00000      -11.00000
    61.00000       27.00000       34.00000
    75.00000       38.00000       37.00000
```

```
    83.00000       27.00000       56.00000
    60.00000       78.00000      -18.00000
    93.00000       20.00000       73.00000
    86.00000       68.00000       18.00000
    68.00000       72.00000      -4.000000
    53.00000       60.00000      -7.000000
    87.00000       65.00000       22.00000
    63.00000       80.00000      -17.00000
    47.00000       62.00000      -15.00000
    52.00000       36.00000       16.00000
    38.00000       43.00000      -5.000000
    90.00000       57.00000       33.00000

|_*STAT komutu, listelenen verilerin ortalamalarını, standart sapmalarını,
|_*varyanslarını ve maksimumlarını hesap eder. PCOR opsiyonu ise listelenen
|_*verilerin korelasyon matrisini yazar. Opsiyonlar "/" işaretinden sonra
|_*komuta ilave edilir.
|_*
|_STAT X Y XMINY / PCOR
 NAME         N     MEAN      ST. DEV        VARIANCE      MINIMUM      MAXIMUM
 X           30   62.933      21.156          447.58       23.000       93.000
 Y           30   52.067      21.749          473.03       20.000       97.000
 XMINY       30   10.867      30.345          920.81      -70.000       73.000

  CORRELATION MATRIX OF VARIABLES -       30 OBSERVATIONS
 X           1.0000
 Y         -0.21483E-03   1.0000
 XMINY      0.69734     -0.71689       1.0000
               X            Y            XMINY
|_*DELETE komutu ve ALL opsiyonu oluşturulan bütün değişkenleri siler.
|_DELETE / ALL
 ALL VARIABLES HAVE BEEN DELETED
|_*Verilerin okunmasının kullanışlı ve genellikle daha pratik yolu, ayrı
|_*bir veri dosyası kullanmaktır. Veriler, değişkenlerin her biri bir
|_*sütuna yerleştirilerek düzenlenebilir. Ayrıca veriler, sıra halinde de
|_*veri dosyasına yazılabilir. Fakat bu durumda READ komutuna BYVAR
|_*opsiyonu ilave edilmelidir.
|_SAMPLE 1 30
|_READ (MTABLE2) X Y
 UNIT 88 IS NOW ASSIGNED TO: MTABLE2
    2 VARIABLES AND       30 OBSERVATIONS STARTING AT OBS       1
|_GENR XMINY=X-Y
|_STAT X Y XMINY / PCOR
 NAME         N     MEAN      ST. DEV        VARIANCE      MINIMUM      MAXIMUM
 X           30   62.933      21.156          447.58       23.000       93.000
 Y           30   52.067      21.749          473.03       20.000       97.000
 XMINY       30   10.867      30.345          920.81      -70.000       73.000

  CORRELATION MATRIX OF VARIABLES -       30 OBSERVATIONS
 X           1.0000
 Y         -0.21483E-03   1.0000
 XMINY      0.69734     -0.71689       1.0000
               X            Y            XMINY
|_* Bütün SHAZAM dosyaları STOP komutu ile sonlandırılır.
|_STOP
```

III. BASİT REGRESYON

3.1. Giriş

Ekonometri çalışmalarında en fazla kullanılan araçlardan birisi regresyon analizidir. Bu nedenle regresyon analizinin ana hatları çizilerek konulara başlanılacaktır. Gelecek konularda eldeki verilerin analizinde ihtiyaç duyulan değişikliklerle bu temel yöntemin farklı biçimleri ve geliştirilmiş modelleri ele alınacaktır.

Konuya regresyon analizinin ne olduğu sorusu ile işe başlayalım. Regresyon analizi, verilen açıklanan (bağımlı) bir değişkenle bir veya daha çok açıklayıcı (bağımsız) değişken arasındaki ilişkiyi açıklama ve değerlendirme ile ilgilenir. Açıklanan değişken y ile, açıklayıcı değişkenler ise x_1, x_2, ... , x_k ile ifade edilecektir. Burada k = 1 ise, x değişkenlerinden sadece bir tane var demektir. Böyle tek bir açıklayıcı değişkene sahip ilişkiye basit regresyon denilmektedir. Ders notlarının bu kısmında basit regresyon tartışılacaktır. Eğer k > 1 ise, yani birden çok x değişkenli bir ilişki söz konusu ise buna çoklu regresyon denilmektedir. Çoklu regresyon ise sonraki bölümlerde ele alınacaktır.

Örnek 1 y: satışlar
x: reklam masrafları

Burada reklam masrafları ile satışlar arasındaki ilişki bulunmaya çalışılmaktadır.

Örnek 2 y: bir ailenin tüketim masrafları
x_1: ailenin aylık geliri
x_2: ailenin serveti
x_3: ailedeki fert sayısı

Burada, bir taraftaki tüketim masrafları ile diğer taraftaki ailenin aylık geliri, ailenin serveti ve ailedeki fert sayısı arasındaki ilişki belirlenmeye çalışılmaktadır. Bu ilişkileri belirlemenin birçok amaçları vardır. Bu amaçlar aşağıdaki gibi sıralanabilir.

- Bir x değişkeni ile ifade edilebilecek politikaların etkilerinin analizi
- Verilen bir dizi x değişkeni için y değerinin kestirilmesi
- x'in y üzerine önemli bir etkisinin olup olmadığının tahmini

Örnek 2'de ailedeki fert sayısının aile tüketim harcamalarına etkisinin olup olmadığı belirlenebilir. Burada ifade edilen "önemli" kelimesinin istatistiksel olarak gerçek manası ilerideki konularda etraflıca ele alınacaktır.

Burada da y ve x değişkenlerinin benzer yapıda olmadığı belirtilmektedir. Daha açık ifade edilirse, x değişkenlerinin y'yi etkilediği veya x değişkenlerini değiştirilebilen veya kontrol edilebilen değişkenler olduğu, y'nin ise etkilenen değişken olduğu

varsayılmıştır. Literatürde y ve x değişkenleri için değişik isimler kullanılmaktadır. Bunlardan bazıları şöyledir.

Y	**$x_1, x_2, ... , x_3$**
(a) Kestirilen değişken	Kestiren değişken
(b) Açıklanan değişken	Açıklayan değişken
(c) Bağımlı değişken	Bağımsız değişken
(d) Etkilenen değişken	Sebep değişkeni
(e) Dâhili değişken	Harici değişken
(f) Hedef değişkeni	Kontrol değişkeni

Buradaki her bir kavram, regresyon analizinin belli bir amaçla kullanımı ile ilgilidir. (a) terimi daha çok kestirme amacına yönelik olarak kullanılır. Örneğin, satışlar kestirilen, reklam harcamaları ise kestiren değişkenlerdir. Yukarıdaki şıklardan b ve c'deki terminoloji değişik insanlar tarafından regresyon analizi ile ilgili tartışmalarda kullanırlar. Bu iki terim aynı manayı ifade ederler. (d) daha çok sebep sonuç çalışmalarında kullanılır. (e) ekonometriye ait bir terminolojidir. Bu terim çok denklemli modellerde kullanılmaktadır. (f) terimi ise daha çok kontrol problemlerinde kullanılır. Örneğin amaç, bir alışveriş merkezinin satışlarının belli bir seviyeye çıkarılması olabilir. Bunun için de reklam harcamalarının bu amaca ulaşmak için ne kadar olması gerektiğinin belirlenmesi istenebilir.

Bu bölümde ve gelecek bölümde yani basit ve çoklu regresyon konuları işlenirken, (b) ve (c)'deki terimler kullanılacaktır. Bu bölümde sadece bir açıklanan (bağımlı) değişken ve bir açıklayıcı (bağımsız) değişken durumu dikkate alınacaktır.

3.2. İlişkilerin Belirlenmesi

Önceden de belirtildiği gibi, bir açıklanan bağımlı değişken ve bir açıklayıcı bağımsız değişken durumu basit regresyon olarak burada tartışılacaktır. y ile x arasındaki ilişki denklem 3.1'deki gibi ifade edilebilir.

$$y = f(x) \qquad (3.1)$$

Burada f (x) x'in bir fonksiyonudur. Bu fonksiyonda iki farklı ilişkinin izah edilmesi gerekmektedir.

- Belirleyici ve matematiksel bir ilişki
- Verilen x değişkeni için y'nin tek bir değerini vermeyen, fakat ihtimal terimleri ile tam olarak açıklanabilen istatistiksel bir ilişki

Regresyon analizlerinde kullanacak olan ilişki 2. tip bir ilişkidir. Örnek olarak satışlarla reklam harcamaları arasındaki ilişki gösterilebilir. Bu durumda,

$$y = 2500 + 100\,x - x^2$$

belirleyici bir ilişkidir. Farklı reklam harcamaları için satış miktarları birebir belirlenebilir. Bu ilişkiyi kullanarak verilen x'ler için y değerleri aşağıdaki gibidir.

x	y
0	2500
20	4100
50	5000
100	2500

Diğer taraftan reklam harcamaları ile satışlar arasındaki ilişkinin aşağıdaki gibi olduğu varsayılabilir.

$$y = 2500 + 100\,x - x^2 + u$$

Burada u = ½ ihtimalle +500 ve ½ ihtimalle -500 değerlerine sahip olduğunu varsayalım.

Bu durumda x'in değişik değerleri için y'nin değeri tam olarak değil de ihtimalli olarak belirlenebilir. Örneğin, eğer reklam harcamaları 50 ise, satışlar ½ ihtimalle 5500, ½ ihtimalle 4500 olur. Bu durumda her bir x değeri için elde edilen y değerleri aşağıdaki gibi olur.

x	y
0	2000 veya 3000
20	3600 veya 4600
50	4500 veya 5500
100	2000 veya 3000

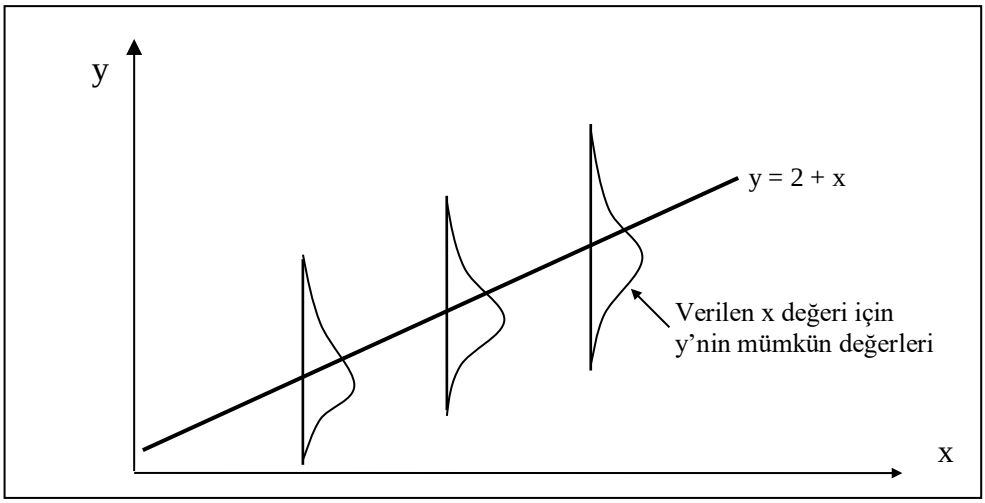
Verilen x değerleri için y değerleri yukarıdaki sekiz muhtemel durumdan herhangi biri olabilir. Örneğin y için aşağıdaki değerlere sahip olunabilir.

x	y
0	2000
20	4600
50	5500
100	2000

Örneğin, devamlı bir dağılıma sahip hata terimi ortalaması 0 ve varyansı 1 olan normal bir dağılım olsun. Bu durumda her bir x değeri için y'nin normal dağılımı söz konusudur. Gözlenen y değeri bu dağılımdan herhangi bir gözlem olabilir. Örneğin, eğer x ve y'nin ilişkisi

$$y = 2 + x + u$$

şeklinde ise burada hata terimi u, ortalaması 0 ve varyansı 1 olan normal bir dağılım gösterir. Bu dağılım N(0,1) şeklinde gösterilir. Dolayısıyla bu ilişki, x ve y'nin her bir değeri için normal bir dağılıma sahip olacaktır. Bu durum şekil 3.1'de gösterilmiştir. Şekilde çekilen çizgi y = 2 + x formunda belirleyici bir ilişkidir. Her bir x için y'nin gerçek değerleri dikey olarak çizilen çizgilerin herhangi bir noktasında olabilir. Bu gibi durumlarda x ve y arasındaki ilişki stokastik / istatistiksel ilişki olarak adlandırılır.



Şekil 3.1. Bir stokastik/tesadüfi ilişki

İlk baştaki 3.1 fonksiyonuna gidilirse, f (x) fonksiyonunun x'de doğrusal olduğu kabul edildiğinde aşağıdaki şekilde yazılabilir.

$$f(x) = \alpha + \beta x$$

Bu fonksiyona ilave olarak ilişkinin stokastik olduğunu kabul edilsin. Bu durumda denklem 3.2'deki fonksiyon elde edilir.

$$y = \alpha + \beta x + u \qquad (3.2)$$

Burada hata terimi olarak bulunan u bilinen bir ihtimal dağılımına sahiptir. Yukarıdaki denklemde $\alpha + \beta x$ y'nin belirleyici kısmı, u ise stokastik veya şansa bağlı kısmıdır. Eldeki x ve y değişkenleri kullanılarak tahmin edilen α ve β, regresyon katsayıları veya regresyon parametreleri olarak bilinir.

Deterministik ve stokastik kısımların toplanır olması gerektiğine dair hiçbir sebep yoktur. Fakat burada modele basitten başlanıp, daha sonra karmaşık konulara girilecektir. Bu yüzden ilk başta modeli doğrusal ve hata terimini de toplanır olarak kabul edilmektedir. Bazı basit alternatif fonksiyonel formlar ileride tartışılacaktır.

Hata terimi neden modele ilave edilmesi gerekmektedir. Yukarıda verilen denklemdeki u hata teriminin kaynakları nelerdir? Bu kaynakları üç ana başlıkta toplanabilir.

1. *İnsan tepkilerindeki beklenmeyen şansa bağlı unsurlar*: Örneğin; eğer y bir hane halkının tüketim harcamaları, x hanenin kullanılabilir geliri ise, her bir hanenin tüketiminde beklenmeyen şansa bağlı unsurlar vardır. Hane halkı bir makine gibi hareket etmez. Hane halkı bir ay tüketiminde çok müsrif olurken diğer bir ayda çok cimri davranabilir.
2. *Çok sayıda dikkate alınmayan değişkenlerin etkisi*: Yine örneğimizdeki gibi y'yi etkileyen tek değişken x değildir. Ailedeki fert sayısı, ailenin zevkleri, harcama

alışkanlıkları ve benzeri faktörler y'yi etkilerler. Hata terimi bütün bu değişkenlerin etkisini içine alır. Bu değişkenlerin bazıları rakamsal olarak ifade edilemez ve diğer bazıları belirlenememiş olabilir.

3. *y'deki ölçüm hatası*: Bu unsur yukarıda verilen örnekteki aile tüketimi ile ilgili ölçüm hatasına tekabül etmektedir. Yani elde edilen bir değer tam doğru olarak ölçülmeyebilir. Modeldeki hata terimi u için yapılan bu tartışmayı haklı çıkarmak, özellikle x'de, örneğin tüketilebilir gelirde, hiç bir ölçüm hatası olmadığının söylenmesi oldukça zordur. Her iki değişkende de (y, x) ölçüm hatası olması durumu ilerideki konularda tekrar ele alınacaktır.

Bütün bu karmaşık durumları hep birden ele almak yerine yeri geldiğinde sırayla bu konuları irdelemek daha doğru olacaktır. Burada şimdilik eşitlik 3.3'deki y'nin ölçümünde hata olduğu ve fakat x'de olmadığı kabul edilecektir.

Eğer x ve y ile ilgili n tane gözlem var ise, bu durumda yukarıdaki denklem aşağıdaki gibi yazılabilir.

$$y_i = \alpha + \beta x_i + u_i \qquad i = 1, 2, ..., n \qquad (3.3)$$

Burada amaç α ve β gibi bilinmeyen parametreleri, verilen n sayıdaki x ve y gözlemlerini kullanarak tahmin etmektir. Bunu yapmak için hata terimi u_i hakkında bazı varsayımların ileri sürülmesi gerekmektedir. Bu varsayımlar aşağıdaki gibi sıralanabilir.

1. *Sıfır ortalama*, $E(u_i) = 0$, bütün i'ler için
2. *Ortak varyans*, $var(u_i) = \sigma^2$, bütün i'ler için
3. *Bağımsızlık*, u_i ve u_j birbirinden tüm $i \neq j$'ler için bağımsızdırlar.
4. *x_j'nin bağımsızlığı*, u_i ve x_j bütün i ve j'ler için bağımsızdırlar. Eğer x_j şansa bağlı bir değişken olarak kabul ediliyorsa bu varsayım otomatik olarak gelmektedir. Yani u'nun dağılımı x'in değerine bağlı değildir.
5. *Normallik*, u_i'ler bütün i'ler için normal dağılmıştır. Varsayım 2, 3 ve 4'e bağlı olarak bu kabul, u_i'nin sıfır ortalama ve σ^2 varyans ile bağımsız ve normal olarak dağılmış olduğunu ifade eder. Bu $u_i \sim IN(0, \sigma^2)$ olarak yazılmaktadır.

Bunlar başlangıçta kullanılan varsayımlardır. İlerde bu varsayımlardan bazılarını gevşetilecektir. Normallik varsayımının, normal, *t* ve F dağılımları ile çıkarımlar yapılacağı için gevşetilmesi mümkün değildir. Fakat 2, 3 ve 4. varsayımlar ileride ki konularda gevşetilecektir. Birinci varsayım ise sürekli olarak kullanılacaktır.

$E(u_i) = 0$ olduğundan yukarıdaki denklem 3.4'deki eşitlik gibi yazılabilir.

$$E(y_i) = \alpha + \beta x_i \qquad (3.4)$$

Bu çoğu zaman ana kitlenin regresyon fonksiyonu olarak adlandırılır. Buna α ve β parametrelerinin tahminleri ikame edildiğinde örnek regresyon fonksiyonu elde edilmiş

olur. Basit regresyon modelindeki α ve β parametrelerinin tahmini için üç çeşit yöntem vardır.

1. Momentler metodu
2. En küçük kareler metodu
3. Maksimum Benzerlik metodu

Bunlardan şimdilik sadece birinci ve ikincisi, yani Momentler ve En Küçük Kareler Metodu burada tartışılacaktır. Basit regresyon durumunda bu üç metot da aynı sonucu vermektedir. Fakat tahmin genelleştirildiğinde metotlar değişik sonuçlar vermektedir.

3.3. Momentler Metodu

Daha önce bahsedilen hata terimi hakkındaki varsayımlara göre E (u) = 0 ve cov (x, u) = 0'dır. Momentler metodunda bu varsayımların yerine örnek karşılıkları yerleştirilerek gerekli hesaplamalar yapılır. Burada α' ve β', sırası ile α ve β için tahminciler olduğu kabul edilsin. Hata terimi u_i'nin örnek karşılığı ise tahmin edilen hata terimi $\hat{u}_i$'dir. Aynı zamanda rezidual olarak da adlandırılan hata terimi aşağıdaki gibi ifade edilir.

$$\hat{u}_i = y_i - \alpha' - \beta' x_i$$

α' ve β' yı belirlemek için yukarıdaki iki denklem, örnek karşılıkları ile aşağıdaki gibi yazılabilir.

Ana kitle varsayımı	**Örnek karşılığı**
$E(u) = 0$	$1/n \sum \hat{u}_i = 0$ veya $\sum \hat{u}_i = 0$
$Cov (x, u) = 0$	$1/n \sum x_i \hat{u}_i = 0$ veya $\sum x_i \hat{u}_i = 0$

Bu ve takip eden denklemlerde $\sum$, $\sum_{i=1}^{n}$'i temsil etmektedir. Bu durumda iki denkleme sahip olunmaktadır.

$$\sum \hat{u}_i = 0 \text{ veya } \sum (y_i - \alpha' - \beta' x_i) = 0$$

$$\sum x_i \hat{u}_i = 0 \text{ veya } \sum x_i (y_i - \alpha' - \beta' x_i) = 0$$

$\sum \alpha' = n\alpha'$ şeklinde alındığında bu denklemler aşağıdaki gibi yazılabilir.

$$\sum y_i = n\alpha' + \beta' \sum x_i$$

$$\sum x_i y_i = \alpha' \sum x_i + \beta' \sum x_i^2$$

Bu denklemler çözüldüğünde α' ve β' değerleri elde edilir. Bu denklemler "normal denklemler" olarak adlandırılır.

Uygulamalı Örnek 3.1

5 aylık dönem için spor giyimleri satan firmaya ait reklam harcamaları (x) ve satış geliri (y) dikkate alındığında elde edilen gözlemler çizelge 3.1'deki gibi olsun.

Çizelge 3.1. Firmanın aylara göre reklama harcamaları ve satış gelirleri

Aylar	satış gelirleri (y) (milyon TL)	reklam harcamaları (x) (bin TL)
1	3	1
2	4	2
3	2	3
4	6	4
5	8	5

α' ve β' değerlerini elde etmek için $\sum x$, $\sum x^2$, $\sum xy$, $\sum y$'leri aşağıda Çizelge 3.2'deki gibi hesaplanabilir.

Çizelge 3.2. Firma ile ilgili verilerin işlenişi

gözlem	x	Y	x^2	xy	$\hat{U}$
1	1	3	1	3	0,8
2	2	4	4	8	0,6
3	3	2	9	6	-2,6
4	4	6	16	24	0,2
5	5	8	25	40	1,0
Toplam	15	23	55	81	0

n = 5 olduğundan dolayı aşağıdaki normal denklemler elde edilir.

$$5\alpha' + 15\beta' = 23$$

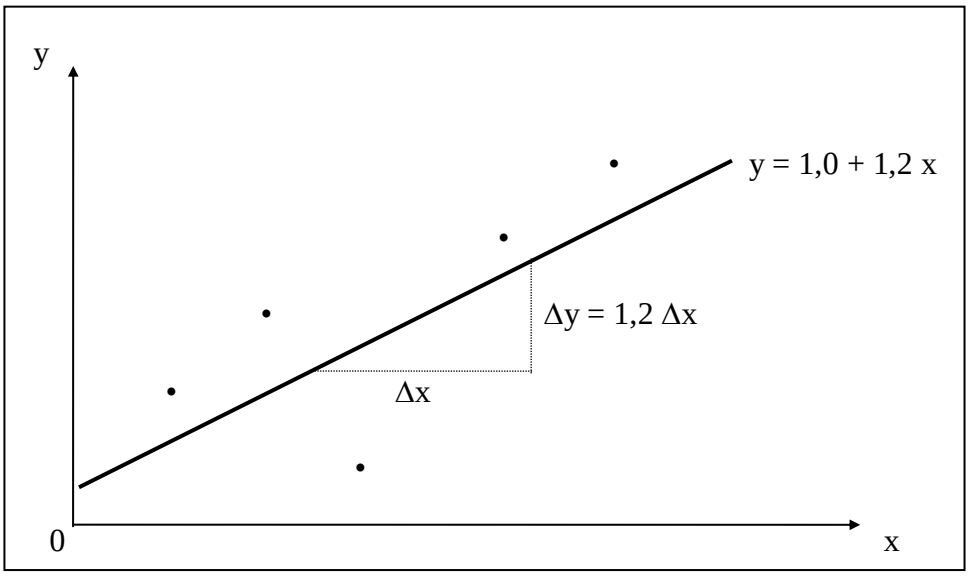
$$15\alpha' + 55\beta' = 81$$

Bu denklemler bize $\alpha' = 1{,}0$, $\beta' = 1{,}2$ değerlerini verir. Bu yüzden örnek regresyon denklemi aşağıdaki gibi olur.

$$y = 1{,}0 + 1{,}2x$$

Örnek gözlemleri ve tahmin edilen regresyon çizgisi Şekil 3.2'de gösterilmiştir. Kesişme noktası, x = 0 olduğunda y'nin değerini verir. Bu, reklam harcamaları sıfır olduğunda satış gelirlerinin 1 milyon olacağını ifade eder. Eğim katsayısı olan 1,2 x'in değişmesi karşısında y'nin değişmesinin 1,2 Δx oranında olduğunu ifade eder. Örneğin reklam harcamalarındaki 1 bin TL'lik artış, firmanın satış gelirlerinde ortalama 1,2 milyon TL artış sağlar.

Tahmin edilen 1,0 ve 1,2 değerlerinde belli ölçüde belirsizlik söz konusudur. Bu durum, ihtimal dağılımları dikkate alınarak ileride tartışılacaktır. Diğer önemli bir hususta y'nin değerini kestirirken örnek aralığından fazla uzaklaşılmamasıdır. Aksi takdirde firma sahibi, örneğin örnek aralığının çok dışında 100 bin TL'lik reklam harcaması ile 120 milyon satış geliri hesap edebilir ki bu mümkün değildir.



Şekil 3.2. Örnek değerleri ve tahmin edilen regresyon çizgisi

Daha önce bahsedilen tahmin hataları aşağıdaki gibi ifade edilebilir.

$$\hat{u}_i = y_i - (1{,}0 - 1{,}2\ x_i)$$

Bunlar y'nin verilen x değerleri için tahmin edilen regresyon çizgisi ile kestirildiğinde yapılacak hatalardır. Burada dikkat edilecek husus, y'nin herhangi bir gelecek değerinin tahmin edilmiyor olmasıdır. Sadece örneğe ait kestirme hataları dikkate alınmaktadır. Burada $x_i = 0$ için veya $x_i = 6$ için u_i değeri elde edilemez, çünkü bunlara karşılık y_i değerleri yoktur. $\sum\hat{u}_i = 0$ değeri başlangıçta varsayım olarak ifade edilmiştir. Burada örneğe ait hata kareler toplamı aşağıdaki gibi hesap edilebilir.

$$\sum\hat{u}_i^2 = (0{,}8)^2 + (0{,}6)^2 + (-2{,}6)^2 + (0{,}2)^2 + (1{,}0)^2 = 8{,}8$$

Gelecek konuda açıklanacak olan en küçük kareler yöntemi $\sum\hat{u}_i^2$'nin minimum olduğu α' ve β' değerlerini veren bir yöntemdir. Yani tahmin edilen hata kareler toplamı minimumdur. Bu metotla da aynı α ve β değerleri tahmin edilir. Çünkü elde edilen normal denklemler aynıdır.

3.4. En Küçük Kareler Yöntemi

En Küçük Kareler yöntemi, modern istatistiksel analizin otomobilidir. Onun sınırlı kullanımlarına, bazen meydana gelen kazalarına ve başka etkilerden dolayı kirlenmesine

rağmen bu yöntemin çok sayıdaki formları, açılımları ve unsurları birçok istatistiksel analizin yapılmasını sağlar ve herkes tarafından bilinir ve değer verilir.

En Küçük Kareler yöntemi α' ve β''yı sırasıyla α ve β'nın tahmincisi olarak seçilmesini gerekli kılmaktadır. Böylece 3.5 denklemi

$$Q = \sum (y_i - \alpha' - \beta' x_i)^2 \qquad (3.5)$$

bir minimum olur. Q aynı zamanda verilen x_i değeri ile kestirdiğimiz y_i'nin örnek içindeki kestirme ve tahmin edilen regresyon denkleminin hata kareler toplamıdır.

En küçük kareler yönteminin arkasındaki mana, x_i ve y_i değerlerine ait noktaların grafiğini veren şekil 3.2'deki gibi grafiksel olarak açıklanabilir. Bir regresyon çizgisi, bu noktalara mümkün olacak en yakın bir şekilde geçirilir. Burada, şu soru ile karşı karşıya kalınmaktadır. Yakın demekle kastedilen nedir? Denklem 3.5'de Q'nun minimum yapılması yöntemi, noktaların çizgiden dikey uzaklıklarının karelerinin toplamının minimum yapıldığını ima etmektedir.

Çok kolay ve hızlı bilgisayar programları ile insanlar sadece sonuçları sağlamayı istemekte ve bu tahmincilerin nasıl elde edildiğini bile bilmeye ihtiyaç duymamaktadırlar. Ancak en küçük kareler tahmincilerinin nasıl elde edildikleri hakkında biraz bilgi sahibi olunması, sonuçları daha iyi anlayabilmek ve yorumlayabilmek açısından çok büyük önem arz etmektedir.

Yukarıdaki denklem 3.5'deki Q'yu minimum yapmak için onun α' ve β''ya göre birinci dereceden türevi alıp sıfıra eşitlenir. Burada $\sum^{n}_{i=1}$, $\sum$ olarak verilecektir

$$\partial Q / \partial \alpha' = 0 \;\Rightarrow\; \sum 2 (y_i - \alpha' - \beta' x_i)(-1) = 0$$

y_i eşitliğin soluna çekilir. Burada 2'ler birbirini götürür ve aşağıdaki denklem elde edilir.

$$\sum y_i = n\alpha' + \beta' \sum x_i$$

her iki tarafı n'e bölünüp sadeleştirildiğinde,

$$y_{ort} = \alpha' + \beta' x_{ort} \qquad (3.6)$$

denklem 3.6 elde edilir.

Buna ilaveten β''ya göre Q 'nun birinci dereceden türevi alındığında,

$$\partial Q / \partial \beta' = 0 \;\Rightarrow\; 2 \sum (y_i - \alpha' - \beta' x_i)(-x_i) = 0$$

elde edilir. Bu da yukarıdaki gibi sadeleştirilirse,

$$\sum y_i x_i = \alpha' \sum x_i + \beta' \sum x_i^2 \qquad (3.7)$$

3.7 denklemi elde edilir. Bu 3.6 ve 3.7 denklemleri **normal denklemler** olarak adlandırılırlar. Denklem 3.6'daki α''nın değeri ($\alpha' = y_{ort} - \beta' x_{ort}$) denklem 3.7'de yerine konulduğunda ve sadeleştirildiğinde sırasıyla aşağıdaki denklem 3.8 elde edilir.

$$\sum y_i x_i = \sum x_i (y_{ort} - \beta' x_{ort}) + \beta' \sum x_i^2$$

$$\sum y_i x_i = n x_{ort} (y_{ort} - \beta' x_{ort}) + \beta' \sum x_i^2$$

$$\sum y_i x_i = n x_{ort} y_{ort} - \beta' n x_{ort}^2 + \beta' \sum x_i^2$$

En son denklemi β' parantezine alıp β' çekildiğinde aşağıdaki sonuçlar elde edilir.

$$\sum y_i x_i = n x_{ort} y_{ort} + \beta' (- n x_{ort}^2 + \sum x_i^2)$$

$$\beta' = (\sum y_i x_i - n x_{ort} y_{ort}) / (\sum x_i^2 - n x_{ort}^2) \qquad (3.8)$$

Diğer taraftan aşağıdaki ifadeleri notasyonla ifade ederek kolaylık sağlamak mümkündür.

$$S_{yy} = \sum (y_i - y_{ort})^2 = \sum y_i^2 - n y_{ort}^2$$

$$S_{xy} = \sum (x_i - x_{ort})(y_i - y_{ort}) = \sum y_i x_i - n x_{ort} y_{ort}$$

$$S_{xx} = \sum (x_i - x_{ort})^2 = \sum x_i^2 - n x_{ort}^2$$

Bu durumda yukarıdaki denklem, yani β' aşağıdaki gibi yazılır.

$$\beta' S_{xx} = S_{yy} \text{ veya } \beta' = S_{xy} / S_{xx} \qquad (3.9)$$

Böylece en küçük kareler yönteminin tahminleri aşağıdaki gibi ifade edilir.

$$\boxed{\beta' = S_{xy} / S_{xx} \text{ ve } \alpha' = y_{ort} - \beta' x_{ort}} \qquad (3.10)$$

Tahmin edilen hata terimleri ise aşağıdaki gibi yazılır.

$$\hat{u}_i = y_i - \alpha' - \beta' x_i$$

Her iki normal denklemde de bu hataların karşıladığı görülmektedir.

$$\sum \hat{u}_i = 0 \quad \text{ve} \quad \sum x_i \hat{u}_i = 0$$

Hata kareler toplamı (RSS) da aşağıdaki gibi elde edilir.

$$RSS = \sum (y_i - \alpha' - \beta' x_i)^2 \text{ burada } \alpha' \text{ yerine yukarıdaki denklem konulursa,}$$

$$= \sum [y_i - y_{ort} - \beta' (x_i - x_{ort})]^2$$

$$= \sum (y_i - y_{ort})^2 + \beta'^2 \sum (x_i - x_{ort})^2 - 2\beta' \sum (y_i - y_{ort})(x_i - x_{ort})$$

$$= S_{yy} + \beta'^2 S_{xx} - 2\beta' S_{xy} \quad \text{burada } \beta' = S_{xy} / S_{xx} \text{ olur ise, o zaman}$$

$$= S_{yy} - (S^2_{xy} / S_{xx})$$

$$= S_{yy} - \beta' S_{xy}$$

S_{yy} genellikle toplam kareler toplamı TSS ve $\beta' S_{xy}$ açıklanan kareler toplamı olarak ifade edilir ve böylece aşağıdaki eşitlik elde edilir.

$$\underset{\text{(toplam)}}{TSS} = \underset{\text{(açıklanan)}}{ESS} + \underset{\text{(hata)}}{RSS}$$

Bazı yazarlar RSS'yi regresyon kareler toplamı ve ESS'yi de hata kareler toplamı olarak kullanmayı isterler. Buradaki karmaşıklık, "explained" (açıklanan) ve "error" (hata) kelimelerinin "e" harfi ile başlaması ve aynı zamanda "regression" (regresyon) ve "residual" (artakalan) kelimelerinin "r" harfi ile başlamasıdır. Bu ders notlarında RSS hata kareler toplamı, ESS de açıklayıcı kareler toplamı olarak gösterilecektir. Diğer taraftan tahmin edilen örnek hatası olan $\hat{u}_i = y - \alpha' - \beta' x_i$ "rezidual" olarak adlandırılacaktır.

Açıklayıcı kareler toplamının toplam kareler toplamına oranı r^2_{xy} ile ifade edilir. Burada r_{xy} ise korelasyon katsayısı olarak ifade edilmektedir. Bu yüzden r^2_{xy} = ESS / TSS ve 1 - r^2_{xy} = RSS / TSS dir. Eğer r^2_{xy} yüksek ise yani 1'e yakınsa, x'e y'nin iyi bir açıklayıcı değişkeni denir. r^2_{xy} terimi determinasyon katsayısı olarak adlandırılmakta ve 1 ile 0 arasında bir değer alması gerekmektedir. Eğer r^2_{xy} 0'a yakınsa, x değişkeni y'nin değişmesini az açıklamaktadır denir. Ayrıca r^2_{xy} determinasyon katsayısı aşağıdaki gibi değişik şekillerde ifade edilebilir.

Özetlersek;

$$\boxed{r^2_{xy} = ESS / TSS = (TSS - RSS) / TSS = \beta' S_{xy} / S_{yy}}$$

regresyon katsayıları için,

$$\boxed{\beta' = S_{xy} / S_{xx} \quad \text{ve} \quad \alpha' = y_{ort} - \beta' x_{ort}}$$

rezidual kareler toplamı için,

$$\boxed{RSS = S_{yy} - (S^2_{xy} / S_{xx}) = S_{yy} - \beta' S_{xy} = S_{yy} (1 - r^2_{xy})}$$

ve determinasyon katsayısı için ise,

$$\boxed{r^2_{xy} = S^2_{xy} / (S_{xx} S_{yy}) = \beta' S_{xy} / S_{yy}}$$

tahmincileri böylece belirlenmiş olur.

En Küçük Kareler yönteminin tahmincileri olan β' ve α', tahmin edilebilecek başka herhangi bir doğrudan daha küçük RSS'ye sahip tahmin edilen doğruyu bize vermektedir.

Uygulamalı Örnek 3.2

Çizelge 3.3'deki verilerin 10 işçiye ait olduğunu varsayalım.

Çizelge 3.3. İşçilere ait üretim gözlemleri

Gözlem	X	Y	x^2	y^2	xy
1	10	11	100	121	110
2	7	10	49	100	70
3	10	12	100	144	120
4	5	6	25	36	30
5	8	10	64	100	80
6	8	7	64	49	56
7	6	9	36	81	54
8	7	10	49	100	70
9	9	11	81	121	99
10	10	10	100	100	100
Toplam	80	96	668	952	789

Burada, x: işçilerin çalışma saati, y: üretim miktarıdır. Be verileri kullanarak üretim ve çalışma saati arasındaki ilişkiyi tespit edilmek istiyoruz.

$$x_{ort} = 80/10 = 8$$

$$y_{ort} = 96/10 = 9,6$$

$$S_{xx} = \sum x_i^2 - nx_{ort}^2 = 668 - 10(8)^2 = 668 - 640 = 28$$

$$S_{xy} = \sum y_i x_i - nx_{ort}y_{ort} = 789 - 10(8)\,(9,6) = 789 - 768 = 21$$

$$S_{yy} = \sum y_i^2 - ny_{ort}^2 = 952 - 10(9,6)^2 = 952 - 921,6 = 30,4$$

$$r^2_{xy} = S^2_{xy} / (S_{xx}\, S_{yy}) = 21^2 / (28*30,4) = 0,52$$

$$\beta' = S_{xy} / S_{xx} = 21 / 28 = 0,75$$

$$\alpha' = y_{ort} - \beta' x_{ort} = 9,6 - 0,75\,(8) = 3,6$$

Böylece y'nin x üzerine regresyonuna ait denklem aşağıdaki gibi yazılır.

$$y = 3,6 + 0,75x \qquad r^2 = 0,52$$

Burada x işçi çalışma saatini, y üretim seviyesini temsil ettiğinden, eğim katsayısı 0.75 işgücünün marjinal verimliliğini ölçmektedir. Kesişme noktası olarak 3.6'nın manası da işçi çalışma saati 0 olduğunda üretim 3.6 olacağı manasına gelmektedir. Şurası açık ki bu mantıklı bir açıklama değildir. Ancak bu daha önceden bahsettiğimiz noktaya işaret etmektedir. Burada y'nin kestirilecek değerini örnek sınırlarının çok dışında kestirmemek gerekir. Çünkü x değişkeninin örnek aralığı sadece 5'den 10'a kadar değişmektedir.

3.5. Basit Regresyon Modelinde İstatistiksel Çıkarım

Önceki kısımda en küçük karelerin tahmincilerinin nasıl elde edileceği ile ilgili yöntemler verilmişti. En küçük kareler yönteminde α ve β katsayılarının tespitinde hata

terimi için herhangi bir spesifik ihtimal dağılımı gerekmiyor. Fakat parametrelerin tahmin aralığını tespit etmek için ve onlar hakkında herhangi bir hipotez testini kurmak için hata teriminin normal dağılıma sahip olduğunu varsaymak gerekiyor. En küçük kareler yöntemi tahmincileri;

1. sapmasızdırlar
2. bütün sapmasız tahminciler arasında en küçük varyansa sahiptirler.

Daha önce yaptığımız diğer varsayımların karşılandığı kabul edilmesi durumunda hata terimi u_i normal dağılıma sahip olmasa bile bu özellikler geçerlidir. Daha önce yaptığımız bu varsayımları hatırlayalım.

1. $E(u_i) = 0$
2. $V(u_i) = \sigma^2$ bütün i'ler için
3. u_i ve u_j bütün $i \neq j$'ler için bağımsız
4. x_j'ler u_i'lerden bağımsız

Şimdi, bunlara ilave olarak u_i'nin normal dağılıma sahip olduğu varsayımı ile α ve β'nın güven aralığının nasıl belirleneceği gösterilerek α ve β hakkındaki hipotez test edilecektir. İlk önce α ve β'nın örnek dağılımına ihtiyaç vardır. Bunların nasıl belirlendiği değil, sadece sonuçları burada verilecektir. Bu eşitlikleri ve tanımlamaları aşağıdaki gibi sıralayabiliriz.

$E(\alpha') = \alpha$	$Var(\alpha') = \sigma^2 (1/n + x^2_{ort} / S_{xx})$
$E(\beta') = \beta$	$Var(\beta') = \sigma^2 / S_{xx})$ ve
$E(Cov(\alpha', \beta')) = 0$	$Cov(\alpha' , \beta') = \sigma^2 (-x_{ort} / S_{xx})$

Eğer hata varyansı σ^2 bilinseydi bu sonuçlar kullanışlı olacaktı. Fakat uygulamada σ^2 bilinmemekte ve tahmin edilmesi gerekmektedir. Eğer RSS, hata kareler toplamı ise bu durumda, σ^2 için sapmasız bir tahmin edici olan $s^2 = RSS / (n-2)$'dir. Burada RSS / σ^2, (n-2) serbestlik derecesi ile χ^2 dağılımına sahiptir ve RSS'nin dağılımı α' ve β'nın dağılımından bağımsızdır.

Bu sonuçlar σ^2'nin güven aralığını tespit etmede ve σ^2 hakkındaki hipotez testini test etmede kullanılabilir. Fakat hâlâ α ve β hakkında yorum yapma problemi ile karşı karşıya olma durumu söz konusudur. Bu amaçla *t*-dağılımı kullanılmaktadır.

Eğer iki değişkenden $x_1 \sim N(0,1)$ ve $x_2 \sim \chi^2$ ise, k serbestlik derecesi ile x_1 ve x_2 bağımsızdır. Bu durumda, $x = x_1 / (x_2 / k)^{1/2}$ = standart normal / (ortalanmış bağımsız $\chi^2)^{1/2}$ k serbestlik derecesi ile *t*-dağılımına sahiptir.

Bu durumda $(\beta'- \beta) / (\sigma^2 / S_{xx})^{1/2} \sim N(0,1)$ olmaktadır. Yani parametre değeri ortalamadan çıkarılıp sonra standart sapmaya bölünmektedir. Aynı zamanda $RSS / \sigma^2 \sim \chi^2_{n-2}$ olur ve iki dağılım bağımsızdırlar. Bu durumda $[(\beta'-\beta) / (\sigma^2 / S_{xx})^{1/2}] / [RSS / (n-2) \sigma^2]^{1/2}$ oranı hesap edilir ve bu oran (n-2) serbestlik derecesi ile normal dağılıma sahiptir. Şimdi σ^2'ler birbirini götürür ve s^2 yerine RSS / (n -2) şeklinde yazıldığında, sonuç olarak n-2 serbestlik derecesine sahip $(\beta'-\beta) / (s^2/S_{xx})^{1/2}$ formülü elde edilir. Burada dikkat edilmeli ki β''nın varyansı σ^2/S_{xx}'dir ve σ^2 bilinmediğine göre, sapmasız bir tahminci olarak RSS/n-2 kullanılır. Bu yüzden s^2/S_{xx} β''nın tahmin edilmiş varyansıdır ve onun karekökü $SE(\beta')$ ile gösterilen standart hatadır.

Aynı yöntemi α' için de kullanabiliriz. Eğer α''nın varyansında s^2'yi σ^2 yerine ikâme edilirse ve karekökü alınırsa α''nın standart hatası $SE(\alpha')$ hesap edilir. Böylece aşağıdaki sonuçlar elde edilir.

Bu durumda $(\alpha'-\alpha) / SE(\alpha')$ ve $(\beta'-\beta) / SE(\beta')$'nın her biri n-2 serbestlik derecesi ile *t*-dağılımına sahiptir. Bu dağılımlar α ve β için güven aralıklarını belirlemede ve aynı zamanda α ve β hakkındaki hipotez testlerinde kullanılırlar.

Uygulamalı Örnek 3.3

Standart hatanın hesaplamasını göstermek için önceki Çizelge 3.3'deki veriler ve elde edilen regresyon sonuçları dikkate alınacaktır. Bu tahmin edilen model $\hat{y} = 3,6 + 0,75$ x dir. Şimdi α ve β'nın standart hatası 3 aşamada hesap edilecektir.

1. α' ve β' 'nın varyanslarının σ^2 cinsinden hesaplanması
2. σ^2 için $s^2 = RSS / (n - 2)$'nın ikâme edilmesi
3. Elde edilen ifadenin karekökünün alınması

Burada,

$$V(\alpha') = \sigma^2 (1 / n + x_{ort}^2 / S_{xx}) = \sigma^2 (1 / 10 + 64 / 28) = 2,39 \sigma^2$$

$$V(\beta') = (\sigma^2 / S_{xx}) = \sigma^2 / 28 = 0,036 \sigma^2$$

$$s^2 = 1 / (n-2) (S_{yy} - (S^2_{xy} / S_{xx})) = 1 / 8 [30,4 - (21)^2 / 28] = 1,83$$

$$SE(\alpha') = [(2,39) (1,83)]^{1/2} = 2,09$$

$$SE(\beta') = [(0,036)(1,83)]^{1/2} = 0,256$$

Standart hatalar genellikle parantez içinde regresyon katsayılarının altında ve daha küçük puntoyla ifade edilmelidir.

α, β ve σ^2 İçin Güven Aralıkları

(α' - α) / SE(α') ve (β' - β) / SE (β') n-2 serbestlik derecesi ile *t*-dağılımına sahip olduğundan n-2 = 8 serbestlik derecesinde *t*-dağılımı tablosunu kullanarak aşağıdaki sonuçlar elde edilir.

$$P [-2,306 < (\alpha' - \alpha) / SE (\alpha') < 2,306] = 0,95 \text{ ve}$$

$$P [-2,306 < (\beta' - \beta) / SE (\beta') < 2,306] = 0,95$$

Bunlar α ve β için güven aralıklarını verir. Burada α', β', SE(α') ve SE(β') değerleri yerine konulup sadeleştirildiğinde α ve β için %95 güven sınırları bulunur. Bu değerler α için (-1,22, 8,42), ve β için (0,16 ve 1,34) olarak bulunur. Burada dikkat edilmesi gereken husus, %95 güven sınırlarının α için α' ± 2,306 SE (α') ve β için β' ± 2,306 SE (β') şeklinde de yazılabilir olmasıdır.

Ekonometride çok sık kullanılmamasına rağmen χ^2 dağılımı ile test yöntemi, σ^2'nin güven aralığının hesap edilmesinde kullanılacaktır. RSS/σ^2, n-2 serbestlik derecesi ile χ^2 dağılımına sahip olduğundan gerekli güven aralığının tespit edilmesi için χ^2 dağılımının tablosu kullanılmaktadır. Burada güven aralığının %95 olduğu ve RSS = (n-2) s^2 olduğu kabul edilirse, 8 serbestlik derecesi ile χ^2 dağılımı tablosunda 2,18'den küçük bir değer elde etme ihtimali 0,025 ve 17,53 den büyük bir değer elde etmenin ihtimali 0.025'dir. Bu yüzden,

$$P [2.18 < 8\, s^2 / \sigma^2 < 17.53] = 0,95 \text{ veya}$$

$$P [8\, s^2 / 17,53 < \sigma^2 < 8\, s^2 / 2,18] = 0,95$$

s^2 = 1,83 olduğundan, bu değer yerine ikame edilirse, σ^2 için %95 güven sınırları 0,84 ve 6,72 olarak belirlenir.

Burada dikkat edilirse α ve β için güven aralıları α' ve β' etrafında simetriktir çünkü *t*-dağılımı simetrik bir dağılıma sahiptir. Bu durum σ^2 için aynı değil, yani s^2 etrafındaki değerler simetrik değildir.

Yine burada α, β ve σ^2 için elde edilen güven aralıkları oldukça geniştir. Bunu daraltmak için güven katsayısı %80 olarak alınırsa, güven sınırları daralmış olur. Örneğin β için %80 güven sınırları β' ± 1,397 SE (β')'dır. Çünkü 8 serbestlik derecesi ile *t* cetvelinden elde edilen değerler, P [-1,397 < *t* < 1,397] = 0,80 ihtimalle güven aralığının alt ve üst sınırlarını temsil etmektedirler. Böylece %80 ihtimalle β için güven sınırları 0,75 ± 1,397 (0,256) = (0,39 - 1,11) olmaktadır.

Bu güven sınırları iki taraflı olarak belirlenmiştir. Bazen sadece β için üst veya alt sınırların bulunması istenir. Örneğin, *t*-dağılımı tablosunda 8 serbestlik derecesiyle %95 ihtimalle güven aralığı P (*t* < 1,86) = 0,95 ve P (*t* > -1,86) = 0,95 şeklinde ifade edilir. Tek taraflı bir aralık tahmini olduğundan β için üst sınır, β' + 1,86 SE(β') = 0,75 + 1,86

(0,256) = 0,75 + 0,48 = 1,23 olarak hesap edilir ve % 95 güven aralığı (-∞ - 1,23) şeklinde ifade edilir. Aynı şekilde β için alt sınır, β' – 1,86 SE (β') = 0,75 – 0,48 = 0,27 olarak hesap edildiğinden %95 güven aralığı (0,27 - +∞) şeklinde ifade edilir.

Hipotez Testleri

Şimdi hipotez testlerine dönülebilir. Hipotez testinde H_0: β = 1 olsun. Bilinmektedir ki, t_0 = β' - β / SE (β'), n-2 serbestlik derecesi ile *t*-dağılıma sahiptir. Burada t_0 hesaplanan *t* değeri olsun. Eğer alternatif hipotez H_1: β ≠ 1 ise bu durumda $|t_0|$, test istatistiği olarak kabul edilir. Böylece eğer β'nın gerçek değeri 1 ise t_0 = (0,75 -1,0) / (0,256) = -0,98 ve dolayısıyla $|t_0|$ = 0,98 olur ve 8 serbestlik derecesi için *t* tablosuna bakıldığında,

P (*t* > 0,706) = 0,25 önem seviyesinde reddedilirken,

P (*t* < 1,397) = 0,10 önem seviyesinde kabul edilir.

Burada P(t > 0,98), çift taraflı testte yaklaşık olarak 0,19 veya P (|*t*| > 0,98) ise tek taraflı testte yaklaşık olarak 0,38'dir. Bu ihtimal çok düşük değildir ve bu durumda β = 1 hipotezi reddedilmez. Geleneksel olarak % 5 önem seviyesi düşük ihtimal olarak kullanılır.

Burada dikkat edersek 8 serbestlik derecesi için % 5 ihtimal noktaları iki taraflı test için ± 2,306, tek taraflı test için ± 1,86'dır. Bu yüzden hem düşük ve hem de yüksek *t* değeri, ileri sürülen hipotez için karşı bir delil olarak dikkate alındığında, eğer gözlenen t_0 > 2,306 veya t_0 < -2,306 ise hipotez reddedilir. Diğer tarafta sadece en yüksek veya en düşük *t* değeri karşı hipotez olarak dikkate alınıyorsa bu durumda eğer t_0 > 1,86 veya t_0 < -1,86 sapmanın teklif edilen yönüne bağlı olarak reddedilir.

İleri sürülen hipotezin reddi için %5 ihtimal seviyesi geleneksel olarak kullanılsa da bu rakamla ilgili kutsal herhangi bir şey söz konusu değildir. l890-1962 yılları arasında yaşayan ve modern istatistik metotlarının babası İngiliz iktisatçı R. A. Fisher analizlerde 0,05 ve 0,01 önem seviyelerini kullandığı için bu rakamlar evrensel olarak kullanılır hale gelmiştir.

Dikkat edilecek diğer bir husus test edilen hipotez yani β = 1 H_0 hipotezi olarak adlandırılır. Fakat terminolojide β = 0 sadece null hipotezi olarak kabul edilmektedir. Bu ders notları kapsamında ilk olarak ileri sürülen ve test edilen hipotez null hipotezi olarak kabul edilecek ve önem seviyesinde standart terminoloji olan 0,05 ve 0,01 seviyeleri kullanılacaktır.

Ayrıca burada dikkat edilmesi gereken bir husus da, güven aralığı ile hipotez testi arasında karşılıklı ilişkinin olmasıdır. Örneğin β için daha önce belirlenen güven aralığı 0,16 ve 1,34'dür. Bu durumda β = β_0 diyen ve β_0'ın bu aralıkta olduğu herhangi bir hipotez iki taraflı testte ve %5 önem seviyesinde reddedilmez. Eğer hipotez β = 1,27 ise

reddedilmez, fakat eğer $\beta = 1{,}35$ veya $\beta = 0{,}1$ ise reddedilir. Bir taraflı testte de sadece tek taraflı aralık ele alınır.

Ayrıca *t* değerine bağlı olarak bazı regresyon katsayılarının önemli veya önemsiz olduğunu ifade etmek geleneksel bir metottur. Katsayı üzerine yıldız işareti konularak önemli olduğu belirtilir. Örneğin önceden tahmin edilen üretim işgücü ilişkisini veren denklem aşağıdaki gibi yazılabilir.

$$y = \underset{(2,09)}{3{,}6} + \underset{(0,256)}{0{,}75^{*}}\, x \qquad r^2 = 0{,}52$$

Burada yıldız işareti, eğim katsayısının %5 seviyesinde önemli olduğunu gösterir. Fakat bu ifade sıfırdan önemli derecede farklıdır manasındadır. Bu ifade sadece hipotez $\beta = 0$ olduğunda geçerlidir. Bu tip bir hipotezde bazı durumlarda önemsiz olabilir. Önceki örnekte olduğu gibi, x çalışma saatlerini y'de üretim miktarını ifade eder ise, $\beta = 0$ hipotezi mantıklı değildir. Tabii ki iş saatlerinin üretime hiç etkisi yok diye hipotez kurmak anlamsız olur.

Uygulamalı Örnek 3.4

Amerikan Üniversitelerinde yüksek lisans yapmak isteyen öğrencilerin girmesinin gerekli olduğu GRE ve GMAT sınavlarından elde edilen sonuçların karşılaştırılması ile ilgili bir örnek verilecektir. Florida Üniversitesi İktisadi ve İdari Bilimler Fakültesinde öğrencilerin yüksek lisans çalışmasına uygunluğunun ölçülmesi için iki değişik test kullanılıyor. İktisat bölümü GRE notuna göre diğer bölümler ise GMAT değerlerine göre karar veriyorlar. Yüksek lisans ve doktora öğrencilerine asistanlık ve burs vermek için bu iki testin sonuçları dikkate alınarak karar verilecektir. Fakat bütün öğrencilerin her iki sınava ait notu yoktur. Bu yüzden bu notlar arasındaki ilişkiyi tahmin etmek hangi öğrencilerin alınıp alınmayacağı hususunda karar vermek için önem arz etmektedir. Başparmak kuralına göre bu ilişkinin aşağıdaki gibi olabileceği belirtiliyor.

$$\text{GRE notu} = 2 * (\text{GMAT notu}) + 100$$

Buradaki soru şudur. Bu ilişki veya kural ne kadar gerçekçidir? Bu soruya cevap vermek için Florida Üniversitesindeki bu iki testi de alan 262 öğrenciden veriler toplanmaktadır. GRE ve GMAT notları arasında yüksek derecede korelasyon olduğu yapılan testlerden görülmektedir. Korelasyon katsayısı Amerikalı öğrencileri için 0,71 olurken yabancı öğrenciler için ise 0,80'dir.

Eğer elde öğrencilerin bazılarının başarısını ölçen GRE ve diğer bazılarının GMAT notu varsa, bu durumda bütün notları GRE'ye çevrilebilir. Bu durumda yapılacak olan şey GRE notunun GMAT üzerine regresyonunu hesaplamaktır. Alternatif olarak ta bütün notları GMAT'a çevrilir. Bu sefer GMAT skorunun GRE üzerine regresyonunu tahmin etmek gerekir. Bu iki regresyon sonuçları çizelge 3.4 ve 3.5'de verilmiştir.

Çizelge 3.4. GRE'nin GMAT üzerine regresyonu ile ilgili modellerin tahmin sonuçları

Öğrenci	**Sayı**	**GRE'nin GMAT Üzerine Regresyonu**	**R^2**
Bütün Öğrenciler	255	GRE = 333 + 1,470 GMAT (0,087)	0.53
ABD'li Öğrenciler	211	GRE = 336 + 1,458 GMAT (0,101)	0.50
Yabancı Öğrenciler	44	GRE = 284 + 1,606 GMAT (0,183)	0.64

Hem ABD'li ve hem de yabancı öğrenciler için tahmin yapılması mümkün olduğu halde yani her iki notu da birbirine çevrilebilmesine rağmen burada sadece bir set tahmini kullanacak ve kestirmeler, GRE'nin GMAT üzerine tüm öğrenciler için olan regresyonuna göre yapıldığında aşağıdaki sonuçlar elde edilecektir.

Çizelge 3.5. GMAT'ın GRE üzerine regresyonu ile ilgili modellerin tahmin sonuçları

Öğrenci	**Sayı**	**GMAT'ın GRE Üzerine Regresyonu**	**R^2**
Bütün Öğrenciler	255	GMAT = 125 + 0,363 GRE (0,021)	0,53
ABD'li Öğrenciler	211	GMAT = 151 + 0,345 GRE (0,024)	0,50
Yabancı Öğrenciler	44	GMAT = 60 + 0,400 GRE (0,046)	0,64

Sonuç olarak, öğrencilerin kabul edilmediği düşük notlar hariç belirlediğimiz kural, GMAT'a bağlı GRE notlarını yüksek tahmin etmektedir. Bu yüzden GRE notuna sahip öğrencilerin aleyhine bir durum söz konusu olmaktadır. Bu yüzden de ekonomi bölümündeki öğrenciler GRE skoru ile alındığından ekonomi bölümü aleyhine bir durum söz konusu olacaktır.

Çizelge 3.5. Model kestirmelerinin başparmak kuralı ile karşılaştırılması

GMAT	**GRE**	**Başparmak Kuralı**
400	921	900
450	995	1000
500	1068	1100
550	1142	1200
600	1215	1300
650	1289	1400
700	1362	1500

YAZILIM UYGULAMASI

```
*******************************************************************************
  SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
  ** Copyright (C)  1999  by K.J. White - All Rights Reserved **

  FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
  AT: ATATURK UNIVERSITESI - ERZURUM

  FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
*******************************************************************************
```

Uygulamalı Örnek 3.1, sayfa: 37

```
*******************************************************************************
|_READ (MTABLE3) Y X
 UNIT 88 IS NOW ASSIGNED TO: MTABLE3
    2 VARIABLES AND        6 OBSERVATIONS STARTING AT OBS       1

|_SAMPLE 1 5
|_*Satış geliri, STAT komutunu SUMS= opsiyonu ile toplanabilir. Bu opsiyon
|_*belli bir örnek aralığındaki değişkenleri toplar. Burada komutun
|_*önündeki ? işareti, işlem sonunda yazılacak olan outputu yazdırmaz.
|_?STAT X / SUMS=SUMX
|_?STAT Y / SUMS=SUMY
|_*Çift yıldız (**) üssel gücü, tek yıldız ise (*) çarpma işlemini yapar.
|_GENR X2=X**2
|_GENR XY=X*Y
|_?STAT X2 / SUMS=SUMX2
|_?STAT XY / SUMS=SUMXY
|_* GEN1 komutu bir sabit oluşturur.
|_GEN1 AHAT=1.0
|_GEN1 BHAT=1.2
|_* Şimdi tahmin edilen hata terimi (UHAT) üretilip ve toplanacaktır.
|_GENR UHAT=Y-AHAT-BHAT*X
|_?STAT UHAT / SUMS=SUMUHAT
|_PRINT X Y X2 XY UHAT
       X              Y             X2             XY             UHAT
    1.000000       3.000000       1.000000       3.000000      0.8000000
    2.000000       4.000000       4.000000       8.000000      0.6000000
    3.000000       2.000000       9.000000       6.000000      -2.600000
    4.000000       6.000000       16.00000       24.00000      0.2000000
    5.000000       8.000000       25.00000       40.00000       1.000000

|_PRINT SUMX SUMY SUMX2 SUMXY SUMUHAT
     SUMX          SUMY         SUMX2        SUMXY        SUMUHAT
    15.00000      23.00000     55.00000     81.00000     0.6661338E-15

|_STAT Y X
 NAME         N     MEAN         ST. DEV       VARIANCE      MINIMUM      MAXIMUM
 Y            5     4.6000       2.4083        5.8000        2.0000       8.0000
 X            5     3.0000       1.5811        2.5000        1.0000       5.0000
|_* MEAN opsiyonu belirlenen ortalamayı (M), CPDEV opsiyonu ise ortalamadan
|_*sapmaların kareler toplamını (SXX) kaydeder.

|_STAT X / MEAN=M CPDEV=SXX
 NAME         N     MEAN         ST. DEV       VARIANCE      MINIMUM      MAXIMUM
 X            5     3.0000       1.5811        2.5000        1.0000       5.0000
```

```
|_*OLS comutu en küçük kareler yoluyla regresyon tahmini yapar. COEF=
|_*opsiyonu, tahmin edilen katsayıyı (C), STDERR= opsiyonu ise standar
|_*hayatı (S) kaydeder. LIST opsiyonu ise tahmin edilen hataların
|_*(rezidual) grafiyini çizer.

|_OLS Y X / LIST COEF=C STDERR=S
        5 OBSERVATIONS     DEPENDENT VARIABLE= Y
 ...NOTE..SAMPLE RANGE SET TO:      1,      5
  R-SQUARE =   0.6207     R-SQUARE ADJUSTED =   0.4943
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   2.9333
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.7127
 SUM OF SQUARED ERRORS-SSE=   8.8000
 MEAN OF DEPENDENT VARIABLE =   4.6000
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR       3 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
     X      1.2000     0.5416     2.216     0.114 0.788    0.7878      0.7826
 CONSTANT  1.0000     1.796      0.5567    0.617 0.306    0.0000      0.2174
     OBS.   OBSERVED      PREDICTED   CALCULATED
      NO.     VALUE         VALUE       RESIDUAL
       1     3.0000        2.2000      0.80000                     I *
       2     4.0000        3.4000      0.60000                     I *
       3     2.0000        4.6000      -2.6000             *       I
       4     6.0000        5.8000      0.20000                     I*
       5     8.0000        7.0000       1.0000                     I  *
|_*FC komutu, bir OLS regresyonunu takibeden bir kestirmeyi yapmada
|_*kullanılır. BEG= ve END= opsiyonları kestirilecek değerler aralığını
|_*ifade eder. Burada 6. bir x değeri ve karşılığı olarak 0 y değeri
|_*verilere eklenmiştir.

|_FC / LIST BEG=6 END=6
 DEPENDENT VARIABLE = Y                1 OBSERVATIONS
 REGRESSION COEFFICIENTS
    1.20000000000       1.00000000000
     OBS.   OBSERVED     PREDICTED   CALCULATED  STD. ERROR
      NO.     VALUE         VALUE       RESIDUAL
        6    0.0000        8.2000      -8.2000       2.482
|_*Şimdi SHAZAM kullanılarak kitapta gösterildiği şekli ile tahmin ve
|_*kestirmeler yapılacaktır. Burada C(1) eğim parametresini, C(2) kesişme
|_*parametresini temsil etmektedir. Şimdi YHAT üretilecek ve yazılacaktır.

|_GEN1 YHAT=C(2)+C(1)*6
|_PRINT YHAT
    YHAT
   8.200000
|_*$SIG2 ve $N, varyans ve örnek büyüklüğünü temsil eden geçici
|_*değişkenlerdir. Farklı komutlardan sonra kullanılabilecek geçici
|_*değişkenler vardır. İhtiyaç duyulan yerlerde bunlar kullanılır. Bütün
|_*geçici değişkenlerin başında $ işareti vardır.
|_GEN1 VYHAT=$SIG2*(1+1/$N+((6-M)**2)/SXX)
..NOTE..CURRENT VALUE OF $SIG2=   2.9333
..NOTE..CURRENT VALUE OF $N   =   5.0000
|_GEN1 SEYHAT=SQRT(VYHAT)
|_PRINT SEYHAT VYHAT
      SEYHAT           VYHAT
      2.481935         6.160000
```

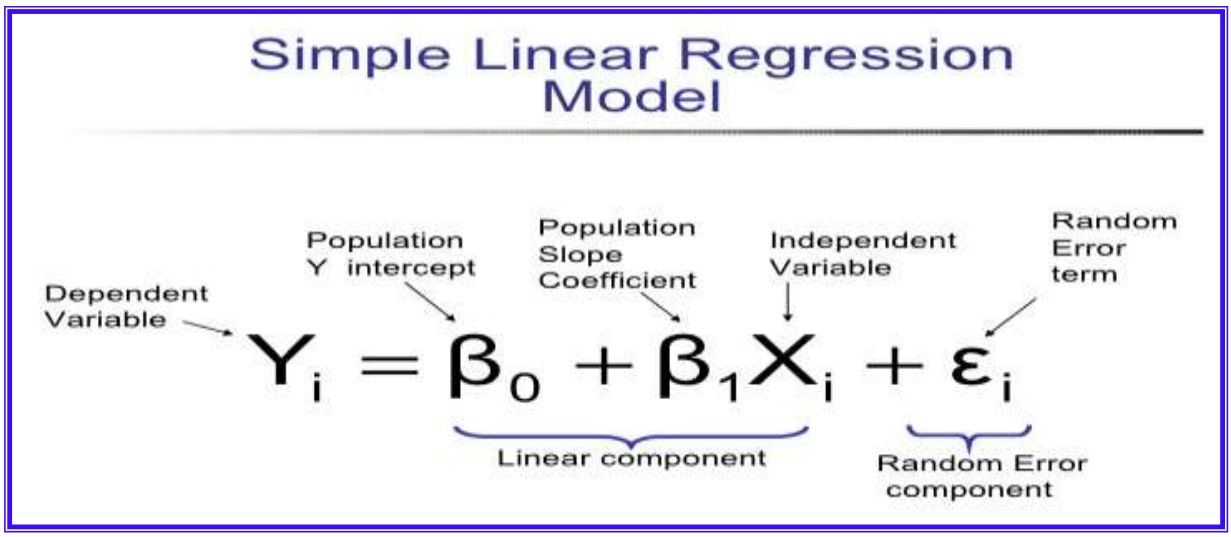
```
|_*Kestirmenin bir %90 güven aralığını oluşturalım. 3 serbestlik
|_*derecesiyle t dağılımının kritik değeri 2.353'dür. Burada LBND: alt
|_*sınır, UPBDN: üst sınırı göstermektedir.
|_GEN1 LBND=YHAT-2.353*SEYHAT
|_GEN1 UPBND=YHAT+2.353*SEYHAT
|_PRINT LBND C UPBND
    LBND          C                              UPBND
   2.360008       1.200000       1.000000       14.03999
|_* Ortalama satış tutarı için bir %90 güven aralığı oluşturalım.
|_GEN1 VARYA=$SIG2*(1/$N+((6-M)**2)/SXX)
..NOTE..CURRENT VALUE OF $SIG2=   2.9333
..NOTE..CURRENT VALUE OF $N   =   5.0000
|_GEN1 SEYA=SQRT(VARYA)
|_PRINT SEYA VARYA
    SEYA          VARYA
   1.796292       3.226667
|_GEN1 LBNDA=YHAT-2.353*SEYA
|_GEN1 UPBNDA=YHAT+2.353*SEYA
|_PRINT LBNDA C UPBNDA
    LBNDA         C                              UPBNDA
   3.973324       1.200000       1.000000       12.42668
|_DELETE / All
ALL VARIABLES HAVE BEEN DELETED
|_STOP
*******************************************************************************
```

Uygulamalı Örnek 3.3, sayfa: 44

```
*******************************************************************************
|_OLS Y X / COEF=C STDERR=S
        10 OBSERVATIONS     DEPENDENT VARIABLE= Y
 ...NOTE..SAMPLE RANGE SET TO:      1,     10
  R-SQUARE =   0.5181     R-SQUARE ADJUSTED =   0.4579
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   1.8312
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.3532
 SUM OF SQUARED ERRORS-SSE=   14.650
 MEAN OF DEPENDENT VARIABLE =   9.6000
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR     8 DF    P-VALUE CORR. COEFFICIENT  AT MEANS
   X       0.75000   0.2557    2.933    0.019 0.720    0.7198     0.6250
 CONSTANT   3.6000   2.090     1.722    0.123 0.520    0.0000     0.3750
|_*CONFID komutunun kullanılarak güven aralığı oluşturalım. NOFPLOT
|_*opsiyonu, iki katsayı belirlendiği zaman ortak güven bölgesini
|_*silmektedir. Bu nedenle t dağılımına göre sadece tek boyulu aralıklar
|_*yazılacaktır.

|_CONFID X CONSTANT / NOFPLOT
USING 95% AND 90% CONFIDENCE INTERVALS
 CONFIDENCE INTERVALS BASED ON T-DISTRIBUTION WITH   8 D.F.
      - T CRITICAL VALUES =   2.306 AND   1.860
 NAME   LOWER 2.5%   LOWER 5%   COEFFICIENT   UPPER 5%   UPPER 2.5%   STD. ERROR
 X       0.1603    0.2743     0.75000     1.226      1.340       0.256
CONSTANT -1.220   -0.2877     3.6000      7.488      8.420       2.090
|_*SHAZAM otomatik olarak %95 güven aralığını hesap eder. TCRIT= opsiyonu,
|_*güven aralığını farklı önem seviyelerinde kullanılmak üzere kullanılır.
|_*Uygun bir kritik değer belirleyelim. Örneğin 1.397 gibi bir kritik
|_*değer ile %80 güven aralığını belirleyelim.
```

```
|_CONFID X CONSTANT / NOFPLOT TCRIT=1.397
 CONFIDENCE INTERVALS BASED ON T-DISTRIBUTION WITH   8 D.F.
      - T CRITICAL VALUE =   1.397
 NAME         LOWER     COEFFICIENT     UPPER       STD. ERROR
  X         0.39273       0.75000       1.1073      0.25574
  CONSTANT  0.68002        3.6000       6.5200       2.0902
|_*Şimdi 1.86 gibi bir kritik değer ile tek taraflı %95 güven aralığını
|_*belirleyelim.
|_GENR ONELOW=C-1.86*S
 WARNING.. FINISHED AT OBS    2 THE LENGTH OF S
|_PRINT ONELOW C
    ONELOW
  0.2743278     -0.2877290
    C
  0.7500000       3.600000
|_GENR ONEUP=C+1.86*S
WARNING.. FINISHED AT OBS    2 THE LENGTH OF S
|_PRINT C ONEUP
     C
   0.7500000       3.600000
     ONEUP
    1.225672       7.487729
|_*Şimdi 2.1797 ve 17.5346 kritik değerleri ile χ² dağılımını kullanarak σ²
|_*için %95 güven aralığını belirleyelim. Burada $DF serbestlik derecesini
|_*ifade eden geçici bir değişkendir.
|_GEN1 LOWS=$DF*$SIG2/17.5346
 ..NOTE..CURRENT VALUE OF $DF  =   8.0000
 ..NOTE..CURRENT VALUE OF $SIG2=   1.8312
|_GEN1 UPS=$DF*$SIG2/2.1797
 ..NOTE..CURRENT VALUE OF $DF  =   8.0000
 ..NOTE..CURRENT VALUE OF $SIG2=   1.8312
|_PRINT LOWS $SIG2 UPS
     LOWS         $SIG2       UPS
   0.8354910      1.831250    6.721108
|_DELETE / ALL
ALL VARIABLES HAVE BEEN DELETED
|_STOP
```



IV. BASİT REGRESYON ANALİZLERİ

4.1. Kesişme Katsayısı Olmaksızın Regresyon

Bazen regresyon denklemi kesişme katsayısı kullanılmaksızın tahmin edilir. Bu da orijinden geçen regresyon denklemi olarak adlandırılır. Buna ihtiyaç duyulmasının sebeplerinin ilki ekonomik teori böyle olmasını isteyebilir. İkincisi de değişmelerdeki bir takım dönüşümlerden dolayı kesişme katsayısı olmayan bir sonuca varılmasıdır.

Bu durumda normal denklemler ve diğer formüller aynı olacak ve fakat ortalama düzeltmeleri olmayacaktır. Yani, $S_{xx} = \sum x_i^2$, $S_{xy} = \sum x_i\, y_i$, $S_{yy} = \sum y_i^2$ şeklinde hesap edilecektir. Önceki bölümdeki uygulamalı örnek 3.2'deki veriler kullanılarak aşağıdaki sonuçları elde ederiz.

$S_{xx} = 668, \quad S_{xy} = 789, \quad S_{yy} = 952$

$\beta' = S_{xy} / S_{xx} = 789 / 668 = 1{,}181$

$s^2 = 1/(n-1)\,(S_{yy} - S^2_{xy} / S_{xx}) = 1/9\,[\,952 - ((789)^2 / 668)\,] = 2{,}23$

$V(\beta') = s^2 / S_{xx} = 2{,}23 / 668 = 0{,}00334$

$SE\,(\beta') = 0{,}058$

$r^2 = S^2_{xy} / (S_{xx} S_{yy}) = 0{,}979$

Böylece regresyon denklemi aşağıdaki gibidir.

$$y = \underset{(0{,}058)}{1{,}181\, x} \qquad r^2 = 0{,}98$$

Kesişme katsayısı olan regresyonla karşılaştırdığımızda sonuçlar daha iyi görülmektedir. β' için t değeri 3 den 20 ye yükselirken, r^2 değeri 0,52 'den 0,98 'e yükselmiştir. Yüksek r^2 değeri, her zaman model iyi tahmin ediliyor anlamına gelmez. Hangi modelin daha iyi kestirme yaptığına bakmak gerekir. Üretim işgücünün bir fonksiyonu olduğundan kesişme katsayısı olmayan regresyon modeli daha anlamlıdır.

Bazı bilgisayar programları, kesişme katsayısı olmadan da regresyon hesap etmekte fakat doğru r^2 vermemektedirler. Dikkat edilecek olursa, $1 - r^2 = RSS / S_{yy}$'dir. Eğer RSS sabit katsayı olmadan hesap edilir ve S_{yy} ortalama düzeltilmesi yapılırsa negatif r^2 bile elde edilebilir. Negatif r^2 elde etmesek bile çok küçük r^2 elde edilmesi söz konusu olabilir. Yukarıdaki örnekte sabit katsayı olmadan tahmin edilen regresyonda RSS = 20,08 ve ortalaması düzeltilmiş S_{yy} = 30,40 bulunur. Bu durumda r^2 = 0,34 olarak bulunur ki bu değer olması gereken r^2 değerinin çok altında hesaplanmış olur.

4.2. Ters Regresyon

Şimdiye kadar hep y'nin x üzerine regresyonundan bahsedildi. Buna doğru regresyon denilebilir. Bazen x'in y üzerine regresyonu da dikkate alınabilir. Bu da ters regresyon olarak adlandırılır. Ters regresyon, uygun bir örnek olması açısından ücretlerdeki cinsiyet ayrımı analizlerinde kullanıla gelmiştir. Örneğin; y: ücret, x: yetenekler olarak

kabul edilir ve ücretlerin belirlenmesinde cinsiyet ayrımının var olup olmadığının belirlenmesi isteniyorsa aşağıdaki hususlar dikkate alınabilir.

1. Aynı yeteneklere (x değeri) sahip kadın ve erkeğin aynı miktarda ücret (y değeri) alıp almadığının araştırılmasıdır. Bu soru direk regresyonla yani y'nin x üzerine regresyonu tahmin edilerek cevaplandırılabilir.
2. Aynı ücrete (y değeri) sahip kadın ve erkeğin aynı yeteneklere (x değeri) sahip olup olmadığının araştırılmasıdır. Bu soruda ters regresyonla yani x 'in y üzerine regresyonu tahmin edilerek cevaplandırılabilir.

Burada verdiğimiz örnekte, her iki soruda anlamlıdır ve bu yüzden her iki regresyon sonucuna da bakmamız gerekir. Ters regresyon için regresyon denklemi aşağıda verildiği gibi yazılabilir.

$$x_i = \alpha'' + \beta'' y_i + v_i$$

Burada v_i, daha önceden u_i için bahsedilen benzer varsayımları içinde bulunduran hatalardır. Regresyon modelinde x ve y'nin yer değiştirmesi, parametrelerin aşağıdaki biçimde tahmin edilmesine neden olmaktadır.

$$\beta'' = S_{xy} / S_{yy} , \qquad \alpha'' = x_{ort} - \beta'' y_{ort}$$

Eğer örnek hata kareler toplamını RSS" ile ifade edilirse,

$$RSS'' = S_{xx} - S^2_{xy} / S_{yy}$$

denklemi elde edilir. Bu durumda aşağıdaki gibi bir sonuçla karşılaşılır.

$$\beta' * \beta'' = S^2_{xy} / S_{xx} S_{yy} = r^2_{xy}$$

Böylece, r^2_{xy} değeri 1'e ne kadar yakın olursa her iki regresyon doğrusu birbirine o kadar daha yakın olur.

Uygulamalı Örnek 4.1

Ters regresyonda ise yine bir önceki bölümdeki uygulamalı örnek 3.2'deki veriler kullanılır ise aşağıdaki tahminler elde edilir.

$$\beta'' = S_{xy} / S_{yy} = 21 / 30,4 = 0,69 \text{ ve}$$

$$\alpha'' = x_{ort} - \beta'' y_{ort} = 8,0 - 0,69*(9,6) = 1,37$$

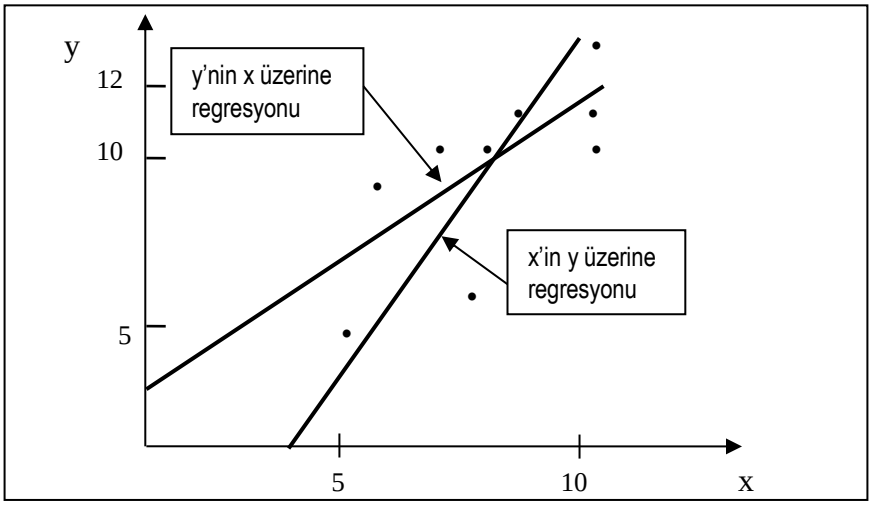
Böylece x 'in y üzerine regresyonu aşağıdaki gibi yazılabilir.

$$x = 1,37 + 0,69\ y \qquad r^2 = 0,52$$

Dikkat edilecek olursa, $\beta' * \beta''$ katsayılarının çarpımı (0,75*(0,69) = 0,52), r^2_{xy}'ye eşit olmaktadır. Bu iki regresyon çizgisi Şekil 4.1'de verilmiştir.

Gözlemlerin dağılımını dikkate alarak hatayı en az yapma aşağıdaki eşitlikteki gibi olur.

$$\sum (y_i - \alpha' - \beta' x_i)^2$$



Şekil 4.1. y'nin x üzerine, x'in y üzerine olan regresyon çizgileri

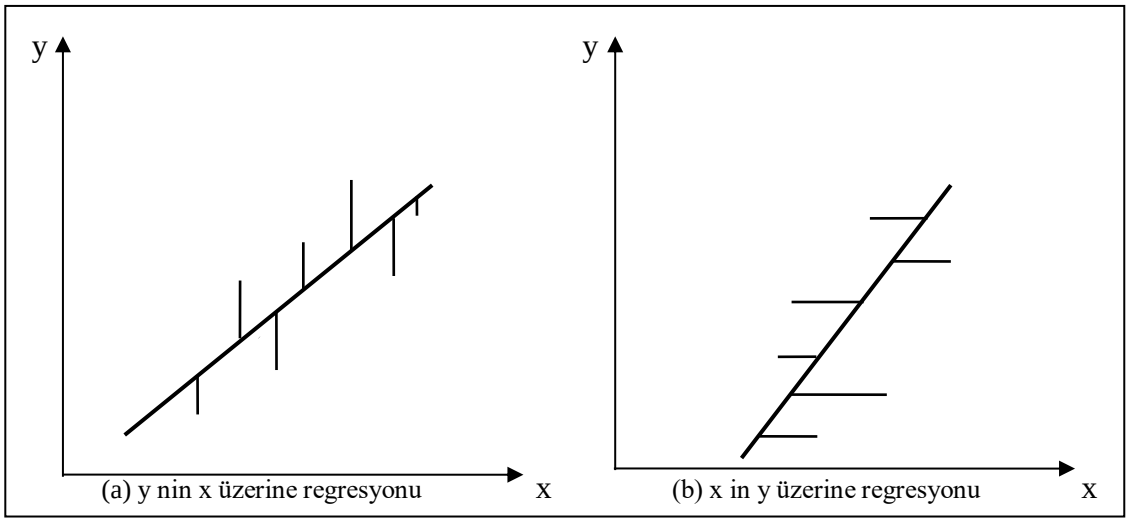
Örneklere ait noktalar arasından geçen doğruya, bu noktaların dikine uzaklıklarının karelerinin toplamını minimum yapar. Bu şekil 4.2 (a) de gösterilmiştir. Çizgi y'nin x üzerine regresyonunu gösterir.

Diğer taraftan ters regresyon için hatayı en az yapma aşağıdaki eşitlikteki gibi olur.

$$\sum (x_i - \alpha'' - \beta'' y_i)^2$$

Örneklere ait noktalar arasından geçen doğruya, bu noktaların yatay olan uzaklıklarının karelerinin toplamını minimum yapar. Bu şekil 4.2 (b) de gösterilmiştir. Çizgi x'in y üzerine regresyonunu gösterir.

Burada y'nin x üzerine mi yoksa x'in y üzerine mi regresyonu daha doğru sorusu ile karşılaşılmaktadır. Bu probleme ışık tutacak bilgiler aşağıdaki gibi sunulabilir.



Şekil 4.2. y'nin x üzerine ve x'in y üzerine olan regresyonunda rezidual kareler toplamının en küçük yapılması

- Örneğin reklam harcamalarının satış miktarlarını etkilemesi gibi, eğer neden olma yönü biliniyorsa burada reklam harcamaları açıklayıcı, satış miktarları da açıklanan değişken olarak kabul edilebilir. Burada problem çok açık ve cevabı da çok kolaydır.
- Bazı durumlarda bu problem açık değildir. Eğer x ve y ortak bir normal dağılışa sahip ise, hem x'in y üzerine hem de y'nin x üzerine regresyonu mantıklı olmaktadır. Bu durumda her iki şekilde de tahmin yapılarak verilen x değeri için y'nin ve verilen y değeri için x'in kestirmesi yapılabilir.
- Bazı modellerde y ve x'in ikisi de hata ile ölçülmüş olabilir. Bu durumda hem y'nin x üzerine hem de x'in y üzerine regresyonunun tahmini yapılmalı ki, β'nın sınırları belirlenebilsin. Bu husus ileri seviyede ekonometri konularında ele alınmaktadır.
- Ücretlerde cinsiyet ayrımcılığı örneğinde olduğu gibi, problemin ortaya konuş şekline göre her iki yöndeki regresyon da mantıklı olabilir.
- Bazen de hangi regresyonun daha mantıklı olduğu hususu, verilerin nasıl elde edildiğine de bağlıdır. Örneğin, önceki örneğimizde olan çalışma saatleri ile üretim arasındaki ilişkiyi ele alalım. Burada hangi regresyonun mantıklı olacağı verilerin nasıl elde edildiğine bağlıdır. Eğer işçilere belli bir çalışma süresi tanınarak elde ettikleri üretim gözleniyor ise, bu durumda y'nin x üzerine regresyonu doğru olandır. Bu durumda x kontrol edilen değişkendir. Ancak, eğer işçilere belli bir işi yapma görevi verilir ve o işi ne kadar iş zamanında yaptığı gözleniyorsa bu durumda x'in y üzerine regresyonu mantıklıdır. Bu durumda y kontrol edilen değişkendir. Burada değişkenlerden sadece birinin kontrol edildiği deneme ortamı söz konusudur. Deneme ortamından elde edilen verilerde hangisinin bağımlı ve hangisinin bağımsız değişken olduğu açıktır. Fakat çoğu ekonomik veride bu tam açık değildir.

Yumurta mı civcivden çıkar yoksa civciv mi yumurtadan çıkar? sorusu da benzer bir ilişkiyi ifade etmektedir. Cevap ilişkiyi nasıl kurguladığınıza bağlı olarak değişir.

4.3. Basit Regresyon Modeli İçin Varyans Analizi

Regresyon modeli ile ilgili diğer bir konuda çok sık olarak kullanılan varyans analizidir. Bu analiz, toplam kareler toplamının TSS, açıklanan (regresyon) kareler toplamı ESS ve hata (rezidual) kareler toplamı RSS olarak ayrılmasıdır. Burada tablo yapmanın amacı açılanan kareler toplamının önemini test edebilmektir. Bu durumda basit regresyon söz konusu olduğundan β'nın öneminin test edilmesi anlamına gelir. Çizelge 4.2 bize toplam kareler toplamının parçalanmış halini göstermektedir.

Çizelge 4.2. Basit regresyon modeli için varyans analizi

Varyasyonun kaynağı	Kareler toplamı	Serbestlik derecesi	Ortalama kareler
X	$ESS = \beta' S_{xy}$	1	ESS / 1
Hata	$RSS = S_{yy} - \beta' S_{xy}$	n-2	RSS / (n-2)
Toplam	$TSS = S_{yy}$	n-1	

Burada, $\beta = 0$ olması varsayımında, F istatistiğine sahip olunur çünkü σ^2'ler iptal olur ve F = (ESS / 1) / RSS / (n-2) elde edilir ki, bu da 1 ve (n-2) serbestlik derecesiyle F-dağılımına sahiptir. Bu F-istatistiği $\beta = 0$ hipotez testinde kullanılabilir.

Uygulama örnek 3.2'deki veriler için varyans analizi çizelge 4.3'de hazırlanmıştır. Burada F istatistiği, F = 15,75 / 1,83 = 8,6'dır. Yüzde 5 önem seviyesinde ve 1 ve 8 serbestlik derecesinde F cetvel değeri 5,32 olduğundan model açıklayıcıdır denilebilir. Burada dikkat edilmesi gereken husus, β'nın önemliliği için kullanılan t-istatistiği, $\beta' / SE(\beta') = 0,75 / 0,256 = 2,93$, ve tablodaki varyans analizinden sağlanan F-istatistiği arasındaki ilişkidir. Bu ilişki, F-istatistiğinin t-istatistiğinin karesi olmasıdır.

Çizelge 4.3. Örnekte kullanılan veriler İçin varyans analizi

Varyasyonun kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Ortalama kareler**
X	15,75	1	15,75
Hata	14,65	8	1,83
Toplam	30,4	9	

Bu yüzden, basit regresyon modeli, varyans analizi tablosu fazla bilgi vermez. İlerideki derslerde de görüleceği gibi, çoklu regresyonda durum aynı değildir. Varyans analizi tablosu, bütün regresyon parametrelerinin hepsine önemlilik testi sağlar. Bu noktada r^2, F ve t arasındaki ilişki aşağıdaki gibi ifade edilebilir.

$$ESS = \beta' S_{xy} = r^2 S_{yy} \quad \text{ve} \quad RSS = S_{yy} - \beta' S_{xy} = (1-r^2) S_{yy}$$

Böylece F-istatistiği aşağıdaki gibi yazılabilir.

$$F = r^2 / [(1-r^2) / (n-2)] = (n-2) r^2 / 1-r^2$$

Burada $F = t^2$ bu durumunda

$$r^2 = t^2 / [t^2 + (n-2)]$$

Örnekteki verilerle bunu kontrol edildiğinde $t^2 = 8,6$ ve (n-2) = 8 ise bu durumda

$$r^2 = 8,6 / (8,6+8) = 8,6 / 16,6 = 0,52$$

önceden de elde edildiği gibi $r^2 = t^2 / [t^2 + (n-2)]$ eşitliği bize $\beta = 0$ hipotezi testi için t oranı ile r^2 arasındaki ilişkiyi vermektedir.

4.4. Basit Regresyon Modeli İle Kestirme

Tahmin edilen regresyon denklemi $\hat{y} = \alpha' + \beta' x$, verilen x değerleri için y'nin kestirmesini yapmak için kullanılırken, tahmin edilen denklem x' = α"+β"y, verilen y değeri için x'in değerini vermektedir. Burada verilen x için y'nin değerinin nasıl belirleneceği açıklanacaktır.

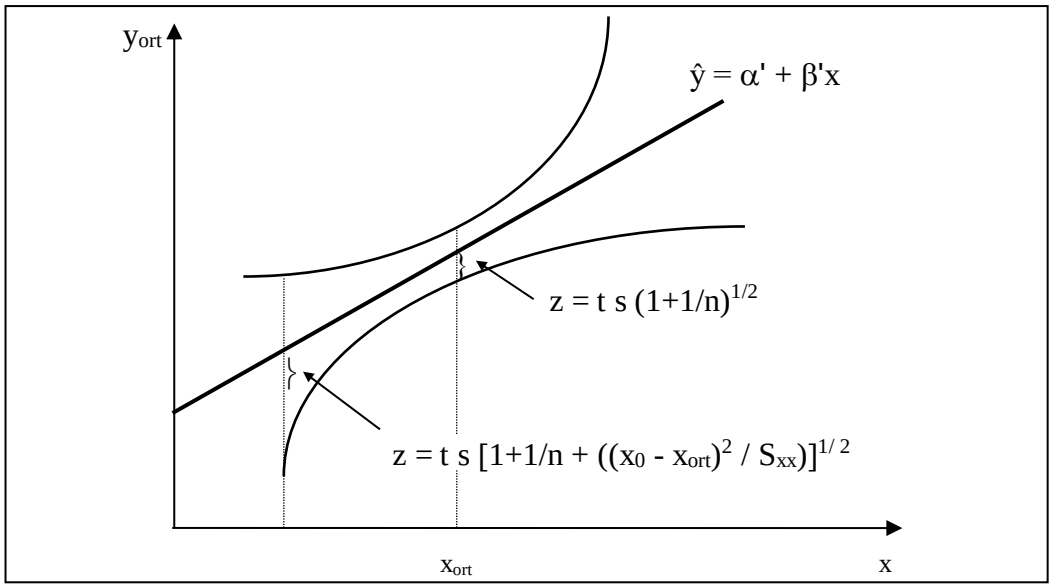
Varsayalım ki, x_0, x'in verilen değeridir. Daha sonra y'nin karşılığı olan $\hat{y}_0$'ın değeri, $\hat{y}_0 = \alpha' + \beta' x_0$ denklemi ile bulunmaya çalışılacaktır. Burada y_0'ın gerçek değeri, $y = \alpha + \beta x_0 + u_0$ denklemiyle verilmektedir. Bu durumda u_0 hata terimidir. Dolayısıyla tahmin etme hatası $\hat{y}_0 - y_0 = (\alpha' - \alpha) + (\beta' - \beta) x_0 - u_0$ olur. Bu nedenle $E(\alpha' - \alpha) = 0$, $E(\beta' - \beta) = 0$ ve $E(u_0) = 0$ olduğundan $E(\hat{y}_0 - y_0) = 0$ olur. Bunlar, ilk başta verdiğimiz denklemin sapmasız bir tahminci olduğunu gösterir. Burada $\hat{y}_0$ ve y_0 değişkenleri şansa bağlı değişkenler olduğundan $E(\hat{y}_0) = E(y_0)$ olması sebebiyle kestirici sapmasızdır. Kestirme hatasının varyansı ise aşağıdaki gibi ifade edilir.

$$V(\hat{y} - y_0) = V(\alpha' - \alpha) + x_0^2 V(\beta' - \beta) + 2 x_0 cov(\alpha' - \alpha, \beta' - \beta) + V(u_0)$$

$$= \sigma^2 (1/n + x_{ort}^2 / S_{xx}) + \sigma^2 (x_0^2 / S_{xx}) - 2x_0 \sigma^2 (x_{ort} / S_{xx}) + \sigma^2$$

$$= \sigma^2 (1 + 1/n + (x_0 - x_{ort})^2 / S_{xx})$$

Yukarıdaki denkleme göre, x_0 x'den uzaklaştıkça varyans artmaktadır. Burada gözlemlerin ortalamasıyla α' ve β''nın hesap edilmiş olması temeline dayanmaktadır. Bu durum aşağıdaki şekil 4.3'de gösterilmektedir.



Şekil 4.3. Verilen x değeri için y'nin tahmin değerleri için güven kesitleri
[Güven şeritleri z ile gösterilmiştir, burada z=tSE (ŷ) ve t kullanılan t değeridir.]

Eğer x_0 değeri x ile ilgili örnek gözlemleri aralığında ise buna örnek içindeki kestirme denilmektedir. Eğer x_0 örnek dışında bir aralıktan geliyorsa bu tip kestirmeye örnek dışı kestirme denilebilir. Örnek olarak 12 örnek sayısı temeline dayalı olarak tahmin edilen bir talep fonksiyonu ele alınmış olsun.

$$y = 10{,}0 + 0{,}90\ x$$

burada: y: tüketim harcamaları

x: harcanabilir gelir

Bu modelle ilgili olarak $s^2 = 0,01$, $x_{ort} = 200$, $S_{xx} = 4000$ ve $x_0 = 250$ olarak verildiği kabul edilir ise, kestirmenin değeri aşağıdaki gibi olur.

$$y_0 = 10,0 + 0,9\ (250) = 235$$

$$SE\ (y_0) = [\ 0,01\ (1+1/12 + 2500/4000)\ ]^{1/2} = 0,131$$

10 serbestlik derecesiyle *t* tablosundan t = 2,228 olduğundan, y_0 değeri için %95 güven aralığı $235 \pm 2,228\ (0,131) = 235 \pm 0,29$, yani (234,71, 235,29) olur.

Beklenen Değerlerin Kestirmesi

Bazen verilen x_0 için, y_0 ile değil de E (y_0) değeriyle ilgileniliyor olunabilir. Yani y_0 ile değil de y_0'ın ortalama değeriyle ilgileniliyor olabilir. Bununla ilgili bir örnek verilecektir. E (y_0) = $\alpha + \beta\, x_0$ olduğundan bu, E' (y_0) = $\alpha' + \beta'\, x_0$ tarafından tahmin edilecektir ki buda önceden de dikkate aldığımız gibi y_0'ın aynısıdır. Bu yüzden kestirme, kestirmek istenilen y_0 veya E (y_0) da olsa aynı olacak, fakat kestirmenin hatası farklı ve daha küçük olacaktır. Yani, elde edilecek güven aralığı farklı olacaktır. Bu durumda tahmin hatası aşağıdaki gibidir.

$$E'\ (y_0) - E\ (y_0) = (\ \alpha' - \alpha) + (\beta' - \beta)\ x_0$$

Dikkat edilirse bu, $-u_0$ hariç hata terimi olan $y_0' - y_0$ ile aynıdır. Burada u_o'ın varyansı σ^2 olduğundan, önceden tahmin edilen hata varyansından σ^2 yi çıkarmak gerekir. Bu yüzden,

$$Var\ (\ E'\ (y_0) - E\ (y_0)\ ) = \sigma^2\ (1/n + (x_0 - x_{ort})^2 / S_{xx})$$

olur. Tahminin standart hatası, σ^2 yerine s^2'yi ikame ettikten sonra bu ifadenin kara kökü alınarak tespit edilir. Güven aralıkları ise E' (y_0) $\pm\ t_c$ * SE ($\hat{y}$) olarak verilir. Burada t_c, *t* tablosundan bulunan cetvel değeridir.

Uygulamalı Örnek 4.2

Daha önceki uygulamalı örnek 3.2'de ele alınan atletik spor giysileri mağazası ile ilgili tahmin edilen regresyon denklemi aşağıdaki gibidir.

$$\hat{y} = 1,0 + 1,2\ x\ ,\qquad x_{ort} = 3,0,\qquad RSS = 8,8$$

Reklam harcamaları 6 milyon TL'ye çıktığı zaman satış gelirinin ne olacağı satışlardan sorumlu müdür tarafından istenmektedir. Ayrıca y'nin kestirmesi için güven aralığının da tahmin edilmesi isteniyor.

Burada $x_0 = 6$ olduğunda, $\hat{y}_o = 1,0 + 1,2\ (6) = 8,2$, olur. Kestirme hatasının varyansı, σ^2 $[\ 1 + 1 / 5 + (6 - 3)^2 / 10\ ] = 2,1\sigma^2$ ve s^2 = RSS / s.d. = 8,8 / 3 = 2,93 olduğundan kestirmenin standart hatası $(2,1*2,93)^{1/2} = 2,48$ olur. Burada, % 10 önem seviyesinde ve 3 serbestlik derecesiyle *t* dağılımındaki t değeri 2,353 dür. Verilen $x_0 = 6$ için y_0'ın güven aralığı $8,2 \pm 2,48*(2,353)$ = (2,36 ve 14,04) olduğundan, reklam harcamaları 6 milyon TL'ye çıktığında satış gelirinin % 90 güven aralığı, 2,36 milyar TL ile 14,04 milyar TL arasında olabileceği belirlenmiş olur.

Reklam harcamaları 6 milyon TL iken satış müdürü, gelecek iki yıl için aylık ortalama satış gelirinin kestirilmesini ve aynı zamanda bu kestirme için %90 güven aralığını istiyor. Bu durumda, y_0 ile değil de E (y_0) ile ilgilenilmektedir. Kestirme yine aynı şekilde tespit edilir ve 1,0 + 1,2 (6) = 8,2 denklemi kullanılır. Fakat tahmin hatasının varyansı $\sigma^2 (1/5 + (6-3)^2 / 10) = 1,1 \sigma^2$ olduğunda $s^2 = 2,93$ yerine konulur ve karekökü alınırsa standart hata elde edilmiş olur.

SE (E' (y_0)) = 1,795 %90 güven aralığı ise

$8,20 \pm 2,353 (1,795) = 8,20 \pm 4,22$

Böylece ortalama satışlar için %90 güven aralığı 3,98 milyar TL ile 12,42 milyar TL arasında olacaktır. Bu güven aralığı y_0 için hesapladığımızdan daha dardır.

4.5. Anormal Gözlemler

Regresyon parametre tahminlerinin birkaç anormal gözlem tarafından etkilendiği sık sık karşılaşılan bir durumdur. Anormal gözlemler (outlier) diğer gözlemlerin çok uzağında kalan gözlemlerdir. Bu tip gözlemler genellikle anormal faktörlerin etkisiyle oluşmaktadırlar. Fakat, EKK metodunu kullanıldığında, bu tek gözlem tahmin edilen regresyon denklemi üzerinde önemli değişiklikler yapmaktadır. Basit regresyon durumunda gözlemlerin grafiği çizilerek bu ekstrem gözlemler saptanabilir. Halbuki çoklu regresyon durumunda bu mümkün olmadığından rezidual ($\hat{u}_i$) analizi yapmalıdır.

Tek başına regresyon denklemi ve r^2 bize ilişki ile ilgili tüm bilgileri vermez. Bu durum çizelge 4.4'deki veriler kullanılarak bir örnekle açıklanabilir.

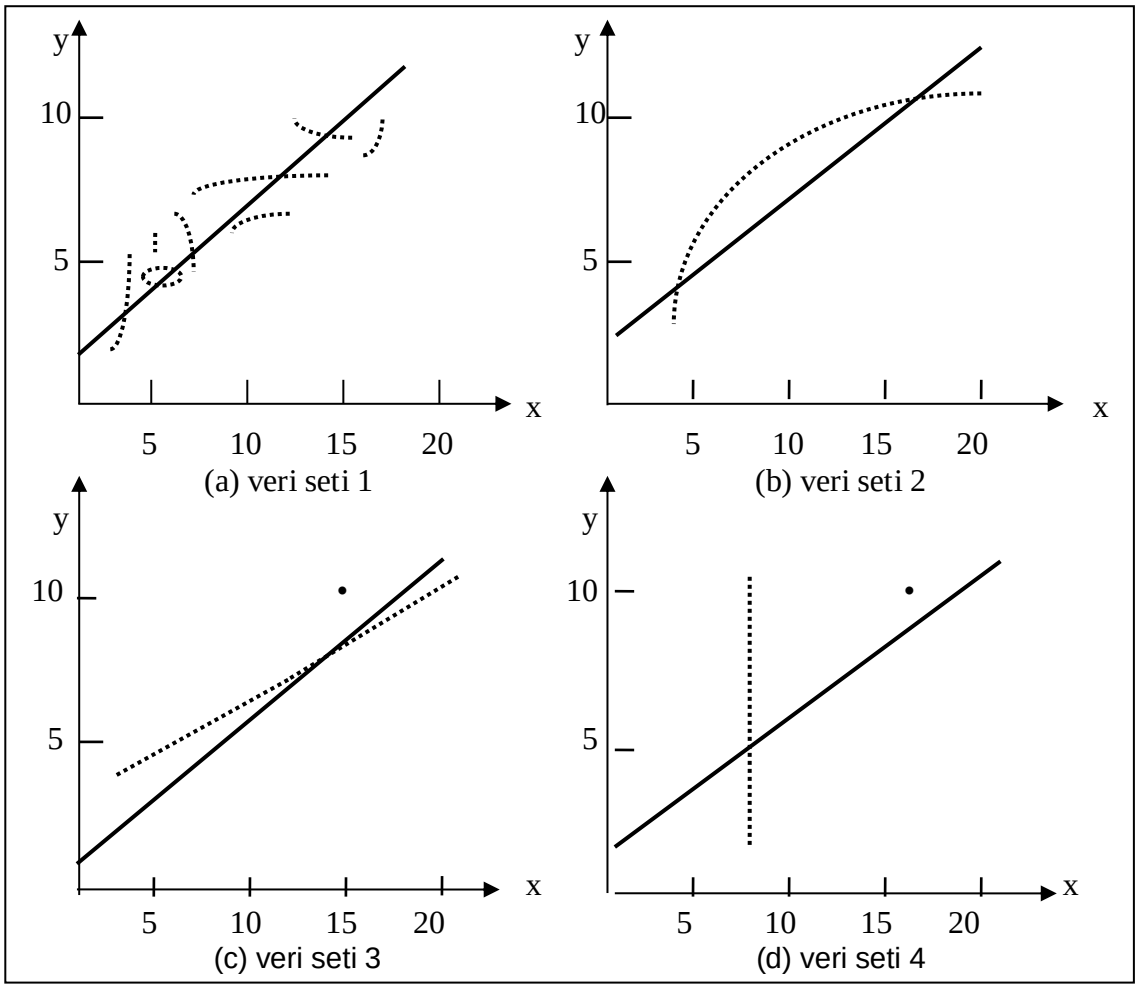
Çizelge 4.4. Benzer regresyon modelini veren dört veri seti

veri seti:	**1-3**	**1**	**2**	**3**	**4**	**4**
değişken:	**x**	**y**	**y**	**y**	**x**	**y**
1	10	8,04	9,14	7,46	8	6,58
2	8	6,95	8,14	6,77	8	5,76
3	13	7,58	8,74	12,74	8	7,71
4	9	8,81	8,77	7,11	8	8,00
5	11	8,33	9,26	7,81	8	8,47
6	14	9,96	8,10	8,84	8	7,04
7	6	7,24	6,13	6,08	8	5,25
8	4	4,26	3,10	5,39	19	12,5
9	12	10,84	9,13	8,15	8	5,56
10	7	4,82	7,26	6,42	8	7,91
11	5	5,68	4,74	5,73	8	6,89

Çizelge 4.4'deki veriler kullanılarak her bir y'nin verilen x'ler regresyonunu tahmin için hesaplamalar yapıldığında n = 11, $x_{ort} = 9,0$, $y_{ort} = 7,5$, $S_{xx} = 110,0$, $S_{yy} = 41,25$, $S_{xy} = 55,0$ olarak bulunur. Regresyon denklemi ise

$\hat{y} = 30 + 0,5 x$ $r^2 = 0,667$ olarak tahmin edilir.
(0,118)

Regresyon kareler toplamı 1 s.d. ile 27,5, örnek hatası kareler toplamı ise 9 s.d. ile 13,75 olarak bulunur. Burada regresyon denklemleri benzer olsalar bile bu dört veri grubu çok farklı özellikler göstermektedirler. Bu gözlemlerin şekil 4.4'deki gibi grafiğinin çizilmesi durumunda bu farklılıklar ortaya çıkmaktadır.



Şekil 4.4. Dört değişik veri grubu için benzer regresyon çizgileri

Birinci veri grubu ile ilgili herhangi bir problem yoktur. İkinci veri grubu regresyon çizgisinin doğrusal olmadığını gösteriyor. Üçüncü veri grubunda ise bir tek gözlemin regresyon doğrusunu nasıl yukarı çevirdiğini gösteriyor. Eğer bir tek gözlem dikkate alınmasaydı regresyon doğrusu noktalı doğru gibi biraz daha farklı olacaktı. Dördüncü veri grubunun ise, tek bir gözlem sebebiyle regresyon doğrusu tamamen dik olması gerekirken ne kadar farklı bir şekilde tahmin edildiğini göstermektedir. Eğer bu aşırı gözlem dikkate alınmazsa ortaya tamamen dik bir doğru çıkmaktadır. Bu şekilde örneklenen durumlarla gerçek hayatta da karşılaşılmaktadır. Bazı yıllar ve aylar arasındaki önemli farklılıklar nedeniyle farklı veriler elde edilebilir. Örneğin, tüketimle ilgili veriler savaş yıllarında çok farklı olabilir. Et tüketimi ile ilgili analizlerde kurban bayramının bulunduğu ayda anormal gözlemler elde edilebilir. Normal şartlardaki

parametreleri tahmin ederken böyle anormal durumdaki verileri veri setinden çıkarmak gerekebilir. Bu verileri veri setinden çıkarmak için yeterli gerekçeler olabilir. Veriler içinden anormal rakamları veya genel seyrin dışındaki gözlemleri sebepsiz olarak veri setinden çıkarmak doğru olmaz. Bunun yerine regresyon modelinin fonksiyonel formunda değişiklikler yaparak anormal gözlemlerin etkilerini azaltmak veya model içinde doğru algılanmasını sağlamak mümkündür.

Uygulamalı Örnek 4.3

Bu örnekte 1929-1984 dönemindeki ABD'nin tüketim fonksiyonunun tahmini yapılmaktadır (Çizelge 4.5). Aşağıdaki tabloda kişi başına tüketilebilir gelir (Y) ve kişi başına tüketim harcamaları (C) ABD sabit doları ile verilmiştir. Veriler devamlı olmayıp 1929, 1933 ve 1939'dan itibaren devamlıdır.

Çizelge 4.5. ABD'de yıllara göre tüketim ve gelir seviyeleri (1972 doları)

Yıllar	C	Y	Yıllar	C	Y	Yıllar	C	Y
1929	1765	1883	1953	2277	2501	1969	3245	3564
1933	1356	1349	1954	2278	2483	1970	3277	3665
1939	1678	1754	1955	2384	2582	1971	3355	3752
1940	1740	1847	1956	2410	2653	1972	3511	3860
1941	1826	2083	1957	2416	2660	1973	3623	4080
1942	1788	2354	1958	2400	2645	1974	3566	4009
1943	1815	2429	1959	2487	2709	1975	3609	4051
1944	1844	2483	1960	2501	2709	1976	3774	4158
1945	1936	2416	1961	2511	2742	1977	3924	4280
1946	2129	2353	1962	2583	2813	1978	4057	4441
1947	2122	2212	1963	2644	2865	1979	4121	4512
1948	2129	2290	1964	2751	3026	1980	4093	4487
1949	2140	2257	1965	2868	3171	1981	4131	4561
1950	2224	2392	1966	2979	3290	1982	4146	4555
1951	2214	2415	1967	3032	3389	1983	4303	4670
1952	2230	2441	1968	3160	3493	1984	4490	4941

C'nin Y üzerine regresyonunu tahmin edildiğinde

$$C = \underset{(58124)}{-24,944} + \underset{(0,018)}{0,911}\ Y \qquad r^2 = 0,9823$$

modeli elde edilir. Burada β için t değeri 0,911 / 0,018 = 50,47'dir.

Bu durumda yapılacak ikinci iş rezidualları hesap etmektir. Bu yapıldığında 6, 7, 8 ve 9. verilerin büyük ve negatif olduğu çizelge 4.6'da görülmektedir. Bu rakamlar 1942-1945 arasındaki II. dünya savaşı yıllarına tekabül etmektedir ki bu yıllarda tüketici harcamaları savaş şartlarından olumsuz etkilenmektedir. Bu anormal veriler, veri setinden çıkarıldığında aşağıdaki sonuçlar elde edilir.

$$C = \underset{(22,353)}{85,725} + \underset{(0,007)}{0,885}\ Y \qquad r^2 = 0,9975$$

Önceki modelde kesişme katsayısı sıfırın çok altında negatif bir sayı iken şimdi pozitif olmuştur. Ayrıca marjinal tüketim eğilimi de daha küçülmüştür. Burada β'nın standart

hatasının daha küçük olması modelin istatistik açıdan, kesişme katsayısının pozitif olması ekonomik teori açısından modelin geliştiğini göstermektedir.

Çizelge 4.6. Önceki tablodaki verilerden tahmin edilen tüketim fonksiyonu için reziduallar

Gözlem	Rezidual	Gözlem	Rezidual	Gözlem	Rezidual
1	75,0	17	24,2	33	24,1
2	152,4	18	41,6	34	-35,9
3	105,5	19	57,4	35	-37,1
4	82,8	20	18,8	36	20,5
5	-46,1	21	18,4	37	-67,9
6	-330,9	22	16,0	38	-60,2
7	-372,2	23	44,8	39	-55,4
8	-392,4	24	58,8	40	12,1
9	-239,4	25	38,7	41	51,0
10	11,0	26	46,0	42	37,4
11	132,4	27	59,7	43	36,7
12	68,4	28	20,1	44	31,5
13	109,4	29	5,0	45	2,1
14	70,5	30	7,6	46	22,5
15	39,5	31	-29,5	47	74,8
16	31,8	32	3,7	48	15,0

Uygulamalı Örnek 4.4

Bu örnekteki veriler, çizelge 4.7'de verilmiştir. Bu tablo 1962-1980 yılları arasında Avustralya'daki işsizlik oranlarını ve işsizlik ücretlerini kapsıyor Burada işsizlik oranının işsizlik ücreti üzerine basit regresyonu yapılacaktır.

y_1 : Delikanlılar için işsizlik oranı
y_2 : Kızlar için işsizlik oranı
x : İşsizlik geliri (sabit dolar)

Regresyon sonuçları aşağıdaki gibi tahmin edilmektedir.

$$y_1 = 2{,}478 + 0{,}212\ x \qquad r^2 = 0{,}690$$
$$(1{,}234)\quad (0{,}035)$$
$$y_2 = 3{,}310 + 0{,}226\ x \qquad r^2 = 0{,}660$$
$$(1{,}410)\quad (0{,}039)$$

Bu denklemler işsizlik geliri arttıkça işsizlik oranının da arttığını gösteriyor. Fakat Çizelge 4.8'de verilen reziduallar kontrol edildiğinde 1973, 1974 ve 1977–1980 yılları arasındaki verilerin büyük olduğu ortaya çıkmaktadır (Çizelge 4.8). Eğer bu verileri elemine edilirse aşağıdaki sonuçlar elde edilir.

$$Y_1 = 2{,}156 + 0{,}203\ x \qquad r^2 = 0{,}876$$
$$(0{,}610)\quad (0{,}023)$$
$$Y_2 = 2{,}799 + 0{,}225\ x \qquad r^2 = 0{,}955$$
$$(0{,}394)\quad (0{,}015)$$

Burada x'in katsayıları fazla değişmemiş fakat r^2 değerlerine bakıldığında 13 gözlemle tahmin edilen modellerin daha iyi olduğu görülmektedir.

Çizelge 4.7. Gençler arasındaki yıllara göre işsizlik oranları ve işsizlik ücretleri

	Geçlerde İşsizlik Oranı (%)		16-17 Yaş Arası İşsizlik Ücreti	
Yıllar	**Erkek**	**Kadın**	**Cari Fiyatlarla Dolar**	**Sabit Fiyatlarla (1981) Dolar**
1962	4,5	5,9	3,50	13,3
1963	3,4	4,5	3,50	13,2
1964	2,5	4,4	3,50	12,7
1965	4,4	5,1	3,50	12,2
1966	3,1	5,5	3,50	11,9
1967	4,1	5,7	3,50	11,6
1968	4,7	6,0	3,50	11,3
1969	5,8	6,3	4,50	14,1
1970	5,7	5,1	4,50	13,4
1971	6,9	6,1	4,50	12,5
1972	8,6	9,3	7,50	20,0
1973	7,6	7,8	23,00	54,1
1974	11,9	12,3	31,00	62,7
1975	14,6	16,7	36,00	63,8
1976	13,6	15,6	36,00	55,8
1977	17,1	18,8	36,00	51,1
1978	16,4	18,8	36,00	47,4
1979	15,9	18,9	36,00	43,1
1980	15,7	17,6	36,00	39,5

Fakat burada 6 gözlem dikkate alınmadan yapılan tahmin doğru değildir. İlk örnekte savaş yıllarını silmek için bir gerekçemiz vardı. Çünkü II. dünya savaşı yılları, tüketimi normalin dışında etkilemiştir. Fakat bu örnekte böyle bir anormallik söz konusu değil ve büyük reziduallar normal şartların etkisiyle oluştuğundan bu örnekte çözüm bu anormal gözlemlerin dikkate alınmaması değildir. Bu zaman serisi verilerinde birçok sebepten dolayı önceki yılların verileri sonraki yıllar üzerine etkisini elemine etmek için modelin yeniden yapılandırılması gerekli olabilir. Bunun çözümü şu anki konunun dışındadır. Fakat burada verilmek istenen mesaj bütün anormal gözlemlerin elemine edilmemesi gerektiğini vurgulamaktır.

Çizelge 4.8. İşsizlik oranlarının işsizlik gelirleri üzerine regresyonunu rezidualları

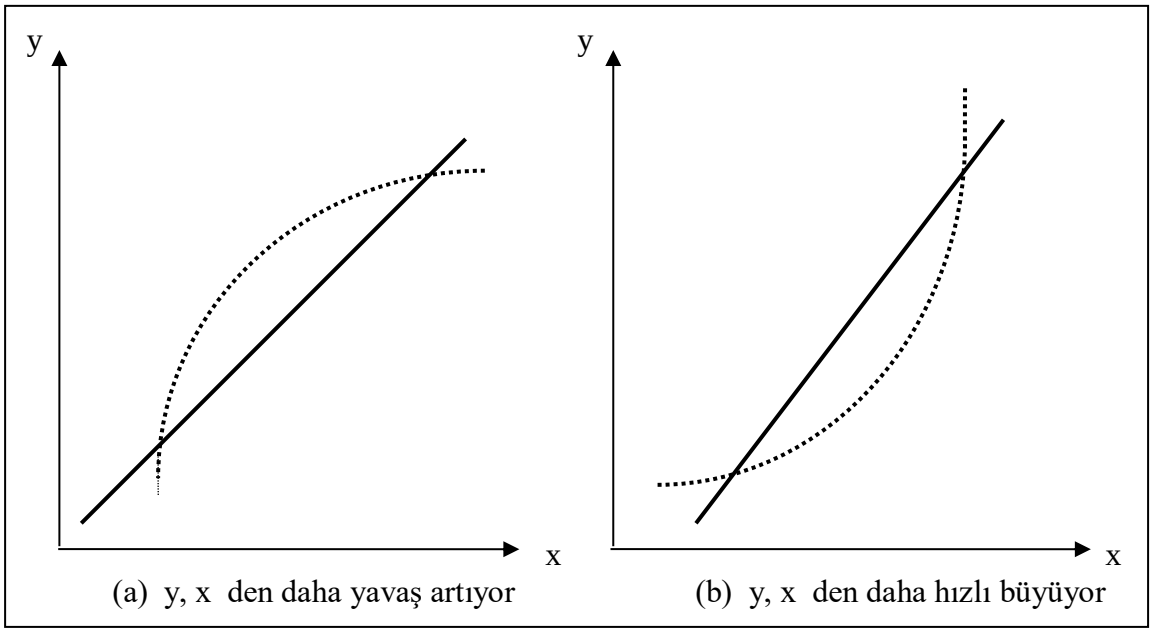
Yıllar	Y_1'in Regresyonu	Y_2'nin Regresyonu	Yıllar	Y_1'in Regresyonu	Y_2'nin Regresyonu
1962	-0,80	-0,42	1972	1,88	1,47
1963	-1,88	-1,80	1973	-6,34	-7,73
1964	-2,67	-1,78	1974	-3,89	-5,20
1965	-0,67	-0,97	1975	-1,42	-1,04
1966	-1,90	-0,50	1976	-0,72	-0,33
1967	-0,84	-0,23	1977	3,78	3,93
1968	-0,18	0,13	1978	3,86	4,77
1969	0,33	-0,20	1979	4,27	5,84
1970	0,38	-1,24	1980	5,04	5,35
1971	1,77	-0,04			

4.6. Regresyon Denklemleri İçin Alternatif Fonksiyonel Formlar

Bazen x ve y arasındaki ilişki doğrusal değil de eğrisel olabilir. Bu durumda bu ilişki için uygun fonksiyonel form kullanmak gerekir. Bir takım transformasyonlar sonucu doğrusal hale getirilerek kullanabilecek birçok fonksiyonel formlar vardır.

Örneğin, aşağıdaki şekil 4.5 a'da görüldüğü gibi eğer y değeri x'den daha yavaş artıyorsa, kullanılması ihtimal dâhilinde olan fonksiyon $y = \alpha + \beta \log x$'dir. Burada x ve y gibi iki değişkenden birinin logaritması alındığından bu form semilog form olarak adlandırılır. Bu durumda eğer bir değişken yeniden belirlenirse $X = \log x$, ve dolayısıyla $y = \alpha + \beta X$ şeklini alır. Böylece açıklanan değişkenin y ve açıklayıcı değişkenin ise $X = \log x$ olduğu bir doğrusal regresyon modeli elde edilmiş olur.

Yine başka bir örnek, şekil 4.5 b'de gösterilmiştir. Burada y, x'den daha hızlı arttığından kullanılması ihtimal dâhilindeki fonksiyonel form $y = A\, e^{\beta x}$'dir. Bu durumda her iki tarafında logaritmalarını alıp diğer türlü bir semilog fonksiyonu $\log y = \log A + \beta x$ elde edilir. Burada $Y = \log y$ ve $\alpha = \log A$ olarak alınırsa, $Y = \alpha + \beta x$ olur.



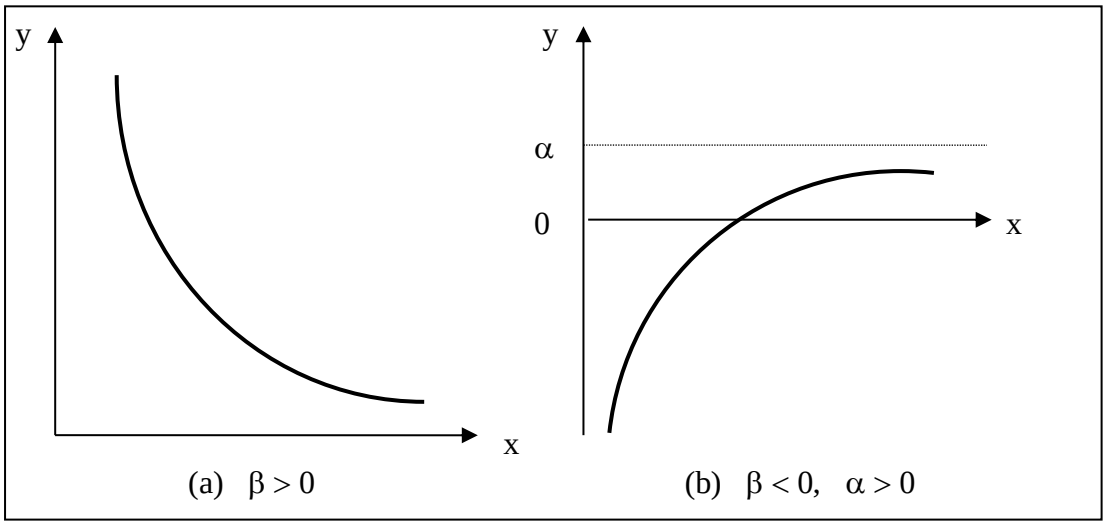
(a) y, x den daha yavaş artıyor (b) y, x den daha hızlı büyüyor

Şekil 4.5. Doğrusal regresyonun uygun olmadığı veri grupları

Elde edilen bu yeni form, doğrusal bir regresyon denklemi olur. Kullanılacak diğer alternatif model $y = A\, x$'dir. Bu durumda her iki tarafın logaritması alındığında, $\log y = \log A + \beta \log x$ olur. Bu durumda β katsayısı aynı zamanda elastikiyet olarak ifade edildiğinden bu form ekonometride çok yaygın olarak kullanılır. Hem x ve hem de y'nin logaritması alındığı için bu modele double-log adı verilmektedir. Şimdi $Y = \log y$ ve $X = \log x$ ve $\alpha = \log A$ şeklinde belirlenirse yine $Y = \alpha + \beta X$ formu elde edilir ki bu doğrusal bir regresyon denklemidir.

Elimizdeki veriler Şekil 4.6'da ki gibi bir görünümde ise aşağıda verilen bazı diğer fonksiyonel formlar kullanılışlı olabilir:

$$Y = \alpha + \beta / x \quad \text{veya} \quad Y = \alpha + \beta / \sqrt{x}$$



Şekil 4.6. (a) Artan x ile y azalıyor, (b) artan x ile y artıyor.

Birinci durumda $X = 1 / x$ olarak, İkinci durumda ise $X = 1 / \sqrt{x}$ olarak ifade edilmektedir. Her iki durumda da denklem transformasyondan sonra $y = \alpha + \beta x$. şeklinde doğrusal olmaktadır. Eğer $\beta > 0$ ise, ilişki Şekil 4.4 (a)'daki gibi, yok eğer $\beta < 0$ ise, bu durumda ilişki Şekil 4.4(b)'deki gibi ortaya çıkmaktadır.

Uygulamalı Örnek 4.5

ABD'ye ait 1929-1967 arasındaki 39 yıllık üretim, işgücü ve sermaye verilerinin değişkenleri aşağıdaki gibidir.

X: Sabit dolarla GSMH indeksi

L: Kişi sayısı, çalışma şartları ve eğitim seviyesine göre ayarlanmış işgücü girdi indeksi

K: Faydalanma oranına göre adapte edilmiş sermaye girdi indeksi

Normalde üretim işgücü ve sermayeye bağlı olduğundan X'in L ve K üzerine regresyonu tahmin edilmelidir. Burada basit regresyon analizi yapılacağından sadece üretim ile işgücü arasındaki ilişki tahmin edilecektir.

Verilerde X'in L ve K'dan daha hızlı arttığı, yani Şekil 4.5 (b)'deki duruma benzer olduğu ifade edilmektedir. Bu durum da tek veya çift taraflı logaritmik formların kullanılmasını öngörmektedir. Katsayılar doğrudan elastikiyet olarak alınabildiği için ikinci metot ilave bir kolaylık daha sağlamaktadır. Ayrıca X L'ye göre öyle çok fazla

hızlı artmadığından tek tarafı logaritmik yerine aşağıdaki gibi çift taraflı logaritmik fonksiyonel form kullanılabilir.

$$\log X = \alpha + \beta \log L_1 + u$$

X'in L_1 üzerine doğrusal regresyonu aşağıda verilmiştir. Parantez içinde standart hatalar verilmiştir.

$$X = \underset{(23,55)}{-338,01} + \underset{(0,106)}{3,052}\, L_1 \qquad r^2 = 0,9573$$

Rezidualları kontrol edildiğinde ilk 12 gözlem pozitif, ortadaki 17 gözlem negatif ve son 10 gözlem tekrar pozitiftir. Bu da Şekil 4.5 (b)'deki ilişkiyi göstermektedir. Bu durumda çift taraflı logaritmik formda tahmin edildiğinde,

$$\log X = \underset{(0,226)}{-5,480} + \underset{(0,084)}{2,084} \log L_1 \qquad r^2 = 0,9851$$

elde edilmektedir.

Logaritmik doğrusal formun yorumu kolay olduğundan yaygın olarak kullanılmaktadır. Burada üretimin işgücü elastikiyeti yaklaşık olarak 2'dir. Doğal olarak bu değerin çok yüksek olduğu düşünülebilir. Bunun sebebi sermayenin modele koyulmamasından kaynaklanabilir.

Basit regresyonda fonksiyonun tipini belirlemenin en kolay yolu veri setindeki örnek noktalarını grafik üzerinde görüp en uygun fonksiyonu seçmektir. Bunu çoklu regresyonda yapmak mümkün olmadığından yerine rezidual analizi kullanılır.

YAZILIM UYGULAMASI

```
**************************************************************************
  SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
  ** Copyright (C)  1999  by K.J. White - All Rights Reserved **

  FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
  AT: ATATURK UNIVERSITESI - ERZURUM
  FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
**************************************************************************
```

Uygulamalı Örnek 4.3, sayfa: 62

```
**************************************************************************
|_SAMPLE 1 48
|_READ (MTABLE3.6) YEAR C Y

 UNIT 88 IS NOW ASSIGNED TO: MTABLE3.6
    3 VARIABLES AND       48 OBSERVATIONS STARTING AT OBS       1
|_OLS C Y / LIST
        48 OBSERVATIONS     DEPENDENT VARIABLE= C
 ...NOTE..SAMPLE RANGE SET TO:      1,     48
  R-SQUARE =   0.9823     R-SQUARE ADJUSTED =   0.9819
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   13011.
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   114.07
 SUM OF SQUARED ERRORS-SSE=  0.59852E+06
 MEAN OF DEPENDENT VARIABLE =   2788.4
```

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 46 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
Y	0.91074	0.1805E-01	50.47	0.000	0.991	0.9911	1.0089
CONSTANT	-24.944	58.12	-0.4291	0.670	-0.063	0.0000	-0.0089

OBS. NO.	OBSERVED VALUE	PREDICTED VALUE	CALCULATED RESIDUAL
1	1765.0	1690.0	75.029
2	1356.0	1203.6	152.36
3	1678.0	1572.5	105.51
4	1740.0	1657.2	82.815
5	1826.0	1872.1	-46.118
6	1788.0	2118.9	-330.93
7	1815.0	2187.2	-372.23
8	1844.0	2236.4	-392.41
9	1936.0	2175.4	-239.39
10	2129.0	2118.0	10.983
11	2122.0	1989.6	132.40
12	2129.0	2060.6	68.359
13	2140.0	2030.6	109.41
14	2224.0	2153.5	70.464
15	2214.0	2174.5	39.518
16	2230.0	2198.2	31.838
17	2277.0	2252.8	24.194
18	2278.0	2236.4	41.588
19	2384.0	2326.6	57.425
20	2410.0	2391.2	18.763
21	2416.0	2397.6	18.387
22	2400.0	2384.0	16.048
23	2487.0	2442.2	44.761
24	2501.0	2442.2	58.761
25	2511.0	2472.3	38.707
26	2583.0	2537.0	46.045
27	2644.0	2584.3	59.687
28	2751.0	2730.9	20.058
29	2868.0	2863.0	5.0016
30	2979.0	2971.4	7.6241
31	3032.0	3061.5	-29.539
32	3160.0	3156.3	3.7448
33	3245.0	3220.9	24.083
34	3277.0	3312.9	-35.902
35	3355.0	3392.1	-37.136
36	3511.0	3490.5	20.505
37	3623.0	3690.9	-67.857
38	3566.0	3626.2	-60.195
39	3609.0	3664.4	-55.446
40	3774.0	3761.9	12.106
41	3924.0	3873.0	50.996
42	4057.0	4019.6	37.368
43	4121.0	4084.3	36.705
44	4093.0	4061.5	31.474
45	4131.0	4128.9	2.0794
46	4146.0	4123.5	22.544
47	4303.0	4228.2	74.809
48	4490.0	4475.0	15.000

```
|_*Şimdi büyük sapmalı hataları, SKIPIF komutu ile dikkate almayalım.
|_*Aşağıdaki mantık ilişkilerinin hepsini SKIPIF (ve GENR) komutu ile
|_*kullanmak mümkündür. .EQ., .NE., .GE., .LE., .LT., .AND., .OR., .NOT.

|_SKIPIF ((YEAR.EQ.1942).OR.(YEAR.EQ.1943).OR.(YEAR.EQ.1944).OR.(YEAR.EQ.1945))
 OBSERVATION     6 WILL BE SKIPPED
 OBSERVATION     7 WILL BE SKIPPED
 OBSERVATION     8 WILL BE SKIPPED
 OBSERVATION     9 WILL BE SKIPPED


|_OLS C Y / LIST
        44 OBSERVATIONS     DEPENDENT VARIABLE= C
 ...NOTE..SAMPLE RANGE SET TO:      1,     48
  R-SQUARE =   0.9975     R-SQUARE ADJUSTED =   0.9975
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   1760.4
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   41.957
 SUM OF SQUARED ERRORS-SSE=   73936.
 MEAN OF DEPENDENT VARIABLE =   2874.1

VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      42 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
    Y       0.88523  0.6806E-02 130.1    0.000 0.999   0.9988     0.9702
 CONSTANT   85.725    22.35    3.835    0.000 0.509   0.0000     0.0298
|_* SKIPIF komutu DELETE SKIP$ komutu ile sonlandırılır.

|_DELETE SKIP$
 VARIABLE SKIP$    IS DELETED        48 WORDS RELEASED

|_DELETE / ALL
ALL VARIABLES HAVE BEEN DELETED

|_STOP
```

****

Uygulamalı Örnek 4.5, sayfa: 66

****

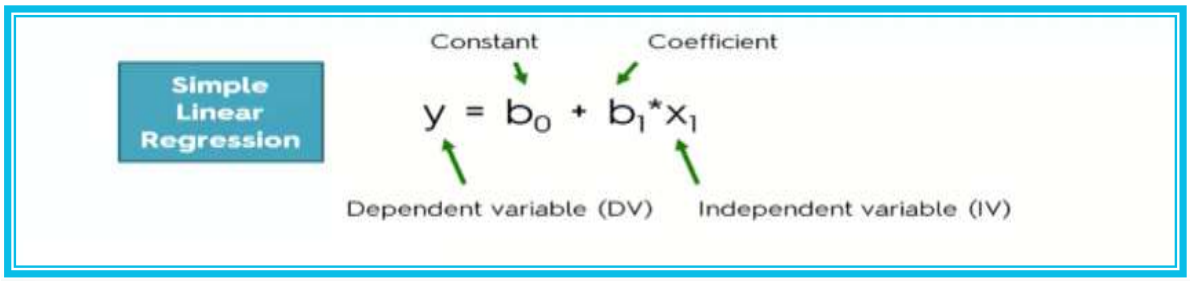
```
|_SAMPLE  1 39
|_READ (MTABLE3.11) YEAR X L K
 UNIT 88 IS NOW ASSIGNED TO: MTABLE3.11
    6 VARIABLES AND       39 OBSERVATIONS STARTING AT OBS       1
|_OLS X L
        39 OBSERVATIONS     DEPENDENT VARIABLE= X
 ...NOTE..SAMPLE RANGE SET TO:      1,     39
  R-SQUARE =   0.9573     R-SQUARE ADJUSTED =   0.9561
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   897.32
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   29.955
 SUM OF SQUARED ERRORS-SSE=   33201.
 MEAN OF DEPENDENT VARIABLE =   325.94

 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      37 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
    L       3.0519    0.1060     28.80   0.000  0.978  0.9784    2.0370
 CONSTANT -338.01    23.55     -14.35   0.000 -0.921  0.0000   -1.0370
```

```
|_PLOT X L
                    *=X
                    M=MULTIPLE POINT
    800.00        |
    757.89        |
    715.79        |
    673.68        |
    631.58        |
    589.47        |                              * *
    547.37        |                            *
    505.26        |                          **
    463.16        |                        **
    421.05        |                      M*
    378.95        |                     M*
    336.84        |                    M*
    294.74        |                 *M
    252.63        |              MM**
    210.53        |           *
    168.42        |       *MM*
    126.32        |    M***
    84.211        |
    42.105        |
    0.0000        |
                  _____________________________________________
              120.000    180.000    240.000    300.000    360.000
                                    L
|_*Logaritması alınmış değişkenler üretip regresyonu tekrar çalıştıralım.
|_*LOGLOG opsiyonu düzeltilmiş log-benzerlilik fonksiyonunu ve regresyonun
|_*elastikiyetlerini hesaplar.

|_GENR LNX=LOG(X)
|_GENR LNL=LOG(L)
|_OLS LNX LNL / LOGLOG
        39 OBSERVATIONS     DEPENDENT VARIABLE= LNX
 ...NOTE..SAMPLE RANGE SET TO:      1,     39
R-SQUARE =   0.9851     R-SQUARE ADJUSTED =   0.9847
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.32533E-02
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.57038E-01
 SUM OF SQUARED ERRORS-SSE=  0.12037
 MEAN OF DEPENDENT VARIABLE =   5.6874

VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      37 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
  LNL       2.0837  0.4214E-01  49.45     0.000  0.993  0.9925     2.0837
 CONSTANT -5.4802     0.2260    -24.24     0.000 -0.970  0.0000    -5.4802
|_STOP
```



V. ÇOKLU REGRESYON

5.1. Giriş

Basit regresyonda bir açıklanan değişken y ile açıklayıcı değişken x arasındaki ilişkinin nasıl tahmin edildiği anlatıldı. Çoklu regresyonda ise y ile çok sayıdaki açıklayıcı değişkenler x_1, x_2, x_3, ..., x_k arasındaki ilişki ifade edilmeye çalışılacaktır. Örneğin talep tahmini çalışmalarında bir malın talep edilen miktarı ile malın fiyatı, ikame malların fiyatları ve tüketici geliri arasındaki ilişki ifade edilmeye çalışılır. Bunun gibi çoklu regresyon modelleri aşağıdaki gibi yazılabilir.

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i \qquad i = 1, 2, 3, ..., n$$

Hata terimleri olarak ifade edilen u_i ler y'deki ölçüm hataları ve y ile x'ler arasındaki ilişkilerin belirlenmesinde yapılan hataları içermektedir. Basit regresyonda yapılan varsayımların aynıları burada da yapılabilir. Bunlar;

1. $E(u_i) = 0$
2. $V(u_i) = \sigma^2$, bütün i'ler için
3. u_i ve u_j'ler birbirinden bağımsız bütün $i \neq j$'ler için
4. u_i ve x_i'ler birbirinden bağımsız bütün i ve j'ler için
5. u_i'ler bütün i'ler için normal dağılıma sahiptir

İlk dört varsayım ile en küçük kareler (EKK) yöntemi minimum varyansa sahip ve sapmasız olan α, β_1, β_2, ..., β_k tahminlerini üretir. Varsayımların beşincisi olan normal dağılım varsayımı ise güven aralığı ve önemlilik testlerini yapmak için gerekli görülmektedir. Bu varsayım EKK tahmin yönteminin etkin sonuç veren özelliklerini temin için gerekli değildir.

Bu varsayımlara ilave olarak x_1, x_2, ..., x_n'lerin aralarında eş doğrusallık olmadığı, yani bu bağımsız değişkenler arasında belirleyici doğrusal bir ilişkinin bulunmadığı varsayılacaktır. Örneğin; aşağıdaki regresyon modelimizin olduğunu kabul edelim.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$$

Fakat burada x_1 ve x_2 arasında $2x_1 + x_2 = 4$ şeklinde belirleyici doğrusal bir ilişki mevcut ise bu durumda x_2, x_1 kullanılarak yazılırsa, $x_2 = 4 - 2x_1$ olur ve bu durumda regresyon denklemi aşağıdaki hali alır.

$$y = \alpha + \beta_1 x_1 + \beta_2 (4 - 2x_1) + u$$

$$= (\alpha + 4\beta_2) + (\beta_1 - 2\beta_2) x_1 + u$$

Bu durumda α, β_1 ve β_2'yi ayrı ayrı tahmin etmeyi değil, $(\alpha + 4\beta_2)$ ve $(\beta_1 - 2\beta_2)$'yi tahmin etmek gerekmektedir.

Açıklayıcı değişkenler arasında tam bir doğrusal ilişki olması durumuna tam eş doğrusallık deniliyor. İki değişkenin dikkate alınması durumunda, tam ilişki x_1 ile x_2 arasındaki korelasyon katsayısının +1 veya -1 olduğunu ima etmektedir. Buradaki analizlerde tam eş doğrusallık dikkate alınmayacaktır. Yani x_1, x_2, ..., x_k gibi birçok değişkenin y üzerine olan etkisinin analizlerinde kısmi ve birleşik etkilerin ayırt edilmesi gerekmektedir. Örneğin; talep edilen miktara gelir ve fiyatın etkisinin tahmin edilmesi durumunda gelir ve fiyatın birleşik ve kısmi etkileri dikkate alınmalıdır.

1- Gelir sabit tutulduğunda talep edilen miktar üzerine fiyatın etkisi

2- Fiyat sabit tutulduğunda talep edilen miktar üzerine gelirin etkisi

Bu kısımda bu ve benzeri problemler çözülmeye çalışılacaktır. Eğer fiyat ile gelir arasında çok yüksek bir korelasyon var ise, bu iki değişkenin ayrı ayrı etkilerini tespit etmek zor olacaktır. Burada iki açıklayıcı değişken ile başlanılacak ve daha sonra k sayıda açıklayıcı değişken için modeller yazılacak ve açıklanacaktır.

5.2. İki Açıklayıcı Değişkene Sahip Bir Model

Aşağıda iki değişkenli bir model verilmektedir.

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \qquad i = 1, 2, ..., n$$

Hata terimi hakkında yapılan varsayımlar aşağıdaki denklemleri ima etmektedir;

$$E(u) = 0 \qquad Cov(x_1, u) = 0 \qquad Cov(x_2, u) = 0$$

Burada α', β'_1 ve β'_2, α, β_1 ve β_2 'nin tahmincileri olduğu kabul edildiğinde u_i hata teriminin örnek karşılığı aşağıdaki gibidir.

$$\hat{u}_i = y_i - \alpha' - \beta'_1 x_{1i} - \beta'_2 x_{2i}$$

Bu durumda α', β'_1 ve β'_2'yi tahmin etmek için kullanılan üç denklem, popülasyan (ana kitle) ana kitle varsayımları yerine onların örnek karşılıkları konularak elde edilir.

Ana kitle Varsayımları	**Örnek Karşılıkları**
$E(u) = 0$	$(1/n) \sum \hat{u}_i = 0$ veya $\sum \hat{u}_i = 0$
$Cov(u, x_1) = 0$	$(1/n) \sum x_{1i} \hat{u}_i = 0$ veya $\sum x_{1i} \hat{u}_i = 0$
$Cov(u, x_2) = 0$	$(1/n) \sum x_{2i} \hat{u}_i = 0$ veya $\sum x_{2i} \hat{u}_i = 0$

Normal denklemler olarak da ifade edilen bu denklemler aynı zamanda EKK metodu kullanılarak da sağlanabilir.

En Küçük Kareler Metodu

Bu metot, popülasyon parametreleri α, β_1, β_2'nin örnek temsilcileri olan α', β'_1, β'_2 parametrelerinin tahmin edicilerini aşağıdaki denklemi minimum yaparak seçer.

$$Q = \sum (y_i - \alpha' - \beta'_1 x_{1i} - \beta'_2 x_{2i})^2$$

Q'nun α', β'_1 ve β'_2'ye göre türevi alınıp sıfıra eşitlendiğinde aşağıdaki sonuçlar alınır:

$$\partial Q / \partial \alpha' = 0 \Rightarrow \sum 2 (y_i - \alpha' - \beta'_1 x_{1i} - \beta'_2 x_{2i}) (-1) = 0 \qquad \text{(i)}$$

$$\partial Q / \partial \beta'_1 = 0 \Rightarrow \sum 2 (yi - \alpha' - \beta'_1 x_{1i} - \beta'_2 x_{2i}) (-x_{1i}) = 0 \qquad \text{(ii)}$$

$$\partial Q / \partial \beta'_2 = 0 \Rightarrow \sum 2 (y_i - \alpha' - \beta'_1 x_{1i} - \beta'_2 x_{2i}) (-x_{2i}) = 0 \qquad \text{(iii)}$$

Önceden de belirtildiği gibi bu üç denklem normal denklemler olarak bilinmektedir. Bu denklemlerden birincisi (i) aşağıdaki gibi sadeleştirilebilir:

$$\sum y_i = n \alpha' + \beta'_1 \sum x_{1i} + \beta'_2 \sum x_{2i} \quad \text{veya}$$

$$y_{ort} = \alpha' + \beta'_1 x_{1ort} + \beta'_2 x_{2ort}$$

burada, $y_{ort} = 1/n\sum y_i$, $x_{1ort} = 1/n\sum x_{1i}$, $X_{2ort} = 1/n\sum x_{2i}$ olduğundan yukarıdaki ikinci (ii) eşitlik aşağıdaki gibi yazılabilir.

$$\sum x_{1i} y_i = \alpha' \sum x_{1i} + \beta'_1 \sum x_{1i}^2 + \beta'_2 \sum x_{1i} x_{2i}$$

Yukarıdaki denklemden α elde edilir ve burada yerine ikame edilirse, aşağıdaki denklem elde edilir

$$\sum x_{1i} y_i = n x_{1ort} (y_{ort} - \beta_1 x_{1ort} - \beta_2 x_{2ort}) + \beta'_1 \sum x_{1i2} + \beta'_2 \sum x_{1i} x_{2i}$$

Aşağıdaki notasyonlarla bu denklem sadeleştirilebilir.

$$S_{11} = \sum x_{1i}^2 - n x_{1ort}^2 \qquad S_{1y} = \sum x_{1i} y_i - n x_{1ort} y_{ort}$$

$$S_{12} = \sum x_{1i} x_{2i} - n x_{1ort} x_{2ort} \qquad S_{2y} = \sum x_{2i} y_i - n x_{2ort} y_{ort}$$

$$S_{22} = \sum x_{2i}^2 - n x_{2ort}^2 \qquad S_{yy} = \sum y_i^2 - n y_{ort}^2$$

Böylece yukarıdaki denklem aşağıdaki gibi yazılabilir:

$$S_{1y} = \beta'_1 S_{11} + \beta'_2 S_{12},$$

Yine aynı yolla üçüncü denklemde (iii) aşağıdaki gibi yazılabilir:

$$S_{2y} = \beta'_1 S_{12} + \beta'_2 S_{22},$$

şimdi β'_1 ve β'_2'yi elde etmek için bu iki denklem aşağıdaki gibi çözülebilir.

$$\beta'_1 = S_{22} S_{1y} - S_{12} S_{2y} \; / \; \Delta \qquad \beta_2 = S_{11} S_{2Y} - S_{12} S_{1Y} \; / \; \Delta$$

Burada $\Delta = S_{11} S_{22} - S_{12}^2$'dir. β'_1 ve β'_2'yi elde ettikten sonra α' yukarıdaki α ile ilgili denklem kullanılarak hesap edilebilir.

$$\alpha' = y_{ort} - \beta'_1 x_{1ort} - \beta'_2 x_{2ort}$$

Özetlenirse, hesaplama işlemleri aşağıdaki gibi sıralanabilir.

1- Bütün ortalamaların (y_{ort}, x_{1ort}, x_{2ort}) hesap edilmesi

2- Bütün kareler ve çarpımlar toplamının ($\sum x_{1i}^2$, $\sum x_{2i}^2$, $\sum x_{1i}$, $\sum x_{2i}$ vs.) hesaplanması

3- S_{11}, S_{12}, S_{22},S_{1y}, S_{2y} ve S_{yy}'nin hesaplanması

4- β'_1 ve β'_2'nin tahmin edilmesi

5- α''nın tahmin edilmesi

Basit regresyondakine benzer bir yolla diğer hesaplamalar da yapılabilir:

$$RSS = S_{yy} - \beta'_1 S_{1y} - \beta'_2 S_{2y}$$

$$ESS = \beta'_1 S_{1y} + \beta'_2 S_{2y}$$

$$R^2_{y.12} = (\beta'_1 S_{1y} + \beta'_2 S_{2y}) / S_{yy}$$

Üç açıklayıcı değişkenin olması durumundaki tahmin yöntemi de buna çok benzerdir. Bu üç açıklayıcı değişken için normal denklemler aşağıdaki gibidir.

$$\alpha' = y_{ort} - \beta'_1 x_{1ort} - \beta'_2 x_{2ort} - \beta'_3 x_{3ort}$$

$$S_{1y} = \beta'_1 S_{11} + \beta'_2 S_{12} + \beta'_3 S_{13}$$

$$S_{2y} = \beta'_1 S_{21} + \beta'_2 S_{22} + \beta'_3 S_{23}$$

$$S_{3y} = \beta'_1 S_{31} + \beta'_2 S_{32} + \beta'_3 S_{33}$$

Bu denklemler peş peşe eleme yapılarak çözülebilir, fakat bu çok zaman alır. Veriler yüklendiğinde bu tip hesapları çok kısa zamanda çözecek bilgisayar programları mevcuttur. Bu nedenle bu eşitliklerin çözülmesinden ziyade bu katsayıların yorumu üzerinde durulacaktır. Üç değişken olması durumunda;

$$RSS = S_{yy} - \beta'_1 S_{1y} - \beta'_2 S_{2y} - \beta'_3 S_{3y} \quad \text{ve}$$

$$R^2_{y.123} = (\beta'_1 S_{1y} + \beta'_2 S_{2y} + \beta'_3 S_{3y}) / S_{yy}\text{'dir.}$$

Dikkat edilirse RSS'nin her zaman $S_{yy}(1 - R^2)$'ye eşit olduğu görülür.

Uygulamalı Örnek 5.1

Çoklu regresyonu çözmek için bilgisayar programları mevcuttur ve bu yüzden burada bütün ayrıntıları tek tek ele almak gereksizdir. Fakat, bu bilgisayar programlarının hangi hesapları yaptığını bilmek gerçekten eğitici olmaktadır. Aşağıdaki örnek bu hesaplamaları göstermektedir. Yine aşağıdaki çizelgede bir firmadan şansa bağlı olarak seçilmiş beş kişinin aldıkları ücret, liseden sonraki eğitim yılları ve firmadaki tecrübelerine ait rakamlar verilmiştir.

y	x_1	x_2	$y-y_{ort}$	x_1-x_{1ort}	x_2-x_{2ort}
30	4	10	0	-1	0
20	3	8	-10	-2	-2
36	6	11	6	1	1
24	4	9	-6	-1	-1
40	8	12	10	3	2

Burada;

y : Günlük ücret (TL)

x_1 : Liseden sonra okunan yıl sayısı

x_2 : Firmada çalışılan yıllar

Ortalamalar $y_{ort} = 30$, $x_{1ort} = 5$, $x_{2ort} = 10$ olarak hesap edilir. Ortalamadan sapmaların karelerinin toplamları $S_{11} = 16$, $S_{12} = 12$, $S_{22} = 10$, $S_{1y} = 62$, $S_{2y} = 52$, $S_{yy} = 272$ olur.

Normal denklemler:

$S_{11}\,\beta'_1 + S_{12}\,\beta'_2 = S_{1y} \Rightarrow 16\,\beta'_1 + 12\,\beta'_2 = 62$

$S_{21}\,\beta'_1 + S_{22}\,\beta'_2 = S_{2y} \Rightarrow 12\,\beta'_1 + 10\,\beta'_2 = 52$ olarak hesap edilir.

Bu denklemler çözülürse, $\beta'_1 = -0.25$, $\beta'_2 = 5{,}5$ bulunur. Bu değerler α eşitliğinde yerine koyulursa, $\alpha' = y_{ort} - \beta'_1 x_{1ort} - \beta'_2 x_{2\,ort} = 30 - (-1{,}25) - 55 = -23{,}75$ elde edilir.

$R^2 = (\beta'_1 S_{1y} + \beta'_2 S_{2y}) / S_{yy} = 271{,}5 / 272 = 0{,}998$

Böylece regresyon denklemi aşağıdaki gibi belirlenebilir.

$y = -23{,}75 - 0{,}25\,x_1 + 5{,}5x_2 \qquad R^2 = 0{,}998$

Bu denklem, firmadaki deneyim yıllarının eğitim yıllarından daha önemli olduğunu ifade etmektedir. Denklem, eğitim sabit tutulduğunda, firmadaki 1 yıllık deneyim günlük ücreti 5,5 TL artırmakta olduğunu belirtmektedir. Yani biraz daha açılırsa, aynı eğitim seviyesindeki iki kişiden birinin firmadaki tecrübesi ötekinden bir yıl fazla ise, onun günlük ücreti 5,5 TL daha fazladır. Aynı şekilde iki kişi aynı tecrübeye sahip ise ve bunlardan birinin eğitimi diğerinden bir yıl fazla ise 0,25 TL az günlük ücret alır, çünkü katsayı negatiftir.

Sabit katsayı olan -23,75 deneyimin ve liseden sonraki eğitimin sıfır olması durumunda alınan günlük ücreti ifade eder. Ücretin negatif olması mümkün değilken burada x_1 ve x_2'nin sıfır olması durumunda ücret negatif olmaktadır. Bu durum modelde yanlış bir şeylerin olduğunu düşündürmektedir.

Buradan çıkarılacak sonuç, eldeki örneğin, firmada çalışan bütün insanları temsil etmediğidir. Örnekteki veriler bir alt gruptan alındığından, bu örnekteki insanların firmadaki tecrübesi 8 ile 12 yıl arasında değişmektedir. Dolayısıyla bu örnek aralığının

çok uzağından rakamlar kullanarak doğru kestirme yapmak pek mümkün olmaz. Yani yeni işe başlamış bir kişinin gelirini bu eşitliği kullanarak bulmak mümkün değildir.

Tecrübe ve eğitim değişkenlerinin etkisi basit regresyon olarak ayrı ayrı tahmin edildiğinde aşağıdaki ilgi çekici sonuç ortaya çıkmaktadır.

$y = 10{,}625 + 3{,}875\ x_1 \qquad R^2 = 0{,}883$

$y = -22{,}0 + 5{,}2\ x_2 \qquad R^2 = 0{,}994$

Basit regresyon denklemi, eğitim yıllarındaki bir yıllık artışın günlük ücrette 3,875 TL'lik gelir sağladığı görülmektedir. Fakat çoklu regresyonda tecrübe model içine alındığında, eğitim yıllarının artışı ücret artışına sebep olmamaktadır. Bu yüzden, tecrübeyi dikkate almayışımız bize, eğitim yıllarının ücret üzerine olan etkisi hakkında sapmalı sonuçlar vermektedir.

5.3. Çoklu Regresyonda İstatistiksel Yorum

Burada önceki örnekte kullanılan iki değişkenli modelin sonuçları dikkate alınacaktır. Yapılan diğer varsayımlarla birlikte u_i'nin normal dağılım gösterdiği dikkate alınırsa, $u_i \sim IN(0, \sigma^2)$ ima edilir ve aşağıdaki sonuçlar ortaya çıkar.

1. α', β'_1 ve β'_2, ortalamaları α, β_1 ve β_2 olan normal dağılıma sahiptirler.
2. x_1 ve x_2 arasında korelasyonu r_{12} olarak ifade edilirse,

   $\text{Var}(\beta'_1) = \sigma^2 / S_{11}(1 - r^2_{12})$

   $\text{Var}(\beta'_2) = \sigma^2 / S_{22}(1 - r^2_{12})$

   $\text{Cov}(\beta'_1, \beta'_2) = -\sigma^2 r^2_{12} / S_{12}(1 - r^2_{12})$

   $\text{Var}(\alpha') = \sigma^2 / n + x_1{}^2_{ort}\ \text{var}(\beta'_1) + 2x_{1ort}x_{2ort}\ \text{Cov}(\beta'_1, \beta'_2) + x_2{}^2_{ort}\ \text{Var}(\beta'_2)$

   $\text{Cov}(\alpha', \beta'_1) = -[\, x_{1ort}\text{var}(\beta'_1) + x_{2ort}\ \text{cov}(\beta'_1, \beta'_2)\,]$

   $\text{Cov}(\alpha', \beta'_2) = -[\, x_{1ort}\text{cov}(\beta'_1, \beta'_2) + x_{2ort}\ \text{var}(\beta'_2)\,]$ elde edilir.

Yorumlar

1. Diğer şartlar aynı kaldığında r_{12}'nin değeri yükseldikçe β'_1 ve β'_2'nin varyansları yüksek olmaktadır. Eğer r_{12} çok yüksek ise β'_1 ve β'_2, çok kesin bir şekilde tahmin edilmesi mümkün olmaz.
2. $S_{11}(1-r^2_{12})$ x_1'in x_2 üzerine regresyonunun hata kareler toplamını tanımlamaktadır. Aynı şekilde $S_{22}(1-r^2_{12})$ x_2'nin x_1 üzerine regresyonunun hata kareler toplamıdır. Burada, basit regresyondaki $\text{Var}(\beta_1) = \sigma^2 / RSS_1$'in aslında bu iki değişkenli regresyonda diğer değişkenin etkisi çıkarıldıktan sonraki varyans olduğu ortaya çıkmaktadır. Bu durum çok sayıdaki açıklayıcı değişken için de aynıdır.

3. Eğer RSS hata kareler toplamı ise RSS / σ^2, (n-3) serbestlik derecesiyle χ^2 dağılımına sahiptir. Bu sonuç σ^2 hakkındaki ifadelerin güven aralığını belirlemek için kullanılabilir.

4. Eğer σ^2 = RSS / (n-3) ve E (s^2) = σ^2 ise s^2, σ^2 için sapmasız bir tahmincidir.

5. Eğer s^2 ikinci ifadedeki σ^2 yerine ikame edilirse tahmin edilmiş varyanslar ve kovaryanslar elde edilir. Tahmin edilen varyansın karekökü standart hata olarak adlandırılır ve SE ile ifade edilir. Bu durumda (α' - α) / SE (α'), (β'_1 - β_1) / SE (β'_1) ve (β'_2 - β_2) / SE (β'_2)'nin her biri (n-3) serbestlik derecesiyle *t*-dağılımına sahiptir.

Burada 3'den 5'e kadar ki sonuçlar basit regresyonda karşılığa sahiptirler. Çoklu regresyonda ilave olarak parametreler için güven bölgesi ve ortak testleri mevcuttur. Bu husus aşağıdaki şekilde ifade edilir.

6. $F = \frac{1}{2} s^2 [S_{11} (\beta'_1 - \beta_1)^2 + 2S_{12} (\beta'_1 - \beta_1) (\beta'_2 - \beta_2) + S_{22} (\beta'_2 - \beta_2)^2] 2$ ve (n-3) serbestlik derecesiyle bir F-dağılımına sahiptir. Bu sonuç β_1 ve β_2 için ortak bir güven bölgesi oluşturmak ve β_1 ve β_2 yi birlikte test etmek için kullanılır.

Buradaki 1'den 6'ya kadar olan sonuçlar k sayıdaki değişken için de ifade edilebilir. Bu ifadenin nasıl olacağını uygulama örneğinden sonra verelim.

Uygulamalı Örnek 5.2

Bir üretim fonksiyonu aşağıdaki gibi belirlenmiş olsun.

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \qquad u_i \sim IN (0, \sigma^2)$$

Burada:

y: Üretim (log y)

x_1: İşgücü girdisi (log x_1)

x_2: Sermaye girdisi (log x_2)'dir.

Örnek büyüklüğü n = 23 olan verilerle ilgili istatistik özetler aşağıdaki gibi verilebilir.

x_{1ort} =10	S_{11} = 12	S_{1y} = 10
x_{2ort} =5	S_{12} = 8	S_{2y} = 8
y_{ort} = 12	S_{22} = 12	S_{yy} = 10

Burada yapılacak işler sırasıyla şunlardır.

a). α', β'_1, β'_2, s^2'nin tahminleri yapılır ve standart sapmaları hesaplanarak ve regresyon denklemi yazılır.

b). α, β_1, β_2 ve σ^2 değerleri için % 95 güven aralığı bulunur ve % 5 önem seviyesinde $\beta_1 = 0$ ve $\beta_2 = 0$ hipotez testleri ayrı ayrı test edilir.

c). β_1 ve β_2 için %95 güven bölgesi bulunur ve şekille gösterilir.

d). $\beta_1 = 1$, $\beta_2 = 0$ hipotezleri % 5 önem seviyesinde F testi yapılır.

Çözüm

a)- Bu istatistik özetlerden normal denklemler aşağıdaki gibi hesaplanabilir.

$$12\ \beta'_1 + 8\ \beta'_2 = 10$$

$$8\ \beta'_1 + 12\ \beta'_2 = 8$$

Bu denklemler, $\beta'_1 = 0,7$ ve $\beta'_2 = 0,2$ ve $\alpha' = 4$ değerlerini verir.

$$R^2_{y.12} = (\beta'_1 S_{1y} + \beta'_2 S_{2y}) / S_{yy} = (0,7 * 10 + 0,2 * 8) / 10 = 0,86$$

$$RSS = S_{yy}(1-R^2) = 10\ (1 - 0,86) = 1,4$$

$$s^2 = RSS / n-3 = 1,4 / 20 = 0,07$$

$$r^2_{12} = s^2_{12} / S_{11}S_{22} = 64 / 144$$

$$S_{11}(1-r^2_{12}) = 12\ (80 /144) = 80 / 12 = 20 / 3$$

Buradan aşağıdaki sonuçlar da elde edilir.

$$Var\ (\beta'_1) = \sigma^2 / (S_{11}\ (1-r^2_{12})) = (3 / 20)\ \sigma^2$$

$$Var\ (\beta'_2) = \sigma^2 / (S_{22}\ (1-r^2_{12})) = (3 / 20)\ \sigma^2 \text{ ve}$$

$$Cov\ (\beta'_1, \beta'_2) = (-\sigma^2\ r^2_{12}) / (S_{12}\ (1-r^2_{12})) = (-\sigma^2\ (64/144)) / (8\ (80) / 144) = \sigma^2 / 10$$

Buradan da $x_{1ort} = 0$ ve $x_{2ort} = 5$ olduğunda aşağıdaki sonuçlar hesaplanabilir.

$$V\ (\alpha') = \sigma^2\ [\ (1/23) + (10)^2\ (3/20) - (2 * 10 * 5) / 10 + (5^2 * 3) / 20\ ] = 8,7935\ \sigma^2$$

0.07 olan σ^2'yi yerine ikame edip ve karekökünü alarak aşağıdaki sonuçlar elde edilir.

$$SE\ (\beta'_1) = SE\ (\beta'_2) = \sqrt{(0,21 / 20)} = 0,102$$
$$SE\ (\alpha') = 0,78$$

Böylece regresyon denklemi aşağıdaki gibi ifade edilebilir

$$y = 4,0 + 0,7\ x_1 + 0,2\ x_2 \qquad R^2 = 0,86$$
$$(0.78)\ \ (0.102)\ \ (0.102)$$

Parantez içindeki değerler o katsayıların standart sapmalarını vermektedir.

b)- 20 serbestlik derecesindeki *t*-dağılımını kullanarak α, β_1 ve β_2 için güven aralığı tespit edilebilir.

$$\alpha' \pm 2,086\ SE\ (\alpha') = 4,0 \pm 1,63 = (2,37 - 5,63)$$

$$\beta'_1 \pm 2,086\ SE\ (\beta'_1) = 0,7 \pm 0,21 = (0,49 - 0,91)$$

$$\beta'_2 \pm 2,086\ SE\ (\beta'_2) = 0,2 \pm 0,21 = (-0,01 - 0,41)$$

Burada % 5 önem seviyesinde $\beta_1 = 1,0$ hipotezi reddedilecektir. Çünkü 1,0 değeri, β_1 için olan % 95 güven aralığının dışındadır. Burada $\beta_2 = 0$ hipotezi reddedilemeyecek,

çünkü 0 değeri β_2 için belirlenen % 95 güven aralığının içine girmektedir. Eğer χ^2 dağılımında 20 serbestlik derecesi kullanılırsa,

$$P\,(9{,}59 < 20\,s^2 / \sigma^2 < 34{,}2) = 0{,}95 \text{ veya}$$

$$P\,(20\,s^2 / 34{,}2 < \sigma^2 < 20\,s^2 / 9{,}95) = 0{,}95 \text{ elde edilir.}$$

Böylece σ^2 için % 95 güven aralığı [(20 * 0,07) / 34,2 - (20 * 0,07) / 9,59] = (0,041 – 0,146) şeklinde elde edilir.

c)- Önem seviyesinin %5, serbestlik derecesinin 2 ve 20 olması durumunda F-dağılımı tablo değeri 3,49'dur. Altıncı yorum kullanılarak β_1 ve β_2 için % 95 güven bölgesi aşağıdaki gibi yazılabilir.

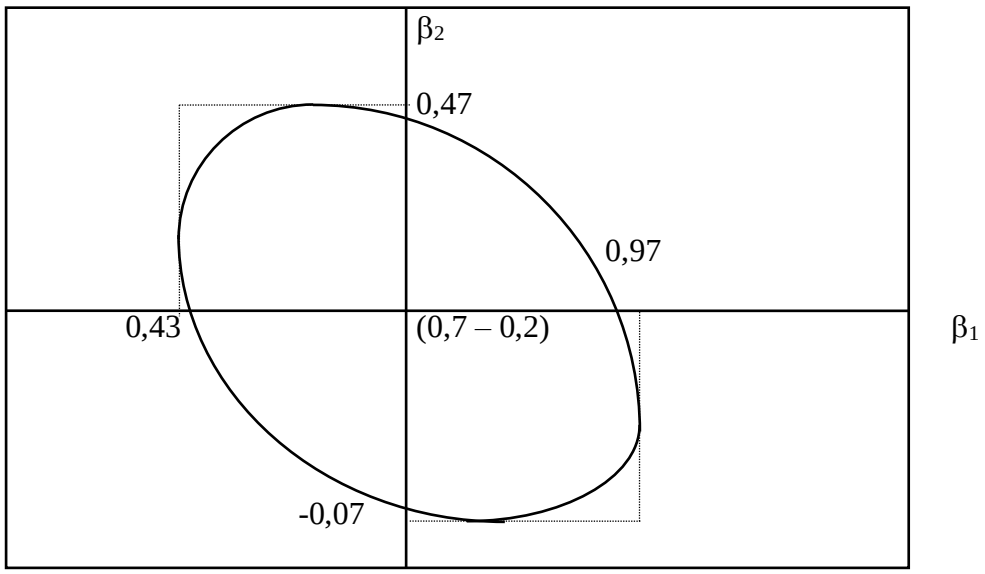
$$[S_{11}(\beta'_1 - \beta_1)^2 + 2S_{12}(\beta'_1 - \beta_1)(\beta'_2 - \beta_2) + S_{22}(\beta'_2 - \beta_2)^2] \leq (3{,}49 * 2 * s^2) \text{ veya}$$

$$[12(0{,}7-\beta_1)^2 + 16(0{,}7- \beta_1)(0{,}2-\beta_2) + 12(0{,}2-\beta_2)^2] \leq (3{,}49 * 2 * 0{,}07)$$

Her iki taraf 12'ye bölünüp $\beta'_1 - \beta_1$'den $\beta_1 - \beta'_1$'e ve $\beta'_2 - \beta_2$'den $\beta_2 - \beta'_2$'ye değiştirilerek aşağıdaki sonuç elde edilir.

$$(\beta_1 - 0{,}7)^2 + 3/4\,(\beta_1 - 0{,}7)(\beta_2 - 0{,}2) + (\beta_2 - 0{,}2)^2 \leq 0{,}041$$

Bu bir (0,7 ve 0,2) merkezli elipstir. Bu elips Şekil 5.1'de gösterilmiştir.



Şekil 5.1. Çoklu regresyonda β_1 ve β_2 için güven elipsi

Burada dikkat edilecek iki önemli husus vardır. Birincisi; şekilde yani örnekte olduğu gibi eğer cov (β_1, β_2) < 0 ise elips sola doğru yatmaktadır. Eğer cov (β_1, β_2) > 0 ise elips bu sefer sağa doğru meyillidir. İki bağımsız değişken durumunda bu S_{12}'nın işaretine bağlı olacaktır. Dikkat alınacak ikinci husus elipsten elde edilen β_1 ve β_2'nin sınırları β_1 ve β_2 için önceden ayrı ayrı hesap edildiğinde % 95 güven sınırlarından farklı olacaktır.

Limitler β_2 için (-0,07 ve 0,47) ve β_1 için (0,43 ve 0,97) olacaktır. Bunun sebebi ise bu durumda bizim ortak güven sınırlarına sahip olmamızdan kaynaklanmaktadır. Eğer β'_1 ve β'_2 bağımsız ise ortak aralık için güven katsayısı, β_1 ve $\beta 2$'nin ayrı ayrı hesap edilen güven katsayılarının basitçe çarpımı olacaktır. Aksi halde ortak aralık için güven katsayısı farklı olacaktır.

Ayrı aralıklar, ortak aralıklar, ayrı testler ve ortak testler arasındaki bu farklılık, çoklu regresyon modelindeki istatistiksel yorumdan bahsedildiğinde akılda tutulmalıdır. Bundan ileride daha detaylı bahsedilecektir.

d)- Burada % 5 önem seviyesinde $\beta_1 = 1$ ve $\beta_2 = 0$ hipotezlerini test etmek için, bu noktanın % 95 güven elips içinde olup olmadığı kontrol edilebilir. Fakat güven alanı elipsi her zaman çizilemediğinden daha önce yorum 6'da tartışılan F değeri uygulanacaktır.

Bu durumda,

$$F = (\tfrac{1}{2} s^2) [S_{11} (\beta'_1 - \beta_1)^2 + 2 S_{12} (\beta'_1 - \beta_1) (\beta'_2 - \beta_2) + S_{22} (\beta'_2 - \beta_2)^2]$$

2 ve n-3 serbestlik derecesi ile F dağılımına sahiptir. Buradan, gözlenen F_0 değeri

$$F_o = (1/ (2 * 0,07)) [12 (0,7-1,0)^2 + 2 (8) (0,7 - 1,0) (0,2 - 0) + 12 (0,2 - 0)^2] = 4,3$$

olur. Buradan % 5 güven seviyesinde 2 ve 20 serbestlik derecesiyle F tablosundaki değer 3,49'dur. Bu durumda $F_0 > 3,49$ olduğundan % 5 güven seviyesindeki hipotez reddedilir.

K Sayıdaki Açıklayıcı Değişken Durumunda Formüller

İki açıklayıcı değişken için kesin ifadeler kullanarak basit ve çoklu regresyon arasındaki fark gösterilmeye çalışılmıştır. Genel durum için ifadeler matriks notasyonu halinde düzgün bir şekilde ifade edilebilir. Eğer çoklu regresyon denklemi k sayıdaki değişkene sahip ise aşağıdaki gibi yazılabilir.

$$y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i$$

Böylece daha önceki sonuçlarda aşağıdaki değişiklikler yapılabilir. Yorum 2, k sayıdaki bağımsız değişken durumunda aşağıdaki gibi ifade edilebilir.

$$Var (\beta_i) = \sigma^2 / RSS_i \quad i = 1, 2, ..., k \quad \text{için}$$

Burada RSS_iX_i'nin diğer k-1 x değişkenlerinin tümü üzerine olan regresyonunun hata kareler toplamıdır.

Hâlbuki gerçek hayatta bu kadar değişkenli regresyon analizi yapılmaz. Bu formül sadece basit ve çoklu regresyon arasındaki farkı göstermek için kullanılmaktadır.

Yorum 3-5 için ise serbestlik derecesini n-3 den n-k-1'e çevrilebilir. Basit regresyonda k = 1 ve serbestlik derecesi (sd) = n-2, iki değişkenli modelde de k = 2 ve sd = n-3'dür. Bütün durumlarda α'nın tahmini için s.d. 1'dir. Altıncı yorum için;

$$F = (1 / k\, s^2) * [\ \sum^{k}_{i=1} \sum^{k}_{j=1} S_{ij}\ (\beta'_i - \beta_i)\ (\beta'_j - \beta_j)\ ]$$

k ve (n-k-1) serbestlik derecesiyle F dağılımına sahiptir.

Burada çoklu regresyon modelinin kullanılacağı birkaç örnek verilecektir. Bu örneklerin tahminlerini yapmak için farklı veri setlerinden yararlanılacaktır.

Uygulamalı Örnek 5.3

Eldeki veri seti bir şehir merkezinin yakınındaki kırsal arazilerin satış fiyatlarıyla ilgilidir. Değişkenler; Pt: arazi fiyatı, WL: arazinin ağaçlık oranı, DA: arazinin hava alanından uzaklığı, D75: arazinin anayola olan uzaklığı, A: arazi parselinin büyüklüğü, MO: arazinin belli bir dönemde kaçıncı ayda satıldığıdır. Zamanla arazi kıymetleri değer kazandığından, arazinin satıldığı ay olan MO en önemli değişken olarak dikkate alınabilir. Ağaçlı olan, havaalanına yakın olan ve ana yola yakın olan arazilerin fiyatlarının yüksek olması da ayrıca beklenir. Son olarak parsel alanındaki değişmeye karşılık, arazi fiyatlarında daha az değişme olduğundan, parsel alanındaki büyüme fiyatta düşmeye neden olabilir. Dolayısıyla A'nın yüksek değişkenliği sorununu çözmek için logaritması alınmıştır. Tahmin edilen denklem aşağıdaki gibidir Parantez içindeki değerler standart hatlardır.

$$\underset{}{\text{Log } P_t} = \underset{}{9{,}213} + \underset{(0.204)}{0{,}143}\ \text{WL} - \underset{(0.099)}{0{,}33}\ \text{DA} - \underset{(0{,}011)}{0{,}0057}\ \text{D75} - \underset{(0{,}0142)}{0{,}203}\ \text{LogA} + \underset{(0{,}0345)}{0{,}0141}\ \text{MO}$$
$$\underset{(0{,}0039)}{}$$

$R^2 = 0{,}559 \quad n = 67$

Tahmin edilen bütün katsayılar % 5 önem seviyesinde istatistikî olarak önemli olması yanında katsayılar da beklenen işarete sahiptirler.

Uygulamalı Örnek 5.4

Birinci örnekteki regresyon denkleminde değişkenlerin aralarındaki mantıksal ilişki açısından hangi işaretlere sahip olması gerektiği hususuna önem verilerek analiz yapılmıştır. Bazen işaret doğru olsa bile, katsayının büyüklüğü ekonomik manada anlamsız olmaktadır. Bu durumda kişi modelini veya verilerini gözden geçirmelidir. Bu örnekte katsayıların büyüklüğü üzerinde durulacaktır.

Gayrimenkullere uygulanan bir vergi değişikliğini halk oylayarak kabul ediyor. Bu yeni kanun gayrimenkul vergilerinde değişikliğe ve önemli bir düşüşe sebep olduğu ve dolayısıyla ev fiyatlarının yükselmesine neden olduğu belirtiliyor. Bu verginin ev fiyatları üzerine olan etkisi bu çalışmada tespit edilmeye çalışılıyor. Modelin ilgili piyasayı daha iyi temsil etmesi açısından gayrimenkul vergileri dışında ev fiyatlarını etkileyen diğer faktörler de dikkate alınmıştır. Bu yüzden evin özelliklerinden örneğin; m^2 olarak evin alanı, ömrü ve kalitesi modelde değişken olarak yer almıştır. Ayrıca evin bulunduğu bölgede yaşayanların ortalama geliri ve evin şehre uzaklığı gibi birtakım ekonomik faktörler de model içinde dikkate alınmıştır. Tahmin edilen denklem aşağıdaki gibidir Parantez içindeki rakamlar t değerleridir.

$$Y = 0{,}171 + \underset{(2.97)}{7{,}275}\ X_1 + \underset{(2.32)}{0{,}547}\ X_2 + \underset{(1.34)}{0{,}00073}\ X_3 + \underset{(3.26)}{0{,}0638}\ X_4 - \underset{(2.24)}{0{,}0043}\ X_5 + \underset{(1.80)}{0{,}857}\ X_6$$

$R^2 = 0{,}897, \quad n = 64$

Burada,

Y: Vergideki değişmeden sonraki ortalama ev fiyatları

X_1: Vergi değişimi sonrası ortalama ev başına gayrimenkul vergisindeki düşme

X_2: Ortalama olarak evin alanı

X_3: O bölgede yaşayanların ortalama aile geliri

X_4: Evin yaşı

X_5: Evin şehir merkezine olan uzaklık

X_6: Evin kalitesi

Bütün değişkenlerin katsayıları beklenen işarete sahiplerdir. Burada X_1'in katsayısı, yani gayrimenkul vergisindeki 1 TL değişme evin değerini 7 TL değiştirmektedir. Burada soru; bu büyüklüğün gerçekten anlamlı olup olmadığıdır. Gayrimenkullere yapılan bu vergilendirmenin gelecek yıllarda da aynı seviyede kalması beklendiği kabul edilirse, yıllık 1 TL'lik gelirin bugünkü değeri 1/r olur. Burada r aynı seviyede kalacağı beklenen faiz oranıdır. Eğer bu faiz oranı % 14.29 ise, 1 TL'lik gelirin bugünkü değeri 7 TL'ye eşit olur. Araştırmanın yapıldığı sırada faiz oranının bu seviyede olduğu belirtilmektedir. Bu denklemle tahmin edilen kapitalizasyon değeri X_1'in katsayısı 7 civarındadır ki bu değer % 12-15'lik faiz oranı ile beklenebilecek bir büyüklüktür. Sonuç olarak elde edilen katsayının anlamlı olduğu ortaya çıkmaktadır.

Uygulamalı Örnek 5.5

ABD'deki benzin tüketiminin 1947–1969 yılları arasındaki durumuna ait veriler bu örnekte kullanılmıştır. Burada kullanılan veriler; K: araba başına yıllık gidilen yol (mil), C: araba sayısı (milyon), M: bir galon benzinle alınan yol uzunluğu (mil), Pg: 1953 yılı baz alınan benzin fiyat indeksi, POP: nüfus (milyon), L: işgücü (milyon) ve Y: 1958 fiyatlarıyla harcanabilir gelirdir (ABD doları). Bu verilerden kişi başına galon olarak benzin tüketimi aşağıdaki formülle belirlenebilir.

$$G = (K * C) / (M * POP)$$

Eğer G'nin P_g ve Y üzerine regresyon analizi yapılırsa, aşağıdaki sonuçlar elde edilir. Parantez içindeki rakamlar t değerleridir.

$$G = \underset{(-1,14)}{-117{,}70} + \underset{(0,44)}{0{,}373}\ P_g + \underset{(12,02)}{0{,}156}\ Y \qquad R^2 = 0{,}953$$

P_g'nin katsayısının önemli olmaması yanında yanlış işarete de sahiptir. Sadece kişi başına gelir, benzin tüketiminin en önemli belirleyicisi olarak görülmektedir. Denklem logaritmik formda yazılırsa aşağıdaki sonuçlar elde edilir;

$$\underset{}{\text{Log } G = -8{,}72 + 0{,}535 \text{ Log } P_g + 1{,}541 \text{ Log } Y} \qquad R^2 = 0{,}940$$
$$\underset{(-3{,}00)\quad (1{,}23)\qquad\qquad (11{,}21)}{}$$

Burada da yine P_g yanlış işarete sahiptir. G'yi belirlerken nüfus (POP) yerine işgücü kullanılırsa (G = K * C / M * L) belki bu yanlış düzeltilebilir. Nüfustan ziyade işgücü mil olarak araba başına gidilen yol ile daha yakından alakalıdır düşüncesi böyle bir değişikliğin gerekçesi olarak kabul edilebilir. Bu durumda elde edilen sonuçlar aşağıdaki gibidir.

$$\underset{}{G = -248{,}44 + 0{,}258\ P_g + 0{,}406\ Y} \qquad R^2 = 0{,}925$$
$$\underset{(-2{,}18)\quad (0{,}92)\quad (9{,}15)}{}$$

P_g hala yanlış işarete sahiptir. Modelin logaritmik formlu sonuçları aşağıdaki gibidir.

$$\underset{}{\text{Log } G = -8{,}53 + 0{,}541 \text{ Log } P_g + 1{,}636 \text{ Log } Y} \qquad R^2 = 0{,}907$$
$$\underset{(-2{,}18)\quad (0{,}92)\qquad\qquad (8{,}82)}{}$$

Bu durumda yine doğru işaretler elde edilmediğinden belli bir aralıktaki veriler kullanılabilir. Bunun için 1947-1960 arasındaki veriler kullanıldığında modelimiz aşağıdaki hali almaktadır.

$$\underset{}{G = -\ 439{,}4 - 2{,}241\ P_g +\ 0{,}662\ Y} \qquad R^2 = 0{,}979$$
$$\underset{(-2{,}92)\quad (-1{,}64)\qquad (22{,}25)}{}$$

Bunun da logaritmik formu tahmin edilirse aşağıdaki gibi bir modele sahip olunur.

$$\underset{}{\text{Log } G = -10{,}447 - 0{,}387 \text{ Log } P_g + 2{,}472 \text{ Log } Y} \qquad R^2 = 0{,}974$$
$$\underset{(-5{,}67)\quad (-1{,}16)\qquad\qquad (20{,}04)}{}$$

Özet olarak, 1947-1960 döneminde talebin fiyat elastikiyeti negatif, talebin gelir elastikiyeti pozitif ve önemli derecede birim elastikiyetten büyüktür. Aynı regresyon denkleminin 1947-1969 arası için tahmin edilmesiyle elde edilen sonuç fiyata tepki göstermezken, gelir elastikiyeti yine 1'den çok büyüktür. Benzin tüketicilerinin 1960-1969 arasında fiyat değişimlerinden etkilenmediği ve bu nedenle modeli etkilediği görülmektedir.

Uygulamalı Örnek 5.6

Gıda tüketimi ile ilgili 1922-1941 yıllarına ait zaman serisi verileri kullanılarak gıda talebinin fiyat elastikiyeti ve gelir elastikiyeti, bu örnekte tahmin edilmeye çalışılacaktır. Q_D'nin P_D ve Y üzerine regresyonu aşağıdaki sonuçları vermektedir. Parantez içindeki t değerleridir.

$$\underset{}{Q_D = 92{,}05 - 0{,}1421\ P_D + 0{,}236\ Y} \qquad R^2 = 0{,}7813$$
$$\underset{(15{,}76)\quad (-2{,}13)\qquad (7{,}56)}{}$$

Tüm katsayılar doğru işaretlere sahip olmalarına rağmen birisi çıksa ve derse ki bütün değişkenler trende sahipler. Bu düşünce dikkate alınıp T, yani trend ilave bir değişken olarak modele koyulabilir. Trend değişkeninin modele koyulması durumda sonuçlarımız aşağıdaki gibi hepsi doğru işarete sahip olur.

$$Q_D = \underset{(18,47)}{105,1} - \underset{(-4.58)}{0,325}\, P_D + \underset{(9,85)}{0,315}Y - \underset{(3,67)}{0,224}T \qquad R^2 = 0,8812$$

Şimdi fiyatın katsayısı öncekine göre daha önemlidir. Ayrıca, Q_D'nin P_D üzerine basit regresyonu P_D için yanlış işaret vermektedir. Gelir olmadan fiyat yalnız başına modeli açıklamada yeterli olmamaktadır. Ayrıca R^2 değerinin daha yüksek olduğu, fakat genelde trend değişkeninin model üzerinde fazla etkili olmadığı görülmektedir.

Burada dört uygulama örneği verilmiştir. İlk iki örnek yatay-kesit (cross-sectional) verileri, diğer iki örnekte ise zaman serisi (time-series) verileri kullanılmıştır. Yatay kesit verileri ile ilgili tahminlerde doğru sonuçlar elde etmemize rağmen zaman serisi verilerinden anlamlı sonuçlar elde etme de problemler yaşadık. Bu çok yaygın olan bir durumdur. Burada problem serilerin çoğunun birlikte hareket etmesidir.

YAZILIM UYGULAMASI

```
**************************************************************************
   SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
   ** Copyright (C)  1999  by K.J. White - All Rights Reserved **
   FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
   AT: ATATURK UNIVERSITESI - ERZURUM
   FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
**************************************************************************
```

Uygulamalı Örnek 5.3, sayfa 81

```
**************************************************************************
|_SAMPLE 1 67
|_READ (MTABLE4.7) OBS PRICE WL DA D75 A MO
 UNIT 88 IS NOW ASSIGNED TO: MTABLE4.7
    7 VARIABLES AND       67 OBSERVATIONS STARTING AT OBS       1
|_GENR LNP=LOG(PRICE)
|_GENR LNA=LOG(A)
|_OLS LNP WL DA D75 LNA MO
        67 OBSERVATIONS     DEPENDENT VARIABLE= LNP
 ...NOTE..SAMPLE RANGE SET TO:      1,     67
  R-SQUARE =   0.5588     R-SQUARE ADJUSTED =   0.5226
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.13040
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.36111
 SUM OF SQUARED ERRORS-SSE=   7.9545
 MEAN OF DEPENDENT VARIABLE =   8.3930

VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      61 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
   WL       0.14841  0.9903E-01  1.499   0.139  0.188   0.1283     0.0056
   DA  -0.33451E-01 0.1098E-01 -3.046   0.003 -0.363  -0.4170    -0.0630
  D75  -0.57478E-02 0.1422E-01 -0.4042  0.687 -0.052  -0.0552    -0.0039
  LNA  -0.20286     0.3467E-01 -5.852   0.000 -0.600  -0.5015    -0.0764
   MO   0.14116E-01 0.3918E-02  3.603   0.001  0.419   0.3095     0.0400
CONSTANT 9.2126     0.2044      45.08   0.000  0.985   0.0000     1.0977
|_*Bu örnek, TEST ve CONFID komutlarını göstermek için kullanılacaktır.
|_*Ortak testler için kullanılacak TEST formu aşağıdaki gibidir.
|_TEST
|_TEST DA=0.5
```

```
|_TEST WL=0
|_END
 F STATISTIC =   1182.1334     WITH    2 AND   61 D.F.  P-VALUE= 0.00000
 WALD CHI-SQUARE STATISTIC =  2364.2668    WITH   2 D.F.  P-VALUE= 0.00000
 UPPER BOUND ON P-VALUE BY CHEBYCHEV INEQUALITY = 0.00085
|_*Açıkca görülmektedir ki 1182.13 F değeri, 2, 62 serbestlik derecesi ve
|_*%95 güven aralığının F kritik değeri olan 3.15'den büyük olduğundan Ho
|_*hipotezi reddedilmelidir. Şimdi bir güven elipsi çizelim.
|_CONFID DA WL
 USING 95% AND 90% CONFIDENCE INTERVALS
 CONFIDENCE INTERVALS BASED ON T-DISTRIBUTION WITH  61 D.F.
      - T CRITICAL VALUES =   1.994 AND   1.667
 NAME   LOWER 2.5%   LOWER 5%   COEFFICIENT   UPPER 5%   UPPER 2.5%   STD. ERROR
 DA  -0.5535E-01 -0.5176E-01 -0.33451E-01 -0.1515E-01 -0.1156E-01   0.011
 WL  -0.4906E-01 -0.1667E-01  0.14841      0.3135      0.3459       0.099
 CONFIDENCE REGION PLOT FOR DA       AND WL
  USING F DISTRIBUTION WITH 2 AND   61 D.F.     F-VALUE =   3.130
                   M=MULTIPLE POINT
  -0.13878E-16    |
  -0.42105E-02    |
  -0.84211E-02    |            M* *  M  ****
  -0.12632E-01    |      + MMM*             *MMM   +
  -0.16842E-01    |      MM*                    MMM
  -0.21053E-01    |    MM                         MM*
  -0.25263E-01    |   MM                            MM
  -0.29474E-01    |   M                              MM
  -0.33684E-01    |   M               *               M
  -0.37895E-01    |   M                               M
  -0.42105E-01    |   MM                              M
  -0.46316E-01    |    MM                            M
  -0.50526E-01    |      *MM                        MM*
  -0.54737E-01    |        *MM                    *MM
  -0.58947E-01    |      +   MMM**            **MMM +
  -0.63158E-01    |               ** *  M  * M
  -0.67368E-01    |
  -0.71579E-01    |
  -0.75789E-01    |
  -0.80000E-01    |
                   ____________________________________________

              -0.150     0.000     0.150     0.300     0.450
                                   WL
|_DELETE / ALL
ALL VARIABLES HAVE BEEN DELETED
|_STOP
*******************************************************************************
```

Uygulamalı Örnek 5.5, sayfa 82

```
*******************************************************************************
|_SAMPLE 1 23
|_READ (MTABLE4.8) YEAR K C M PG PT POP L Y
 UNIT 88 IS NOW ASSIGNED TO: MTABLE4.8
    9 VARIABLES AND       23 OBSERVATIONS STARTING AT OBS       1
|_GENR G=(K*C)/(M*POP)
|_OLS G PG Y
        23 OBSERVATIONS     DEPENDENT VARIABLE= G
```

```
 ...NOTE..SAMPLE RANGE SET TO:      1,     23
  R-SQUARE =   0.9531     R-SQUARE ADJUSTED =   0.9485
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   110.24
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   10.500
 SUM OF SQUARED ERRORS-SSE=   2204.9
 MEAN OF DEPENDENT VARIABLE =   216.88
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      20 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PG         0.37265   0.8372     0.4451   0.661  0.099  0.0373     0.1675
 Y          0.15557   0.1294E-01  12.02   0.000  0.937  1.0065     1.3752
 CONSTANT  -117.70     102.9     -1.144   0.266 -0.248  0.0000    -0.5427
|_GENR LNG=LOG(G)
|_GENR LNPG=LOG(PG)
|_GENR LNY=LOG(Y)
|_OLS LNG LNPG LNY / LOGLOG
        23 OBSERVATIONS     DEPENDENT VARIABLE= LNG
 ...NOTE..SAMPLE RANGE SET TO:      1,     23
  R-SQUARE =   0.9396     R-SQUARE ADJUSTED =   0.9336
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.31921E-02
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.56498E-01
 SUM OF SQUARED ERRORS-SSE=  0.63841E-01
 MEAN OF DEPENDENT VARIABLE =   5.3569
 LOG OF THE LIKELIHOOD FUNCTION(IF DEPVAR LOG) = -88.1450
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      20 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPG       0.53528    0.4364    1.227    0.234  0.264  0.1162     0.5353
 LNY        1.5410     0.1375    11.21    0.000  0.929  1.0617     1.5410
 CONSTANT -8.7246      2.907    -3.001    0.007 -0.557  0.0000    -8.7246

|_GENR NEWG=(K*C)/(M*L)
|_OLS NEWG PG Y
        23 OBSERVATIONS     DEPENDENT VARIABLE= NEWG
 ...NOTE..SAMPLE RANGE SET TO:      1,     23
  R-SQUARE =   0.9250     R-SQUARE ADJUSTED =   0.9175
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   1296.7
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   36.009
 SUM OF SQUARED ERRORS-SSE=   25934.
 MEAN OF DEPENDENT VARIABLE =   555.61

VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      20 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PG         0.25811    2.871   0.8989E-01 0.929  0.020   0.0095     0.0453
 Y          0.40627  0.4439E-01   9.153   0.000  0.898   0.9695     1.4019
 CONSTANT  -248.44     352.9    -0.7040   0.490 -0.156   0.0000    -0.4472

|_GENR LNNEWG=LOG(NEWG)
|_OLS LNNEWG LNPG LNY / LOGLOG
 REQUIRED MEMORY IS PAR=       5 CURRENT PAR=    1000
        23 OBSERVATIONS     DEPENDENT VARIABLE= LNNEWG
 ...NOTE..SAMPLE RANGE SET TO:      1,     23
  R-SQUARE =   0.9068     R-SQUARE ADJUSTED =   0.8975
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.58070E-02
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.76204E-01
 SUM OF SQUARED ERRORS-SSE=  0.11614
 MEAN OF DEPENDENT VARIABLE =   6.2940
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
```

```
  NAME    COEFFICIENT   ERROR      20 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPG       0.54063    0.5886    0.9184    0.369  0.201    0.1081     0.5406
 LNY        1.6360     0.1855    8.820     0.000  0.892    1.0383     1.6360
 CONSTANT -8.5291      3.921    -2.175     0.042 -0.437    0.0000    -8.5291
|_* Now omit the observations from 1961 to 1969.
|_SAMPLE 1 14
|_OLS NEWG PG Y
        14 OBSERVATIONS     DEPENDENT VARIABLE= NEWG
 ...NOTE..SAMPLE RANGE SET TO:      1,     14
  R-SQUARE =   0.9787     R-SQUARE ADJUSTED =   0.9748
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   183.28
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   13.538
 SUM OF SQUARED ERRORS-SSE=   2016.0
 MEAN OF DEPENDENT VARIABLE =   476.68
 LOG OF THE LIKELIHOOD FUNCTION = -54.6539
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      11 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PG         -2.2411    1.368    -1.639     0.130 -0.443  -0.0724     -0.4727
 Y          0.66248  0.2981E-01  22.22     0.000  0.989   0.9815      2.3945
 CONSTANT  -439.39     150.2    -2.925     0.014 -0.661   0.0000     -0.9218
|_OLS LNNEWG LNPG LNY / LOGLOG
        14 OBSERVATIONS     DEPENDENT VARIABLE= LNNEWG
 ...NOTE..SAMPLE RANGE SET TO:      1,     14
  R-SQUARE =   0.9737     R-SQUARE ADJUSTED =   0.9689
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.10843E-02
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.32928E-01
 SUM OF SQUARED ERRORS-SSE=  0.11927E-01
 MEAN OF DEPENDENT VARIABLE =   6.1511
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      11 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPG       -0.38738   0.3355    -1.155    0.273 -0.329  -0.0565    -0.3874
 LNY         2.4719    0.1233    20.04     0.000  0.987   0.9816     2.4719
 CONSTANT  -10.477     1.847     -5.671    0.000 -0.863   0.0000    -10.4767
|_DELETE / ALL
 ALL VARIABLES HAVE BEEN DELETED
|_STOP
*******************************************************************************
```

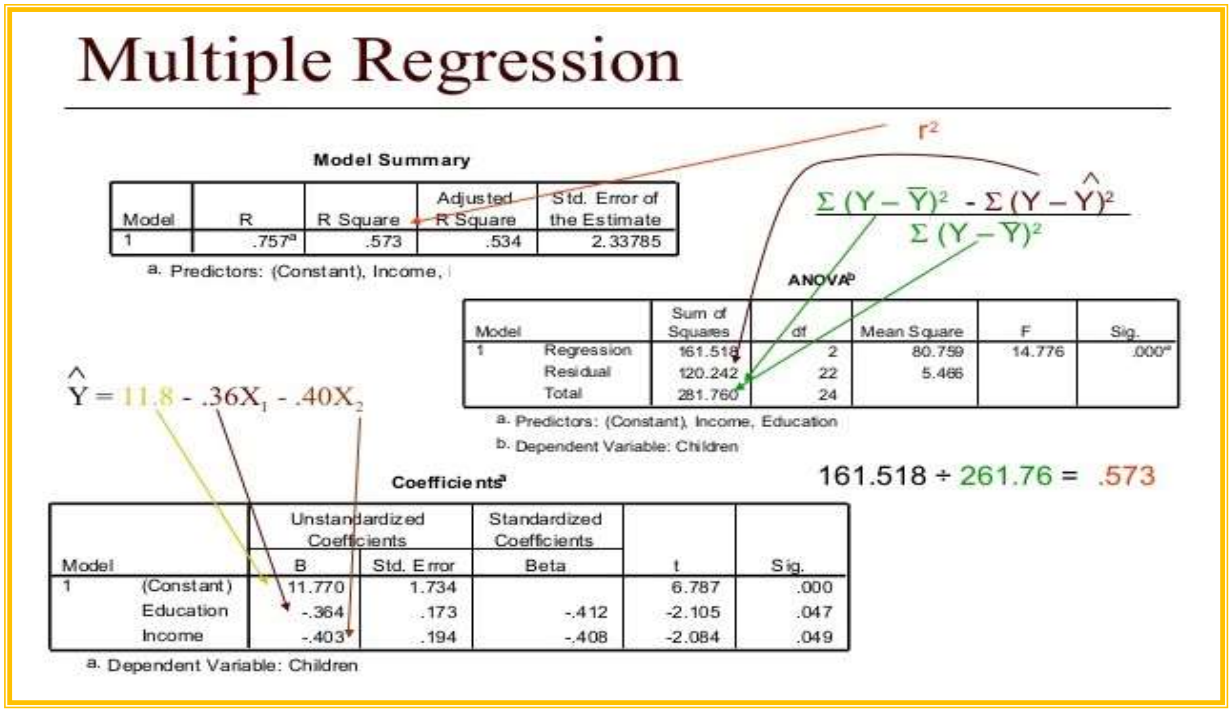
Uygulamalı Örnek 5.6, sayfa: 83

```
*******************************************************************************
|_SAMPLE 1 20
|_READ (MTABLE4.9) YEAR QD PD Y QS PS T
 UNIT 88 IS NOW ASSIGNED TO: MTABLE4.9
    7 VARIABLES AND       20 OBSERVATIONS STARTING AT OBS       1
|_OLS QD PD Y
        20 OBSERVATIONS     DEPENDENT VARIABLE= QD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.7813     R-SQUARE ADJUSTED =   0.7556
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   1.9517
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.3970
 SUM OF SQUARED ERRORS-SSE=   33.178
 MEAN OF DEPENDENT VARIABLE =   100.75
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PD        -0.14217    0.6656E-01  -2.136     0.048-0.460    -0.2787    -0.1421
 Y          0.23600    0.3117E-01   7.571     0.000 0.878     0.9880     0.2285
 CONSTANT   92.052      5.839       15.76     0.000 0.967     0.0000     0.9136
```

```
|_OLS QD PD Y T

        20 OBSERVATIONS     DEPENDENT VARIABLE= QD
 ...NOTE..SAMPLE RANGE SET TO:       1,     20
  R-SQUARE =    0.8812     R-SQUARE ADJUSTED =    0.8589
 VARIANCE OF THE ESTIMATE-SIGMA**2 =    1.1270
 STANDARD ERROR OF THE ESTIMATE-SIGMA =    1.0616
 SUM OF SQUARED ERRORS-SSE=    18.031
MEAN OF DEPENDENT VARIABLE =    100.75
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PD        -0.32506     0.7104E-01  -4.576     0.000-0.753    -0.6373    -0.3248
 Y          0.31524     0.3207E-01   9.831     0.000 0.926     1.3197     0.3052
 T         -0.22443     0.6122E-01  -3.666     0.002-0.676    -0.4698    -0.0234
 CONSTANT   105.09       5.687       18.48     0.000 0.977     0.0000     1.0431

|_OLS QD PD
        20 OBSERVATIONS     DEPENDENT VARIABLE= QD
 ...NOTE..SAMPLE RANGE SET TO:       1,     20
  R-SQUARE =    0.0441     R-SQUARE ADJUSTED =   -0.0090
 VARIANCE OF THE ESTIMATE-SIGMA**2 =    8.0576
 STANDARD ERROR OF THE ESTIMATE-SIGMA =    2.8386
 SUM OF SQUARED ERRORS-SSE=    145.04
 MEAN OF DEPENDENT VARIABLE =    100.75
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PD         0.10712     0.1175      0.9114     0.374 0.210     0.2100     0.1070
 CONSTANT    89.969      11.85       7.591     0.000 0.873     0.0000     0.8930
|_DELETE / ALL
ALL VARIABLES HAVE BEEN DELETED
|_STOP
```



VI. ÇOKLU REGRESYON ANALİZLERİ

6.1. Regresyon Katsayılarının Yorumu

Basit regresyonda açıklayıcı değişkenin açıklanan değişken üzerine olan etkisini ölçmekle ilgilenilmiştir. Regresyon denklemi aşağıdaki gibi de yazılabilir.

$$y - y_{ort} = \beta (x - x_{ort}) + \hat{u}_i$$

bu etki β' tarafında ölçülmektedir ve burada,

$$\beta' = S_{xy} / S_{xx} \quad \text{veya} \quad \text{cov}(x,y) / V(x)$$

Çoklu regresyonda x_1 ve x_2 gibi iki açıklayıcı değişkenli bir denklemde x_1 ve x_2'nin y üzerine birlikte ve kısmi etkisi üzerinde konuşulabilir. Bu çoklu regresyon denklemi aşağıdaki gibi yazılabilir.

$$y - y_{ort} = \beta'_1 (x_1 - x_{1ort}) + \beta'_2 (x_2 - x_{2ort}) + \hat{u}_i$$

Burada x_1'in kısmi etkisi β'_1 tarafından ve x_2'nin kısmi etkisi β'_2 tarafından ölçülmektedir. Kısmi etkiden anlatılmak istenen, bir değişkenin diğer değişken sabit tutulması veya diğer değişkenin etkisi elemine edildikten sonraki etkidir. Bu yüzden β'_1, x_2'nin x_1 üzerine etkisi elemine edildikten sonraki x_1'in y üzerine etkisinin ölçümü olarak yorumlanabilir. Aynı şekilde β'_2 de x_1'in x_2 üzerine olan etkisi elemine edildikten sonra x_2'nin y üzerine olan etkisinin ölçümü olarak yorumlanabilir.

Bu yorum β_1'in tahmincisi olan β'_1'nın iki ayrı basit regresyonla çıkarılabileceğini ifade etmektedir.

1. *Basamak*: x_1 in x_2 üzerine regresyonunu, yani $x_1 = b_{12}\, x_2$ hesaplanır. Regresyon katsayısının b_{12} tarafından ifade edildiği varsayıldığında bu denklemdeki rezidualleri w_i ile ifade edilebilir.

$$w_i = x_{1i} - x_{1ort} - b_{12} (x_{2i} - x_{2ort})$$

Dikkat edilirse w_i, x_2'nin x_1 üzerine etkisi elemine edildikten sonra kalan x_1'in etkilediği kısımdır.

2. *Basamak*: Şimdi y_i'nin w_i üzerine regresyonu tahmin edilebilir. Burada regresyon katsayısı β'_1'den başka bir şey değildir. Bu katsayıyı çoklu regresyonda tahmin etmiştik. Burada ise aynı katsayıyı iki aşamalı basit regresyonla elde ettik.

6.2. Kısmi Korelasyon ve Çoklu Korelasyon

Eğer bir açıklanan değişken y ve üç açıklayıcı değişken x_1, x_2, x_3'e sahip olursak, r^2_{y1}, r^2_{y2}, r^2_{y3} sırasıyla y ile x_1, x_2, x_3 arasındaki basit korelasyonların karesidirler. Bu yüzden r^2_{y1}, r^2_{y2} ve r^2_{y3} sırasıyla y'deki varyansın sadece x_1 sadece x_2 ve sadece x_3 tarafından açıklanan kısmını ölçer. Diğer tarafta $R^2_{y.123}$, x_1, x_2 ve x_3'ün birlikte açıkladığı y'deki varyasyonu ölçer. Bunların dışında ayrıca başka bir şey daha ölçmek isteyebiliriz. O da

örneğin; x_1 regresyon denklemi içerisinde varken x_2 ilave edildiğinde ne kadar ve modelde x_1 ve x_2 varken x_3 ilave edildiğinde ne kadar açıklama yapar? Bu soruların cevabı, $r^2_{y2.1}$ ve $r^2_{y3.12}$ gibi kısmi determinasyon katsayıları ile ölçülür. Burada noktadan sonraki değişkenler önceden modelde olan değişkenlerdir. Üç açıklayıcı değişken ile $r_{y1.2}$, $r_{y1.3}$, $r_{y2.1}$, $r_{y2.3}$, $r_{y3.1}$ ve $r_{y3.2}$ kısmi korelasyon katsayılarına sahip olunur. Bunlar birinci dereceden kısmi korelasyon diye adlandırılır. Aynı zamanda üç adet ikinci dereceden kısmi korelasyon katsayısı $r_{y1.23}$, $r_{y2.13}$ ve $r_{y3.12}$ olarak verilebilir. Yine noktadan sonraki değişkenler daha önceden model içerisinde mevcuttur. Kısmi korelasyon katsayısının derecesi noktadan sonraki değişken sayısına bağlıdır. Burada notasyon olarak kullanılan en yaygın ve kullanışlı yol basit ve kısmi korelasyonda küçük r, çoklu korelasyonda ise büyük R'nin kullanılmasıdır. Örneğin $R^2_{y.12}$, $R^2_{y.13}$, $R^2_{y.23}$ ve $R^2_{y.123}$ çoklu determinasyon katsayılarıdır.

Kısmi determinasyon katsayıları nasıl tahmin edilir? Bunun için r^2 ve t^2 arasındaki ilişki kullanılır. Örneğin, $r^2_{y2.3}$'ü hesap etmek için y'nin x_2 ve x_3 üzerine regresyonu dikkate alınmalıdır. Regresyon denklemi aşağıdaki gibi olsun.

$$y' = \alpha' + \beta'_2 x_2 + \beta'_3 x_3 \text{ ve}$$

$$t_2 = \beta'_2 / SE(\beta'_2) \text{ olsun. Bu durumda}$$

$$r^2_{y2.3} = t^2_2 / (t^2_2 + s.d.) \text{ olur.}$$

Burada s.d. gözlem sayısından tahmin edilen parametre sayısı çıkarılarak bulunur. Çoklu regresyon denkleminde bulunan değişkenler kısmi korelasyon katsayısında bahsedilen değişkenlerin hepsidir. Örneğin eğer, $r^2_{y4.237}$ kısmi korelasyon katsayısı hesap edilmek isteniyorsa, y'nin x_4, x_2, x_3 ve x_7 üzerine regresyonun hesap edilmelidir. Önce t değeri $t_4 = \beta'_4 / SE(\beta'_4)$ olarak hesap edilir. Burada β_4'ün x_4'ün katsayısı olduğu kabul edilirse $r^2_{y4.237}$ aşağıdaki gibi hesap edilir.

$$r^2_{y4.237} = t^2_4 / (t^2_4 + s.d.)$$

Burada s.d. = n-5'dir. Çünkü bu modelde 1 adet α ve 4 adet β vardır.

Kısmi korelasyonlar daha fazla açıklayıcı değişkenin modele ilave edilip edilmeyeceğine karar vermede çok önemlidir. Örneğin; x_1 ve x_2 diye iki açıklayıcı değişkene sahip olduğunu, r^2_{y2}'nin 0,95 gibi çok yüksek olduğu ve fakat $r^2_{y2.1}$'nin 0,1 gibi çok düşük olduğunu varsayalım. Bunun manası eğer x_2 y'yi açıklamada yalnız başına kullanılırsa iyi iş görmektedir. Fakat eğer x_1 ilave edilirse x_2 artık y'yi açıklamada fazla iş görememektedir. Çünkü x_1 aynı işi yapabilmektedir. Bu durumda x_2'yi model içinde bulundurmanın bir anlamı kalmıyor. Halbuki burada aşağıda verilen değerler gibi bir durum söz konusudur.

$$r^2_{y1} = 0,95 \quad \text{ve} \quad r^2_{y2} = 0,96$$

$$r^2_{y1.2} = 0,1 \quad \text{ve} \quad r^2_{y2.1} = 0,1$$

Bu durumda her iki değişken de y ile yüksek bir korelasyona sahip ve kısmi korelasyon her ikisinde de çok düşüktür. Bu durum çoklu eş doğrusallık olarak adlandırılmakta olup ileride ele alınacaktır. Bu örnekte sadece x_1 veya sadece x_2 kullanılabilir. Veya iki değişkenin herhangi bir kombinasyonunu açıklayıcı bir değişken gibi kullanabilir. Örneğin, x_1'in vasıflı işçi, x_2'nin vasıfsız işçi girdisi ve y'nin çıktı olduğu kabul edildiğinde kısmi korelasyon, işgücünün iki kısma ayrılmasıyla elde edilen x_1 ve x_2, y'yi açıklamada ilave katkı sağlamamaktadır. Bu nedenle $x_1 + x_2$ yeni bir değişken yani toplam işgücünü temsil eden bir açıklayıcı değişken olarak kullanılması daha doğru sonuçlar verebilecektir.

6.3. Basit, Kısmi ve Çoklu Korelasyon Katsayıları Arasındaki İlişkiler

Değişik tip korelasyon katsayıları arasındaki ilişkiyi analiz etmek için, her bir basamaktaki hata kareler toplamı $RSS = S_{yy}(1-R^2)$ ilişkisinin kullanılması gerekir. Burada x_1 ve x_2 açıklayıcı değişkenlerine sahip olunması durumunda $S_{yy}(1-R^2_{y.12})$, x_1 ve x_2 ilave edildikten sonraki RSS'yi, $S_{yy}(1-r^2_{y1})$ ise sadece x_1'in ilavesinden sonraki RSS'yi ve $r^2_{y2.1}$ ise RSS'nin x_2 tarafından açıklanan kısmını ölçer. Bu yüzden, x_2'nin ilavesinden sonra açıklanamayan RSS de aşağıdaki gibi ifade edilebilir.

$$(1 - r^2_{y2.1})\, S_{yy}\, (1 - r^2_{y1})$$

Burada $S_{yy}(1 - R^2_{y.12})$'ye eşittir. Bu yüzden aşağıdaki sonucu elde ederiz.

$$1 - R^2_{y.12} = (1 - r^2_{y1})\,(1 - r^2_{y2.1})$$

Eğer üç değişkenli bir model söz konusu ise aşağıdaki sonuç elde edilir.

$$(1- R^2_{y.123}) = (1 - r^2_{y1})\,(1 - r^2_{y2.1})\,(1 - r^2_{y3.12})$$

Buradaki 1, 2, 3 indisleri gerekli olan sıralamaya göre yer değiştirebilirler. Buradaki düzen 3, 1, 2 şeklinde olması gerekiyorsa, bu durumda aşağıdaki şekilde yazılabilir.

$$(1- R^2_{y.123}) = (1 - r^2_{y3})\,(1 - r^2_{y1.3})\,(1 - r^2_{y2.31})$$

Burada not edilmesi gereken husus, kısmi r^2, basit r^2'den büyük veya küçük olabilir. Örneğin, x_2 değişkeni y'nin varyansını % 20 açıklayabilir, fakat x_1 ilave edildikten sonra rezidual varyansını % 50 açıklayabilir. Bu durumda basit $r^2 = 0{,}2$ ve kısmi $r^2 = 0{,}5$ olabilir. Bu yüzden burada farklı varyansların açıklanan oranlarından bahsedilmektedir. Diğer tarafta, basit r^2 ve kısmi r^2 hiç bir zaman çoklu korelasyon katsayısının karesi olan R^2'den büyük olamaz. Bu çıkarım yukarıda izah edilen formüllerden de yapılabilir.

Uygulamalı Örnek 6.1

Konuyu açıklamak için kullanılacak olan örnek basit r^2, kısmi r^2 ve çoklu R^2'yi göstermektedir. Ayrıca bu örnek açıklayıcı değişkenler arasında bir oransallık yani çoklu eşdoğrusallık olduğu zaman, çoklu regresyon katsayılarının yorumunda ortaya çıkan bazı problemlere ışık tutmaktadır.

Burada hastane masraflarıyla alakalı bir analiz üzerinde durulmaktadır. Bunun için 177 hastane ile ilgili veriler kullanılmaktadır. Açıklayıcı değişkenler cerrahi, göz, dahiliye gibi 9 bölümün her birindeki hastaların tüm hastalara oranlarıdır. Bunların regresyon katsayıları, standart hataları, *t* değerleri, kısmi r^2'leri, basit r^2'leri ve her kategorideki ortalama masraflar çizelge 6.1'de verilmiştir.

Çizelge 6.1. Hastane masrafları regresyonu

Değişken	Regresyon katsayısı	Standart hata	t değeri	Kısmi r^2	Basit r^2	Her bir kategori için ort. Maliyet
1	44,97	18,89	2,38	0,0326	0,1423	114,48
2	-44,54	28,51	-1,56	0,0143	0,0074	24,97
3	-36,81	14,88	-2,47	0,0350	0,0343	32,70
4	-54,26	16,52	-3,28	0,0602	0,0947	15,25
5	-29,82	17,18	-1,74	0,0177	0,0062	39,69
6	28,51	20,27	1,41	0,0117	0,0478	98,02
7	-10,74	21,47	-0,50	0,0015	0,0099	58,72
8	-34,63	16,34	-2,12	0,0261	0,0011	34,88
9	0	-	-	-	-	69,51
Sabit = 69,51	R^2 = 0,3076					

Bu çizelgedeki sonuçlar birtakım açıklamaları gerektiriyor. Buradaki t değeri sadece regresyon katsayılarının standart hatalara bölünmesiyle elde ediliyor. Kısmi r^2, $t^2 / (t^2 +$ s.d.) formülü ile hesap edilir. Buradan hareketle birinci değişken için kısmi r^2 aşağıdaki şekilde hesap edilebilir.

$$\text{Kısmi } r^2 = (2,38)^2 / [\ (2,38)^2 + 168\ ] = 0,0326\text{'dır.}$$

Basit r^2 ise hastanenin her bir bölümündeki ortalama maliyetin ve o bölümdeki hastaların toplam hasta sayısına oranının korelasyonunun kareköküdür. Bu ikisinin bir arada verilmesinin sebebi, kısmi r^2'nin basit r^2'den nasıl daha yüksek veya düşük olduğunu göstermek içindir. Ancak hem basit hem de kısmi r^2, R^2'den büyük değildir.

Bu örnekteki regresyon katsayılarının değişik bir yolla izah edilmesi gerekmektedir. Normal durumda her bir regresyon katsayısı diğer şeyler değişmediğinde bağımsız değişkenin her bir birim değişmesi sonucu bağımlı değişken y'deki değişmeyi ölçer. Bu örnekte; bağımsız değişkenler her kategorideki oranlar olduğundan, bir değişkende herhangi bir değişme olduğunda diğerlerini sabit tutmak mümkün değildir. Bu durumda katsayıların en güzel yorumu birinci değişkenin değerini 1'e eşit ve diğerlerini sıfır kabul edersek, bu durumda bağımlı değişkenin tahmin edilen değeri sabit + birinci değişkenin katsayısı = 69,51 + 44,97 = 114,48'dir. Bu birinci bölümdeki bir hastaya yapılan muamelenin ortalama maliyetidir. Aynı şekilde, sabit + ikinci değişkenin katsayısı = 69,51 – 44,54 = 24,97 olur. Bu değer ikinci kategorideki bir hastaya yapılan muamelenin maliyetidir.

Uygulamalı Örnek 6.2

Bu örnekte kullanılacak olan 1927-1941 ve 1948-1962 dönemlerini içeren kişi başına gıda tüketimi, gıda fiyatı ve kişi başına gelir verileri çizelge 6.2'de verilmiştir.

Çizelge 6.2. ABD gıda tüketimi, gıda fiyatı ve gelir indeksleri

Yıl	Tüketim	Fiyat	Gelir	yıl	Tüketim	fiyat	Gelir
1927	88,9	91,7	57,7	1948	96,7	105,3	82,1
1928	88,9	92,0	59,3	1949	96,7	102,0	83,1
1929	89,1	93,1	62,0	1950	98,0	102,4	88,6
1930	88,7	90,9	56,3	1951	96,1	105,4	88,3
1931	88,0	82,3	52,7	1952	98,1	105,0	89,1
1932	85,9	76,3	44,4	1953	99,1	102,6	92,1
1933	86,0	78,3	43,8	1954	99,1	101,9	91,7
1934	87,1	84,3	47,8	1955	99,8	100,8	96,5
1935	85,4	88,1	52,1	1956	101,5	100,0	99,8
1936	88,5	88,0	58,0	1957	99,9	99,8	99,9
1937	88,4	88,4	59,8	1958	99,1	101,2	98,4
1938	88,6	83,5	55,9	1959	101,0	98,8	101,8
1939	91,7	82,5	60,3	1960	100,7	98,4	101,8
1940	93,3	83,0	64,1	1961	100,8	98,8	103,1
1941	95,1	86,2	73.7	1962	101,0	98,4	105,5

Bu veriler kullanılarak aşağıdaki iki model tahmin edilmiştir.

$$\text{Denklem 1:} \quad \log q = \alpha + \beta_1 \log p + \beta_2 \log y$$

$$\text{Denklem 2:} \quad \log q = \alpha + \beta_1 \log p + \beta_2 \log y + \beta_3 \log p \log y$$

İkinci eşitlikteki en son ifade, fiyat ve gelir elastikiyetlerinin değişmesine izin veren bir interaksiyon terimidir. Birinci denklemdeki katsayılar sırasıyla fiyat ve gelir elastikiyetini doğrudan vermektedir.

Çizelge 6.2'deki verilerden yararlanılarak hem belirlenen iki döneme ait hem de tüm gözlemler için elde edilen α, β_1, β_2 ve β_3 tahminleri çizelge 6.3'de verilmiştir. Yorum yapması daha kolay olduğu için parantez içinde verilen değerler *t* değerleridir. Eğer önemlilik derecesi ile ilgileniyorsak *t* değerleri daha kullanışlıdır. Eğer güven aralığı ile ilgileniyorsak o zaman standart hata değerleri daha kullanışlıdır. Zaten biri verildiğinde diğerini tahmin etmek mümkündür.

İnteraksiyonun *t* değeri, her iki dönem için ayrı ayrı dikkate alınarak analiz edildiğinde çok düşük *t* değerine sahip olduğu görülmektedir. Bu nedenle biz sadece tüm veriler kullanılarak tahmin edilen modeli dikkate alabiliriz. Bu denklem dikkate alındığında ilk bakışta gelirin *t* değerinin düşük olduğu ve hatta yanlış işarete sahip olduğu görülmektedir. Fakat bu denklem pozitif gelir elastikiyeti vermektedir.

Çizelge 6.3. Gıda talep denklemlerinin tahminleri

	Denklem 1			Denklem 2		
	1927-41	1948-62	Tüm Göz.	1927-41	1948-62	Tüm Göz.
α	4,555 (22,67)	5,052 (5,61)	4,050 (29,65)	4,058 (0,54)	16,632 (0,61)	8,029 (4,47)
β_1 (fiyat)	-0,235 (-4,41)	-0,237 (-1,54)	-0,120 (-2,95)	-0,123 (-0,07)	-2,745 (-0,47)	-0,996 (-2,51)
β_2 (gelir)	0,245 (10,63)	0,141 (3,03)	0,242 (17,85)	0,368 (0,20)	-2,416 (-0,40)	-0,718 (-1,66)
β_3 (interaksiyon)				-0,028 (-0,07)	0,554 (0,43)	0.211 (2,22)
N	15	15	30	15	15	30
R^2	0,9066	0,8741	0,9731	0,9066	0,8762	0,9775
F^c	58,2	41,7	489	35,6	26,0	374,5
10^2 x RSS	0,1151	0,0544	0,2866	0,1151	0,0535	0,2412
d.f.	12	12	27	11	11	26
10^4 x s^{2d}	0,9594	0,4534	1,0613	1,0462	0,4866	0,9278

Bu denklemdeki tahminlere göre gelir elastikiyeti aşağıdaki gibi hesaplanabilir.

$$\eta_{qy} = d \log q / d \log y$$
$$= \beta_2 + \beta_3 \log p = - 0,718 + 0,211 \log p$$

Bu verilerde log p, 4.33 ve 4.66 aralığında olduğundan gelir elastikiyeti de 0.195 ile 0.265 arasında değişmektedir.

Fiyat elastikiyeti ise aşağıdaki gibi hesap edilebilir.

$$\eta_{qp} = d \log q / d \log p$$
$$= \beta_1 + \beta_3 \log y = - 0,996 + 0,211 \log y$$

Bu nedenle gelir artarken gıda talebinin fiyat elastikiyeti azalmaktadır. Çizelge 6.2'deki verilere ait log y değerleri 3,78 ile 4,66 arasında yer almaktadır. Buna bağlı olarak gıda talebinin fiyatı elastikiyeti de - 0,198 ile - 0,013 arasında yer almaktadır.

6.4. Çoklu Regresyon Modelinde Kestirme

Çoklu regresyonda kestirme formülü basit regresyon modelindekine çok benzerdir. Sadece bir fark vardır ki oda kestirilen değerin standart hatasını belirlemek için bütün regresyon katsayılarının varyans ve kovaryansına ihtiyaç vardır. Burada iki değişken ve k sayıdaki değişken durumunda nasıl ifade edileceği gösterilmeye çalışılacaktır. Ancak k değişkenli ifade şeklini açıklamak bu noktada çok gerekli görülmemektedir.

Tahmin edilen regresyon denklemi aşağıdaki gibi olsun.

$$\hat{y} = \alpha' + \beta'_1 x_1 + \beta'_2 x_2$$

Şimdi verilen x_1'in x_{10} ve x_2'nin x_{20} değerleri için y'nin y_0 değerinin saptanması dikkate alındığında bu değişkenlerin geleceğe ait bazı değerleri olabilir. Bu durumda,

$$y_0 = \alpha + \beta_1 x_{10} + \beta_2 x_{20} + u_0$$

şeklinde bir modele sahip olunursa, bu modelin örnek tahmini,

$$\hat{y}_0 = \alpha' + \beta'_1 x_{10} + \beta'_2 x_{20}$$

olur. Bu durumda y değerinin kestirme hatası aşağıdaki gibidir.

$$\hat{y}_0 - y_0 = \alpha' - \alpha + (\beta'_1 - \beta_1) x_{10} + (\beta'_2 - \beta_2) x_{20} - u_0$$

Burada $E(\alpha' - \alpha)$, $E(\beta'_1 - \beta_1)$, $E(\beta'_2 - \beta_2)$ ve $E(u_0)$ sıfıra eşit olduğu için $E(\hat{y}_0 - y_0) = 0$ olur. Bu yüzden $\hat{y}_0$, doğru bir tahmincidir. Burada dikkat edilmesi gereken husus, $E(\hat{y}_0) = E(y_0)$ olmasıdır. Çünkü hem $\hat{y}_0$ ve hem de y_0 şansa bağlı değişkenlerdir. Bu durumda ŷ'nin kestirme hatasının varyansı aşağıdaki gibi yazılabilir.

$$\sigma^2 (1 + 1/n) + (x_{10} - x_{1ort})^2 \text{ var } (\beta'_1) + 2 (x_{10} - x_{1ort}) (x_{20} - x_{2ort}) \text{ cov } (\beta'_1, \beta'_2) + (x_{20} - x_{2ort})^2 \text{ var } (\beta'_2)$$

Bu durumda k değişkenli modelin kestirme hatasının varyansı aşağıdaki gibidir.

$$\sigma^2 (1 + 1/n) + \sum_{i=1}^{k} \sum_{j=1}^{k} (x_{i0} - x_{iort}) (x_{j0} - x_{jort}) \text{ cov } (\beta'_i, \beta'_j)$$

Burada σ^2, iki açıklayıcı değişken durumunda RSS/(n-3) ile ve genel durumda ise RSS / (n-k-1) ile hesap edilir.

Uygulamalı Örnek 6.3

Bir önceki konuda verilen uygulamalı 5.2 örneğindeki

$$y = 4{,}0 + 0{,}7\, x_1 + 0{,}2\, x_2$$

regresyon modeli üzerinde çalışılırsa ve $x_{10} = 12$ ve $x_{20} = 7$ olduğu durumda,

$$y = 4{,}0 + 0{,}7 (12) + 0{,}2 (7) = 13{,}8 \text{ olur.}$$

Burada dikkat edilecek husus x_{10} ve x_{20}'nin ortalamadan sapması birbirine eşittir.

$$x_{10} - x_{1ort} = 12 - 10 = 2$$

$$x_{20} - x_{2ort} = 7 - 5 = 2$$

Daha önce hesaplanan var (β'_1), var (β'_2), cov (β'_1, β'_2) ve s^2 için olan ifadeleri kullanarak y'nin kestirme hatasının tahmin edilen varyansı aşağıdaki gibi elde edilir.

$$0{,}07 (1 + 1 / 23) + 4 (3 / 20 + 3 / 20 – 2 / 10) (0{,}07) = 0{,}101$$

Kestirmenin standart hatası bu durumda, 0,318 olarak bulunur. Buradan y'nin kestirmesinin % 95 güven aralığı t dağılımı değeriyle aşağıdaki gibi hesap edilir.

$$13{,}8 \pm 2{,}086 (0{,}318) \Rightarrow 13{,}8 \pm 0{,}66 \Rightarrow 13{,}14 \text{ ve } 14{,}46\text{'dır.}$$

6.5. Varyans Analizi ve Hipotez Testleri

Bölüm 5'de β_1 ve β_2 hakkındaki hipotezlerin birlikte F testi uygulamalı örnek 5.2'de yapıldı. Bu test için alternatif bir hesaplama aşağıdaki formülle yapılabilir.

$$F = [(RRSS - URSS) / r] / [URSS / (n-k-1)]$$

burada,

URSS: sınırlanmamış hata kareler toplamı

RRSS: hipotez sınırlamaları koyularak elde edilen kareler toplamı.

r: hipotez tarafından koyulan sınırlama sayısı

Bir örnek olarak $\beta_1 = 1$ ve $\beta_2 = 0$ hipotezlerini, yukarıda ifade edilen örnekteki modele iki sınırlama olarak koyulsun. Sınırlanmamış RSS = 1.4 olarak önceden belirlenmişti. Sınırlanmış RSS'yi elde etmek için, aşağıdaki genel formülü minimum yapmak gerekir.

$$\Sigma (y_i - \alpha - 1{,}0\, x_{1i} - 0{,}0 x_{2i})^2$$

β_1 ve β_2 bilindiği için sadece α'nın hesap edilmesi gerekiyor.

$$\alpha' = y_{ort} - 1{,}0\, x_{1ort}$$

Buradan,
$$RSS = \Sigma [y_i - y_{ort} - (x_{1i} - x_{1ort})]^2$$
$$= S_{yy} + S_{11} - 2S_{1y}$$
$$= 10{,}0 + 12{,}0 - 2\,(10{,}0) = 2{,}0$$

Burada r = 2, n = 23, k = 2 olduğundan

$$F = [(2{,}0 - 1{,}4) / 2] / (1{,}4 / 20) = 0{,}3 / 0{,}07 = 4{,}3$$

olur. Bu değer daha önce belirlenen değerin aynısıdır. F cetvel değeri 2 ve 20 serbestlik derecesi ve % 5 önem seviyesinde 3,49'dur. Hesaplanan F değeri olan 4,3, cetvel değeri olan 3.49'dan büyük olduğundan hipotez reddedilir. Özel bir durum olan,

$$\beta_1 = \beta_2 = ... = \beta_k = 0$$

hipotezine sahip olunması durumunda,

$$URSS = S_{yy}(1- R^2), \text{ ve } RRSS = S_{yy}$$

eşitliklerine sahip olunur ve bu durumda F testi,

$$F = [(S_{yy} - S_{yy}(1-R^2)) / k] / [S_{yy}(1-R^2) / (n-k-1)] = [R^2 / (1-R^2)] * [(n-k-1) / k]$$

olur ki bu test k ve n-k-1 serbestlik derecesiyle F-dağılımına sahiptir. Bu testin sağladığı bilgi eğer hipotez kabul edilirse x'lerin hiçbirinin y'yi etkilemediği, yani regresyon denkleminin kullanışsız olduğu, fakat hipotezin reddi durumunda x'lerin açıklayıcı

olduğudur. Fakat hangisinin açıklamada daha kullanışlı olduğu konusunda bize herhangi bir bilgi vermediğinden her bir değişken için t testi yapmak gerekir.

Geleneksel olarak daha önce basit regresyonda dikkate alınan çizelgenin benzeri bir varyans analizi çizelgesi çoklu regresyonda da hazırlanır. Bu form Çizelge 6.4'de gibidir. Burada yapılan y'nin varyansı iki kısma ayrılmaktadır. Bunlar açıklayıcı (regresyon) kısım ve hata (rezidual) kısmıdır.

Çizelge 6.4. Çoklu regresyon modeli için varyans analizi

Varyasyonun Kaynağı	**Kareler Toplamı (SS)**	**Serbestlik Derecesi (s.d.)**	**Ortalama Kareler SS / s.d.**	**F**
Regresyon	$R^2 S_{yy}$	k	$R^2S_{yy} / k = MS_1$	
Rezidual	$(1- R^2) S_{yy}$	n - k - 1	$(1-R^2) S_{yy} / n-k-1 = MS_2$	$F = MS_1 / MS_2$
Toplam	S_{yy}	n - 1		

Örneğin; hastane masraflarına yönelik Çizelge 6.1'deki veriler kullanıldığında varyans analizi sonuçları Çizelge 6,5'deki gibi elde edilir. F hesap değeri olan 9,3 % 1 önem seviyesi ve 8 -168 serbestlik derecesi ile F tablo değeri olan 2,51'den çok yüksektir.

Çizelge 6.5. Hastane masraflarının regresyonu için varyans analizi

Varyasyonun Kaynağı	**Kareler Toplamı (SS)**	**Serbestlik Derecesi (s.d.)**	**Ortalama Kareler SS / s.d**	**F**
Regresyon	10.357	8	1.294,6	
Rezidual	23.311	168	138,8	F = 1.294,6 / 138,8 = 9,3
Toplam	33,668	176	---	

Bu sonuca göre, $\beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$ hipotezi reddediliyor. Bütün bunların manası, açıklayıcı değişkenlerin hastaneler arasındaki her bir hastaya ait ortalama maliyet varyasyonunu açıklamada önemli olmaktadır. Fakat hangi değişkenin önemli olduğu hakkında bilgi verilmemektedir.

6.6. İlgili Değişkenlerin İhmali ve İlgisiz Değişkenlerin Mevcudiyeti

Şu ana kadar tahmin edilen regresyon modellerinin bulundurduğu açıklayıcı değişkenlerin hepsinin ilgili olduğu varsayılmıştır. Pratikte böyle bir varsayım çok az durum için doğrudur. Bazen bazı ilgili değişkenler ölçüm eksikliği veya dikkatsizlik sonucu model içine alınmazlar. Diğer bazı zamanlarda ise ilgisiz değişkenler model içinde mevcut olabilir. Burada öğrenmeğe çalışacağımız husus, bu problemler mevcut olduğunda model hakkındaki yorumun nasıl değişeceğidir.

İlgili Değişkenlerin İhmali

Burada ilk olarak ilgili değişkenin ihmali dikkate alındığında doğru denklemin aşağıdaki gibi olduğu kabul edilsin.

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

Bunun yerine, x_2'yi ihmal ederek aşağıdaki denklem tahmin edilsin.

$$y = \beta_1 x_1 + u$$

Bu yanlış tanımlanmış bir model olarak ortaya çıkmaktadır. β_1'in tahmini için aşağıdaki denklem kullanılmaktadır.

$$\beta'_1 = \sum x_1 y / \sum x_1^2$$

Yukarıdaki ilk denklemdeki y burada y'nin yerine yazılırsa sonuç aşağıdaki gibi olur.

$$\beta'_1 = \sum x_1 (\beta_1 x_1 + \beta_2 x_2 + u) / \sum x_1^2 = \beta_1 + \beta_2 (\sum x_1 x_2 / \sum x_1^2) + (\sum x_1 u / \sum x_1^2)$$

Burada $E(\sum x_1 u) = 0$ olduğundan $E(\beta'_1) = \beta_1 + b_{21} \beta_2$ olur ve $b_{21} = \sum x_1 x_2 / \sum x_1^2$ x_2'nin x_1 üzerine olan regresyonundaki regresyon katsayısıdır. Bu sebeple β'_1, β_1'in doğru olmayan bir tahmincisidir ve bu sapma aşağıdaki gibi ifade edilir.

sapma (bias) = [Model dışındaki değişkenin katsayısı] * [Model dışındaki değişkenin model içindeki değişken üzerine regresyonun regresyon katsayısı]

β_1'in tahmincisi olarak $\beta_1\hat{}$ kabul edilirse, $\beta_1\hat{}$'nın varyansı aşağıdaki gibi verilir.

$$Var(\beta_1\hat{}) = \sigma^2 / S_{11}(1 - r_{12}^2) \text{ burada, } S_{11} = \sum x_1^2 \text{ dir.}$$

Diğer tarafta, $Var(\beta'_1) = \sigma^2 / S_{11}$ dir.

Bu yüzden, $\beta_1\hat{}$ sapmalı bir tahmincidir. Ancak β'_1'den daha küçük bir varyansa sahiptir. Gerçekte eğer r_{12}^2 çok yüksekse, varyans dikkat çeker bir ölçüde küçük olacaktır. Fakat β'_1'nın tahmin edilen standart hatası $\beta_1\hat{}$'nın standart hatasından küçük olması gerekmeyebilir. Bu durum hatanın tahmin edilen varyansı olan σ^2'den dolayıdır, çünkü bu varyans yanlış belirlenen modelde yüksek olabilir. Bu σ^2 rezidual kareler toplamının serbestlik derecesine bölünmesiyle bulunur. Bu yanlış belirlenen model için yüksek veya düşük olabilir.

Uygulamalı Örnek 6.4

Daha öncede örnek verilen fiyat ve gelir verilerine göre yapılan talep tahminini dikkate aldığımızda talep doğrusu denklemi aşağıdaki gibidir.

$$Q_D = \alpha + \beta_1 P_D + \beta_2 Y + u$$

Eğer, gelir değişkeni dikkate alınmazsa model aşağıdaki gibi tahmin edilir.

$$Q_D = \underset{(11,85)}{89,97} + \underset{(0,118)}{0,107}\ P_D \qquad s^2 = 2,338$$

Parantez içindeki değerler standart hatalardır. Burada P_D'nin katsayısı yanlış işarete sahiptir. Acaba bu gelir değişkeninin dikkate alınmamasıyla ilişkilendirilebilir mi? Cevap evettir. Çünkü P_D'nin katsayısı, sapması aşağıda verilen sapmalı bir tahmindir.

sapma (bias) = [gelir değişkeninin katsayısı] * [gelirin fiyat üzerine olan regresyon katsayısı]

Gelirin katsayısının pozitif olması beklenir. Aynı zamanda, kullanılan veriler zaman serisi verileri olduğundan P_D ile Y arasında pozitif bir korelasyon beklenir. Dolayısıyla yapılan yanlışın pozitif olması beklenir. Bu durum negatif bir katsayıyı pozitif bir katsayıya çevirebilir.

Y modele ilave edildiğinde regresyon denklemi aşağıdaki sonucu verir.

$$Q_D = \underset{(5,84)}{92,05} - \underset{(0,067)}{0,142}P_D + \underset{(0,031)}{0,236}Y \quad s^2 = 1,952$$

Parantez içindeki değişkenler standart hatalardır. P_D bu sefer negatif olmuştur. Aynı zamanda, P_D katsayısının standart hatası yanlış belirlenen modelde gerçek modelden daha yüksektir. P_D'in katsayısına ait varyansı yanlış belirlenen modelde daha küçük olmasının beklenmesi gerçeğine rağmen sonuç böyledir.

<u>İlgisiz Değişkenlerin Mevcudiyeti</u>

İlgisiz değişkenlerin mevcudiyetinin dikkate alınması durumunda gerçek denklemin aşağıdaki gibi olduğu varsayılsın.

$$y = \beta_1 x_1 + u$$

Fakat tahmin edilen model aşağıdaki gibi olsun,

$$y = \beta_1 x_1 + \beta_2 x_2 + v$$

Bu yanlış belirlenen denklemden elde edilen en küçük kareler tahminleri $\beta_1\hat{}$ ve $\beta_2\hat{}$ aşağıdaki gibi verilmiş olsun.

$$\beta_1\hat{} = (S_{22} S_{1y} - S_{12} S_{2y}) / (S_{11} S_{22} - S^2_{12})$$

$$\beta_2\hat{} = (S_{11} S_{2y} - S_{12} S_{1y}) / (S_{11} S_{22} - S^2_{12})$$

Burada $S_{11} = \sum x_1^2$, $S_{1y} = \sum x_1 y$, $S_{12} = \sum x_1 x_2$ ve böyle devam ediyor olsun. $y = \beta_1 x_1 + u$ olduğundan $E(S_{2y}) = \beta_1 S_{12}$ ve $E(S_{1y}) = \beta_1 S_{11}$ olur. Böylece aşağıdaki sonuç elde edilir.

$$E(\beta_1\hat{}) = \beta_1 \text{ ve } E(\beta_2\hat{}) = 0$$

Bu yüzden, her iki parametre içinde sapmasız tahminler elde edilir. İlgili değişkenin göz ardı edilmesi ile ortaya çıkan sapma ile ilgili önceki sonuçlarla bir araya gelen bu sonuç, şüphe olması durumunda, değişkeni model dışında tutmaktansa model içerisine almanın daha iyi olacağına inanılmasını sağlayabilir. Fakat bu böyle değildir. Çünkü ilgisiz

değişkenin parametre üzerine etkisi olmasa bile bu parametrenin varyansını etkilemektedir.

Doğru denklemdeki β_1'ın tahmincisi olan β'_1'nın varyansı, aşağıdaki gibi verilmektedir.

$$V(\beta'_1) = \sigma^2 / S_{11}$$

Diğer tarafta, yanlış belirlenen denklemden aşağıdaki varyansa sahibiz.

$$Var(\beta_1\hat{}) = \sigma^2 / (1 - r^2_{12}) S_{11}$$

Burada r_{12}, x_1 ve x_2 arasındaki korelasyondur. Bu yüzden $r_{12} = 0$ olmadığı müddetçe var $(\beta_1\hat{})$, var (β'_1)'den büyük olmaktadır. Bu yüzden modele ilgisiz değişkenler ilave edildiğinde sapmasız fakat etkin olmayan tahminler elde edilir. Kullanılan rezidual varyansının tahmincisinin, σ^2'nin sapmasız bir tahmincisi olduğu gösterilebilir. Bu yüzden yanlış belirlenmiş denklemden tahmin edilen varyansın kullanımından, yeni bir sapma ortaya çıkmaz.

6.7. Serbestlik Derecesi ve R^2

Eğer n sayıda gözlemin kullanıldığı 3 parametresi olan bir model tahmin ediliyorsa, çoklu regresyon analizinin başında belirlenen normal denklemlerden görülecek ki tahmin edilen rezidual u_i üç doğrusal sınırlamayı yerine getirmektedir.

$$\Sigma u_i = 0 \qquad \Sigma x_{1i} u_i = 0 \qquad \Sigma x_{2i} u_i = 0$$

veya gerçekte sadece n-3 değişen rezidual vardır denilebilir. Çünkü verilen herbir n-3 rezidual için kalan üç rezidual yukarıdaki denklemleri çözerek bulunabilir. Bu nokta n-3 serbestlik derecesi olarak ifade edilebilir.

Önceden de görüldüğü gibi rezidual varyansı σ^2'nin tahmini s^2 = RSS / (serbestlik derecesi) şeklinde verilmişti. Açıklayıcı değişkenlerin sayısını artırdıkça RSS düşer fakat bunun yanında serbestlik derecesinde bir düşme de olmaktadır. Burada s^2'ye ne olacağı pay ve paydadaki oransal düşmeye bağlıdır. Bu yüzden, daha fazla açıklayıcı değişken ilave edildikçe s^2'nin artmaya başlayacağı bir nokta olacaktır. Dolayısıyla s^2'nin en düşük olduğundaki değişkenlerin seçilmesi çoğu zaman tavsiye edilmektedir.

Bu aynı zamanda, çoklu regresyonda düzeltilmiş $\acute{R}^2$'nin kullanılmasının gerekçesidir. Önceden tanımlanan R^2 modele değişken ilave edildikçe 1,0'e kadar artmaya devam eder ve bu yüzden serbestlik derecesi problemini dikkate almaz. Düzeltilmiş $\acute{R}^2$ basitçe R^2'nin serbestlik derecesinin ayarlanmış şeklidir ve aşağıdaki ilişki ile belirlenir.

$$1 - \acute{R}^2 = [(n -1) / (n - k - 1)] * (1- R^2)$$

Buradaki k bağımsız değişken sayısıdır ve k+1 n'den çıkarılır, çünkü k sayıdaki bağımsız değişkenin katsayısına (β) ilave olarak bir de sabit katsayı (α) tahmin edilmektedir. Yukarıdaki denklem, aşağıdaki gibi de yazılabilir.

$$(1- \acute{R}^2)\, S_{yy} \,/\, (n - 1) \;=\; (1- R^2)\, S_{yy} / (n - k - 1) \;=\; s^2$$

S_{yy} ve n sabit olduğundan, denklem içindeki bağımsız değişken sayısını artırdıkça s^2 ve $(1- \acute{R}^2)$ aynı yönde, s^2 ve $\acute{R}^2$ ters yönde hareket etmektedir. Bu yüzden denilebilir ki s^2'yi minimum yapan değişkenler aynı zamanda R^2'yi de en yüksek yapmaktadır.

Şimdi kısa bir hatırlatma yapılırsa aşağıdaki sonuçlar elde edilir.

TSS: Toplam kareler toplamı

ESS: Açıklanan kareler toplamı

RSS: Hata kareler toplamı

TSS = ESS + RSS

$$R^2 \;=\; ESS / TSS \;=\; (TSS - RSS) / TSS \;= 1 - (RSS / TSS)$$

Buradan da görüldüğü gibi açıklayan kareler toplamı ne kadar büyük ve hata kareler toplamı ne kadar küçük olursa R^2'de o kadar büyük olur ve bu da modelin bağımlı değişkeni iyi açıkladığını ifade eder. Çoklu değişkenlerde ise $\acute{R}^2$'yi kullanarak bir yargıya varılmaktadır. Yukarıdaki formülden de anlaşılacağı gibi R^2 0 ile 1 arasında bir değerdir ve 1 'e ne kadar yaklaşırsa model o kadar iyi açıklayıcıdır denir. Fakat yukarıda ifade edildiği gibi, bağımsız değişken sayısı arttığında R^2 otomatik olarak artacağından, R^2 iyi bir gösterge olmaz. Bunun yerine serbestlik derecesini dikkate alan düzeltilmiş $\acute{R}^2$'yi kullanmak gerekir.

6.8. İstikrarlılık (Stability) Testleri

Çoklu regresyon denklemi tahmin edildiğinde ve bu denklem gelecek zaman dilimleri için y'nin kestirmesinde kullanıldığı zaman, hem modelin hem de y'nin tahmini için tüm zaman periyodundaki parametrelerin sabit olduğu kabul edilmektedir. Bu parametrelerin sabitliği hipotezini doğrulamak için bazı testler geliştirilmiştir. Bunlardan biri aşağıda izah edilmeye çalışılacaktır.

Varyans Analizi Testi

Büyüklüğü n_1 ve n_2 olan iki veri setinin regresyon denklemleri aşağıdaki gibidir.

$$Y = \alpha_1 + \beta_{11} x_1 + \beta_{12} x_2 + \ldots\ldots\ldots + \beta_{1k} x_k + u \quad \text{birinci set için}$$

$$Y = \alpha_2 + \beta_{21} x_1 + \beta_{22} x_2 + \ldots\ldots\ldots + \beta_{2k} x_k + u \quad \text{ikinci set için}$$

Buradaki β'lar için ilk indis veri setini ikincisi ise değişkeni ifade etmektedir. İki veri setinden elde edilen ana kitle parametrelerinin istikrar testi aşağıdaki hipotezin testidir.

$$H_0: \beta_{11} = \beta_{21},\ \beta_{12} = \beta_{22},\ \ldots\ldots\ldots,\ \beta_{1k} = \beta_{2k},\ \alpha_1 = \alpha_2$$

Eğer bu hipotezler doğru ise, iki veri seti bir araya getirilerek elde edilen veri seti için bir tek denklem tahmin edilebilir.

Kullanılacak F testi, önceki bölümlerde URSS ve RSS temeline dayalı olarak açıklanan F testidir. URSS'yi elde etmek için her bir veri seti için ayrı bir denklem tahmin edilmekte ve burada RSS_1, birinci veri seti için hata kareler toplamını, RSS_2 ise ikinci veri seti için hata kareler toplamıdır.

RSS_1/σ^2 $(n_1\text{-}k\text{-}1)$ serbestlik derecesiyle ve RSS_2/σ^2, $(n_2\text{-}k\text{-}1)$ serbestlik derecesiyle bir χ^2 dağılımına sahiptir. İki veri seti birbirinden bağımsız olduğu için $(RSS_1+RSS_2)/\sigma^2$ $(n_1+n_2\text{-}2k\text{-}2)$ serbestlik derecesiyle χ^2 dağılımına sahiptir. Biz burada (RSS_1+RSS_2)'yi URSS ile ifade edeceğiz. Sınırlanmış hata kareler toplamı (RRSS) bir araya getirilen verilerin regresyonundan elde edilir. Burada modele koyulan sınırlama parametrelerin sayısı aynıdır.

Böylece, $(RSS_1+RSS_2)/\sigma^2$ ve $(n_1+n_2\text{-}2k\text{-}1)$ serbestlik derecesiyle χ^2 dağılımına sahiptir. Bu durumda,

$$F = [\ (RRSS - URSS) / (k+1)\ ] / [\ URSS / (n_1 + n_2 - 2k - 2)\ ]$$

Bu denklem (k+1) ve $(n_1+n_2\text{-}2k\text{-}2)$ serbestlik derecesiyle bir F-dağılımına sahiptir.

Uygulamalı Örnek 6.5

Elimizde iki ayrı periyoda ait aynı talep modeli ile ilgili zaman serisi verileri (Çizelge 6.2) ve bu 1927-1941 ve 1948-1962 dönemleri ve bir de bütün olarak bu iki dönemi içeren dönemle ilgili tahminler vardır. Buna göre,

URSS = İki ayrı regresyona ait RSS'lerin toplamı

= 0,1151+ 0,0544 = 0,1695 30 – 4 - 2 = 24 serbestlik deresiyle

RRSS = Bir araya getirilen verilerden elde edilen RSS

= 0,2866 30 - 2 - 1 = 27 serbestlik derecesiyle

Bir araya getirilen verilerle yapılan regresyon, iki dönem içindeki parametrelerin birbirine eşit olduğu sınırlamasını koymaktadır.

$$F = [\ (0{,}2866 - 0{,}1695) / 3\ ] / (0{,}1695 / 24) = 5{,}53$$

F tablosundan 3 ve 24 serbestlik derecesi ile ve %5 önem seviyesi ile 3,01 ve %1 önem seviyesi ile de 4,72 olarak F tablo değeri bulunmaktadır. Böylece %1 önem seviyesinde bile, istikrarlılık (stability) hipotezi reddedilmektedir. Sonuç olarak verilerin birleştirilmesinin doğru olmayacağına inanılmaktadır. Parametrelerin ayrı ayrı tahmin edilmesinin doğru olacağı sonucu ortaya çıkmaktadır.

YAZILIM UYGULAMASI

****

Uygulamalı Örnek 6.2, sayfa: 93

****

```
|_SAMPLE 1 30
|_READ (MTABLE4.3) YEAR Q P Y
 UNIT 88 IS NOW ASSIGNED TO: MTABLE4.3
    4 VARIABLES AND       30 OBSERVATIONS STARTING AT OBS       1
|_GENR LNQ=LOG(Q)
|_GENR LNP=LOG(P)
|_GENR LNY=LOG(Y)
|_GENR INTER=LNP*LNY
|_* Önce tüm gözlemleri kullanarak tahmin yapalım
|_OLS LNQ LNP LNY / LOGLOG
        30 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     30
  R-SQUARE =   0.9731     R-SQUARE ADJUSTED =   0.9711
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.10628E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.10309E-01
 SUM OF SQUARED ERRORS-SSE=  0.28695E-02
 MEAN OF DEPENDENT VARIABLE =   4.5419
VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      27 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNP      -0.11889     0.4036E-01  -2.946     0.007-0.493    -0.1880    -0.1189
 LNY       0.24115     0.1344E-01   17.95     0.000 0.961     1.1455     0.2412
 CONSTANT   4.0473     0.1360       29.76     0.000 0.985     0.0000     4.0473
|_*Gözlem sayısı ($N), kesişme katsayısı dahil katsayıların sayısı ($K) ve
|_*hata kareler toplamı ($SSE) gibi geçici değişkenlerin OLS komutundan
|_*sonra kullanılması mümkündür. Karışıklıktan kaçınmak için, ders
|_*notlarındaki RSS ile SHAZAM'daki SSE aynı şeyi ifade ettiğini bu noktada
|_*belirtmek gerekir.
|_GEN1 N1=$N
 ..NOTE..CURRENT VALUE OF $N   =   30.000
|_GEN1 RSS1=$SSE
 ..NOTE..CURRENT VALUE OF $SSE =  0.28695E-02
|_*Bir sonraki denklem tahmin edilirken LOGLOG opsiyonu kullanılmamaktadır
|_*çünkü, INTER değişkeni basit log formunda değildir.
|_OLS LNQ LNP LNY INTER
        30 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     30
  R-SQUARE =   0.9774     R-SQUARE ADJUSTED =   0.9748
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.92779E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.96322E-02
 SUM OF SQUARED ERRORS-SSE=  0.24122E-02
 MEAN OF DEPENDENT VARIABLE =   4.5419
VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      26 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNP      -0.99625     0.3970      -2.509     0.019-0.442    -1.5754    -0.9948
 LNY      -0.71839     0.4324      -1.661     0.109-0.310    -3.4123    -0.6781
 INTER     0.21118     0.9513E-01   2.220     0.035 0.399     5.8047     0.9051
 CONSTANT   8.0290      1.798       4.465     0.000 0.659     0.0000     1.7677
|_GEN1 N2=$N
 ..NOTE..CURRENT VALUE OF $N   =   30.000
|_GEN1 RSS2=$SSE
 ..NOTE..CURRENT VALUE OF $SSE =  0.24122E-02
|_*Şimdi 1927-1941 dönemi için her iki denklemi de hesap edelim.
```

```
|_SAMPLE 1 15
|_OLS LNQ LNP LNY / LOGLOG
        15 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     15
  R-SQUARE =   0.9066     R-SQUARE ADJUSTED =   0.8910
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.95944E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.97951E-02
 SUM OF SQUARED ERRORS-SSE=  0.11513E-02
 MEAN OF DEPENDENT VARIABLE =   4.4872
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      12 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNP      -0.23520     0.5338E-01  -4.406     0.001-0.786    -0.4699    -0.2352
 LNY       0.24324     0.2289E-01   10.63     0.000 0.951     1.1331     0.2432
 CONSTANT   4.5549     0.2009       22.67     0.000 0.989     0.0000     4.5549
|_GEN1 N11=$N
 ..NOTE..CURRENT VALUE OF $N   =   15.000
|_GEN1 K11=$K
 ..NOTE..CURRENT VALUE OF $K   =   3.0000
|_GEN1 RSS11=$SSE
 ..NOTE..CURRENT VALUE OF $SSE =  0.11513E-02
|_OLS LNQ LNP LNY INTER
        15 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     15
  R-SQUARE =   0.9066     R-SQUARE ADJUSTED =   0.8812
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.10462E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.10229E-01
 SUM OF SQUARED ERRORS-SSE=  0.11509E-02
 MEAN OF DEPENDENT VARIABLE =   4.4872
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      11 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNP      -0.12261      1.704     -0.7197E-01 0.944-0.022    -0.2450    -0.1216
 LNY       0.36789      1.885      0.1951     0.849 0.059     1.7138     0.3301
 INTER    -0.28217E-01 0.4267     -0.6612E-01 0.948-0.020    -0.7311    -0.1127
 CONSTANT   4.0577      7.522      0.5395     0.600 0.161     0.0000     0.9043
 |_GEN1 N21=$N
 ..NOTE..CURRENT VALUE OF $N   =   15.000
 |_GEN1 K21=$K
 ..NOTE..CURRENT VALUE OF $K   =   4.0000
 |_GEN1 RSS21=$SSE
 ..NOTE..CURRENT VALUE OF $SSE =  0.11509E-02
 |_STOP
```

****

Uygulamalı Örnek 6.5, sayfa 102

```
**************************************************************************
|_*DIAGNOS ve CHOWONE=n opsiyonu (n= verilerin ilk setindeki en son
|_*gözlemin numarasıdır) model istikrarı için test yapılmasında
|_*kullanılacaktır. Daha genel olarak, CHOWTEST opsiyonu test
|_*istatistiklerini elde etmek için her bir gözlem kesinti noktasında
|_*kullanılabilir. Bu opsiyonlar sadece OLS regresyonunu takiben
|_*kullanılmalıdır.
|_SAMPLE 1 30
|_OLS LNQ LNP LNY / LOGLOG
        30 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     30
  R-SQUARE =   0.9731     R-SQUARE ADJUSTED =   0.9711
```

```
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.10628E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.10309E-01
 SUM OF SQUARED ERRORS-SSE=  0.28695E-02
 MEAN OF DEPENDENT VARIABLE =   4.5419
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      27 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNP      -0.11889     0.4036E-01  -2.946     0.007-0.493    -0.1880    -0.1189
 LNY       0.24115     0.1344E-01   17.95     0.000 0.961     1.1455     0.2412
 CONSTANT   4.0473     0.1360       29.76     0.000 0.985     0.0000     4.0473
 |_DIAGNOS / CHOWONE=15
 DEPENDENT VARIABLE = LNQ             30 OBSERVATIONS
 REGRESSION COEFFICIENTS
  -0.118894327503      0.241153816769       4.04725346116
SEQUENTIAL CHOW AND GOLDFELD-QUANDT TESTS
    N1   N2    SSE1        SSE2       CHOW    PVALUE    G-Q       DF1  DF2 PVALUE
    15   15 0.11513E-02 0.54408E-03  5.5400     0.005  2.116       12   12 0.104
            CHOW TEST - F DISTRIBUTION WITH DF1=   3 AND DF2=  24
 |_OLS LNQ LNP LNY INTER
        30 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     30
  R-SQUARE =   0.9774     R-SQUARE ADJUSTED =   0.9748
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.92779E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.96322E-02
 SUM OF SQUARED ERRORS-SSE=  0.24122E-02
 MEAN OF DEPENDENT VARIABLE =   4.5419
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      26 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNP      -0.99625     0.3970     -2.509     0.019-0.442    -1.5754    -0.9948
 LNY      -0.71839     0.4324     -1.661     0.109-0.310    -3.4123    -0.6781
 INTER     0.21118     0.9513E-01   2.220     0.035 0.399     5.8047     0.9051
 CONSTANT   8.0290      1.798       4.465     0.000 0.659     0.0000     1.7677
|_DIAGNOS / CHOWONE=15
DEPENDENT VARIABLE = LNQ             30 OBSERVATIONS
 REGRESSION COEFFICIENTS
  -0.996251146596   -0.718392980215     0.211184982127     8.02897714188
SEQUENTIAL CHOW AND GOLDFELD-QUANDT TESTS
    N1   N2    SSE1        SSE2       CHOW    PVALUE    G-Q       DF1  DF2 PVALUE
    15   15 0.11509E-02 0.53533E-03  2.3682     0.084  2.150       11   11 0.110
            CHOW TEST - F DISTRIBUTION WITH DF1=   4 AND DF2=  22
|_*SHAZAM tarafından otomatik olarak hesaplanan Chow istatistiği Ders
|_*notlarında hesap edilen F istatistiğinin benzeridir.
|_STOP
```

****

Tarım ile ilgili Örnek

****

```
|_* BU PROGRAM BUĞDAY VERIMI UZERINE AZOTLU VE FOSFORLU GUBRELERIN ETKISINI ANALIZ YAPAR.
|_* TÜRKİYENİN 75 İLİNE AİT ORTALAMA DEKARA BUĞDAY VERİMİ, AZOTLU VE FOSFORLU GÜBRE KUL.
|_*alttaki üç satırdaki komut, bilgisayarın veri dosyasını bulması, örnek sayısını bilmesi ve değişkenleri isimlendirerek okuması içindir.
|_FILE 1 A:azot.txt
 UNIT  1 IS NOW ASSIGNED TO: A:azot.txt
|_SAMPLE 1 75
|_READ (1) V A F G
  4 VARIABLES AND      75 OBSERVATIONS STARTING AT OBS      1
|_GENR D1 = DUM(G.EQ.2)
|_GENR D2 = DUM(G.EQ.3)
|_*aşağıdaki ols en küçük kareler yöntemiyle basit regresyon analizi yapar (ilk değişken bağımlı) ayrıca varyans analizini anavo
|_*komutundan dolayı yapar.
```

|_OLS V A F / ANOVA
75 OBSERVATIONS DEPENDENT VARIABLE = V
...NOTE..SAMPLE RANGE SET TO: 1, 75
R-SQUARE = 0.6189 R-SQUARE ADJUSTED = 0.6083
VARIANCE OF THE ESTIMATE-SIGMA**2 = 2416.0
STANDARD ERROR OF THE ESTIMATE-SIGMA = 49.153
SUM OF SQUARED ERRORS-SSE= 0.17395E+06
MEAN OF DEPENDENT VARIABLE = 205.87
ANALYSIS OF VARIANCE - FROM MEAN

	SS	DF	MS	F
REGRESSION	0.28253E+06	2.	0.14127E+06	58.471
ERROR	0.17395E+06	72.	2416.0	
TOTAL	0.45648E+06	74.	6168.7	

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 72 DF	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
A	2.8509	0.47557	5.9947	0.5770	0.64914	0.38216
F	1.5468	0.96283	1.6065	0.1860	0.17396	0.14600
CONSTANT	97.139	13.811	7.0332	0.6382	0.00000E+00	0.47184

|_*aşağıdaki komutlar a=1 ve f=0.5 hipotezlerini test eder.
|_TEST
|_TEST A=1
|_TEST F=0.5
|_END
F STATISTIC = 25.032270 WITH 2 AND 72 D.F.
WALD CHI-SQUARE STATISTIC = 50.064540 WITH 2 D.F.
|_*aşağıdaki komutlar kuadratik fonksiyon için değişkenleri transform ederek yeni değişkenler oluşturur.
|_GENR AA = A*A
|_GENR FF = F*F
|_GENR AF = A*F
|_*kuadratik fonksiyonu tahmin eder

|_OLS V A F AA FF AF
75 OBSERVATIONS DEPENDENT VARIABLE = V
...NOTE..SAMPLE RANGE SET TO: 1, 75
R-SQUARE = 0.6399 R-SQUARE ADJUSTED = 0.6138
VARIANCE OF THE ESTIMATE-SIGMA**2 = 2382.5
STANDARD ERROR OF THE ESTIMATE-SIGMA = 48.811
SUM OF SQUARED ERRORS-SSE= 0.16439E+06
MEAN OF DEPENDENT VARIABLE = 205.87

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 69 DF	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
A	1.9077	1.3774	1.3850	0.1645	0.43438	0.25573
F	6.6990	3.2598	2.0551	0.2402	0.75340	0.63231
AA	-0.16876E-01	0.27939E-01	-0.60402	-0.0725	-0.31073	-0.88294E-01
FF	-0.20926	0.10977	-1.9064	-0.2237	-0.93696	-0.46207
AF	0.10087	0.91034E-01	1.1080	0.1322	0.83645	0.31930
CONSTANT	70.621	22.709	3.1098	0.3506	0.00000E+00	0.34303

|_*aşağıdaki genr komutu verileri logaritmasını alarak transform sonucu yeni veriler elde eder.
|_GENR LV = LOG(V)
|_GENR LA = LOG(A)
|_GENR LF = LOG(F)
|_*iki taraflı logaritmik fonksiyon

|_OLS LV LA LF
75 OBSERVATIONS DEPENDENT VARIABLE = LV
...NOTE..SAMPLE RANGE SET TO: 1, 75
R-SQUARE = 0.5830 R-SQUARE ADJUSTED = 0.5715
VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.70463E-01
STANDARD ERROR OF THE ESTIMATE-SIGMA = 0.26545
SUM OF SQUARED ERRORS-SSE= 5.0733
MEAN OF DEPENDENT VARIABLE = 5.2512

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 72 DF	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
LA	0.36200	0.72194E-01	5.0143	0.5087	0.66447	0.21280
LF	0.71044E-01	0.80189E-01	0.88596	0.1038	0.11740	0.37985E-01
CONSTANT	3.9343	0.13835	28.438	0.9583	0.00000E+00	0.74922

|_*aşağıdaki stop komutu programı sona erdirir.
|_STOP

VII. EKONOMETRİ PROBLEMLERİ

*7.1. Değişen Varyanslılık (Heteroskedasticity)

Şimdiye kadar yaptığımız varsayımlardan biri regresyon analizinde hata teriminin ortak bir varyansa (σ^2) sahip olduğu idi. Bu sabit varyanslılık (homoskedasticity) olarak kabul edilir. Eğer hata terimleri sabit bir varyansa sahip değil ise bu duruma da değişen varyanslılık denilmektedir. Böyle sabit varyansa sahip olmadığında sorulabilecek birçok soru vardır. Bunlar aşağıdaki gibi sıralanabilir

- Bu problem nasıl tespit edilir?
- EKK tahmincilerinin özellikleri üzerine etkileri nelerdir? EKK kullanıldığında tahmin edilen standart hatalar üzerine etkileri nelerdir?
- Bu problemin çözüm yolları nelerdir?

Uygulamalı Örnek 7.1

Elimizde çizelge 7.1'de verilen 20 aileye ait tüketim harcamaları ve gelirleri verileri mevcut olsun. Bu veriler kullanılarak EKK yöntemi ile tahmin edilen denklem de aşağıdaki gibidir. Parantez içindekiler katsayıların standart hatalardır.

Çizelge 7.1. 20 Ailenin tüketim harcamaları (y) ve yıllık gelir seviyeleri (x), bin TL

Aile	y	x	Aile	y	x
1	19,9	22,3	11	8,0	8,1
2	31,2	32,3	12	33,1	34,5
3	31,8	36,3	13	33,5	38,0
4	12,1	12,1	14	13,1	14,1
5	40,7	42,3	15	14,8	16,4
6	6,1	6,2	16	21,6	24,1
7	38,6	44,7	17	29,3	30,1
8	25,5	26,1	18	25,0	28,3
9	10,3	10,3	19	17,9	18,2
10	38,8	40,2	20	19,8	20,1

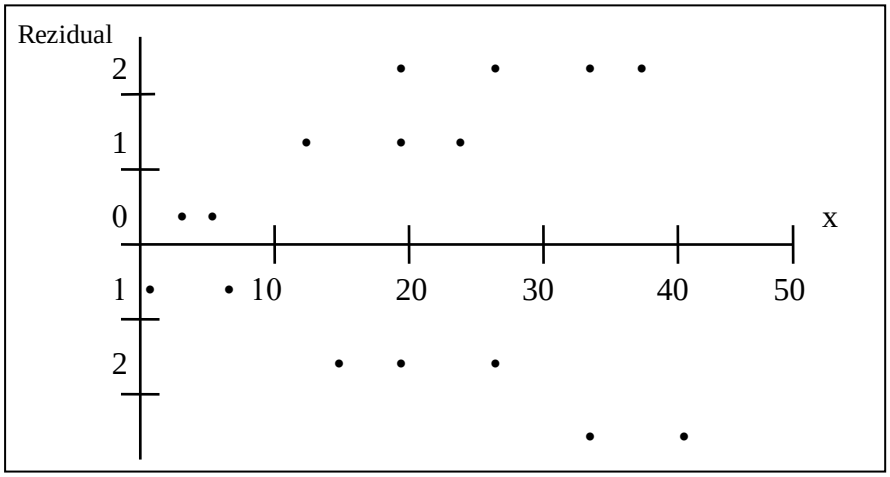
$$y = \underset{(0,703)}{0,847} + \underset{(0,0253)}{0,899}\, x \qquad R^2 = 0,986$$

Önce bu model kullanılarak kestirilen değerler ve peşinden kestirme rezidüallarını hesaplayabiliriz. Çizelge 7.2'de x değerleri küçükten büyüğe kullanılarak kestirilen y değeri ile gözlenen y değeri arasındaki fark olan kestirme rezidualları hesaplanmıştır. Görülebileceği gibi daha büyük x değerleri için bu kestirme rezidulları (hataları) daha büyüktür. Buradaki hataların varyanslarının sabit olmadığı fakat x'in değeri ile arttığı görülmektedir. Şekil 7.1'de ise bu rezidualların grafiği gösterilmiştir. Bu nokta grafiği çok varyanslılık probleminin olduğunu çizelgeden daha iyi bir şekilde göstermektedir.

Çizelge 7.2. Bir önceki çizelgedeki verilerle tahmin edilen tüketim fonksiyonuna ait kestirme rezidualları

Gözlem	x'in değeri	Rezidual	Gözlem	x'in değeri	Rezidual
6	6,2	-0,32	8	26,1	1,18
11	8,1	0,13	18	28,3	-1,30
9	10,3	0,19	17	30,1	1,38
4	12,1	0,37	2	32,3	1,30
14	14,1	-0,43	12	34,5	1,23
15	16,4	-0,80	3	36,6	-1,96
19	18,2	0,69	13	38,0	-1,52
20	20,1	0,88	10	40,2	1,80
1	22,3	-1,00	5	42,3	1,81
16	24,1	-0,92	7	44,7	-2,45

Çizelge 7.2'de verilen kestirme rezidualları grafik üzerinde gösterildiği takdirde Şekil 7.1'deki gibi bir görünüm almaktadır.



Şekil 7.1. Değişen varyanslılık örneği

Hesaplanan bu kestirme rezidualarında ve grafikte gözlenen değişen varyanslılık problemini bazen regresyonu doğrusal logaritmik formda tahmin ederek çözebiliriz. Bu doğrultuda log y'nin logx üzerine regresyonunu aşağıdaki gibi tahmin edebiliriz.

$$\log y = \underset{(0,0574)}{0,0757} + \underset{(0,0183)}{0,9562} \log x \qquad R^2 = 0,9935$$

Parantez içinde gösterilen değerler standart hatalardır. Bağımlı değişkenlerin varyansı farklı olduğundan R^2'lerin karşılaştırılmasına gerek yoktur. Bu problem ileride tartışılacaktır. Bu denklemden elde edilen reziduallar çizelge 7.3'de gösterilmiştir. Bu durumda, x arttığında reziduallarının değerinde görülür bir artış söz konusu değildir. Bu yüzden çok varyanslılık problemi görülmemektedir.

Çizelge 7.3. Logaritmik doğrusal denklemde reziduallar

Gözlem	log x	Rezidual	Gözlem	log x	Rezidual
6	1,82	-0,18	8	3,26	0,44
11	2,09	0,04	18	3,34	-0,53
9	2,33	0,27	17	3,40	0,47
4	2,49	0,34	2	3,48	0,42
14	2,65	-0,33	12	3,54	0,38
15	2,80	-0,56	3	3,60	-0,59
19	2,90	0,35	13	3,64	-0,42
20	3,00	0,41	10	3,69	0,51
1	3,10	-0,54	5	3,74	0,56
16	3,18	-0,46	7	3,80	-0,56

Her halükârda, çok varyanslılık probleminin testine ihtiyaç duyulmaktadır. Bu testlerin yapılışı ileride ele alınacaktır. Önce değişen varyanslılığın belirlenmesini öğrenelim.

Değişen Varyanslılığın Belirlenmesi

Önceki örnekte reziduallar grafik halinde verilerek x ile bir korelasyona sahip olduğu görülmüştür. Normal denklemlere göre $\hat{u}_t$ ile x_t aralarında korelesyon yok, ancak diğer tarafta $\hat{u}_t^2$ ile x_t arasında korelasyon vardır. Bu yüzden eğer regresyon yöntemi kullanarak değişen varyanslılık tespit ediyorsak $\hat{u}_t$'nin x_t^2, x_t^3, ... üzerine veya $|\hat{u}_t|$'nin veya $\hat{u}_t^2$'nin x_t, x_t^2, x_t^3 .... üzerine regresyonunu tahmin edebiliriz. Çoklu regresyon durumunda y_t'nin kestirilen değeri olan $\hat{y}_t$'nin üsleri veya açıklayıcı değişkenlerin tümünün üsleri kullanılmalıdır.

1. Anscombe tarafından önerilen test ve Ramsey tarafından önerilen RESET adı verilen testin her ikisi de $\hat{u}_t$'nin y_t^2, y_t^3 ..... üzerine regresyonunu tahmini ederek katsayıların önemli olup olmadıklarını test eder.
2. White tarafından önerilen test $\hat{u}_t^2$'nin bütün bağımsız değişkenler ve bunların çarpımları üzerine regresyonunu tahmin eder. Örneğin, x_1, x_2 ve x_3 gibi üç bağımsız değişken ile $\hat{u}_t^2$'nin x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_2x_3, ve x_3x_1 üzerine regresyonunu tahmin eder.
3. Glejser ise $|\hat{u}_i| = \alpha + \beta x_i$, $|\hat{u}_i| = \alpha + \beta / x_i$, $|\hat{u}_i| = \alpha + \beta\sqrt{x_i}$, ve benzeri tipte regresyonların tahmin edilmesini ve $\beta = 0$ hipotezinin test edilmesini önerir.

Bütün bu testlerin arkasında olan gizli varsayım, var $(\hat{u}_i) = \sigma_i^2 = \sigma^2 f(z_i)$'dir. Burada z_i bilinmeyen bir değişkendir ve değişik testler bilinmeyen f(z) fonksiyonu etrafında farklı yaklaşımları kullanırlar.

Uygulamalı Örnek 7.2

Çizelge 7.1'deki veriler için çizelge 7.2'deki rezidualleri kullanarak Ramsey, White ve Glejser testleri aşağıdaki sonuçları üretmektedir. Tek bir açıklayıcı değişken x

olduğundan, Ramsey uygulamasında y'nin yerine x kullanılabilir. Bu nedenle test, u' nin x_i^2, x_i^3 ve devamı üzerine regresyonu şeklinde olabilir. Sonuçlar, aşağıdaki gibidir.

$$\hat{u} = - 0,379 + 0,00236\ x^2 + 0,0000549\ x^3 \qquad R^2 = 0,034$$

Buradaki katsayıların hiçbirinin t değeri 1'den büyük olmaması nedeniyle sabit varyanslılık hipotezinin reddedilmesi mümkün görülmemektedir.

White tarafından önerilen test, $\hat{u}^2$'nin x, x^2, x^3 ve devamı üzerine regresyonu şeklindedir. Sonuçlar aşağıdaki gibidir.

$$\hat{u}^2 = \underset{(0,390)\quad(0,014)}{-1,370 + 0,116\ x} \qquad R^2 = 0,7911$$

$$\hat{u}^2 = \underset{(0,620)\quad(0,055)\qquad(0,0011)}{0,493 - 0,071\ x + 0,0037\ x^2} \qquad R^2 = 0,878$$

Her iki durumda da R^2'ler önemli ölçüde yüksektir. Bu nedenle test tek varyanslılık hipotezini reddetmektedir. Çok varyanslılığın düzeltilmesinde önerilen yöntem $\sigma^2 = \gamma_0 + \gamma_1 x + \gamma_2 x^2$ olduğu yerde $V(u_i) = \sigma^2$ varsayılarak regresyon modelini tahmin etmektir.

Glejser testleri aşağıdaki sonuçları vermektedir.

$$|\hat{u}| = \underset{(0,094)\quad(0,0034)}{- 0,209 + 0,0532\ x} \qquad R^2 = 0,927$$

$$|\hat{u}| = \underset{(0,186)\quad(0,037)}{- 1,232 + 0,475\ \sqrt{x}} \qquad R^2 = 0,902$$

$$|\hat{u}| = \underset{(0,155)\quad(2,39)}{1,826 - 1378\ (1/x)} \qquad R^2 = 0,649$$

R^2 değeri dikkate alındığında birinci model (Ramsey) haricinde bütün testler sabit varyanslılık hipotezini reddetmektedir. Önerilen tahmin edilecek model White testinde önerilen modelin aynısıdır.

Katsayılar yeterince istatistiksel olarak önemli olmasa bile logaritmik doğrusal form için, sonuçlar benzerdir. White testi için, çizelge 7.3'deki rezidualleri kullanarak aşağıdaki sonuçlar elde edilmektedir.

$$\hat{u}^2 = -0,211 + 0,129\ x \qquad R^2 = 0,572$$

$$\hat{u}^2 = -0,620 + 0,425\ x - 0,051\ x^2 \qquad R^2 = 0,600$$

Çok kabaca çizelge 7.3'deki reziduallere kabaca bakarak daha önceden hataların sabit varyanslı olduğu sonucuna varılmasına rağmen logaritmik doğrusal formda bile değişen varyanslılık sonucu ortaya çıkmaktadır.

Benzerlilik oranı (likelihood ratio), Goldfeld and Quandt ve Breusch ve Pagan testleri, değişen varyanslılığın test edilmesinde yaygın olarak kullanılan testlerden bazılarıdır. Genellikle ekonometri bilgisayar programları bu testlerden bir veya birkaçını yaparlar. Bu testlerden bazıları χ^2, diğer bazıları F dağılımını kullanmaktadır.

Değişen varyanslılık probleminin çözümü iki ana başlık altında ele alınabilir.

- Birincisi, verilerin logaritmasının alınarak veya başka yollarla transformasyonu yapılarak modelin transformasyonu yapılmış verilerle tahmin edilmesidir.
- İkincisi ise parametrelerin tahmin edilmesinde çoklu varyasyonun değişik yollarla giderilmesidir.

Çoğu bilgisayar programları değişen varyanslılık problemini düzelterek tahmin yapabilmektedir. Burada bahsedilen testler ve değişen varyanslılığın giderilmesi ile ilgili bilgisayar uygulamaları bu bölümün sonunda verilmektedir.

*7.2. Otokorelasyon

Bir önceki kısımda varyans ile ilgili varsayımı kaldırarak değişen varyanslılık başlığı altında ilgili analizleri yaptık. Şimdi ise diğer bir varsayım olan regresyondaki hata terimlerinin bağımsız olma Varsayımının kaldırılması ele alınacaktır. Bu varsayımı kaldırmanın sonuçları burada tartışılmaya çalışılacaktır.

Regresyon modelindeki hata terimlerinin korelasyona sahip olabilmeleri iki durumda ortaya çıkar. Yatay-kesit verilerinin birbirine komşu birimler arasında ortaya çıkar. Örneğin, eğer hane halklarının tüketim davranışları çalışılıyorsa, birbirine komşu olan hanelerin hata terimleri birbiriyle bağlantılı olarak değişebilirler. Bunun böyle olmasının sebebi, hata terimleri ihmal edilen değişkenlerin etkilerini içinde bulundururlar ve bu değişkenler, birbirine komşu olan aileler için korelasyona sahip olmaya meyillidirler. Ayrıca şehirlere ait veriler kullanılıyorsa birbirine, yakın şehirlere ait veriler arasında korelasyon olacaktır. Bu kısımda bu tip korelasyonla ilgilenilmeyecektir. Bu tip korelasyonları meydana getiren faktörler, kukla (dummy) değişkenler olarak sonraki konularda ele alınacaktır.

Bu kısımda zaman serileri ile alakalı olan korelasyonu tartışılacaktır. Bu tip korelasyona otokorelasyon veya serisel korelasyon denilmektedir. Burada t zaman periyodundaki u_t hata terimi, u_{t+1}, u_{t+2}, ...ve u_{t-1}, u_{t-2} .... gibi hata terimleri ile korelasyona sahiptir. Bu tip korelasyon ihmal edilen değişkenleri ihtiva eden regresyon modellerinin hata terimlerinde ortaya çıkar.

Burada, u_t ve u_{t-k} arasındaki korelasyon k derecesinde bir otokorelasyondur diye ifade edilirken, u_t ve u_{t-1} arasındaki korelasyon birinci dereceden bir otokorelasyondur ve ρ_1 ile ifade edilir ve ayni zamanda u_t ve u_{t-2} arasındaki korelasyon ikinci dereceden bir korelasyondur ρ_2 ile ifade edilir. Eğer n tane gözlem varsa (n-1) otokorelasyon var demektir. Fakat bütün bunların verilerden elde edilmesi düşünülemez. Bu yüzden, çoğu zaman bu (n-1) adet olan otokorelasyon bir veya iki parametre ile temsil edilir.

Burada;

1. serisel korelasyonun mevcudiyetini test etmenin nasıl yapılacağı ve
2. hatalarda otokorelasyon varken regresyonun nasıl tahmin edileceği irdelenecektir.

Durbin - Watson Testi

En basit ve en sık kullanılan model u_t'nin u_{t-1} ile korelasyonuna sahip olan hata teriminin bulunduğu modeldir. Bu model için ρ' tarafından temsil edilen ρ hakkındaki hipotezlerin test edilmesi düşünülebilir. Buradaki ρ, u_t ve u_{t-1} en küçük kareler rezidualları arasındaki korelasyondur. Bu amaç için en fazla kullanılan istatistik Durbin-Watson (DW) istatistiğidir ve aşağıdaki şekilde formüle edilebilir.

$$d = \frac{\sum_{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{n} \hat{u}_t^2}$$

Burada u_t, t periyodu için tahmin edilen rezidualdır. Burada d aşağıdaki gibi de yazılabilir.

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1} - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

Eğer örnek çok büyükse $\sum\hat{u}_t^2$ ve $\sum\hat{u}_{t-1}^2$ yaklaşık olarak birbirine eşit olduklarından $d \cong 2(1-\rho')$ olur. Eğer $\rho' = +1$ olursa $d = 0$ ve $\rho' = -1$ olduğunda $d = 4$, olur. Eğer $\rho' = 0$ ise $d = 2$ olur. Eğer d 0'a yakın veya 4'e yakın ise, reziduallar yüksek seviyede korelasyona sahiptirler anlamına gelir.

Burada, d'nin örnek dağılımı, açıklayıcı değişkenlerin değerlerine bağlıdır ve bu yüzden Durbin-Watson d değerinin önemlilik seviyeleri için üst (d_u) ve alt (d_l) sınırlar oluşturmuşlardır. Sıfır otokorelasyona karşı birinci dereceden pozitif otokorelasyon hipotezini test etmek için çizelgeler mevcuttur. Negatif korelasyonun olması durumunda d_l ve d_u'yu yer değiştirilir.

Eğer $d < d_L$, otokorelasyon vardır.

Eğer $d > d_U$, otokorelasyon yoktur.

Eğer $d_L < d < d_U$, otokorelasyon olup olmadığına karar verilemiyor.

Genellikle $\rho = 0$'a karşılık $\rho > 0$ hipotezinin yani pozitif korelasyonun testi için DW çizelgeleri kullanılır. Eğer $d > 2$ ve $\rho = 0$'a karşı $\rho < 0$ hipotezini yani negatif otokorelasyonu test edilmek isteniyorsa, sanki pozitif korelasyon için test yapıyormuş gibi 4 - d değerini alınıp çizelgeler aynen kullanılabilir.

Genelde $d \cong 2(1-\rho')$ şeklinde bir ifadenin doğru olacağının belirtilmesine rağmen, bu yaklaştırma sadece çok sayıdaki örnekler için geçerlidir. $\rho' = 0$ olduğu zaman d'nin ortalaması, yaklaşık bir değer olarak aşağıdaki denklem ile gösterilebilir.

$$E(d) \cong 2 + [\, 2(k-1) / n-k \,]$$

Burada k sabit terim dâhil regresyon parametreleri sayısını, n ise örnek sayısını ifade etmektedir. Bu yüzden 0 serisel otokorelasyon olsa bile, 2'nin üzerindeki istatistik sapmalıdır. Eğer k = 5 ve n = 15 ise, sapma 0,8 kadar büyüktür. Bir örnekle DW testinin gösterilmesi bu noktada uygun olacaktır.

Uygulamalı Örnek 7.3

Üretim, işgücü ve sermaye miktarlarını veren zaman serisi verileri kullanılarak aşağıdaki üretim fonksiyonunu tahmin edilmiştir.

$$\underset{}{\log y} = \underset{}{-3{,}938} + \underset{(0{,}237)}{1{,}451}\ \log L_1 + \underset{(0{,}083)}{0{,}384}\ \log K_2 \qquad \underset{(0{,}048)}{} \qquad R^2 = 0{,}9946 \quad DW = 0{,}88 \quad \rho = 0{,}559$$

Burada k = 2 ve n = 39 ve % 5 önem seviyesi için DW istatistiği $d_L = 1{,}38$ dir. Gözlenen $d = 0{,}88 < d_L$ olduğundan, $\rho = 0$ hipotezi % 5 önem seviyesinde reddedilmektedir.

Serisel korelasyonun testinde DW testi çok yaygın olarak kullanılmasına rağmen, birçok sınırlamalara da sahiptir.

1- Sadece birinci dereceden otokorelasyonu test eder.

2- Hesap edilen değer d_L ve d_U arasında ise test sonuçsuz kalır.

3- Gecikmeli bağımlı değişkeni olan modellerde kullanılmaz.

Bu noktada bütün eleştirileri bu seviyede cevaplamak doğru olmayacaktır. Bunun yerine otokorelasyon probleminin bazı basit çözümleri tartışılabilir.

Eğer DW testi sıfır otokorelasyon testini reddederse bir sonra yapılacak iş, bütün değişkenlerin ρ değeri ile transforme edilerek tahmin edilmesidir. Yani $y_t - \rho' y_{t-1}$'nin $x_t - \rho' x_{t-1}$ üzerine regresyonu tahmin edilebilir. Burada ρ', ρ'nun tahmin edilmiş değeridir. Fakat ρ' örnek hatalarına muhatap olduğundan eğer DW istatistiği olan d değeri küçük ise diğer bir alternatif birinci değişim denklemini kullanmaktır. Aslında kaba bir parmak hesabı ile her ne zaman DW istatistiği $< R^2$ ise birinci dereceden değişim denklemini kullanmak gerekir. Birinci derece değişim denkleminde $(y_t - y_{t-1})$ 'in $(x_t - x_{t-1})$ üzerine regresyonu alınır. Burada ki gizli varsayım, birinci derece değişim denklemlerinin hata terimlerinin bağımsız olmasıdır. Örneğin eğer,

$$y_t = \alpha + \beta x_t + u_t$$

regresyon denklemi ise o zaman,

$$y_{t-1} = \alpha + \beta x_{t-1} + u_{t-1}$$

ve eğer ikinci denklemi birinci denklemden çıkarılırsa;

$$(y_t - y_{t-1}) = \beta\,(x_t - x_{t-1}) + (u_t - u_{t-1})$$'i elde edilir.

Eğer hata terimleri bu denklemde bağımsız ise, En Küçük Kareler Yöntemi ile tahmin yapılabilir. Fakat sabit terim α çıkarma işleminden dolayı kaybolduğundan, regresyon denklemini sabit terim olmaksızın tahmin edilmelidir. Çoğu zaman birinci dereceden

farklılık denklemlerinin regresyonunda bir sabit terim olduğu görülür. Bu yöntem sadece, orijinal denklemde bir doğrusal trend terimi mevcut ise geçerlidir. Eğer regresyon denklemi,

$$y = \alpha + \delta_t + \beta x_t + u_t$$

o zaman,

$$y_{t-1} = \alpha + \delta(t-1) + \beta x_{t-1} + u_{t-1}$$

ve çıkarma işlemini yapıldığında

$$(y_t - y_{t-1}) = \delta + \beta\,(x_t - x_{t-1}) + (u_t - u_{t-1})$$

ki bu sabit değere sahip bir denklemdir.

Ekonometri bilgisayar programları hem otokorelasyon testi yani DW testi yaparak bize DW istatistiğini vermekte hem de modelde otokorelasyon olması durumunda modelde gerekli düzeltmeği yaparak modeli tahmin edebilmektedir. Bunlarla ilgili bilgisayar uygulamaları bu bölümün sonunda verilmektedir.

*7.3. Çoklu Eşdoğrusallık (Multicollinearity)

Çoğu zaman, çoklu regresyon analizinde kullanılan veriler, ortaya atılan sorulara kararlı bir cevap veremezler. Bunun sebebi çok yüksek standart hataların veya çok düşük *t* değerinin elde edilmesidir. Parametrelerin güven aralığı bu yüzden çok geniştir. Bu tip durumlar, açıklayıcı değişkenlerin çok az değişim göstermesi ve/veya aralarında yüksek korelasyon olmasıyla ortaya çıkar. Açıklayıcı değişkenler arasında çok yüksek korelasyon olduğu durum çoklu eşdoğrusallık olarak adlandırılır. Açıklayıcı değişkenler arasında çok yüksek oranda korelasyon olduğu zaman, açıklayıcı değişkenlerin açıklanan değişken üzerine olan etkisini belirlemek çok zor olmaya başlar. Bu hususta sorulacak pratik soru, açıklayıcı değişkenler arasındaki korelasyon ne kadar yüksek olmalı ki parametreler arasındaki yorumda problemler çıksın ve problemle ilgili olarak ne yapılabilir. Açıklayıcı değişkenler arasındaki yüksek korelasyon her zaman problem oluşturması gerekmiyor ve çoklu eşdoğrusallık için ileri sürülen çözümler gerçekte yanlış sonuçlara varmamıza sebep olabilir. Tavsiye edilen tedaviler bazen hastalıktan daha kötü olabilirler.

Çoklu eşdoğrusallık veya açıklayıcı değişkenler arasındaki korelasyonun bir problem olması gerekmez. Bu yüzden çoklu eşdoğrusallık problemi sadece değişkenler arasındaki korelasyon anlamında tartışılamaz. Daha ileri anlamda, değişkenlerin farklı parametrelere sahip olması bu korelasyonların farklı büyüklüklerini bize verecektir. Çoklu eşdoğrusallık problemi ve onun çözüm yolları hakkındaki tartışmaların çoğu, açıklayıcı değişkenler arasındaki korelasyona bağlı kriter temeline dayanmaktadır. Fakat verilecek örnekten anlaşılacağı gibi bu yanlış bir yaklaşımdır.

Uygulamalı Örnek 7.4

İlk önce, açıklayıcı değişkenler arasında mevcut olan yüksek korelasyon belirlenecek ve sonuçlar tartışılacaktır.

Varsayalım ki model aşağıdaki gibidir.

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

Eğer $x_2 = 2x_1$ ise,

$$y = \beta_1 x_1 + \beta_2 (2x_1) + u = (\beta_1 + 2\beta_2) x_1 + u \text{ elde edilir.}$$

Bu durumda sadece $(\beta_1+2\beta_2)$ tahmin edilebilir olmaktadır. β_1 ve β_2 parametrelerini ayrı ayrı tahmin etmek mümkün olmamaktadır. Böyle bir durum tam eş doğrusallık olarak adlandırılır. Çünkü x_1 ve x_2 arasında da tam bir korelasyon $(r^2_{12} = 1)$ vardır. Gerçek uygulamalarda r^2'nin tam 1'e eşit olduğu değil de r^2'nin 1'e yakın olduğu durumlarla karşılaşılır.

Bir örnek olması açısından aşağıdaki hesaplanmış veriler dikkate alınabilir.

$S_{11} = 200$ $\quad S_{1y} = 350$

$S_{12} = 150$ $\quad S_{2y} = 263$

$S_{22} = 113$

Buradan normal denklemlerimiz aşağıdaki gibidir.

$$200\beta'_1 + 150\beta'_2 = 350$$

$$150\beta'_1 + 113\beta'_2 = 263$$

Çözüm $\beta'_1 = 1$ ve $\beta'_2 = 1$. Varsayalım ki, örnekten bir veri çıkarıldı ve yeni değerler,

$S_{11} = 199$ $\quad S_{1y} = 347,5$

$S_{12} = 149$ $\quad S_{2y} = 261,5$

$S_{22} = 112$

ve buradan normal denklemler aşağıdaki gibi yazılabilir.

$$199\beta'_1 + 149\beta'_2 = 347,5$$

$$149\beta'_1 + 112\beta'_2 = 261,5$$

Yeni parametreler ise, $\beta'_1 = -0,5$, $\beta'_2 = 3$ olur.

Bu yüzden verilerde yapılan çok az değişmeler regresyon parametrelerinin tahmininde çok büyük değişmelere neden olmaktadır. Bu iki değişken arasındaki korelasyonun çok yüksek olduğunu aşağıdaki korelasyon denklemi sonucundan görmek mümkündür.

$$r^2_{12} = (150)^2 / 200\ (113) = 0,995$$

Pratikte, gözlem ilavesi veya çıkarılması, varyans ve kovaryanslarda değişmeye sebep olurlar. Bu yüzden x_1 ve x_2 değişkenleri arasındaki yüksek korelasyonun sonuçlarından biri parametre tahminlerinin gözlem ilave ve eksiltmelerine karşı çok duyarlı olmasıdır. Çoklu eş doğrusallık durumu, değişkenlerin ilave ve eksiltilmeleri yolu ile parametre tahminlerin duyarlılığını incelenerek kontrol edilebilir.

Çoklu eş doğrusallık probleminin diğer bir göstergesi sıkça gündeme gelir ve bu da tahmin edilen regresyon katsayılarının standart hatalarının çok yüksek olmasıdır. Fakat yüksek r^2_{12} değerinin, yüksek standart hatayı ima etmesi gerekmez ve aksine r^2_{12}'nin düşük değerleri yüksek standart hataları üretebilir. Bir önceki bölümde iki açıklayıcı değişkene karşılık olacak standart hatalar tahmin edilmişti ve bunlarla ilgili formülleri aşağıdaki gibi idi.

$$V(\beta'_1) = \sigma^2 / S_{11}(1- r^2_{12})$$

$$V(\beta'_2) = \sigma^2 / S_{22}(1- r^2_{12})$$

$$Cov(\beta'_1, \beta'_2) = - [(\sigma^2 r^2_{12}) / (S_{12}(1- r^2_{12}))]$$

burada σ^2 hata teriminin varyansıdır. Bu yüzden eğer; σ^2 yüksek, S_{11} düşük ve r^2_{12} yüksek ise β'_1'in varyansı yüksek olacaktır.

Burada r^2_{12} yüksek olsa bile eğer σ^2 düşük ve S_{11} yüksek ise yüksek standart hata problemine sahip olunmayacaktır. Diğer taraftan r^2_{12} düşük olsa bile σ^2 yüksek ve S_{11} düşük ise standart hataların yüksek olması söz konusu olacaktır. Yani sonuçta r_{12}^2 bize çoklu eş doğrusallık problemi olup olmadığı konusunda fazla bir şey söylememektedir.

Çoklu Eşdoğrusallığın Saptanması

Eğer ikiden fazla değişkene sahip olursak, açıklayıcı değişkenler arasındaki korelasyonu tespit ederken kriterimiz R_i^2 yani bütün değişkenler arasındaki korelasyon dikkate alınmalıdır. Sadece değişkenler arasındaki basit regresyonu dikkate almak doğru olmaz. Doğal olarak sadece R_i^2'yi de dikkate almak doğru olmaz. Çünkü standart hata ve *t* değerleri de ilgili faktörlerdir. Ayrıca R_i^2 model spesifikasyonuna göre değişme de gösterir.

Uygulamalı Örnek 7.5

Çoklu eşdoğrusallığın önemli bir problem olup olmadığının belirlenmesinde ve aynı zamanda tek tek parametreler tahmin edilemese bile parametrelerin bazı doğrusal fonksiyonlarını tahmin etmenin mümkün olacağını göstermede kullanmak için açıklayıcı bir örnek verilmeye çalışılacaktır.

Aşağıda çizelge 7.4'de tüketim (C), gelir (Y) ve kullanılabilir aktifler (L) ile ilgili veriler 1952-1961 yılları arası için dönemlik olarak verilmiştir. Bütün bu bilgiler 1954 yılı cinsinden milyar dolardır.

Çizelge 7.4. Tüketim, gelir ve kullanılabilir aktifler üzerine veriler

Yıl	Dönem	C	Y	L	Yıl	Dönem	C	Y	L
1952	I	220,0	238,1	182,7	1957	I	268,9	291,1	218,2
	II	222,7	240,9	183,0		II	270,4	294,6	218,5
	III	223,8	245,8	184,4		III	273,4	296,1	219,8
	IV	230,2	248,8	187,0		IV	272,1	293,3	219,5
1953	I	234,0	253,3	189,4	1958	I	268,9	291,3	220,5
	II	236,2	256,1	192,2		II	270,9	292,6	222,7
	III	236,0	255,9	193,8		III	274,4	299,9	225,0
	IV	234,1	255,9	194,8		IV	278,7	302,1	229,4
1954	I	233,4	254,4	197,3	1959	I	283,8	305,9	232,2
	II	236,4	254,8	197,0		II	289,7	312,5	235,2
	III	239,0	257,0	200,3		III	290,8	311,3	237,2
	IV	243,2	260,9	204,2		IV	292,8	313,2	237,7
1955	I	248,7	263,0	207,6	1960	I	295,4	315,4	238,0
	II	253,7	271,5	209,4		II	299,5	320,3	238,4
	III	259,9	276,5	211,1		III	298,6	321,0	240,1
	IV	261,8	281,4	213,2		IV	299,6	320,1	243,3
1956	I	263,2	282,0	214,1	1961	I	297,0	318,4	246,1
	II	263,7	286,2	286,2		II	301,6	324,8	250,0
	III	263,4	287,7	287,7		III			
	IV	266,9	291,0	291,0		IV			

Bu 38 gözlemi kullanarak aşağıdaki regresyon denklemleri elde edilmiştir. Parantez içindeki değerler *t* değerleridir.

$$C = \underset{(-1,93)}{-7,160} + \underset{(73,25)}{0,95213}\ Y \qquad r^2 = 0,9933$$

$$C = \underset{(-3,25)}{-10,627} + \underset{(9,60)}{0,68166}\ Y + \underset{(3,96)}{0,37252}\ L \qquad R^2 = 0,9933$$

$$L = \underset{(1,80)}{9,307} + \underset{(37,20)}{0,76207}\ Y \qquad r^2_{LY} = 0,9758$$

Yukarıdaki denklemde L ile Y arasında yüksek derecede korelasyon olduğunu gösteriyor. Gerçekte, L değerini Y cinsinden bu denklemden alıp bir önceki denkleme yerleştirir ve sadeleştirilerek ilk denklem elde edilir. Fakat ortadaki denklemin t değerlerine bakıldığında çoklu eşdoğrusallığın problem olmadığı söylenebilir.

Bu sonuçla model çoklu eşdoğrusallık problemi açısından aklanmış oldu. Katsayıların istikrarlı olup olmadığını bazı gözlemleri çıkararak test etmeye çalışılırsa, sadece 36 gözlem kullanıldığında aşağıdaki sonuçlar elde edilir.

$$C = \underset{(-1,74)}{-6,980} + \underset{(67,04)}{0,95145}\ Y \qquad r^2 = 0,9925$$

$$C = \underset{(-3,71)}{-13,391} + \underset{(8,32)}{0,63258}\ Y + \underset{(4,24)}{0,45065}\ L \qquad R^2 = 0,9951$$

$$L = \underset{(2,69)}{14,255} + \underset{(37,80)}{0,70758}\ Y \qquad r^2_{LY} = 0,9768$$

İlk iki denklem bir önceki denklemlerden ortadaki ile karşılaştırdığında ikinci denklemin birinciye göre daha çok değişiklik gösterdiği ortaya çıkmaktadır. 4. bölümde bahsedilen istikrar testleri uygulanırsa sonuçların % 5 önem seviyesinde önemli olmadığı ortaya çıkacaktır.

Son olarak C'nin kestirme değerini dikkate alınabilir. 1961 yılının iki dönemini alıp yukarıdaki ilk iki denklemi kullanarak kestirmeler, aşağıdaki gibi tahmin edilebilir.

Yıl	Dönem	İlk Denklem	İkinci Denklem
1961	I	295,96	298,93
	II	302,05	304,73

Bu yüzden L'nin modelde bulundurulduğu durumdaki kestirme değerleri L'nin model içinde bulundurulmaması durumundaki modelinden elde edilen değerden daha yüksek olduğu görülmektedir. Eğer kestirme tek kriter olarak ortaya çıkıyorsa L değişkeni model dışında bırakılabilir.

Yukarıdaki örnek, özet olarak çoklu korelasyon problemine bakmanın dört farklı yolunu vermektedir. Bunlar,

1. Açıklayıcı değişkenler arasında korelasyonun çok yüksek olması, çoklu eşdoğrusallığın önemli yani ciddi bir problem olduğunu ima edebilir. Fakat önceden de açıklandığı gibi sadece değişkenler arasındaki korelasyon katsayısına bakmak doğru değildir.
2. Diğer bir kriter *t* değerleri ve standart hata değerleridir. Buradaki örnekte *t* değerlerinin yüksek olması çoklu eşdoğrusallığın önemli bir problem olmadığını göstermektedir.
3. Bazı gözlemlerin çıkarılması ile katsayılarda bir değişmenin olmaması, yani katsayılarda istikrarın olması durumunda çoklu eşdoğrusallığın önemli ve ciddi bir problem olmadığını göstermektedir.
4. Modelden elde edilen kestirmelerin incelenmesi diğer bir kriterdir Eğer çoklu eşdoğrusallık önemli ve ciddi bir problem ise modelden elde edilen kestirmeler bu modelin açıklayıcı değişkenlerinin bir alt grubunu içeren modelden elde edilen kestirmelerden daha kötü olacaktır.

En son kriter, eğer kestirme yapılan analizin amacı ise kullanılabilir. Diğer durumlarda ikinci ve üçüncü kriter daha kullanışlıdır. İlk kriter ise çok kullanışlı değildir ve bu defalarca ifade edilmiştir.

Çoklu Eşdoğrusallığın Çözüm Yolları

Çoklu eşdoğrusallık ile ilgili araştırmacının yapacağı iki temel alternatif vardır:

1. Hiç bir şey yapmamak

Verilerde çoklu eşdoğrusallığın olması, araştırmacının ilgilendiği katsayı tahminlerinin kabul edilemeyecek seviyede yüksek varyasyona sahip olmasına sebep olması

gerekmez. Bunun klasik örneği, Cobb-Douglas üretim fonksiyonunun tahminidir. Sermaye ve işgücü üretim faktörleri yüksek seviyede eş doğrusal olmasına rağmen, parametrelerin iyi tahmin edilmediği söylenemez. Bu söylenenler aşağıdaki sonuçları vermektedir.

a. Eğer R_y^2 değeri yani modelin açıklama gücü, açıklayıcı değişkenler arasındaki korelasyon olan r_i^2'den büyük ise mesele yoktur.

b. *t* istatistikleri 2'den büyük ise çoklu korelasyon hakkında üzülmemek gerekir.

2. İlave bilgileri kullanmaya çalışmak

a. Daha çok veri sağlamak

b. Bazı parametreler arasında bir ilişki belirlemek

c. Bir değişkeni modelden çıkarmak

d. Diğer çalışmalardan elde edilen tahminleri kullanmak

YAZILIM UYGULAMASI

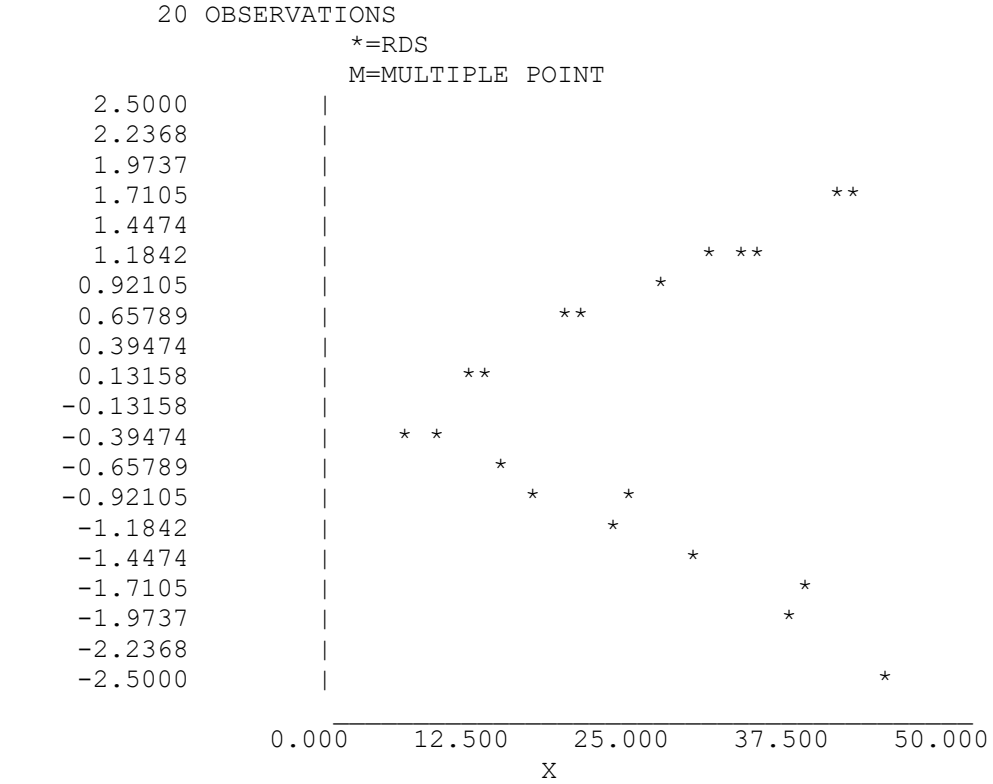
```
**************************************************************************
   SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
   ** Copyright (C)  1999  by K.J. White - All Rights Reserved **

   FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
   AT: ATATURK UNIVERSITESI - ERZURUM

   FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
**************************************************************************
```

Uygulamalı Örnek 7.1, sayfa 107

```
**************************************************************************
|_SAMPLE 1 20
|_READ (MTABLE5.1) OBS Y X
 UNIT 88 IS NOW ASSIGNED TO: MTABLE5.1
    3 VARIABLES AND       20 OBSERVATIONS STARTING AT OBS       1
|_*Şimdi önce OLS'i kullanarak tahminleri yapalım ve ve belirlenen
|_*rezidualleri RSD= opsiyonunu kullanarak RDS değişkeni olarak kaydedelim
|_OLS Y X / RESID=RDS
        20 OBSERVATIONS     DEPENDENT VARIABLE= Y
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9859     R-SQUARE ADJUSTED =   0.9852
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   1.7263
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.3139
 SUM OF SQUARED ERRORS-SSE=   31.074
 MEAN OF DEPENDENT VARIABLE =   23.555
 VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X         0.89932     0.2531E-01   35.53     0.000 0.993     0.9929     0.9640
 CONSTANT  0.84705     0.7034       1.204     0.244 0.273     0.0000     0.0360
|_* Burada şekil 7.1'i Shazamı kullanarak çizelim.
|_PLOT RDS X
 REQUIRED MEMORY IS PAR=       1 CURRENT PAR=    1000
 FOR MAXIMUM EFFICIENCY USE AT LEAST PAR=       2
```




```
|_* Şimdi Çizelge 7.2'yi SHAZAM yoluyla burada oluşturalım
|_PRINT OBS Y X RDS
```

OBS	Y	X	RDS
1.000000	19.90000	22.30000	-1.001992
2.000000	31.20000	32.30000	1.304761
3.000000	31.80000	36.60000	-1.962335
4.000000	12.10000	12.10000	0.3711196
5.000000	40.70000	42.30000	1.811514
6.000000	6.100000	6.200000	-0.3228647
7.000000	38.60000	44.70000	-2.446865
8.000000	25.50000	26.10000	1.180574
9.000000	10.30000	10.30000	0.1899041
10.00000	38.80000	40.20000	1.800096
11.00000	8.000000	8.100000	-0.1315816
12.00000	33.10000	34.50000	1.226247
13.00000	33.50000	38.00000	-1.521390
14.00000	13.10000	14.10000	-0.4275297
15.00000	14.80000	16.40000	-0.7959765
16.00000	21.60000	24.10000	-0.9207766
17.00000	29.30000	30.10000	1.383275
18.00000	25.00000	28.30000	-1.297940
19.00000	17.90000	18.20000	0.6852390
20.00000	19.80000	20.10000	0.8765221

```
|_*SORT komutu listedeki ilk değişkenin derecelemesine göre değişkenleri
|_*artan bir şekilde düzenler. Eğer SORT komutundan önce artan veya azalan
|_*şekilde düzenlenen bir değişken var idiyse, veriler bir kere
|_*düzenlendikten sonra ancak geri orijinal durumuna düzenlenir. Bu
|_*nedenle, verilerin tekrar geri düzenlenebilmesi ihtimali için TIME=(0)
|_*fonksiyonu kullanılarak T değişkeni oluşturulur.
|_GENR T=TIME(0)
|_SORT X OBS RDS Y T
```

```
 DATA HAS BEEN SORTED BY VARIABLE X
|_* Şimdi çizelge 7.2'yi yeniden SHORT komutundan sonra tekrar oluşturalım.
|_PRINT OBS X RDS Y
       OBS             X              RDS             Y
    6.000000       6.200000     -0.3228647       6.100000
    11.00000       8.100000     -0.1315816       8.000000
    9.000000       10.30000      0.1899041       10.30000
    4.000000       12.10000      0.3711196       12.10000
    14.00000       14.10000     -0.4275297       13.10000
    15.00000       16.40000     -0.7959765       14.80000
    19.00000       18.20000      0.6852390       17.90000
    20.00000       20.10000      0.8765221       19.80000
    1.000000       22.30000      -1.001992       19.90000
    16.00000       24.10000     -0.9207766       21.60000
    8.000000       26.10000       1.180574       25.50000
    18.00000       28.30000      -1.297940       25.00000
    17.00000       30.10000       1.383275       29.30000
    2.000000       32.30000       1.304761       31.20000
    12.00000       34.50000       1.226247       33.10000
    3.000000       36.60000      -1.962335       31.80000
    13.00000       38.00000      -1.521390       33.50000
    10.00000       40.20000       1.800096       38.80000
    5.000000       42.30000       1.811514       40.70000
    7.000000       44.70000      -2.446865       38.60000

|_*Şimdi mevcut çok varyanslılık problemini, regresyon modelini iki taraflı
|_*logaritmik formda tahmin ederek ortadan kaldıralım.

|_GENR LNY=LOG(Y)
|_GENR LNX=LOG(X)
|_OLS LNY LNX / LOGLOG RESID=NEWRDS
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNY
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9935     R-SQUARE ADJUSTED =   0.9931
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.20875E-02
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.45689E-01
 SUM OF SQUARED ERRORS-SSE=  0.37575E-01
 MEAN OF DEPENDENT VARIABLE =   3.0339

VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNX       0.95619     0.1825E-01   52.38     0.000 0.997     0.9967     0.9562
 CONSTANT  0.75672E-01 0.5739E-01   1.318     0.204 0.297     0.0000     0.0757
|_*Şimdi yeni reziduallerle 7.1 grafiğini tekrar çizelim

|_PLOT NEWRDS X
FOR MAXIMUM EFFICIENCY USE AT LEAST PAR=       2
        20 OBSERVATIONS
                    *=NEWRDS
                    M=MULTIPLE POINT
   0.60000E-01    |
   0.53684E-01    |
   0.47368E-01    |                                **
   0.41053E-01    |                     *  * *
   0.34737E-01    |               **           *
   0.28421E-01    |          *
   0.22105E-01    |         *
   0.15789E-01    |
```

```
   0.94737E-02    |
   0.31579E-02    |      *
  -0.31579E-02    |
  -0.94737E-02    |
  -0.15789E-01    |    *
  -0.22105E-01    |
  -0.28421E-01    |
  -0.34737E-01    |          *
  -0.41053E-01    |
  -0.47368E-01    |                  *          *
  -0.53684E-01    |                 *    *
  -0.60000E-01    |            *               *     *
                   ____________________________________________

              0.000    12.500    25.000    37.500    50.000
                                   X
|_*Şimdi çizelge 7.2'yi yeni reziduallerle oluşturalım
|_PRINT OBS X NEWRDS Y
       OBS            X             NEWRDS         Y
    6.000000       6.200000     -0.1199287E-01   6.100000
    11.00000       8.100000      0.3557127E-02   8.000000
    9.000000       10.30000      0.2650715E-01   10.30000
    4.000000       12.10000      0.3356382E-01   12.10000
    14.00000       14.10000     -0.3329663E-01   13.10000
    15.00000       16.40000     -0.5576771E-01   14.80000
    19.00000       18.20000      0.3482832E-01   17.90000
    20.00000       20.10000      0.4076192E-01   19.80000
    1.000000       22.30000     -0.5351637E-01   19.90000
    16.00000       24.10000     -0.4576692E-01   21.60000
    8.000000       26.10000      0.4398771E-01   25.50000
    18.00000       28.30000     -0.5319574E-01   25.00000
    17.00000       30.10000      0.4655427E-01   29.30000
    2.000000       32.30000      0.4193349E-01   31.20000
    12.00000       34.50000      0.3804354E-01   33.10000
    3.000000       36.60000     -0.5852347E-01   31.80000
    13.00000       38.00000     -0.4233757E-01   33.50000
    10.00000       40.20000      0.5072226E-01   38.80000
    5.000000       42.30000      0.4984100E-01   40.70000
    7.000000       44.70000     -0.5590332E-01   38.60000
|_STOP
*******************************************************************************
```

Uygulamalı Örnek 7.2, sayfa 109

```
*******************************************************************************
|_* Şimdi burada RAMSEY RESET testini uygulayalım.
|_GENR X2=X**2
|_GENR X3=X**3
|_OLS RDS X2 X3
        20 OBSERVATIONS     DEPENDENT VARIABLE= RDS
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.0337     R-SQUARE ADJUSTED =  -0.0799
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   1.7662
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.3290
 SUM OF SQUARED ERRORS-SSE=   30.026
 MEAN OF DEPENDENT VARIABLE = -0.37748E-15
 VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
```

```
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X2        0.23613E-02 0.3218E-02  0.7338     0.473 0.175     1.1363**********
 X3       -0.54878E-04 0.7210E-04 -0.7612     0.457-0.182    -1.1786**********
 CONSTANT -0.37922     0.7389     -0.5132     0.614-0.124     0.0000**********
|_*DIAGNOS komutu ve RESET opsiyonu, YHAT**2; YHAT**2 and YHAT**3; and
|_*YHAT**2, YHAT**3 and YHAT**4 ile otomatik olarak üç regresyon tahmini
|_*yapmaktadır. F test istatistikleri hesaplanır ve eğer onlar anlamlı ise
|_*yanlış model spesifikasyonu var olduğunu gösterir. (Buradaki örnekte F
|_*istatistikleri önemli değildir). Sadece bir açıklayıcı değişken
|_*olduğundan, ders notlarında mümkün olan x'in kullanılması seçilmiştir.
|_*Bu nedenle, YHAT=a+bX ikamesi yapılabilir. Hatırlayalım ki ? öneki
|_*çıktının yazılmasını engeller. DIAGNOS komutu ve RESET opsiyonu otomatik
|_*olarak üç yardımcıyı çalıştırır.
|_?OLS Y X
|_DIAGNOS / RESET
DEPENDENT VARIABLE = Y                20 OBSERVATIONS
 REGRESSION COEFFICIENTS
   0.899324687854      0.847051631695
 RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
  RESET(2)=  0.34370     - F WITH DF1=   1 AND DF2=  17 P-VALUE= 0.565
  RESET(3)=  0.39957     - F WITH DF1=   2 AND DF2=  16 P-VALUE= 0.677
  RESET(4)=  0.45003     - F WITH DF1=   3 AND DF2=  15 P-VALUE= 0.721
|_* Şimdi WHITE testi uygulayalım.
|_GENR RDS2=RDS**2
|_OLS RDS2 X
        20 OBSERVATIONS     DEPENDENT VARIABLE= RDS2
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.7921     R-SQUARE ADJUSTED =   0.7806
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.52695
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.72591
 SUM OF SQUARED ERRORS-SSE=   9.4851
 MEAN OF DEPENDENT VARIABLE =   1.5537
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X         0.11580     0.1398E-01   8.282     0.000 0.890     0.8900     1.8820
 CONSTANT  -1.3703     0.3886      -3.526     0.002-0.639     0.0000    -0.8820
|_OLS RDS2 X X2
        20 OBSERVATIONS     DEPENDENT VARIABLE= RDS2
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.8781     R-SQUARE ADJUSTED =   0.8638
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.32713
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.57195
 SUM OF SQUARED ERRORS-SSE=   5.5612
 MEAN OF DEPENDENT VARIABLE =   1.5537
 LOG OF THE LIKELIHOOD FUNCTION = -15.5796
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X        -0.70145E-01 0.5481E-01  -1.280     0.218-0.296    -0.5391    -1.1400
 X2        0.36738E-02 0.1061E-02   3.463     0.003 0.643     1.4589     1.8262
 CONSTANT  0.48753     0.6177      0.7893     0.441 0.188     0.0000     0.3138
|_* Şimdi Glejser testini hesap edelim. ABS(x) fonksiyonu mutlak değerleri
|_*hesaplar.
|_GENR ABSRD=ABS(RDS)
|_GENR SQRTX=SQRT(X)
|_GENR INVX=1/X
|_OLS ABSRD X
```

```
        20 OBSERVATIONS     DEPENDENT VARIABLE= ABSRD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9271     R-SQUARE ADJUSTED =   0.9230
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.30865E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.17569
 SUM OF SQUARED ERRORS-SSE=  0.55557
 MEAN OF DEPENDENT VARIABLE =   1.0829
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X         0.51195E-01 0.3384E-02   15.13     0.000 0.963     0.9629     1.1937
 CONSTANT -0.20976     0.9405E-01  -2.230     0.039-0.465     0.0000    -0.1937
|_OLS ABSRD SQRTX
        20 OBSERVATIONS     DEPENDENT VARIABLE= ABSRD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9018     R-SQUARE ADJUSTED =   0.8963
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.41566E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.20388
 SUM OF SQUARED ERRORS-SSE=  0.74818
 MEAN OF DEPENDENT VARIABLE =   1.0829
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 SQRTX     0.47532     0.3697E-01   12.86     0.000 0.950     0.9496     2.1381
 CONSTANT  -1.2325     0.1858      -6.635     0.000-0.842     0.0000    -1.1381
|_OLS ABSRD INVX
        20 OBSERVATIONS     DEPENDENT VARIABLE= ABSRD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.6480     R-SQUARE ADJUSTED =   0.6284
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.14901
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.38601
 SUM OF SQUARED ERRORS-SSE=   2.6821
 MEAN OF DEPENDENT VARIABLE =   1.0829
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 INVX      -13.770      2.392      -5.756     0.000-0.805    -0.8050    -0.6859
 CONSTANT   1.8256     0.1552       11.76     0.000 0.941     0.0000     1.6859
|_* Şimdi Glejser testini iki taraflı logaritmik model üzerinde yapalım.
|_GENR ABSNEWRD=ABS(NEWRDS)
|_GENR SQRTLNX=SQRT(LNX)
|_GENR INVLNX=1/LNX
|_OLS ABSNEWRD LNX
        20 OBSERVATIONS     DEPENDENT VARIABLE= ABSNEWRD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.6560     R-SQUARE ADJUSTED =   0.6369
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.74637E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.86393E-02
 SUM OF SQUARED ERRORS-SSE=  0.13435E-02
 MEAN OF DEPENDENT VARIABLE =  0.41030E-01
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNX       0.20223E-01 0.3452E-02   5.859     0.000 0.810     0.8099     1.5249
 CONSTANT -0.21536E-01 0.1085E-01  -1.984     0.063-0.424     0.0000    -0.5249
|_OLS ABSNEWRD SQRTLNX
        20 OBSERVATIONS     DEPENDENT VARIABLE= ABSNEWRD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.6778     R-SQUARE ADJUSTED =   0.6599
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.69916E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.83616E-02
```

```
 SUM OF SQUARED ERRORS-SSE=  0.12585E-02
 MEAN OF DEPENDENT VARIABLE =  0.41030E-01
VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 SQRTLNX   0.69107E-01 0.1123E-01   6.153     0.000 0.823     0.8233     2.9493
 CONSTANT -0.79978E-01 0.1976E-01  -4.048     0.001-0.690     0.0000    -1.9493
|_OLS ABSNEWRD INVLNX
        20 OBSERVATIONS     DEPENDENT VARIABLE= ABSNEWRD
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.7208     R-SQUARE ADJUSTED =   0.7053
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.60578E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.77832E-02
 SUM OF SQUARED ERRORS-SSE=  0.10904E-02
 MEAN OF DEPENDENT VARIABLE =  0.41030E-01
VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 INVLNX   -0.15881     0.2330E-01  -6.817     0.000-0.849    -0.8490    -1.3022
 CONSTANT  0.94460E-01 0.8029E-02   11.77     0.000 0.941     0.0000     2.3022
```

DIAGNOS komutu ve HET opsiyonu çok varyanslılık için bir Glejser testini de içine alan bir seri test yapar. Eğer χ^2 istatistikleri önemli ise çok varyanslılık için H_0 hipotezi reddedilir.

```
|_?OLS Y X
|_DIAGNOS / HET
DEPENDENT VARIABLE = Y                20 OBSERVATIONS
 REGRESSION COEFFICIENTS
   0.899324687854      0.847051631695
HETEROSKEDASTICITY TESTS
                             CHI-SQUARE     D.F.   P-VALUE
                           TEST STATISTIC
 E**2 ON YHAT:                    15.842     1    0.00007
 E**2 ON YHAT**2:                 17.298     1    0.00003
 E**2 ON LOG(YHAT**2):            12.915     1    0.00033
 E**2 ON X (B-P-G) TEST:
          BASED ON R2:            15.842     1    0.00007
          BASED ON SSR:            7.486     1    0.00622
 E**2 ON LAG(E**2) ARCH TEST:     12.412     1    0.00043
 LOG(E**2) ON X (HARVEY) TEST:     8.160     1    0.00428
 ABS(E) ON X (GLEJSER) TEST:      12.511     1    0.00040
|_*Şimdi iki taraflı logaritmik model üzerine WHITE testi yapalım.
|_GENR NEWRDS2=NEWRDS**2
|_GENR LNX2=LNX**2
|_OLS NEWRDS2 LNX
        20 OBSERVATIONS     DEPENDENT VARIABLE= NEWRDS2
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.5804     R-SQUARE ADJUSTED =   0.5571
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.42079E-06
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.64868E-03
 SUM OF SQUARED ERRORS-SSE=  0.75742E-05
 MEAN OF DEPENDENT VARIABLE =  0.18787E-02
 LOG OF THE LIKELIHOOD FUNCTION =  119.486
VARIABLE   ESTIMATED  STANDARD   T-RATIO       PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      18 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNX       0.12932E-02 0.2592E-03   4.989     0.000 0.762     0.7618     2.1295
 CONSTANT -0.21220E-02 0.8149E-03  -2.604     0.018-0.523     0.0000    -1.1295
|_OLS NEWRDS2 LNX LNX2
        20 OBSERVATIONS     DEPENDENT VARIABLE= NEWRDS2
```

```
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.6074     R-SQUARE ADJUSTED =   0.5612
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.41687E-06
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.64566E-03
 SUM OF SQUARED ERRORS-SSE=  0.70869E-05
 MEAN OF DEPENDENT VARIABLE =  0.18787E-02
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNX       0.41915E-02 0.2693E-02   1.556     0.138 0.353     2.4693     6.9023
 LNX2     -0.50074E-03 0.4631E-03  -1.081     0.295-0.254    -1.7154    -2.6346
 CONSTANT -0.61392E-02 0.3803E-02  -1.614     0.125-0.365     0.0000    -3.2678
|_* Şimdi DIAGNOS komutunu ve HET opsiyonunu iki taraflı logaritmik model
|_*üzerinde kullanalım.
|_?OLS LNY LNX
|_DIAGNOS / HET
DEPENDENT VARIABLE = LNY             20 OBSERVATIONS
 REGRESSION COEFFICIENTS
   0.956186496664      0.756722465062E-01
 HETEROSKEDASTICITY TESTS
                            CHI-SQUARE     D.F.   P-VALUE
                           TEST STATISTIC
 E**2 ON YHAT:                    11.607     1    0.00066
 E**2 ON YHAT**2:                 11.046     1    0.00089
 E**2 ON LOG(YHAT**2):            11.980     1    0.00054
 E**2 ON X (B-P-G) TEST:
           BASED ON R2:           11.607     1    0.00066
           BASED ON SSR:           1.484     1    0.22316
 E**2 ON LAG(E**2) ARCH TEST:      2.854     1    0.09117
 LOG(E**2) ON X (HARVEY) TEST:     3.595     1    0.05795
 ABS(E) ON X (GLEJSER) TEST:       3.753     1    0.05272
|_STOP
*******************************************************************************
```

Uygulamalı Örnek 7.3, sayfa 113

```
*******************************************************************************
|_SAMPLE 1 39
|_READ (MTABLE3.11) YEAR X L1 L2 K1 K2
 UNIT 88 IS NOW ASSIGNED TO: MTABLE3.11
    6 VARIABLES AND       39 OBSERVATIONS STARTING AT OBS       1
|_GENR LNX=LOG(X)
|_GENR LNL1=LOG(L1)
|_GENR LNK1=LOG(K1)
|_*LIST opsiyonu reziduallerin grafiğini çizmek için kullanılır. LIST
|_*opsiyonu otomatik olarak rho'nun tahminini ve otokorelasyon olup
|_*olmadığını belirleyen D-W istatistiğini içeren RSTAT istatistiklerini
|_*hesaplar. Ders notlarında tahmin edilen reziduallerin grafiği olmamasına
|_*rağmen, LIST opsiyonu otokorelasyon problemini daha iyi kavramak için
|_*kullanılabilir.
|_OLS LNX LNL1 LNK1 / LOGLOG LIST
  R-SQUARE =   0.9946     R-SQUARE ADJUSTED =   0.9943
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.12051E-02
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.34714E-01
 SUM OF SQUARED ERRORS-SSE=  0.43382E-01
 MEAN OF DEPENDENT VARIABLE =   5.6874
 LOG OF THE LIKELIHOOD FUNCTION(IF DEPVAR LOG) = -144.524
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
```

```
   NAME    COEFFICIENT   ERROR      36 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNL1       1.4508     0.8323E-01   17.43     0.000 0.946     0.6911     1.4508
 LNK1      0.38381     0.4802E-01   7.993     0.000 0.800     0.3169     0.3838
 CONSTANT  -3.9377     0.2370      -16.61     0.000-0.941     0.0000    -3.9377
     OBS.   OBSERVED     PREDICTED   CALCULATED
      NO.    VALUE        VALUE       RESIDUAL
       1    5.2460       5.2587     -0.12717E-01              * I
       2    5.1481       5.1910     -0.42922E-01           *    I
       3    5.0695       5.1094     -0.39913E-01           *    I
       4    4.9097       4.9227     -0.12966E-01               * I
       5    4.8828       4.9132     -0.30424E-01             *   I
       6    4.9544       4.9743     -0.19924E-01              *  I
       7    5.0363       5.0443     -0.79474E-02                *I
       8    5.1446       5.1342      0.10398E-01                I*
       9    5.2095       5.2121     -0.25697E-02                *
      10    5.1544       5.1090      0.45466E-01                I      *
      11    5.2391       5.1897      0.49349E-01                I      *
      12    5.3254       5.2817      0.43727E-01                I     *
      13    5.4638       5.4454      0.18394E-01                I  *
      14    5.5522       5.5696     -0.17386E-01              *  I
      15    5.6258       5.6246      0.12534E-02                *
      16    5.6737       5.6078      0.65858E-01                I         *
      17    5.6507       5.5396      0.11109                    I          X
      18    5.6131       5.5997      0.13431E-01                I *
      19    5.6344       5.7003     -0.65904E-01        *       I
      20    5.6958       5.7548     -0.59036E-01        *       I
      21    5.6961       5.7166     -0.20556E-01             *  I
      22    5.7958       5.7946      0.11954E-02               *
      23    5.8619       5.8887     -0.26797E-01            *   I
      24    5.8872       5.9196     -0.32365E-01           *    I
      25    5.9373       5.9610     -0.23757E-01             *  I
      26    5.9291       5.9180      0.11042E-01                I *
      27    6.0081       5.9933      0.14802E-01                I *
      28    6.0314       6.0466     -0.15150E-01               * I
      29    6.0469       6.0568     -0.99502E-02                *I
      30    6.0364       6.0221      0.14315E-01                I *
      31    6.0996       6.0993      0.34900E-03                *
      32    6.1253       6.1392     -0.13896E-01               * I
      33    6.1448       6.1410      0.37809E-02                I*
      34    6.2052       6.1930      0.12150E-01                I *
      35    6.2451       6.2201      0.25059E-01                I   *
      36    6.2991       6.2760      0.23108E-01                I  *
      37    6.3616       6.3431      0.18540E-01                I  *
      38    6.4226       6.4220      0.58422E-03                *
      39    6.4475       6.4772     -0.29713E-01             *   I
DURBIN-WATSON = 0.8581    VON NEUMANN RATIO = 0.8807    RHO =  0.57053
 RESIDUAL SUM = -0.17944E-13  RESIDUAL VARIANCE =  0.12051E-02
 SUM OF ABSOLUTE ERRORS=  0.96779
 R-SQUARE BETWEEN OBSERVED AND PREDICTED = 0.9946
 R-SQUARE BETWEEN ANTILOGS OBSERVED AND PREDICTED = 0.9952
 RUNS TEST: 15 RUNS,  20 POS,  0 ZERO,  19 NEG  NORMAL STATISTIC = -1.7821
|_* İlerde kullanmak için bazı sonuçlar kaydedilebilir.

|_GEN1 N=$N
 ..NOTE..CURRENT VALUE OF $N   =   39.000
```

```
|_GEN1 K=$K
 ..NOTE..CURRENT VALUE OF $K    =    3.0000
|_GEN1 DWo=$DW
 ..NOTE..CURRENT VALUE OF $DW   =  0.85808
|_GEN1 RSSo=$SSE
 ..NOTE..CURRENT VALUE OF $SSE =  0.43382E-01
|_*Eğer D-W istatistiği "sonuçsuz alana" düşüyor ise, tam D-W ihtimalini yani
|_*D-W istatistiği için p değerini bulmak için EXACTDW opsiyonukullanılabilir.
********************************************************************************
```

Uygulamalı Örnek 7.5, sayfa 117

```
********************************************************************************
|_* Şimdi üç regresyon tahmin edelim
|_SAMPLE 1 38
|_READ (MTABLE7.1) YEAR C Y L
    4 VARIABLES AND       38 OBSERVATIONS STARTING AT OBS       1
|_OLS C Y
        38 OBSERVATIONS     DEPENDENT VARIABLE= C
 ...NOTE..SAMPLE RANGE SET TO:      1,     38
  R-SQUARE =   0.9933     R-SQUARE ADJUSTED =   0.9931
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   4.3091
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   2.0758
 SUM OF SQUARED ERRORS-SSE=   155.13
 MEAN OF DEPENDENT VARIABLE =   263.07
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      36 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 Y         0.95213     0.1300E-01   73.25     0.000 0.997     0.9967     1.0272
 CONSTANT  -7.1597      3.705      -1.933     0.061-0.307     0.0000    -0.0272
|_OLS C Y L
  R-SQUARE =   0.9953     R-SQUARE ADJUSTED =   0.9951
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   3.1099
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.7635
 SUM OF SQUARED ERRORS-SSE=   108.85
 MEAN OF DEPENDENT VARIABLE =   263.07
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      35 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 Y         0.68166     0.7098E-01   9.604     0.000 0.851     0.7135     0.7354
 L         0.37252     0.9656E-01   3.858     0.000 0.546     0.2866     0.3050
 CONSTANT  -10.627      3.273      -3.247     0.003-0.481     0.0000    -0.0404
|_OLS L Y
  R-SQUARE =   0.9758     R-SQUARE ADJUSTED =   0.9751
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   9.2643
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   3.0437
 SUM OF SQUARED ERRORS-SSE=   333.52
 MEAN OF DEPENDENT VARIABLE =   215.38
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      36 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 Y         0.72607     0.1906E-01   38.09     0.000 0.988     0.9878     0.9568
 CONSTANT   9.3070      5.432       1.713     0.095 0.275     0.0000     0.0432
|_*Dikkat edilirse tahmin edilen katsayılar ders notlarındakilere
|_*uymaktadır fakat t değerleri biraz farklı çıkmaktadır. PCOR opsiyonu
|_*listelenen korelasyon matrisinin yazılmasını sağlar. COR= opsiyonu
|_*korelasyon matrisini, marris değişkeni RLY şeklinde kaydeder.
|_STAT L Y / PCOR COR=RLY
 NAME         N    MEAN        ST. DEV      VARIANCE     MINIMUM     MAXIMUM
 L             38   215.38       19.297      372.36       182.70       250.00
 Y             38   283.82       26.253      689.24       238.10       324.80
```

```
  CORRELATION MATRIX OF VARIABLES -       38 OBSERVATIONS
 L          1.0000
 Y         0.98782       1.0000
              L             Y
|_*MATRIX komutu GNR komutuna benzemektedir fakat burada oluşturulan ve
|_*üzerinde değişiklik yapılabilecek olan bir vektor değil matristir.
|_*Parantez içindeki rakamlar matrisin elamanlarını göstermektedir. Yani
|_*korelasyon katsayısı korelasyon matrisinin ilk sütunu ve ikinci
|_*satırındadır.
|_MATRIX R2LY=RLY(1,2)**2
|_PRINT R2LY
     R2LY   0.9757927
|_*R2LY'ın yukarıda rapor edilen L'nin Y üzerine regresyonundan elde edilen
|_*R² istatistiğ ile benzerlik göstermektedir. Şimdi bazı gözlemleri
|_*çıkararak kestirme performansını değerlendirelim.
|_SAMPLE 1 36
|_OLS C Y
        36 OBSERVATIONS     DEPENDENT VARIABLE= C
 ...NOTE..SAMPLE RANGE SET TO:      1,     36
  R-SQUARE =   0.9925     R-SQUARE ADJUSTED =   0.9923
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   4.5255
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   2.1273
 SUM OF SQUARED ERRORS-SSE=   153.87
 MEAN OF DEPENDENT VARIABLE =   261.06
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      34 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 Y         0.95145     0.1419E-01   67.04     0.000 0.996     0.9962     1.0267
 CONSTANT  -6.9805      4.014      -1.739     0.091-0.286     0.0000    -0.0267
|_OLS C Y L / COEF=BETA2
        36 OBSERVATIONS     DEPENDENT VARIABLE= C
 ...NOTE..SAMPLE RANGE SET TO:      1,     36
  R-SQUARE =   0.9951     R-SQUARE ADJUSTED =   0.9948
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   3.0157
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.7366
 SUM OF SQUARED ERRORS-SSE=   99.519
 MEAN OF DEPENDENT VARIABLE =   261.06
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      33 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 Y         0.63258     0.7600E-01   8.323     0.000 0.823     0.6624     0.6826
 L         0.45065     0.1062       4.245     0.000 0.594     0.3378     0.3687
 CONSTANT  -13.391      3.608      -3.712     0.001-0.543     0.0000    -0.0513
|_OLS L Y
        36 OBSERVATIONS     DEPENDENT VARIABLE= L
 ...NOTE..SAMPLE RANGE SET TO:      1,     36
  R-SQUARE =   0.9768     R-SQUARE ADJUSTED =   0.9761
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   7.8709
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   2.8055
 SUM OF SQUARED ERRORS-SSE=   267.61
 MEAN OF DEPENDENT VARIABLE =   213.56
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      34 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 Y         0.70758     0.1872E-01   37.80     0.000 0.988     0.9883     0.9334
 CONSTANT   14.226      5.294       2.687     0.011 0.419     0.0000     0.0666
|_*Gelecek iki sezon için tüketimi kestirelim. Önce ilk regresyondan elde
|_*edilen parametreler kullanılarak C'nin kestirmesini yapalım. Burada FC
|_*komutu katsayıların tahmininden hemen önce gelmektedir. Bu nedenle
```

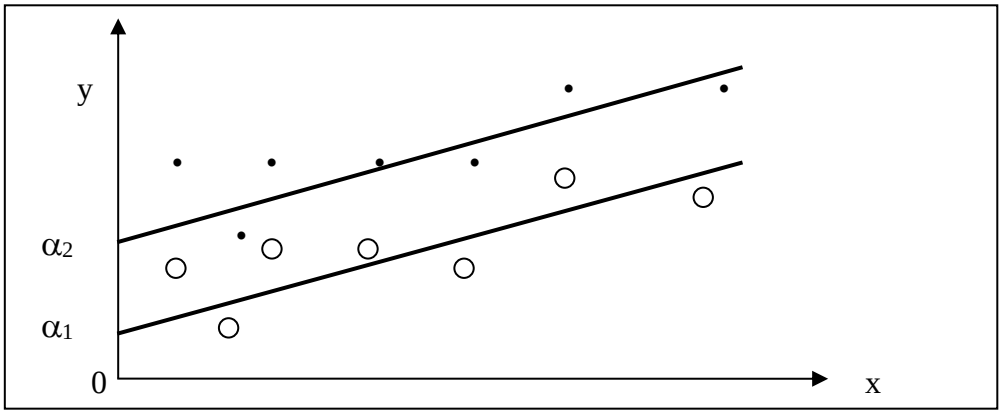
```
|_*tahmin edilen katsayıların orijinal regresyondan kaydedilmesi
|_*gerekmemektedir. Onların isminden de anlaşılacağı gibi, BEG= ve END=
|_*opsiyonları kestirmenin başlangıç ve bitişini gösterir.
|_?OLS C Y
|_FC / BEG=37 END=38 LIST
DEPENDENT VARIABLE = C                2 OBSERVATIONS
 REGRESSION COEFFICIENTS
   0.951448726043      -6.98049540703
     OBS.   OBSERVED     PREDICTED   CALCULATED  STD. ERROR
      NO.    VALUE        VALUE       RESIDUAL
      37   297.00       295.96       1.0392        2.219             I    *
      38   301.60       302.05     -0.45005        2.242          *  I
 SUM OF ABSOLUTE ERRORS=   1.4893
 R-SQUARE BETWEEN OBSERVED AND PREDICTED = 1.0000
 MEAN ERROR =  0.29459
 SUM-SQUARED ERRORS =   1.2825
 MEAN SQUARE ERROR =  0.64126
 MEAN ABSOLUTE ERROR=  0.74464
 ROOT MEAN SQUARE ERROR =  0.80079
 MEAN SQUARED PERCENTAGE ERROR=  0.72351E-01
 THEIL INEQUALITY COEFFICIENT U = 0.098
   DECOMPOSITION
      PROPORTION DUE TO BIAS =  0.13533
      PROPORTION DUE TO VARIANCE =  0.86467
      PROPORTION DUE TO COVARIANCE = -0.81393E-11
      PROPORTION DUE TO REGRESSION =  0.86467
      PROPORTION DUE TO DISTURBANCE = -0.61489E-11
|_*Kestirme için tahmin edilen ikinci regresyondan elde edilen tahmin
|_*edilmiş parametreler üzerine kestirmeyi temellendirelim. Bu sefer
|_*regresyonu tekrar ikinci defa tahmin etmeğe gerek yoktur. Çünkü
|_*katsayılar yukarıda BETA2 olarak kaydedilmişti. Bu katsayıları belirtmek
|_*için opsiyon kullanılmaktadır.
|_FC C Y L / COEF=BETA2 BEG=37 END=38 LIST
DEPENDENT VARIABLE = C                2 OBSERVATIONS
 REGRESSION COEFFICIENTS
   0.632581313166      0.450646562602      -13.3911774236
     OBS.   OBSERVED     PREDICTED   CALCULATED
      NO.    VALUE        VALUE       RESIDUAL
      37    297.00       298.93      -1.9268              *   I
      38    301.60       304.73      -3.1329            *     I
 SUM OF ABSOLUTE ERRORS=   5.0597
 R-SQUARE BETWEEN OBSERVED AND PREDICTED = 1.0000
 MEAN ERROR =  -2.5299
 SUM-SQUARED ERRORS =   13.528
 MEAN SQUARE ERROR =   6.7638
 MEAN ABSOLUTE ERROR=   2.5299
 ROOT MEAN SQUARE ERROR =   2.6007
 MEAN SQUARED PERCENTAGE ERROR=  0.74995
 THEIL INEQUALITY COEFFICIENT U = 0.681
   DECOMPOSITION
      PROPORTION DUE TO BIAS =  0.94624
      PROPORTION DUE TO VARIANCE =  0.53762E-01
      PROPORTION DUE TO COVARIANCE = -0.23115E-12
      PROPORTION DUE TO REGRESSION =  0.53762E-01
      PROPORTION DUE TO DISTURBANCE = -0.18319E-12
|_STOP
```

VIII. KUKLA VE SINIRLI BAĞIMLI DEĞİŞKENLER

8.1. Kesişme Katsayısındaki Değişimler İçin Kukla Değişkeni

Bu başlık altında, daha önceki regresyon analizlerinde kullandığımız sürekli değişkenlerden farklılık gösteren kesikli nitelik değişkenleri incelenecektir. Özel bir çeşit değişken olan ve regresyon analizlerinde karşımıza çıkan ve bir takım problemlere sebep olan bir değişken tipidir. Bazen regresyon analizlerinde bazı değişkenler olur ki bunlar kalitatiftir. Örneğin; üniversite eğitiminin olup olmaması, cinsiyet, ırk ve yaş farklılıkları bu tip değişkenlerdendir. Bu gibi durumlarda bu değişkenlerin etkisi kukla değişkenlerin etkisi ile ölçülür. Buradaki gizli varsayım farklı grup için regresyon doğrusu sadece kesişme noktasında değişir ve doğrunun eğimi aynı kalır. Örneğin; varsayalım ki gelir (y), eğitim görülen yıllar (x) arasındaki ilişki iki grup için Şekil 8.1'de gösterilmiştir. Grafikteki noktalar grup 1 için, daireler ise grup 2 içindir.

Dikkat edilecek olursa burada her iki grupta regresyon doğrularının eğimleri kabaca aynı, kesişme noktaları farklıdır. Bu sebeple regresyon denklemleri aşağıdaki gibi bir görünüm arz eder.



Şekil 8.1 Aynı eğim ve farklı kesişme noktasına sahip regresyon çizgisi

$$Y = \begin{cases} \alpha_1 + \beta x + u & \text{birinci grup için.} \\ \alpha_2 + \beta x + u & \text{ikinci grup için.} \end{cases}$$

Bu iki denklem tek bir denklem haline de getirilebilir.

$$y = \alpha_1 + (\alpha_2 - \alpha_1)\, D + \beta x + u$$

burada,

$$D = \begin{cases} 1 & \text{ikinci grup için.} \\ 0 & \text{birinci grup için.} \end{cases}$$

Buradaki D değişkeni bir kukla değişkenidir. Kukla değişkeninin katsayısı iki kesişme noktası arasındaki farkı ölçer. Eğer çok sayıda grup varsa, daha fazla kukla değişkeni tespit etmemiz gerekir. Bu işlemi üç ayrı grup için yapacak olursak,

$$y = \begin{cases} \alpha_1 + \beta x + u & \text{birinci grup için.} \\ \alpha_2 + \beta x + u & \text{ikinci grup için.} \\ \alpha_3 + \beta x + u & \text{üçüncü grup için} \end{cases}$$

Bu üç denklem aşağıdaki gibi yazılabilir.

$$Y = \alpha_1 + (\alpha_2 - \alpha_1) D_1 + (\alpha_3 - \alpha_1) D_2 + \beta x + u$$

Buradan,

$$D_1 = \begin{cases} 1 & \text{ikinci grup için.} \\ 0 & \text{birinci ve üçüncü grup için.} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{üçüncü grup için.} \\ 0 & \text{birinci ve ikinci grup için.} \end{cases}$$

Burada D_1 ve D_2 değişkenlerini ikame ederek her üç grup için sırasıyla α_1, α_2 ve α_3 kesişme noktalarını elde ettiğimizi kolayca gösterebiliriz. Dikkat edilecek olursa bu üç denklem bir araya getirildiğinde β eğim katsayısının ve u hata teriminin üç grubun hepsi için aynı dağılıma sahip olduğunu varsayıyoruz.

Eğer regresyon denkleminde sabit terim varsa kukla değişkeni kullanmamız gerekiyor. Çünkü sabit terim temel grup için kesişme noktasıdır ve kukla değişkenlerin katsayıları kesişme noktasındaki farklılıkları ölçer ki bu yukarıdaki denklemde görülmektedir. Bu denklemde sabit terim ilk grubun kesişme noktasını, sabit terim artı D_1'in katsayısı ikinci grup için kesişme noktasını ve sabit terim artı D_2'nin katsayısı üçüncü grup için kesişme noktasını ölçer. Burada birinci grup temel grup olarak seçilmiştir. Fakat herhangi bir grup da seçilebilirdi. Kukla değişkenin katsayısı temel gruptaki kesişme noktasından olan farkı ölçer. Eğer regresyon denkleminde kesişme noktası yok ise her grup için bir kukla değişken belirleyebiliriz ve kukla değişkeni de her bir gruba ait kesişme noktasını belirler. Eğer hem sabit terim hem de üç kukla değişkeni model içinde bulundurulursa tam çoklu eş doğrusallık modele sokulmuş olur. Bu yüzden regresyon analizi çalışmaz veya kukla değişkenlerden biri program tarafından dikkate alınmaz.

Belli sayıdaki haneler için tüketime ait veriler ve aynı zamanda gelir seviyelerine ait verilerin olduğunu varsayalım. Bunlara ilave olarak,

1. S: ev reisinin cinsiyetini temsil ediyor.
2. A: ev reisinin yaşı ki bu üç gruba ayrılıyor.

   < 25 yıl 25 ile 50 yıl arası > 50 yıl
3. E: ev reisinin eğitim durumu ki bu da üç gruba ayrılıyor.

   $<$ lise $\geq$ Lise $<$ fakülte $\geq$ fakülte

Bu kalitatif değişkenler model içinde kukla değişkeni formunda aşağıdaki denklem ve tanımlarda izah edildiği gibi bulundurulmaktadır. Görüldüğü gibi her bir kategori için kukla değişkenlerin sayısı sınıflandırma sayısından bir eksik olmaktadır. Bu durumda regresyon denklemi aşağıdaki gibi olur.

$$C = \alpha + \beta Y + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \gamma_4 D_4 + \gamma_5 D_5 + u$$

Kukla değişkeni metodunda yapılan varsayım, her gruptaki değişme sadece kesişme noktasındadır, eğimde değildir.

$$D_1 = \begin{cases} 1 & \text{eğer cinsiyet erkek ise} \\ 0 & \text{eğer cinsiyet kadın ise} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{eğer yaş} < 25 \text{ ise} \\ 0 & \text{diğer durumlar} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{eğer yaş 25 ile 50 arası ise} \\ 0 & \text{diğer durumlar} \end{cases}$$

$$D_4 = \begin{cases} 1 & \text{eğer eğitim} < \text{lise ise} \\ 0 & \text{diğer durumlar} \end{cases}$$

$$D_5 = \begin{cases} 1 & \text{eğer} \geq \text{Lise fakat} < \text{Fakülte ise} \\ 0 & \text{diğer durumlar} \end{cases}$$

Her bir fert için kesişme noktası uygun olan kukla değişkeni değeri konularak sağlanır. Örneğin; 25 yaşından küçük, fakülte mezunu bir erkek için $D_1 = 1$, $D_2 = 1$, $D_3 = 0$, $D_4 = 0$, $D_5 = 0$ ve böylece kesişme noktası, $\alpha + \gamma_1 + \gamma_2$ olacaktır. Yine örneğin; 50 yaşından büyük fakülte mezunu bir bayan için $D_1 = 0$, $D_2 = 0$, $D_3 = 0$, $D_4 = 0$, $D_5 = 0$ olduğundan kesişme noktası α olur.

Kukla değişken metodu aynı zamanda mevsimsel faktörleri dikkate almak için kullanılır. Örneğin; C'nin Y üzerine regresyonunsa mevsimlik veriler mevcut ise, regresyon denklemi aşağıdaki şekilde belirlenebilir.

$$C = \alpha + \beta Y + \lambda_1 D_1 + \lambda_2 D_2 + \lambda_3 D_3 + u$$

burada D_1, D_2 ve D_3 mevsimlere ait kukla değişkenleridir.

$$D_1 = \begin{cases} 1 & \text{mevsim Kış ise} \\ 0 & \text{diğer durumlar için} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{mevsim İlkbahar ise} \\ 0 & \text{diğer durumlar için} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{mevsim Yaz ise} \\ 0 & \text{diğer durumlar için} \end{cases}$$

Eğer aylık veriler mevcut ise, o zaman aşağıdaki kukla değişkenlere sahip olunur.

$D_1 = \begin{cases} 1 & \text{Ocak ayı için} \\ 0 & \text{diğer durumlar için} \end{cases}$

$D_2 = \begin{cases} 1 & \text{Şubat ayı için} \\ 0 & \text{diğer durumlar için} \end{cases}$

Eğer kurban bayramından dolayı kurban bayramının yaşandığı ayda et tüketiminin yoğun olduğu düşünülüyorsa, bu etkiyi görmek için kullanılacak kukla değişkeni aşağıdaki gibi olabilir.

$D = \begin{cases} 1 & \text{Kurban bayramının yaşandığı ay için.} \\ 0 & \text{Diğer aylar için.} \end{cases}$

Uygulamalı Örnek 8.1

Çevre koruma birimi (EPA) araba satın alıcılarının değişik model arabaların nispi yakıt etkinliğini karşılaştırabilmelerini sağlanmak için otomobillere ait bir galon benzin başına mil tahminlerini yayınlamaktadır. Acaba EPA, değişik model arabaların nispi yakıt etkinliğini karşılaştırabilmesi için gerekli tüm bilgileri sağlıyor mu? Bu sorunun cevabını vermek için aşağıdaki regresyon modeli tahmin edilmiştir. Parantez içindeki değerler standart hatalardır.

$$Y = \underset{(1,735)}{7,952} + \underset{(0,061)}{0,693}\ EPA \qquad R^2 = 0,74$$

$$Y = \underset{(5,349)}{22,008} - \underset{(0,001)}{0,002}\ W - \underset{(0,708)}{2760}\ S/A + \underset{(1,413)}{3,280}\ G/D + \underset{(0,097)}{0,41}\ EPA \quad R^2 = 0,82$$

Burada:

Y : Yol testlerine dayalı olarak tüketim birliği tarafından rapor edilen galon başına yol uzunluğu (mil / galon)

W : Arabanın ağırlığı (pound)

S/A : Vites kutusu (otomatik olmayanlar için 0, otomatik olanlar için 1)

G/D: Yakıt kullanımı (benzinli arabalar için 0, dizel motorlar için 1)

EPA: EPA tarafından elde edilen mil tahmini (mil/galon).

W, S/A, G/D değişkenlerinin hepsinin doğru bir işarete sahip olmaması, EPA'nın yakıt etkinliğini tahmin etmede mevcut bilgilerin tümünü kullanmadığı sonucunu EPA tahmin katsayısının düşüklüğü karşımıza çıkarmaktadır.

Kukla değişkenlerin kullanılmasında bazen aşağıdaki gibi problemlerle karşılaşılabilir.

- Çok sayıda kukla değişkenine sahip bazı çalışmalarda katsayıların işaretlerini yorumlamak zorlaşmaya başlar. Çünkü yanlış işarete sahip olmaya başlarlar.

•Bazen kukla değişkenlerin modele sokulması eğim katsayısında önemli değişmelere neden olabilir.

Kalitatif değişkenlerin etkisini ölçmek yanında çok varyanslılık, otokorelasyon ve daha birçok problemini çözmede kullanılan kukla değişkenlerinin kullanımı ekonometri bilgisayar programlarında kolaylıkla yapılabilmektedir. Bu hususla ilgili bir bilgisayar uygulaması bu konunun sonunda verilecektir.

Sınıf Uygulaması:

Derste veri toplayın, regresyon modeline karar verin, modelinizi tahmin edin, tahmin edilen modeli istatistiksel ve mantıksal olarak değerlendirin ve çıkarımlar yaparak bir ekonometri modeli uygulamasının tüm yönlerini çalışınız. Bunun için sınıftaki öğrencilerin ve kendinizin kilo, boy, yaş ve cinsiyet verilerini tam sayımla toplayınız. Buna ikamet (ev/yurt), sigara (içiyor/içmiyor) gibi verileri de ekleyebilirsiniz. Önce bağımlı değişkeni ve bağımsız değişkenleri, sonra hangi bağımsız değişkenlerin kukla (dummy) olup olmadığını belirleyiniz. Sonra modeli yazınız, tahmin ediniz ve değerlendiriniz. Sınıf ortalaması tahmininden hareketle her bir öğrencinin kendi verilerini kullanarak kendiyle ilgili kestirmeler ve çıkarımlar yapmasını sağlayınız.

*8.2. Eğim Katsayısındaki Değişmeler İçin Kukla Değişkeni

Bundan önceki kısımda kesişme katsayısındaki farklılıklara izin veren kukla değişkenini ele aldık. Bu kukla değişkenleri 0 veya 1 olarak varsayılır. Hâlbuki tüm kukla değişkenleri bu formda değildirler. Eğim katsayılarında farklılıklara izin veren kukla değişkenleri de kullanılabilir. Örneğin, regresyon denklemin aşağıdaki gibi olduğunu varsayalım.

$$y_1 = \alpha_1 + \beta_1 x_1 + u_1 \qquad \text{birinci grup için}$$

$$y_2 = \alpha_2 + \beta_2 x_2 + u_2 \qquad \text{ikinci grup için}$$

bu denklemleri birlikte aşağıdaki gibi yazabiliriz.

$$y_1 = \alpha_1 + (\alpha_2 - \alpha_1) * 0 + \beta_1 x_1 + (\beta_2 - \beta_1) * 0 + u_1$$

$$y_2 = \alpha_1 + (\alpha_2 - \alpha_1) * 1 + \beta_1 x_2 + (\beta_2 - \beta_1) * x_2 + u_2$$

veya

$$y = \alpha_1 + (\alpha_2 - \alpha_1) D_1 + \beta_1 x + (\beta_2 - \beta_1) D_2 + u_2 \qquad (8.1)$$

burada $D_1 = \begin{cases} 0 & \text{birinci gruptaki tüm gözlemler için} \\ 1 & \text{ikinci gruptaki tüm gözlemler için} \end{cases}$

$D_2 = \begin{cases} 0 & \text{birinci gruptaki tüm gözlemler için} \\ x_2 & \text{ikinci gruptaki x'in kendi değeri} \end{cases}$

Burada D_1 katsayısı kesişme terimindeki farklılığı, D_2 ise eğimdeki farklılığı ölçmektedir. Eğer hataların benzer bir dağılıma sahip olduğunu varsayarsak 8.1'deki denklemin tahmin edilmesi iki denklemin ayrı ayrı tahmin edilmesine denktir. Eğer

denklem 8.1'deki D_2'yi çıkarırsak bu kesişme katsayısındaki farklılığa izin verir fakat farklı eğime değil. Fakat eğer D_1'i çıkarırsak bu sefer eğimdeki farklılığa izin verir fakat kesişme katsayısında değil.

Uygun kukla değişkenleri eğim ve kesişme katsayılarında farklı zamanlarda değişme olur ise tanımlanabilir. Üç dönem için elimizde veri olduğunu ve ikinci dönemde sadece kesişme katsayısının değiştiğini yani paralel kayma olduğunu varsayalım. Üçüncü dönemde ise kesişme katsayısı ve eğim değişmiştir. Bu durumda aşağıdaki gibi modelleri yazabiliriz.

$$\begin{aligned} y_1 &= \alpha 1 + \beta_1 x_1 + u_1 \quad \text{birinci dönem için} \\ y_2 &= \alpha 2 + \beta_2 x_2 + u_2 \quad \text{ikinci dönem için} \\ y_3 &= \alpha 3 + \beta_3 x_3 + u_3 \quad \text{üçüncü dönem için} \end{aligned} \qquad (8.2)$$

Böylece bu denklemleri birleştirerek modeli aşağıdaki gibi yazabiliriz.

$$y = \alpha_1 + (\alpha_2 - \alpha_1) D_1 + (\alpha_3 - \alpha_1) D_2 + \beta_1 x + (\beta_2 - \beta_1) D_3 + u \qquad (8.3)$$

burada $D_1 = \begin{cases} 1 & \text{ikinci dönemdeki gözlemler için} \\ 0 & \text{diğer dönemler için} \end{cases}$

$D_2 = \begin{cases} 1 & \text{üçüncü dönemdeki gözlemler için} \\ 0 & \text{diğer dönemler için} \end{cases}$

$D_3 = \begin{cases} 1 & \text{bir ve ikinci dönemdeki gözlemler için} \\ x_3 & \text{veya 3. dönemdeki tüm gözlemler için x'in kendi değeri} \end{cases}$

Buradaki örnekte tüm gruplardaki hataların benzer dağılıma sahip olduğu varsayılmaktadır. Bu nedenle bütün gruplardaki hata terimleri birleştirilip 8.2 veya 8.3'deki gibi tek bir hata terimine (u) sahip olacak şekilde yazılan denklem en küçük kareler yöntemi ile tahmin edilebilir.

8.3. Çapraz Denklem Kısıtları İçin Kukla Değişkeni

Önceki kısımda açıklanan yöntem denklemler arasında bazı parametrelerin eşit olması durumu için genişletilebilir.

Uygulamalı Örnek 8.2

Bir uygulamalı örnek olması açısından Çizelge 8.1'deki veriler kullanılarak sığır, domuz ve tavuk eti talebinin birlikte tahmini dikkate alınabilir. Bu tahmin edilecek olan talep denklemleri setinin aşağıdaki gibi olduğu varsayılsın.

$$\begin{aligned} p_1 &= \alpha_1 + \beta_{11} x_1 + \beta_{12} x_2 + \beta_{13} x_3 + \gamma_1 y + u_1 \\ p_2 &= \alpha_2 + \beta_{12} x_1 + \beta_{22} x_2 + \beta_{23} x_3 + \gamma_2 y + u_2 \\ p_3 &= \alpha_3 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{33} x_3 + \gamma_3 y + u_3 \end{aligned} \qquad (8.4)$$

burada p_1: Sığır etinin perakende fiyatı
p_2: Domuz etinin perakende fiyatı
p_3: Tavuk etinin perakende fiyatı
x_1: Kişi başına sığır eti tüketimi
x_2: Kişi başına domuz eti tüketimi
x_3: Kişi başına tavuk eti tüketimi
y: Kişi başına kullanılabilir gelir

Fiyat indeksleri, harcanabilir gelir ve değişik et fiyatları Çizelge 8.1'de verilmiştir. Çizelgedeki perakende fiyatları tüketici fiyat indeksine bölünmüşlerdir. Bu nedenle bu fiyatlar tüketici fiyat indeksi p ile çarpılarak p_1, p_2 ve p_3 fiyatlar elde edilebilir.

Çizelge 8.1. Seçilen et ürünleri için tüketim miktarları ve fiyatlar

			Sığır eti		Domuz eti		Kuzu eti		Dana eti		Tavuk eti	
yıllar	Gelir	İndeks	tüketim	fiyat	tüketim	fiyat	Tüketim	fiyat	Tüketim	fiyat	Tüketim	fiyat
1948	1291	0,838	63,1	82,9	67,8	67,6	5,1	77,8	9,5	77,1	18,3	75,4
1949	1271	0,830	63,9	76,3	67,7	61,5	4,1	82,4	8,9	75,7	19,6	71,8
1950	1369	0,838	63,4	88,3	69,2	60,4	4,0	84,2	8,0	81,1	20,6	68,0
1951	1473	0,906	56,1	90,0	71,9	60,6	3,4	86,7	6,6	87,6	21,7	66,0
1952	1520	0,925	62,2	85,4	72,4	57,3	4,2	86,2	7,2	86,3	22,1	65,0
1953	1582	0,932	77,6	66,2	63,5	62,9	4,7	70,0	9,5	68,7	21,9	62,8
1954	1582	0,936	80,1	64,1	60,0	63,7	4,6	71,0	10,0	65,8	22,8	56,4
1955	1660	0,934	82,0	63,2	66,8	54,6	4,6	69,0	9,4	65,8	21,3	58,7
1956	1742	0,947	85,4	60,9	67,3	51,4	4,5	68,3	9,5	63,6	24,4	50,4
1957	1804	0,981	84,6	63,1	61,1	57,6	4,2	69,9	8,8	65,5	25,5	47,6
1958	1826	1,007	80,5	72,0	60,2	60,5	4,2	74,1	6,7	76,1	28,2	45,8
1959	1904	1,015	81,4	73,3	67,6	52,8	4,8	69,6	5,7	79,8	28,9	41,4
1960	1934	1,031	85,2	70,4	65,2	51,6	4,8	67,6	6,2	77,8	28,2	41,4
1961	1980	1,041	88,0	68,3	62,2	53,3	5,1	63,3	5,7	77,3	30,3	37,0
1962	2052	1,054	89,1	69,8	64,0	52,9	5,1	67,1	5,5	79,5	30,2	38,6

Yukarıda 8.4'deki sistem denklemleri ile ilgili özel durum; β katsayılarındaki simetrik olma özelliğidir. Buradaki özellikler aşağıdaki gibi yazılabilir.

$$dp_1 / dx_2 = dp_2 / dx_1 = \beta_{12}\ ;\ \ dp_1 / dx_3 = dp_3 / dx_1 = \beta_{13}\ ;\ \ dp_2 / dx_3 = dp_3 / dx_2 = \beta_{23}$$

Bu nedenle katsayılar üzerinde denklemler arası kısıtlar mevcuttur. Eğer $V(u_1) = V(u_2) = V(u_3)$ ise, $(\Sigma u_{12} + \Sigma u_{22} + \Sigma u_{32})$'i minimum yapılabilir, normal denklemler elde edebiliriz ve regresyon katsayıları tahmin edebiliriz. Bu metot Waugh tarafından kullanılmıştır. Bunun yerine burada standart regresyon programları ve kukla değişkeni metodu kullanılabilir. Yukarıda 8.4'de verilen denklemler bir denklem halinde aşağıdaki gibi yazılabilir.

$$\begin{vmatrix} P_1 \\ P_2 \\ P_3 \end{vmatrix} = \alpha_1 \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} + \alpha_2 \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix} + \alpha_3 \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix} + \beta_{11} \begin{vmatrix} x_1 \\ 0 \\ 0 \end{vmatrix} + \beta_{12} \begin{vmatrix} x_2 \\ x_1 \\ 0 \end{vmatrix} + \beta_{13} \begin{vmatrix} x_3 \\ 0 \\ x_1 \end{vmatrix} + \beta_{22} \begin{vmatrix} 0 \\ x_2 \\ 0 \end{vmatrix} + \beta_{23} \begin{vmatrix} 0 \\ x_3 \\ x_2 \end{vmatrix} + \beta_{33} \begin{vmatrix} 0 \\ 0 \\ x_3 \end{vmatrix} + \gamma_1 \begin{vmatrix} y \\ 0 \\ 0 \end{vmatrix} + \gamma_2 \begin{vmatrix} 0 \\ y \\ 0 \end{vmatrix} + \gamma_3 \begin{vmatrix} 0 \\ 0 \\ y \end{vmatrix} + \begin{vmatrix} u_1 \\ u_2 \\ u_3 \end{vmatrix}$$

Bu denklem 45 gözlem ve sabit terimin olmaması durumunda 12 kukla değişkeni ile tahmin edilebilir. Kukla değişkenlerinin değerleri kolay bir şekilde üretilebilir. Örneğin; β_{12} için gözlemler seti x_2'nin 15 gözlemini, bunu takip eden x_1'in 15 gözlemini ve 15 sıfırı içerir. Sonuçlar aşağıdaki gibi verilebilir. Parantez içindekiler *t* değerleridir.

$$P_1 = \underset{(12,00)}{118,98} - \underset{(14,55)}{1,534}\, x_1 - \underset{(4,31)}{0,474}\, x_2 - \underset{(3,01)}{0,445}\, x_3 + \underset{(12,61)}{0,0650}\, y$$

$$P_2 = \underset{(9,18)}{149,79} - \underset{(4,31)}{0,474}\, x_1 - \underset{(6,20)}{1,189}\, x_2 + \underset{(1,54)}{0,319}\, x_3 + \underset{(2,83)}{0,0162}\, y$$

$$P_3 = \underset{(7,36)}{131,06} - \underset{(3,01)}{0,445}\, x_1 - \underset{(1,54)}{0,319}\, x_2 - \underset{(4,32)}{2,389}\, x_2 + \underset{(1,66)}{0,0199}\, y$$

Biri çıkıp 8.4'deki denklemler sisteminin uygun olup olmadığı konusunda sorgulama yapabilir. Buradaki asıl ve tek amaç, farklı değişkenlerdeki bazı parametrelerin aynı olması durumunda denklemleri tahmin etmek için kukla değişkeni metotlarının kullanımını göstermektir.

*8.3. Sınırlı Bağımlı Değişken Modelleri

Şimdiye kadar ele aldığımız modellerdeki kukla değişkenler bağımsız veya açıklayıcı değişkenlerdir. Şimdi bağımlı yani açıklanan değişkenlerin kukla değişkenler olduğu modeller dikkate alınacaktır. Bu kukla değişkeni, iki veya daha fazla değer alabilir. Burada sadece 1 ve 0 gibi iki değer aldığı durum analiz edilecektir. Bunun üzerinde değer alan kukla değişkenler daha ileri düzeyde çalışmalarda ele alınmaktadır. Kukla değişkeni sadece iki değer aldığından ikiye ayrılan bir değişken adını almaktadır. Bu tip değişkenlere aşağıdaki gibi çok sayıda örnekler verilebilir.

$$y = \begin{cases} 1 & \text{eğer bir kişi istihdam edilmiş ise} \\ 0 & \text{diğer durum} \end{cases}$$

veya

$$y = \begin{cases} 1 & \text{eğer bir kişi kendi evine sahip ise} \\ 0 & \text{diğer durum} \end{cases}$$

Bağımlı değişkenlerin bir ve sıfır olması durumunun regresyon analizini yapan birçok metot vardır. En basit yöntem normal en küçük kareler yönteminin kullanılmasıdır. Bu durumdaki model, *doğrusal ihtimal modeli* olarak adlandırılır. *Doğrusal diskirminant modeli* olarak adlandırılan diğer bir metot, doğrusal ihtimal modeli ile ilgilidir. Diğer bir alternatifte ise gözlemediğimiz görünmeyen (latent) değişken olduğu ifade edilmektedir. Gözlenen durum aşağıdaki gibi ifade edilir.

$$y = \begin{cases} 1 & \text{eğer } y^* > 0 \\ 0 & \text{diğer durum} \end{cases}$$

Bu, logit ve probit modellerin arkasındaki düşünceyi ifade eder. Önce bu metotlar tartışılıp daha sonra uygulamalı bir örnek verilecektir.

*8.4. Doğrusal İhtimal Modeli

Doğrusal ihtimal modeli ifadesi bağımlı değişken y'nin 1 ve 0 değerlerini alan yani ikiye ayrılan bir değişken olduğu regresyon modelini temsil etmektedir. Basit olması için sadece bir bağımsız "x" değişkenini yani basit regresyonu dikkate alınmaktadır.

Değişken "y" bir olayın olmaması veya olması durumunu temsil eden bir gösterge değişkendir. Örneğin, işsizliğin belirleyicilerini analiz etmede her bir kişinin istihdam edilip edilmediğini gösteren bağımlı ve istihdam durumunu belirleyen bağımsız değişken içeren bir modele sahip olduğumuzu varsayalım. Burada dikkate alınan olay istihdam durumudur. Şimdi sadece iki sonucu ifade eden bu değişkeni tanımlayalım.

$$y = \begin{cases} 1 & \text{eğer kişi istihdam edildi ise} \\ 0 & \text{diğer durum} \end{cases}$$

Benzer şekilde bir bankanın iflasının analizinde de değişken aşağıdaki gibi tanımlanır.

$$y = \begin{cases} 1 & \text{eğer banka iflas etti ise} \\ 0 & \text{diğer durum} \end{cases}$$

Şimdi bu modeli bilinen regresyon düzeni içerisinde yazalım.

$$y_i = \beta\, x_i + u_i \qquad (8.5)$$

Burada $E(u_i) = 0$'dır. Şartlı beklenti $E(y_i / x_i)$ βx_i'ye eşittir. Bu beklenti bağımsız değişken x_i'nin olması durumunda beklenen iki farklı sonuçtan birinin ortaya çıkma ihtimali olarak yorumlanır. Regresyon denklemiyle hesaplanan y'nin değeri, verilen belli bir x değeri durumunda olayın ortaya çıkacağının tahmin edilmiş ihtimalini verir. Uygulamada bu tahmin edilen ihtimaller kabul edilebilir 0 - 1 aralığının dışına çıkabilir.

Burada y_i 1 ve 0 değerlerini aldığından denklem 8.5'deki hata $(1 - \beta x_i)$ ve $(-\beta x_i)$ değerlerini alır. Aynı zamanda denklem 8.5'e atfedilen yorum ve $E(u_i = 0)$ gerekliliği ile bu olayların sırasıyla ihtimali (βx_i) ve $1 - (\beta x_i)$ olacaktır.

Doğrusal ihtimal modelinin kısıtları y = 0 ve y = 1 aralığında belirtilen noktaları gösteren Şekil 8.2'de verilmiştir. Kestirilen değerler (0 - 1) aralığının dışında kolayca yer alabilirler ve kestirilen hatalar çok geniş olabilir.

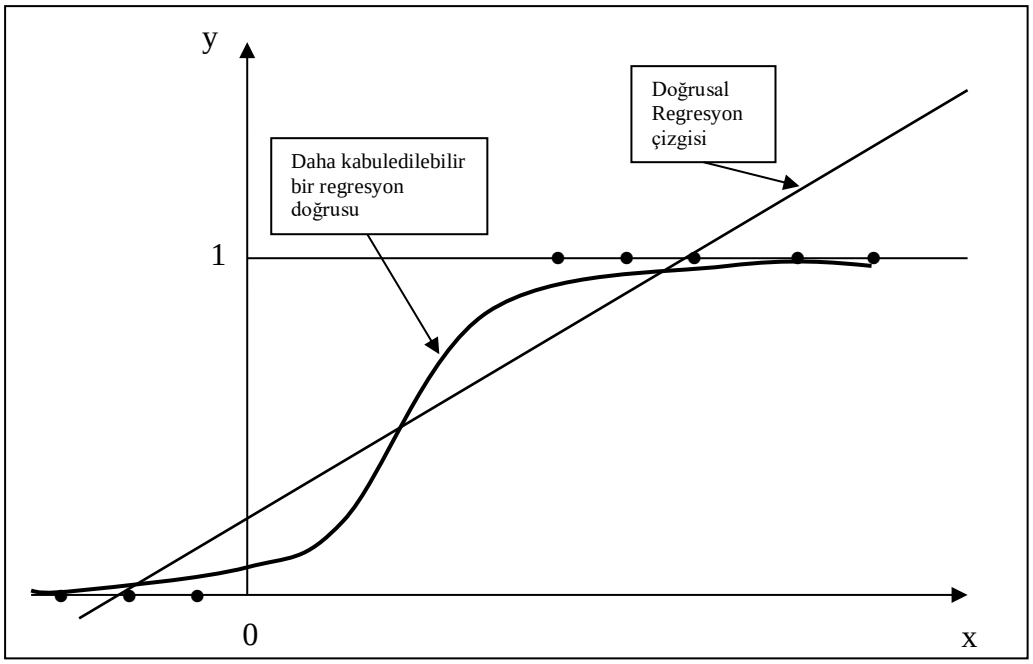
1960 ve 70'li yıllarda doğrusal ihtimal modeli, çoklu regresyon analiziyle kolayca tahmin edilen bir model olduğundan çok yaygın bir şekilde kullanılmıştır.

*8.5. Probit ve Logit Modeller

Diğer bir alternatif de ise 8.6'daki gibi bir modele sahip olduğumuzu varsayalım.

$$y^* = \beta_0 + \Sigma^k_{j=1} \beta_j x_{ij} + u_i \qquad (8.6)$$

burada y^*_i gözlemlenemiyor. Bu çoğu zaman görünmeyen (latent) değişken olarak adlandırılmaktadır. Gözlemlenemeyen kukla değişkeni y_i aşağıdaki gibi tanımlanır.



Şekil 8.2. Doğrusal ihtimal modelinden kestirmeler

$$y_i = \begin{cases} 1 & \text{eğer } y^* > 0 \\ 0 & \text{diğer durumda} \end{cases} \qquad (8.7)$$

Probit ve logit modeller 8.6'daki hata terimi u_i'nin dağılımının spesifikasyonunda farklılık arz ederler. Burada 8.6'daki spesifikasyonla doğrusal ihtimal modeli arasındaki fark, doğrusal ihtimal modelinde ikiye ayrılan (dichotomous) değişkenler olduğu gibi analiz edilmektedir. Halbuki 8.6'da gizlenen görünmeyen bir değişkenin mevcudiyeti varsayılmaktadır. Örneğin, eğer gözlenen kukla değişkeni kişinin istihdam edilip edilmediği ise, y^*_i iş bulma eğilimi veya kabiliyeti olarak tanımlanacaktır. Benzer bir şekilde eğer gözlenen kukla değişkeni kişinin araba satın alıp almadığı ise y^*_i bir araba alma arzusu veya kabiliyeti olarak tanımlanacaktır. Her iki örnekte de "istek" ve "kabiliyet" işin içine girmektedir. Bu nedenle 8.6'daki açıklayıcı değişkenler, bu unsurların her ikisini de açıklayan değişkenler içereceklerdir.

Yapılan çalışmalar probit, logit ve doğrusal ihtimal modellerine ait tahminler arasında yakın bir ilişki olduğunu göstermiştir. Logit modellerle tahmin edilen parametrelerin 1/1,6 = 0,625 katsayısı ile çarpılmasıyla probit modellerden tahmin edilen katsayılar yaklaşık olarak elde edilmektedir. Ayrıca yine logit tahminlerinin 0,25 (eğim katsayısı için) katsayısı ile çarpılarak doğrusal ihtimal tahminleri bulunur. Kesişme katsayısı durumunda logit katsayısı 0,25 ile çarpıldıktan sonra 0,5'in de ilave edilmesi gerekir.

Uygulamalı Örnek 8.3

Bu örnek için çizelge 8.2'deki değişkenler kullanılmaktadır. Bu veriler idam cezasının caydırıcı etkilerini araştırmaya yönelik bir çalışmaya ait yatay kesit verileri olup, ABD'de 1950 yılında 44 eyaletten toplanmıştır. Verilerde iki tane kukla değişken vardır. Güney-Kuzey kuklası olan D_2, açıkça görülmektedir ki bu bir açıklayıcı

değişkendir. Fakat eyalette ölüm cezası olup olmamasını temsil eden D_1 hem açıklayan hem de açıklanan değişken olabilir. Eğer bu açıklanan değişken ise "idam cezasına sahip olma eğilimi" olarak dikkate alınacaktır.

Çizelge 8.2. ABD'de öldürme oranlarının belirleyicileri

N	M	PC	PX	D_1	T	Y	LF	NW	D_2
1	19,25	0,204	0,035	1	47	1,10	51,2	0,321	1
2	7,53	0,327	0,081	1	58	0,92	48,5	0,224	1
3	5,66	0,401	0,012	1	82	1,72	50,8	0,127	0
4	3,21	0,318	0,070	1	100	2,18	54,4	0,063	0
5	2,80	0,350	0,062	1	222	1,75	52,4	0,021	0
6	1,41	0,283	0,100	1	164	2,26	56,7	0,027	0
7	6,18	0,204	0,050	1	161	2,07	54,6	0,139	1
8	12,15	0,232	0,054	1	70	1,43	52,7	0,218	1
9	1,34	0,199	0,086	1	219	1,92	52,3	0,008	0
10	3,71	0,138	0	0	81	1,82	53,0	0,012	0
11	5,35	0,142	0,018	1	209	2,34	55,4	0,076	0
12	4,72	0,118	0,045	1	182	2,12	53,5	0,299	0
13	3,81	0,207	0,040	1	185	1,81	51,6	0,040	0
14	10,44	0,189	0,045	1	104	1,35	48,5	0,069	1
15	9,58	0,124	0,125	1	126	1,26	49,3	0,330	1
16	1,02	0,210	0,060	1	192	2,07	53,9	0,017	0
17	7,52	0,227	0,055	1	95	2,04	55,7	0,166	1
18	1,31	0,167	0	0	245	1,55	51,2	0,003	0
19	1,67	0,120	0	0	97	1,89	54,0	0,010	0
20	7,07	0,139	0,041	1	177	1,68	52,2	0,076	0
21	11,79	0,272	0,063	1	125	0,76	51,1	0,454	1
22	2,71	0,125	0	0	56	1,96	54,0	0,032	0
23	13,21	0,235	0,086	1	85	1,29	55,0	0,266	1
24	3,48	0,108	0,040	1	199	1,81	52,9	0,018	0
25	0,81	0,672	0	0	298	1,72	53,7	0,038	0
26	3,32	0,357	0,030	1	145	2,39	55,8	0,067	0
27	3,47	0,592	0,029	1	78	1,68	50,4	0,075	0
28	8,31	0,225	0,400	1	144	2,29	58,8	0,064	0
29	1,57	0,267	0,126	1	178	2,34	54,5	0,065	0
30	4,13	0,164	0,122	1	146	2,21	53,5	0,065	0
31	3,84	0,128	0,091	1	132	1,42	48,8	0,090	1
32	1,83	0,287	0,075	1	98	1,97	54,5	0,016	0
33	3,54	0,210	0,069	1	120	2,12	52,1	0,061	0
34	1,11	0,342	0	0	148	1,90	56,0	0,019	0
35	8,90	0,133	0,216	1	123	1,15	56,2	0,389	1
36	1,27	0,241	0,100	1	282	1,70	53,3	0,037	0
37	15,26	0,167	0,038	1	79	1,24	50,9	0,161	1
38	11,15	0,252	0,040	1	34	1,55	53,2	0,127	1
39	1,74	0,418	0	0	104	2,04	51,7	0,017	0
40	11,98	0,282	0,032	1	91	1,59	54,3	0,222	1
41	3,04	0,194	0,086	1	199	2,07	53,7	0,026	0
42	0,85	0,378	0	0	101	2,00	54,7	0,012	0
43	2,83	0,757	0,033	1	109	1,84	47,0	0,057	1
44	2,89	0,356	0	0	117	2,04	56,9	0,022	0

İlk olarak M'nin diğer tüm değişkenler üzerine regresyonunu dikkate alındığında sonuç aşağıdaki gibidir. Parantez içindeki değerler standart hata değil, *t* değerleridir.

$$M = \underset{(-0,82)}{-8,50} - \underset{(-1,38)}{3,696}\,PC - \underset{(-0,54)}{3,368}\,PX + \underset{(2,11)}{2,598}\,D_1 - \underset{(2,62)}{0,018}\,T - \underset{(-2,3)}{4,095}\,Y + \underset{(1,82)}{0,400}\,LF + \underset{(1,17)}{6,444}\,NW + \underset{(1,93)}{2,541}\,D_2 \qquad R^2 = 0,7746$$

Burada M: 1950 FBI tahminlerine göre 100,000 kişide öldürme oranı
PC: 1950'deki mahkûmiyet sayısı / öldürülen sayısı
PX: 1946-50 arasında idam edilenlerin sayısı/1950'deki mahkûmiyet sayısı
D_1: Kukla değişkeni, eğer eyalette ölüm cezası var ise 1 yoksa 0
T : 1951'de hapisten çıkan katillerin ortalama hapis süreleri
Y : 1949'daki ailelerin ortalama aile geliri (100 $)
LF: 1950'de iş gücüne katılım oranı
NW: 1950'de beyaz olmayanların toplam nüfusa oranı
D_2: Kukla değişkeni, güney eyaletler 0, diğerleri 1

Modelde de görüldüğü gibi bazı değişkenlere ait katsayıların işaretleri beklenenin yani olması gerekenin tersine çıkmıştır.

Şimdi D_1 kukla değişkenini açıklanan bir değişken olarak dikkate alalım. Bu durumda T, Y, LF, NW ve D_2 açıklayıcı değişkenler olarak kullanılacaktır.

Doğrusal ihtimal modelinin sonuçları aşağıdaki gibidir. Parantez içindeki rakamlar *t* değerleridir. Bu değerler bağımlı değişkenin 0 - 1 özelliğini dikkate almayan normal bir en küçük kareler regresyon tahmininden elde edilmiştir.

$$D_1 = \underset{(1,50)}{1,993} + \underset{(1,46)}{0,00146}\,T + \underset{(2,74)}{0,658}\,Y - \underset{(-1,93)}{0,055}\,LF + \underset{(2,62)}{1,988}\,NW + \underset{(1,81)}{0,343}\,D_2 \qquad R^2 = 0,3376$$

Bu sonuçlara göre güney eyaletleri ve beyaz olmayanların oranının yüksek olduğu eyaletlerde idam cezası olma ihtimali artmaktadır. İstihdam edilen işgücü oranı, idam cezasının olması ihtimali üzerine negatif bir etkisi olduğu görülmektedir. Şaşırtıcı bir sonucu ise pozitif ve istatistiki olarak önemli olan ortalama aile geliri y'nin katsayısıdır. Bu durumu açıklamanın bir yolu New York ve California gibi yüksek gelirli büyük şehirlerde suç oranlarının daha yüksek olmasıdır.

Logit ve probit tahminleri dikkate alındığında durum biraz değişmektedir. Logit modeli aşağıdaki sonuçları vermektedir. Parantez içindeki değerler asimptotik *t* değerleridir.

$$D_1 = \underset{(0,53)}{10,99} + \underset{(1,87)}{0,0194}\,T + \underset{(1,88)}{10,61}\,Y - \underset{(-1,40)}{0,668}\,LF + \underset{(1,95)}{70,99}\,NW + \underset{(0,02)}{13,99}\,D_2$$

Probit model sonuçları aşağıdaki gibidir. Parantez içindekiler, asimptotik *t* değerleridir.

$$D_1 = \underset{(0,61)}{6.92} + \underset{(2,00)}{0,0113}\,T + \underset{(2,05)}{6,46}\,Y - \underset{(-1,59)}{0,409}\,LF + \underset{(2,05)}{42,50}\,NW + \underset{(0,04)}{4,63}\,D_2$$

Logit katsayılarının probit katsayıları ile mukayese edilebilmesi için daha önce de ifade edildiği gibi 1,6'ya bölünmesi gerekmektedir. Bu bölme işlemi sırasıyla 6,87, 0,0121, 6,63, -0,418, 44,37 ve 8,33 değerlerini sağlar. Görüldüğü gibi bu katsayılar probit

katsayılarına çok yakındırlar. Şaşırtıcı olarak D_2 önemli değildir. Fakat diğer tüm katsayılar doğrusal ihtimal modelinde olduğu gibi aynı işaretlere sahiptirler. Y'nin katsayısı ise hala pozitif ve önemlidir.

Daha öncede belirtildiği gibi logit modelin katsayıları doğrusal ihtimal modelinin yaklaşık dört misli olmalıdır. Fakat burada elde edilen katsayılar dört mislinden çok daha fazladır. Bunun böyle olmasının sebeplerinden biri doğrusal ihtimal modelinin sağladığı zayıf model uyumu olabilir. Bu durumu açıklığa kavuşturmak için, R^2'nin değişik ölçümleri hesaplandığında doğrusal ihtimal modelinden elde edilen R^2'lerin, logit ve probit modellerinden elde edilen R^2'lerden çok daha düşük olduğu görülecektir.

Çizelge 8.3'de dört farklı R^2 ölçüm sonucu verilmiştir. İlk ikisinin hesaplaması kolay ve R^2'nin kabul edilebilen ölçümleridir. Diğer ikisi Cragg-Uhler ve McFadden tarafından hesaplanan R^2'lerdir. Sonuçlar logit ve probit arasında seçim yapmaya gerek olmadığı fakat bu ikisinin doğrusal ihtimal modelinden çok daha iyi olduğu görülmektedir. Pratiklik ve kolaylık açısından D_1 ve D_2 arasındaki korelasyonun karesi ve Efron'un hesapladığı R^2 çoğu problemlerin çözümü için yeterlidir.

Çizelge 8.3. Farklı modelle için farklı hesaplanan R^2 değerleri

Farklı yöntemlerle hesaplanan R^2 değerleri	Logit	Probit	Doğrusal İhtimal
D_1 ve D_2 arasındaki korelasyonun karesi	0,6117	0,6099	0,3376
Effron tarafından hesaplanan R^2	0,6116	0,6095	0,3376
Cragg-Uhler tarafından hesaplanan R^2	0,7223	0,7528	0,5273
McFadden tarafından hesaplanan R^2	0,6083	0,6124	0,4029

Probit ve logit modellere karar verildiğinden ve D_2 bu modellerde önemli olmadığından bu değişkeni modelden çıkarıp logit ve probit modellerini yeniden tahmin edilebilir. Yeni tahminler sırasıyla aşağıdaki gibi verilmiştir. Parantez içindeki rakamlar, asimptotik *t* değerleridir.

$$D_1 = \underset{(0,84)}{16,57} + \underset{(1,72)}{0,0165}\,T + \underset{(1,81)}{9,13}\,Y - \underset{(-1,49)}{0,715}\,LF + \underset{(2,38)}{0,715}\,NW$$

$R^2\,(D_1, D_2) = 0,5982$; Effron's $R^2 = 0,5982$;
Craff-Uhler's $R^2 = 0,7077$; cFadden's $R^2 = 0,5914$

$$D_1 = \underset{(0,98)}{10,27} + \underset{(1,86)}{0,0094}\,T + \underset{(1,97)}{5,55}\,Y - \underset{(-1,7)}{0,437}\,LF + \underset{(2,50)}{50,25}\,NW$$

$R^2\,(D_1, D_2) = 0,5950$; Effron's $R^2 = 0,5947$;
Craff-Uhler's $R^2 = 0,7113$; cFadden's $R^2 = 0,5955$

Tekrar, logit katsayılarını probit katsayıları ile kıyaslayabilmek için, önceki katsayılar 1.6'ya bölünmesi gerekmektedir. Bu işlem, sırasıyla 10,36, 0,0103, 5,71, 0,447 ve 53,35 elde edilir ki bu değerler probit katsayılarına yakındır.

YAZILIM UYGULAMASI

```
*******************************************************************************
   SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
   ** Copyright (C)  1999  by K.J. White - All Rights Reserved **
   FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
   AT: ATATURK UNIVERSITESI - ERZURUM
   FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
*******************************************************************************
```

Uygulamalı Örnek 8.2, sayfa 136

```
*******************************************************************************
|_*Burada üç ayrı veri dosyası oluşturulmuştur. MTABLE 8.2 çizelge 8.1'deki
|_*verileri içermektedir. Ayrıca yine çizelge 8.2'deki kullanılabilir gelir
|_*ve fiyat indeksleri sırasıyla DISPOSY ve PRINDEX dosyalarında
|_*bulunmaktadır. MTABLE8.2'de değişkenler sütun halinde değil peş peşe
|_*yazılmıştır. Önce yıllar peş peşe sonra tüketim şeklinde yazılmıştır. Bu
|_*nedenle READ komutundan sonra BYVAR opsiyonu kullanılmıştır.
|_SAMPLE 1 15
|_READ (MTABLE8.2) YEAR X1 PRBF X2 PRPRK X4 PRLMB X5 PRVL X3 PRCHCK / BYVAR LIST
 UNIT 88 IS NOW ASSIGNED TO: MTABLE8.2
   11 VARIABLES AND       15 OBSERVATIONS STARTING AT OBS       1
     YEAR
    1948.000       1949.000       1950.000       1951.000       1952.000
    1593.000       1954.000       1955.000       1956.000       1957.000
    1958.000       1959.000       1960.000       1961.000       1962.000
     X1
    63.10000       63.90000       63.40000       56.10000       62.20000
    77.60000       80.10000       82.00000       85.40000       84.60000
    80.50000       81.40000       85.20000       88.00000       89.10000
      .              .              .              .              .
     PRCHCK
    75.40000       71.80000       68.00000       66.00000       65.00000
    62.80000       56.40000       58.70000       50.40000       47.60000
    45.80000       41.40000       41.40000       37.00000       38.60000
|_* Şimdi diğer iki veri dosyasını okuyalım.
|_READ (DISPOSY) Y / LIST
 UNIT 88 IS NOW ASSIGNED TO: DISPOSY
    1 VARIABLES AND       15 OBSERVATIONS STARTING AT OBS       1
       Y
    1291.000
      .
    2052.000
|_READ (PRINDEX) PRINDEX / LIST
 UNIT 88 IS NOW ASSIGNED TO: PRINDEX
    1 VARIABLES AND       15 OBSERVATIONS STARTING AT OBS       1
       PRINDEX
   0.8380000
      .
    1.054000
|_* Şimdi indeks kullanarak reel haldeki fiyatları nominal hale getirelim.
|_GENR P1=PRBF*PRINDEX
|_GENR P2=PRPRK*PRINDEX
|_GENR P3=PRCHCK*PRINDEX
|_*Şimdi fiyat vektörü oluşturmak için MATRIX komutu kullanılacaktır.
|_*Burada (') sembolü matrisin transpozunu alır ve (|) sembolü iki matrisi
|_*yan yana koyar.
```

```
|_MATRIX P=(P1'|P2'|P3')'
|_PRINT P
     P
    69.47020       63.32900       73.99540       81.54000       78.99500
    61.69840       59.99760       59.02880       57.67230       61.90110
    72.50400       74.39950       72.58240       71.10030       73.56920
    56.64880       51.04500       50.61520       54.90360       53.00250
    58.62280       59.62320       50.99640       48.67580       56.50560
    60.92350       53.59200       53.19960       55.48530       55.75660
    63.18520       59.59400       56.98400       59.79600       60.12500
    58.52960       52.79040       54.82580       47.72880       46.69560
    46.12060       42.02100       42.68340       38.51700       40.68440
|_* Sonraki GENR komutu, 15 gözlem ile sıfırlar ve birler sütunlarını veya
|_*vektörlerini oluştururlar. (örnek aralığına uygun olarak).
|_GENR ONE=1
|_GENR ZERO=0
|_SAMPLE 1 45
|_*Şimdi, sayfa 130'da isimlendirilen bazı vektörleri onların
|_*katsayılarından sonra oluşturalım. Onların kendileri katsayı
|_*olmadığından dolayı bu bir şekilde yanlış yönlendirmektedir. Karışıklığı
|_*önlemek için bu akılda tutulmalıdır.
|_MATRIX ALPHA1=(ONE'|ZERO'|ZERO')'
|_MATRIX ALPHA2=(ZERO'|ONE'|ZERO')'
|_MATRIX ALPHA3=(ZERO'|ZERO'|ONE')'
|_MATRIX BETA11=(X1'|ZERO'|ZERO')'
|_MATRIX BETA12=(X2'|X1'|ZERO')'
|_MATRIX BETA13=(X3'|ZERO'|X1')'
|_MATRIX BETA22=(ZERO'|X2'|ZERO')'
|_MATRIX BETA23=(ZERO'|X3'|X2')'
|_MATRIX BETA33=(ZERO'|ZERO'|X3')'
|_MATRIX GAMMA1=(Y'|ZERO'|ZERO')'
|_MATRIX GAMMA2=(ZERO'|Y'|ZERO')'
|_MATRIX GAMMA3=(ZERO'|ZERO'|Y')'
|_PRINT ALPHA1
     ALPHA1
    1.000000       1.000000       1.000000       1.000000       1.000000
    1.000000       1.000000       1.000000       1.000000       1.000000
    1.000000       1.000000       1.000000       1.000000       1.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
|_PRINT GAMMA3
     GAMMA3
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    0.000000       0.000000       0.000000       0.000000       0.000000
    1291.000       1271.000       1369.000       1473.000       1520.000
    1582.000       1582.000       1660.000       1742.000       1804.000
    1826.000       1904.000       1934.000       1980.000       2052.000
```

```
|_*Şimdi tüm regresyonu çalıştıralım. Hatırlayalım ki & sembolü komutun
|_*bir sonraki satırda devam ettiğini gösterir.
|_SAMPLE 1 45
|_OLS P ALPHA1 ALPHA2 ALPHA3 BETA11 BETA12 BETA13 BETA22 BETA23 BETA33 &
| GAMMA1 GAMMA2 GAMMA3 / NOCONSTANT
        45 OBSERVATIONS     DEPENDENT VARIABLE= P
 ...NOTE..SAMPLE RANGE SET TO:      1,     45
  R-SQUARE =   0.9683     R-SQUARE ADJUSTED =   0.9578
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   4.3303
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   2.0809
 SUM OF SQUARED ERRORS-SSE=   142.90
 MEAN OF DEPENDENT VARIABLE =   58.259
 RAW MOMENT R-SQUARE =   0.9991
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      33 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 ALPHA1     121.39      10.82       11.22     0.000 0.890     5.7161     0.6946
 ALPHA2     149.33      17.79       8.395     0.000 0.825     7.0317     0.8544
 ALPHA3     129.52      19.42       6.670     0.000 0.758     6.0986     0.7410
 BETA11    -1.5462     0.1149      -13.45     0.000-0.920    -5.6279    -0.6739
 BETA12   -0.46804     0.1201      -3.898     0.000-0.562    -1.6065    -0.3802
 BETA13   -0.42763     0.1614      -2.649     0.012-0.419    -1.3860    -0.2457
 BETA22    -1.1835     0.2090      -5.664     0.000-0.702    -3.6758    -0.4455
 BETA23   -0.30130     0.2267      -1.329     0.193-0.225    -0.8230    -0.1553
 BETA33    -2.3473     0.6024      -3.897     0.000-0.561    -2.7328    -0.3259
 GAMMA1   0.63833E-01 0.5623E-02   11.35     0.000 0.892     5.0867     0.6085
 GAMMA2   0.15690E-01 0.6219E-02   2.523     0.017 0.402     1.2503     0.1496
 GAMMA3   0.18725E-01 0.1302E-01   1.438     0.160 0.243     1.4921     0.1785
|_*SYSTEM komutu, RESTRICT opsiyonu ve RESTRICK komutları kullanılarak
|_*yukarıda elde edilen sonuçların aynısı daha kolay bir şekilde elde
|_*edilebilir. SYSTEM komutu, denklemler setinin tahmini için bir prosedür
|_*sağlar. Değişken sayısı, SYSTEM komutunu takiben belirlenir. Buradaki
|_*örnekte bu sayı üçtür. ITER=0 komutu, denklemler arası eşit varyansın
|_*olduğu varsayılır. Yani V(u1) = V(u2) = V(u3) ve denklemler arası
|_*korelasyon mevcut değildir. İlk RESTRICT komutu, ikinci denklemdeki
|_*x1'in katsayısının birinci denklemdeki x2'nin katsayısına eşit olduğunu
|_*modele empoze etmek için kullanılır. Diğer iki RESTRICT komutu benzer
|_*işlemi diğer katsayılar için yapar. Bu simetri kısıtları ders notlarında
|_*ifade edilmiştir. Bir seri RESTRICT komutunu END komutu takip eder.
|_SAMPLE 1 15
|_SYSTEM 3 / RESTRICT ITER=0
|_OLS P1 X1 X2 X3 Y
|_OLS P2 X1 X2 X3 Y
|_OLS P3 X1 X2 X3 Y
|_RESTRICT X1:2=X2:1
|_RESTRICT X1:3=X3:1
|_RESTRICT X2:3=X3:2
|_END
 MULTIVARIATE REGRESSION--    3 EQUATIONS
  12 RIGHT-HAND SIDE VARIABLES IN SYSTEM
 MAX ITERATIONS =       0         CONVERGENCE TOLERANCE =   0.10000E-02
BREUSCH-PAGAN LM TEST FOR DIAGONAL COVARIANCE MATRIX
    CHI-SQUARE =    6.9352     WITH    3 DEGREES OF FREEDOM
  SYSTEM R-SQUARE =  0.9997 ... CHI-SQUARE =   121.08     WITH   9 D.F.
 VARIABLE       COEFFICIENT    ST.ERROR       T-RATIO
 X1          -1.5462        0.15024        -10.292
 X2         -0.46804        0.15694        -2.9823
```

```
 X3          -0.42763        0.21105        -2.0263
 Y            0.63833E-01    0.73503E-02     8.6845
 X1          -0.46804        0.10027        -4.6679
 X2           -1.1835        0.17451        -6.7819
 X3          -0.30130        0.18930        -1.5917
 Y            0.15690E-01    0.51942E-02     3.0206
 X1          -0.42763        0.12440        -3.4374
 X2          -0.30130        0.17466        -1.7251
 X3           -2.3473        0.46419        -5.0568
 Y            0.18725E-01    0.10032E-01     1.8666
  EQUATION  1 OF  3 EQUATIONS
 DEPENDENT VARIABLE = P1                    15 OBSERVATIONS
  R-SQUARE =   0.8990
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   6.2611
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   2.5022
 SUM OF SQUARED ERRORS-SSE=   81.395
 MEAN OF DEPENDENT VARIABLE =   68.786
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      13 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X1         -1.5462     0.1502      -10.29     0.000-0.944    -2.2622    -1.7123
 X2        -0.46804     0.1569      -2.982     0.011-0.637    -0.2420    -0.4477
 X3        -0.42763     0.2110      -2.026     0.064-0.490    -0.2256    -0.1509
 Y          0.63833E-01 0.7350E-02   8.684     0.000 0.924     2.1143     1.5461
 CONSTANT   121.39      14.14       8.585     0.000 0.922     0.0000     1.7648
  EQUATION  2 OF  3 EQUATIONS
 DEPENDENT VARIABLE = P2                    15 OBSERVATIONS
  R-SQUARE =   0.8086
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   2.5556
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.5986
 SUM OF SQUARED ERRORS-SSE=   33.223
 MEAN OF DEPENDENT VARIABLE =   54.640
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      13 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X1        -0.46804     0.1003      -4.668     0.000-0.791    -1.4756    -0.6525
 X2         -1.1835     0.1745      -6.782     0.000-0.883    -1.3184    -1.4251
 X3        -0.30130     0.1893      -1.592     0.135-0.404    -0.3425    -0.1338
 Y          0.15690E-01 0.5194E-02   3.021     0.010 0.642     1.1198     0.4784
 CONSTANT   149.33      14.86       10.05     0.000 0.941     0.0000     2.7331
  EQUATION  3 OF  3 EQUATIONS
 DEPENDENT VARIABLE = P3                    15 OBSERVATIONS
  R-SQUARE =   0.9704
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   2.1756
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   1.4750
 SUM OF SQUARED ERRORS-SSE=   28.282
 MEAN OF DEPENDENT VARIABLE =   51.352
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      13 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 X1        -0.42763     0.1244      -3.437     0.004-0.690    -0.5743    -0.6343
 X2        -0.30130     0.1747      -1.725     0.108-0.432    -0.1430    -0.3860
 X3         -2.3473     0.4642      -5.057     0.000-0.814    -1.1365    -1.1092
 Y          0.18725E-01 0.1003E-01   1.867     0.085 0.460     0.5693     0.6075
 CONSTANT   129.52      14.96       8.656     0.000 0.923     0.0000     2.5221
|_DELETE / ALL
 ALL VARIABLES HAVE BEEN DELETED
|_STOP
```

```
*************************************************************************
```

Uygulamalı Örnek 8.3, sayfa 140

```
*************************************************************************
|_* Önce M'nin tüm değişkenler üzerine regresyonunu tahmin edelim.
|_SAMPLE 1 44
|_READ (MTABLE8.4) M PC PX D1 T Y LF NW D2 / BYVAR
 UNIT 88 IS NOW ASSIGNED TO: MTABLE8.4
    9 VARIABLES AND       44 OBSERVATIONS STARTING AT OBS       1
|_OLS M PC PX D1 T Y LF NW D2
        44 OBSERVATIONS     DEPENDENT VARIABLE= M
 ...NOTE..SAMPLE RANGE SET TO:      1,     44
  R-SQUARE =   0.7746     R-SQUARE ADJUSTED =   0.7230
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   5.5176
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   2.3490
 SUM OF SQUARED ERRORS-SSE=   193.12
 MEAN OF DEPENDENT VARIABLE =   5.4036
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      35 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 PC        -3.6960      2.676      -1.381     0.176-0.227    -0.1174    -0.1782
 PX        -3.5678      6.593     -0.5411     0.592-0.091    -0.0549    -0.0398
 D1         2.5978      1.231       2.110     0.042 0.336     0.2375     0.3824
 T        -0.17941E-01 0.6836E-02  -2.624     0.013-0.406    -0.2476    -0.4533
 Y         -4.0951      1.778      -2.303     0.027-0.363    -0.3634    -1.3496
 LF        0.39973     0.2198       1.819     0.078 0.294     0.2221     3.9255
 NW         6.4439      5.494       1.173     0.249 0.194     0.1646     0.1259
 D2         2.5408      1.316       1.931     0.062 0.310     0.2729     0.1603
 CONSTANT  -8.5011      10.42     -0.8158     0.420-0.137     0.0000    -1.5732
|_* Şimdi doğrusal ihtimal modelini tahmin edelim.
|_OLS D1 T Y LF NW D2
        44 OBSERVATIONS     DEPENDENT VARIABLE= D1
 ...NOTE..SAMPLE RANGE SET TO:      1,     44
 ...WARNING..DEPENDENT VARIABLE IS DUMMY VARIABLE ...TRY LOGIT or PROBIT
  R-SQUARE =   0.3376     R-SQUARE ADJUSTED =   0.2504
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.12480
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.35327
 SUM OF SQUARED ERRORS-SSE=   4.7425
 MEAN OF DEPENDENT VARIABLE =  0.79545
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      38 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 T         0.14615E-02 0.9978E-03   1.465     0.151 0.231     0.2207     0.2508
 Y         0.65773     0.2404       2.736     0.009 0.406     0.6385     1.4726
 LF       -0.54562E-01 0.2826E-01  -1.931     0.061-0.299    -0.3316    -3.6399
 NW         1.9882     0.7595       2.618     0.013 0.391     0.5555     0.2639
 D2        0.34336     0.1892       1.814     0.078 0.282     0.4035     0.1472
 CONSTANT   1.9929      1.326       1.503     0.141 0.237     0.0000     2.5054
|_*
|_*Şimdi LOGIT komutunu kullanarak logit modelini çalıştıralım. LOGIT
|_*komutu bağımlı değişkenin 0 ve 1 değerini alan bir kukla değişkeni
|_*olduğunda kullanılır. Burada BETA(6) sabit üzerine olan katsayıdır.
|_*
|_LOGIT D1 T Y LF NW D2 / COEF=BETA
  LOGIT ANALYSIS     DEPENDENT VARIABLE =D1       CHOICES =  2
       44. TOTAL OBSERVATIONS
       35. OBSERVATIONS AT ONE
        9. OBSERVATIONS AT ZERO
   25 MAXIMUM ITERATIONS
 CONVERGENCE TOLERANCE =0.00100
```

```
 BINOMIAL  ESTIMATE = 0.7955
 ITERATION  0      LOG OF LIKELIHOOD FUNCTION =   -22.292
                                ASYMPTOTIC                          WEIGHTED
 VARIABLE    ESTIMATED      STANDARD     T-RATIO    ELASTICITY      AGGREGATE
   NAME     COEFFICIENT       ERROR                  AT MEANS      ELASTICITY
 T             0.19426E-01  0.10397E-01   1.8685      0.20863E-02  0.24351
 Y               10.610       5.6534      1.8768      0.14865E-01   1.5263
 LF             -0.66763      0.47647     -1.4012     -0.27871E-01  -2.7232
 NW              70.988       36.375       1.9516      0.58967E-02  0.13491
 D2              7.4470       20.744      0.35900      0.19972E-02  0.18596E-03
 CONSTANT        10.993       20.762      0.52950      0.86482E-02  0.83848
 MADDALA R-SQUARE              0.4600
 CRAGG-UHLER R-SQUARE          0.72222
 MCFADDEN R-SQUARE             0.60817
      ADJUSTED FOR DEGREES OF FREEDOM        0.55662
      APPROXIMATELY F-DISTRIBUTED     1.8626      WITH        5  AND     6  D.F.
 CHOW R-SQUARE                 0.61162
          PREDICTION SUCCESS TABLE
                       ACTUAL
                   0            1
           0      7.           1.
 PREDICTED 1      2.          34.

 NUMBER OF RIGHT PREDICTIONS =        41.0
 PERCENTAGE OF RIGHT PREDICTIONS =    0.93182
 EXPECTED OBSERVATIONS AT 0  =          9.0   OBSERVED =      9.0
 EXPECTED OBSERVATIONS AT 1  =         35.0   OBSERVED =     35.0
 SUM OF SQUARED "RESIDUALS" =           2.7805
 WEIGHTED SUM OF SQUARED "RESIDUALS" =     18.808
HENSHER-JOHNSON PREDICTION SUCCESS TABLE
                                                 OBSERVED    OBSERVED
                      PREDICTED  CHOICE           COUNT       SHARE
         ACTUAL            0          1
            0           6.275      2.725       9.000       0.205
            1           2.725     32.275      35.000       0.795
 PREDICTED COUNT        9.001     34.999      44.000       1.000
 PREDICTED SHARE        0.205      0.795       1.000
 PROP. SUCCESSFUL       0.697      0.922       0.876
 SUCCESS INDEX          0.493      0.127       0.202
 PROPORTIONAL ERROR     0.000      0.000
 NORMALIZED SUCCESS INDEX                      0.619
|_*Şimdi probit modelinin tahmini için PROBIT komutunu kullanalım. PROBIT
|_*komutu, LOGIT komutunda olduğu gibi bağımlı değişkenin 0 ve 1
|_*değerlerini alan kukla değişkeni olması durumundaki modellerin
|_*tahmininde kullanılır.
|_PROBIT D1 T Y LF NW D2
 REQUIRED MEMORY IS PAR=       5 CURRENT PAR=    1000
 FOR MAXIMUM EFFICIENCY USE AT LEAST PAR=       7
 PROBIT ANALYSIS     DEPENDENT VARIABLE =D1       CHOICES =  2
       44. TOTAL OBSERVATIONS
       35. OBSERVATIONS AT ONE
        9. OBSERVATIONS AT ZERO
   25 MAXIMUM ITERATIONS
 CONVERGENCE TOLERANCE =0.00100
 BINOMIAL  ESTIMATE = 0.7955
 ITERATION  0      LOG OF LIKELIHOOD FUNCTION =   -22.292
                               ASYMPTOTIC                          WEIGHTED
 VARIABLE    ESTIMATED      STANDARD     T-RATIO    ELASTICITY      AGGREGATE
```

```
  NAME     COEFFICIENT       ERROR                AT MEANS      ELASTICITY
T            0.11312E-01  0.56554E-02   2.0003      0.36724E-03  0.25247
Y              6.4611       3.1480      2.0525      0.27362E-02   1.6413
LF            -0.40930      0.25713     -1.5918     -0.51648E-02  -2.9505
NW             42.496       20.730      2.0500      0.10670E-02  0.14353
D2             3.2825       7.0084     0.46837      0.26610E-03  0.24535E-03
CONSTANT       6.9163       11.316     0.61120      0.16447E-02  0.93214
MADDALA R-SQUARE             0.4623
CRAGG-UHLER R-SQUARE         0.72575
MCFADDEN R-SQUARE            0.61230
     ADJUSTED FOR DEGREES OF FREEDOM        0.56128
     APPROXIMATELY F-DISTRIBUTED     1.8951   WITH     5  AND     6  D.F.
CHOW R-SQUARE                0.60956
         PREDICTION SUCCESS TABLE
                       ACTUAL
                  0            1
           0      7.           2.
PREDICTED 1      2.           33.
NUMBER OF RIGHT PREDICTIONS =         40.0
PERCENTAGE OF RIGHT PREDICTIONS =    0.90909
EXPECTED OBSERVATIONS AT 0  =          9.1   OBSERVED =      9.0
EXPECTED OBSERVATIONS AT 1  =         34.9   OBSERVED =     35.0
SUM OF SQUARED "RESIDUALS" =           2.7952
WEIGHTED SUM OF SQUARED "RESIDUALS" =     18.085

HENSHER-JOHNSON PREDICTION SUCCESS TABLE
                                          OBSERVED    OBSERVED
                    PREDICTED  CHOICE       COUNT       SHARE
        ACTUAL           0          1
           0           6.328      2.672      9.000      0.205
           1           2.744     32.256     35.000      0.795

PREDICTED COUNT         9.073     34.927     44.000      1.000
PREDICTED SHARE         0.206      0.794      1.000
PROP. SUCCESSFUL        0.698      0.924      0.877
SUCCESS INDEX           0.491      0.130      0.204
PROPORTIONAL ERROR      0.002     -0.002
NORMALIZED SUCCESS INDEX                     0.624
|_*Logit modelinden elde edilen katsayılar 1.6 ile bölünerek bu
|_*katsayıların probit katsayılarıyla karşılaştırılabilmesi sağlanır.
|_SAMPLE 1 6
|_GENR CONVBETA=BETA/1.6
|_PRINT CONVBETA
     CONVBETA
   0.1214109E-01   6.631333     -0.4172703       44.36738       4.654377
    6.870800
|_*Logit ve probit komutları otomatik olarak çok sayıda R² değerleri
|_*hesapladığı görülmektedir. Şimdi D₂'yi modelden çıkararak Logit ve
|_*Probit modellerini yeniden tahmin edelim.
|_SAMPLE 1 44
|_LOGIT D1 T Y LF NW / COEF=NBETA
 REQUIRED MEMORY IS PAR=       5 CURRENT PAR=    1000
 FOR MAXIMUM EFFICIENCY USE AT LEAST PAR=       7
  LOGIT ANALYSIS     DEPENDENT VARIABLE =D1       CHOICES =  2
       44. TOTAL OBSERVATIONS
       35. OBSERVATIONS AT ONE
```

```
       9. OBSERVATIONS AT ZERO
  25 MAXIMUM ITERATIONS
 CONVERGENCE TOLERANCE =0.00100
 BINOMIAL  ESTIMATE = 0.7955
 ITERATION  0      LOG OF LIKELIHOOD FUNCTION =   -22.292
                                ASYMPTOTIC                        WEIGHTED
 VARIABLE    ESTIMATED      STANDARD     T-RATIO    ELASTICITY      AGGREGATE
   NAME     COEFFICIENT       ERROR                 AT MEANS      ELASTICITY
 T            0.16516E-01  0.96099E-02   1.7187      0.48732E-02  0.21230
 Y             9.1315       5.0525       1.8073      0.35146E-01   1.3626
 LF           -0.71539      0.47922      -1.4928     -0.82044E-01  -3.0360
 NW            85.362       35.831        2.3823      0.19480E-01  0.16998
 CONSTANT      16.567       19.635       0.84373      0.35804E-01   1.3171
 MADDALA R-SQUARE             0.4508
 CRAGG-UHLER R-SQUARE         0.70773
 MCFADDEN R-SQUARE            0.59144
      ADJUSTED FOR DEGREES OF FREEDOM        0.54954
      APPROXIMATELY F-DISTRIBUTED     1.8096   WITH    4  AND   5  D.F.
 CHOW R-SQUARE                0.59815
          PREDICTION SUCCESS TABLE
                         ACTUAL
                   0             1
            0      7.            1.
 PREDICTED  1      2.           34.
 NUMBER OF RIGHT PREDICTIONS =         41.0
 PERCENTAGE OF RIGHT PREDICTIONS =    0.93182
 EXPECTED OBSERVATIONS AT 0  =          9.0   OBSERVED =      9.0
 EXPECTED OBSERVATIONS AT 1  =         35.0   OBSERVED =     35.0
 SUM OF SQUARED "RESIDUALS" =           2.8769
 WEIGHTED SUM OF SQUARED "RESIDUALS" =     20.201
 HENSHER-JOHNSON PREDICTION SUCCESS TABLE
                                              OBSERVED    OBSERVED
                     PREDICTED  CHOICE         COUNT       SHARE
         ACTUAL           0          1
            0           6.170      2.830      9.000       0.205
            1           2.830     32.170     35.000       0.795
 PREDICTED COUNT        9.000     35.000     44.000       1.000
 PREDICTED SHARE        0.205      0.795      1.000
 PROP. SUCCESSFUL       0.686      0.919      0.871
 SUCCESS INDEX          0.481      0.124      0.197
 PROPORTIONAL ERROR     0.000      0.000
 NORMALIZED SUCCESS INDEX                     0.605
|_PROBIT D1 T Y LF NW
 REQUIRED MEMORY IS PAR=       5 CURRENT PAR=    1000
 FOR MAXIMUM EFFICIENCY USE AT LEAST PAR=       7
 PROBIT ANALYSIS     DEPENDENT VARIABLE =D1        CHOICES =  2
       44. TOTAL OBSERVATIONS
       35. OBSERVATIONS AT ONE
        9. OBSERVATIONS AT ZERO
   25 MAXIMUM ITERATIONS
ITERATION  0      LOG OF LIKELIHOOD FUNCTION =   -22.292
                                ASYMPTOTIC                        WEIGHTED
 VARIABLE    ESTIMATED      STANDARD     T-RATIO    ELASTICITY      AGGREGATE
   NAME     COEFFICIENT       ERROR                 AT MEANS      ELASTICITY
 T            0.94213E-02  0.50703E-02   1.8581      0.84647E-03  0.21721
 Y             5.5511       2.8121       1.9740      0.65059E-02   1.4695
```

```
 LF          -0.43654      0.25642      -1.7025     -0.15245E-01  -3.2898
 NW            50.249       20.087       2.5016      0.34917E-02  0.17744
 CONSTANT      10.267       10.497      0.97812      0.67569E-02   1.4488
 MADDALA R-SQUARE             0.4531
 CRAGG-UHLER R-SQUARE         0.71128
 MCFADDEN R-SQUARE            0.59552
      ADJUSTED FOR DEGREES OF FREEDOM         0.55404
      APPROXIMATELY F-DISTRIBUTED    1.8404    WITH    4  AND     5  D.F.
 CHOW R-SQUARE                0.59467
          PREDICTION SUCCESS TABLE
                          ACTUAL
                   0             1
           0       7.            1.
 PREDICTED 1       2.           34.
NUMBER OF RIGHT PREDICTIONS =         41.0
 PERCENTAGE OF RIGHT PREDICTIONS =     0.93182
 EXPECTED OBSERVATIONS AT 0  =           9.1   OBSERVED =       9.0
 EXPECTED OBSERVATIONS AT 1  =          34.9   OBSERVED =      35.0
 SUM OF SQUARED "RESIDUALS" =            2.9018
 WEIGHTED SUM OF SQUARED "RESIDUALS" =     19.071
 HENSHER-JOHNSON PREDICTION SUCCESS TABLE
                                              OBSERVED    OBSERVED
                      PREDICTED  CHOICE         COUNT       SHARE
        ACTUAL             0          1
           0             6.214      2.786       9.000       0.205
           1             2.873     32.127      35.000       0.795
 PREDICTED COUNT         9.087     34.913      44.000       1.000
 PREDICTED SHARE         0.207      0.793       1.000
 PROP. SUCCESSFUL        0.684      0.920       0.871
 SUCCESS INDEX           0.477      0.127       0.199
 PROPORTIONAL ERROR      0.002     -0.002
 NORMALIZED SUCCESS INDEX                      0.608
 |_SAMPLE 1 5
 |_GENR NCNVBETA=NBETA/1.6
 |_PRINT NCNVBETA
     NCNVBETA
   0.1032279E-01   5.707218     -0.4471184       53.35098       10.35437
 |_STOP
```

$$Y_i = \beta_o + \beta X + \varepsilon_i$$

Y_i=0 if lived, Y_i=1 if died

Prob (Y_i=1) = F(X, β)

Prob (Y_i=0) = 1 − F(X,β)

OLS, also called a linear probability model

ε_i is heteroscedastic, depends on βX

Predictions not constrained to (0,1)

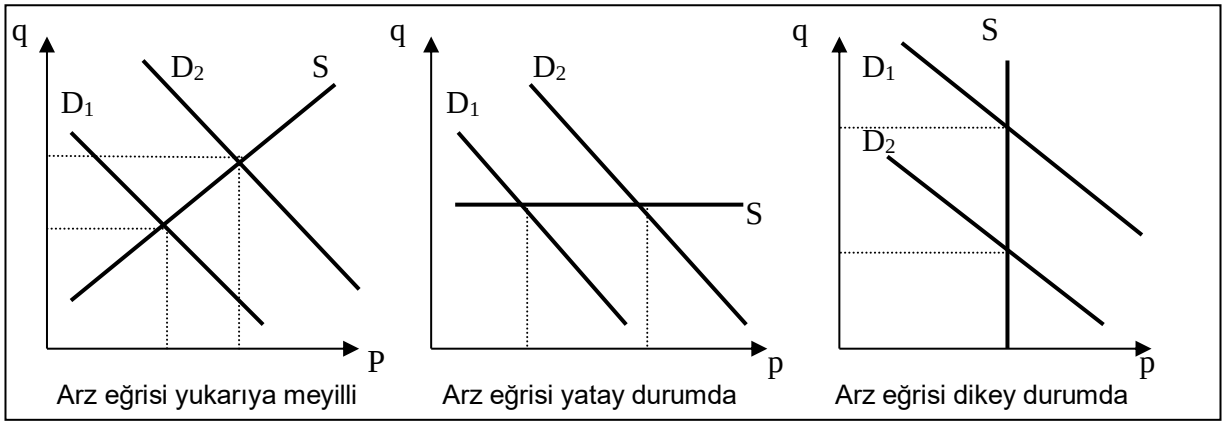
IX. SİMULTANE DENKLEM MODELLERİ

*9.1. Giriş

Bilinen regresyon modelinde eşitliğin solunda y bağımlı veya belirlenen değişkendir. Eşitliğin sağında x_1, x_2, ... x_n ise bağımsız veya belirleyen değişkenlerdir. Burada yapılan önemli varsayım, x'lerin hata teriminden (u) bağımsız olmalarıdır. Bazen bu varsayım örneğin talep ve arz modellerinde olduğu gibi çiğnenir. Arz ve talep modelleri ele alındığında talep fonksiyonu denklem 9.1'deki gibi yazılabilir.

$$q = \alpha + \beta p + u \qquad (9.1)$$

Burada; q talep edilen miktarı, p fiyatı ve u hata terimini temsil etmektedir. Bu ekonometri denklemindeki hata terimi talepteki şansa bağlı kaymaları ifade etmektedir. Aşağıdaki şekil 9.1'de de görüldüğü gibi talep fonksiyonundaki bir değişme, eğer arz eğrisi yukarı doğru meyilli ise hem miktar hem de fiyatta bir değişmeye neden olmaktadır. Eğer arz eğrisi yatay yani tamamen inelastik ise talep eğrisindeki bir kayma sadece fiyatta bir değişme meydana getirir. Eğer arz eğrisi dikey yani tamamen elastik ise talep eğrisindeki değişme sadece miktarda bir değişmeye neden olur.



Şekil 9.1. Talep fonksiyonundaki kaymaların fiyat üzerine etkileri

Bu nedenle eğer arz eğrisi yukarı doğru meyilli veya tamamen yatay ise hata terimi u, p ile korelasyona sahiptir. Bundan dolayı en küçük kareler yöntemi ile denklemin tahmin edilmesi parametrelerin tutarsız tahminlerini verir. Talep denklemi denklem 9.1''deki gibi de yazılmış olabilir.

$$P = \alpha' + \beta' q + \hat{u} \qquad (9.1')$$

Fakat bu durumda da eğer arz eğrisi yukarı doğru meyilli veya tamamen dikey ise û, q ile korelasyona sahip olacaktır. Buradaki soru, talep fonksiyonu ilk denklemdeki gibi mi yoksa ikinci denklemdeki gibi mi yazılmalıdır. Eğer birinci denklemdeki gibi yazılırsa denklem q'ya göre normalleştirilmiş, yok eğer ikinci denkleme göre yazılırsa denklem p'ye göre normalleştirilmiş olur. Fiyat ve miktar arz ve talep denklemlerinin

karşılıklı etkileşimi sonucu belirlendiğinden, denklemlerin normalleştirilmesinin p'ye veya q'ya göre olması fark etmez.

Kullanılan tahmin metotları ve normalleştirme ne şekilde olursa olsun aynı sonuçları verir. Bu durumda yapılacak şey, yukarıda verilen şekildeki miktar ve fiyat arasındaki ilişki çalışılırken talep denkleminin piyasayı oluşturan sistemden ayrı olarak analiz edilmemesidir. Bu durumda çözüm arz fonksiyonunun da dikkate alınması ve arz ve talep denkleminin birlikte tahmin edilmesidir.

Özet olarak en küçük karelerin tek bir denkleme uygulanması, başka şeylerin yanı sıra açıklayıcı değişkenlerin gerçekten dışsal olduklarını, yani bağımlı y değişkeni ile açıklayıcı x'ler arasındaki nedenselliğin tek yönlü olduğunu varsayar. Eğer bu doğru değil ise yani x'ler aynı zamanda y tarafından belirleniyorsa sıradan en küçük kareler yönteminin kullanılması sapmalı ve tutarsız tahminler verir. Eğer bir fonksiyonda iki yönlü nedensellik varsa, bu fonksiyon tek denklemli bir model olarak kendi başına ele alınamaz. İlgili bütün değişkenler arasındaki ilişkileri tanımlayan daha geniş bir denklem takımı dikkate alınmalıdır. Eğer y = f (x) iken x = f (y) ise y ve x arasındaki ilişkiyi tanımlamak için tek denklemli bir model kullanılamaz. Burada çok denklemli bir model kullanmak gerekir. Bu modelde y ve x'in dahili değişkenler olarak kabul edildiği ayrı ayrı denklemler bulunduğu gibi, modelin başka denklemlerinde bunlar açıklayıcı değişken olarak yer alabilirler. Değişkenlerin karşılıklı bağımlılığını gösteren bu denklemler grubuna Simultane/Eşzamanlı Denklem Modelleri denilmektedir.

*9.2. Dahili (Endojen) ve Harici (Eksojen) Değişkenler

Simultane denklem modellerindeki değişkenler dahili ve harici değişkenler olarak sınıflandırılırlar. Geleneksel olarak dahili değişkenler ekonomik model tarafından belirlenen harici değişkenler ise model dışında belirlenen değişkenler olarak tanımlanırlar. Dahili değişkenler müştereken belirlenen, harici değişkenler daha önceden başka bir ortamda belirlenen değişkenler olarak da ifade edilirler. Harici değişkenler bu özelliğinden dolayı modelde hata teriminden bağımsızdırlar.

Bu durum, bir örnekle 9.2'deki gibi gösterilebilir.

$$q = a_1 + b_1 p + c_1 Y + u_1 \qquad \text{talep fonksiyonu} \qquad (9.2)$$

$$q = a_2 + b_2 p + c_2 R + u_2 \qquad \text{arz fonksiyonu}$$

Bu modelde q miktarı, p fiyatı, Y geliri, R yağış miktarını, u_1 ve u_2'de hata terimlerini ifade etmektedir. Bu modelde p ve q dahili değişkenler Y ve R ise harici değişkenlerdir. Gelir ve yağış miktarı hata terimlerinden bağımsızdırlar. Bu model en küçük kareler metoduyla tahmin edilebilir. Ancak bulunacak değerler gerçek değerlerden sapmalı olacaktır. Bunun için yukarıdaki model dolaylı en küçük kareler yöntemiyle tahmin edilir. Fakat *dolaylı en küçük kareler yöntemi* her zaman çalışmaz. İleriki konularda çalışmama durumları ve bu durumda nelerin yapılacağı ve modelin nasıl

basitleştirileceği tartışılacaktır. Bu konuyu daha iyi anlamak için tanımlama kavramını açığa kavuşturmak gerekmektedir.

*9.3. Tanımlama Problemi ve İndirgenmiş Form ile Tahmin Etme

Yukarıdaki denklemlerde u_1 ve u_2'nin p ile korelasyona sahip olması nedeniyle eğer bu denklemler geleneksel en küçük kareler metodu ile tahmin edilirse parametre tahminleri tutarsız olacaktır. Kabaca ifade edilir ise tanımlama kavramı parametrelerin tutarlı tahmini ile ilgilidir. Bu nedenle eğer talep fonksiyonundaki parametrelerin sapmasız bir tahmini bir şekilde gerçekleştirilebilirse talep fonksiyonu tanımlanmıştır denir. Aynı şekilde arz fonksiyonunda parametrelerin sapmasız bir tahmini gerçekleştirilebilirse arz fonksiyonu tanımlanmıştır denir. Tutarlı tahminleri elde etmek tanımlama için gerek şarttır, fakat yeter şart değildir.

Şimdi bu tanımlama ile ilgili hususa açıklık getirilecektir. Bunun için 9.2'deki denklemler dikkate alındığında q ve p, Y ve R cinsinden aşağıdaki gibi tanımlanır.

$$q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} Y - \frac{c_2 b_1}{b_2 - b_1} R + \text{hata}$$

$$p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} Y - \frac{c_2}{b_2 - b_1} R + \text{hata} \qquad (9.3)$$

Bu 9.3'deki denklemlere *indirgenmiş formda denklemler* adı verilir. Yukarıda 9.2'deki ilk denklemler ise ekonomik sistemin yapısını ortaya koydukları için yapısal denklemler olarak adlandırılırlar. Bu indirgenmiş formdaki denklemlerin katsayıları sembolize edilerek ve kısaltma yapılarak denklem 9.4'deki gibi yazılabilirler.

$$q = \pi_1 + \pi_2 y + \pi_3 R + v_1 \qquad (9.4)$$

$$p = \pi_4 + \pi_5 y + \pi_6 R + v_2$$

Bu denklemlerde v_1 ve v_2 hata terimlerini ifade eder ve π'ler ise *indirgenmiş form parametreleridir*. Bu 9.4'deki denklemler modeli normal en küçük kareler yöntemiyle tahmin edilmesi indirgenmiş parametrelerin tutarlı tahminlerini verir. Buradan 9.2'deki denklemlerin yapısal parametreleri tutarlı olarak tahmin edilebilir. Bu *yapısal parametreler* aşağıdaki gibi hesaplanabilir.

$$b'_1 = \pi'_3 / \pi'_6, \qquad b'_2 = \pi'_2 / \pi'_5$$

$$c'_2 = \pi'_6 (b'_1 - b'_2) \qquad c'_1 = -\pi'_5 (b'_1 - b'_2)$$

$$a'_1 = \pi'_1 - b'_1 \pi_4 \qquad a'_2 = \pi'_1 - b'_2 \pi'_4$$

Burada a'_1, a'_2, b'_1, b'_2, c'_1, c'_2 değerleri π'nin tek değerli fonksiyonlarıdır ve yapısal parametrelerin tutarlı tahmincileridir. Bu metot daha öncede belirtildiği gibi *dolaylı en küçük kareler metodu* olarak bilinir.

Her zaman indirgenmiş denklemlerden yapısal katsayıların tahminlerini elde etmek mümkün değildir. Bazen çoklu tahminler elde edildiğinden bunlar arasında seçim yapma problemi ile karşı karşıya kalınır. Denklem 9.5'deki gibi bir talep ve arz denklem sisteminin olduğu varsayılsın.

$$q = a_1 + b_1 p + c_1 y + u_1 \qquad \text{talep fonksiyonu} \qquad (9.5)$$

$$q = a_2 + b_2 p + u_2 \qquad \text{arz fonksiyonu}$$

Bu sistemin indirgenmiş eşitlikleri:

$$q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} Y + v_1$$

$$p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} Y + v_2 \qquad \text{veya}$$

$$q = \pi_1 + \pi_2 y + v_1$$

$$p = \pi_3 + \pi_4 y + v_2$$

Bu durumda $b'_2 = \pi'_2 / \pi'_4$ ve $a'_2 = \pi'_1 - b'_2 \pi'_3$ şeklinde bulunur ancak a_1, b_1 ve c_1'in bulunacağı yol yoktur. Çünkü, yukarıda verilen arz fonksiyonu tanımlı, fakat talep fonksiyonu ise tanımlı değildir. Tanımlama problemine başka bir örnek olarak aşağıdaki model ele alınabilir.

$$q = a_1 + b_1 p + u_1 \qquad \text{talep fonksiyonu}$$

$$q = a_2 + b_2 p + c_2 R + u_2 \qquad \text{arz fonksiyonu}$$

Burada da talep fonksiyonunun tanımlandığı fakat arz fonksiyonunun tanımlanmadığı sonucu çıkmaktadır. Son olarak 9.6 denklem sistemine sahip olunduğu varsayılsın.

$$q = a_1 + b_1 p + c_1 y + d_1 R + u_1 \qquad \text{talep fonksiyonu} \qquad (9.6)$$

$$q = a_2 + b_2 p + u_2 \qquad \text{arz fonksiyonu}$$

Yukarıdaki denklemlerde yağmur, yağışlı günlerde alış veriş yapılmadığı varsayımına dayanarak talep modeli içerisinde yer almıştır. Bu modellere ait indirgenmiş denklemler aşağıdaki gibidir.

$$q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} y - \frac{d_1 b_2}{b_2 - b_1} R + v_1$$

$$p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y - \frac{d_1}{b_2 - b_1} R + v_2 \qquad \text{veya}$$

$$q = \pi_1 + \pi_2 y + \pi_3 R + v_1$$

$$p = \pi_4 + \pi_5 y + \pi_6 R + v_1$$

Bu durumda b_2'nin iki tane tahmini elde edilir. Birisi $b'_2 = \pi'_2 / \pi'_5$ ve diğeri ise $b'_2 = \pi'_3 / \pi'_6$'dır ve bunların birbirine eşit olması gerekmez. Her biri için bir a_2 tahmini yapılır ki bu da $a'_2 = \pi'_1 - b'\pi'_4$ olur.

Yukarıdaki eşitliklerden a_2 ve b_2'yi bulma imkanı varken a_1, b_1, c_1 ve d_1' i talep fonksiyonunda tespit etmek mümkün değildir. Bu durumda arz fonksiyonu fazlasıyla tanımlanmış, ancak talep fonksiyonu tam tanımlanamamış olur. İndirgenmiş form parametrelerinden bir denklemin yapısal parametreleri için tek bir tahmin elde ediliyorsa *tam tanımlanmış* denklem (exactly identified), eğer çoklu tahmin elde ediliyorsa *aşırı tanımlanmış* denklem (over-identified) ve eğer herhangi bir tahmin elde edilmiyorsa *tanımlanmamış denklem* (under-identified) elde dilmektedir.

Doğrusal sistemlerde dikkate alınan ve kullanılan mevcut basit bir hesaplama yöntemi vardır. Bu hesaplama kuralı tanımlama için *düzen şartı* olarak bilinmektedir. Bu kural şöyledir. Bir denklem sisteminde tanımlanma durumunu ortaya koymak için "g" sistemdeki dahili değişkenleri, "k" dikkate alınan denklemde olmayan dahili ve harici değişkenlerin toplamını ifade etmek kaydıyla aşağıdaki sınıflandırma yapılır.

$k = g - 1$, denklem tam tanımlanmıştır.

$k > g - 1$, denklem aşırı tanımlanmıştır.

$k < g - 1$, denklem tanımlanmamıştır.

Bu şart gereklidir, ancak yeterli değildir. Bir sonraki konu başlığı (9.4) altında gerekli şart yanında yeterli şart da verilerek detaylı olarak ele alınacaktır .

Bu noktada burada verilen kuralın bir sisteme uygulanması doğru olacaktır. Denklemler sistemi 9.2'de dahili değişken sayısı "g" 2'dir ve her bir denklemde olmayan değişken sayısı 1'dir. Bu durumda her iki denklem de tam olarak tanımlanmıştır. Denklemler sistemi 9.5'de ise "g" yine 2'dir, ancak birinci denklemde olmayan hiçbir değişken yoktur ve denklem tanımlanmamıştır. İkinci denklemde olmayan sadece bir değişken olduğu için, bu denklem tam olarak tanımlanmıştır. Denklemler sistemi 9.6'daki birinci denklemde tüm değişkenler olduğu için tanımlanmamış iken, ikinci denklemde olmayan iki değişken mevcut olduğundan, yani $k > g-1$ olduğundan denklem aşırı tanımlanmıştır.

Uygulamalı Örnek 9.1

Burada izah edilen dolaylı en küçük kareler yöntemi çok az kullanılan bir yöntemdir. İleriki konularda simultane denklemler modelini tahmin eden daha popüler metotlar ele alınacaktır. Bunlar *iki ve üç basamaklı en küçük kareler* yöntemleridir. Fakat indirgenmiş form denklemlerindeki bazı katsayılar sıfıra yakın ise yapısal denklemlerde hangi değişkenlerin göz ardı edileceği hakkında bilgi vermektedir. Bu noktada en küçük kareler, indirgenmiş formda en küçük kareler ve dolaylı en küçük karelerden elde edilen ve modelin nasıl formüle edileceği konusunda bilgi sağlayan tahminlerin var olduğu iki denklemli bir arz ve talep modeli örneği verilecektir. Örnek aynı zamanda bu bölümüm ilk kısmında ortaya atılan normalleştirme ile ilgili bazı hususlara da değinmektedir. Bu

örnekteki veriler ABD'nin 1922-1941 yılları arasındaki domuz eti arz ve talep miktarlarıdır ve çizelge 9.1'de verilmiştir.

Çizelge 9.1. ABD'nin 1922-1941 yılları arasındaki domuz eti arz ve talep miktarları

Yıllar	P_t	Q_t	Y_t	Z_t	Yıllar	P_t	Q_t	Y_t	Z_t
1922	26,8	65,7	541	74,0	1932	15,6	70,7	390	74,8
1923	25,3	74,2	616	84,7	1933	13,9	69,6	364	73,6
1924	25,3	74,0	610	80,2	1934	18,8	63,1	411	70,2
1925	31,1	66,8	636	69,9	1935	27,4	48,4	459	46,5
1926	33,3	64,1	651	66,8	1936	26,9	55,1	517	57,6
1927	31,2	67,7	645	71,6	1937	27,7	55,8	551	58,7
1928	29,5	70,9	653	73,6	1938	24,5	58,2	506	58,0
1929	30,3	69,6	682	71,2	1939	22,2	64,7	538	67,2
1930	29,1	67,0	604	69,6	1940	19,3	73,5	576	73,7
1931	23,7	68,4	515	68,0	1941	24,7	68,4	697	66,5

P_t: domuz eti perakende fiyatı (cent/pound), Q_t: domuz eti tüketimi (pound/kişi), Y_t: harcanabilir kişisel gelir (dolar/kişi), Z_t: domuz eti üretimi ile ilgili bir değişken.

Tahmin edilen model aşağıdaki gibidir.

$$Q = a_1 + b_1 P + c_1 Y + u \qquad \text{talep fonksiyonu}$$

$$Q = a_2 + b_2 P + c_2 Z + v \qquad \text{arz fonksiyonu}$$

Bu iki denklemli sistemin indirgenmiş form denklemleri tablo 9.1'deki veriler kullanılarak aşağıdaki gibi tahmin edilmiştir.

$$P' = -0,0101 + \underset{(0,1339)}{1,0813}\, Y - \underset{(0,1159)}{0,8320}\, Z \qquad R^2 = 0,893$$

$$Q' = 0,0026 - \underset{(0,0673)}{0,0018}\, Y + \underset{(0,0582)}{0,6339}\, Z \qquad R^2 = 0,898$$

İkinci denklemdeki Y'nin katsayısı sıfıra çok yakın olduğundan denklemden çıkarılabilir. Bu durum $b_2 = 0$ veya arzın fiyata tepkili olmadığını göstermektedir. Her bir durumda indirgenmiş formun yapısal forma çözümü ile yapısal denklem tahminleri aşağıdaki gibi elde edilir.

$$Q = -0,0063 - 0,8220\, P + 0,8870\, Y \qquad \text{talep fonksiyonu}$$

$$Q = 0,0026 - 0,0017\, P + 0,6825\, Z \qquad \text{arz fonksiyonu}$$

Talep fonksiyonunun en küçük kareler tahminleri sırasıyla Q'ya ve P'ye göre normalleştirilmiş şekli aşağıdaki gibidir.

$$Q' = -0,0049 - \underset{(0.0594)}{0,7205}\, P + \underset{(0,0967)}{0,7646}\, Y \qquad R^2 = 0,903$$

$$P' = -0,0070 - \underset{(0,1032)}{1,2518}\, Q + \underset{(0,0861)}{1,0754}\, Y \qquad R^2 = 0,956$$

Yapısal talep fonksiyonunun sırasıyla Q'ya ve P'ye göre normalleştirilmiş şekli aşağıdaki gibi yazılabilir.

$$Q' = - 0,0063 - 0,8220\ P + 0,8870\ Y$$

$$P' = - 0,0077 - 1,2165\ Q + 1,0791\ Y$$

Talep fonksiyonu P'ye göre normalleştirildiğinde talep fonksiyonunun dolaylı en küçük kareler parametre tahminleri doğrudan en küçük kareler tahminleri ile aynıdırlar.

Hangisi doğru normalleştirmedir? Baş tarafta tartışıldığı gibi, eğer arz miktarı fiyata tepkili değil ise, talep fonksiyonu P'ye göre normalleştirilmelidir. Yukarıda görüldüğü gibi, Q için indirgenmiş formdaki denklemde Y'nin katsayısı sıfıra çok yakın idi yani $b_2 = 0$ veya arz miktarı fiyata tepkili değil idi. Bu durum, yanlış bir katsayı işaretini gösteren ve sıfıra yakın bir katsayı olan b_2'nin yapısal tahmini tarafından da doğrulanmış olmaktadır.

Arz fonksiyonundan P'nin çıkarılması ve normal en küçük kareler yöntemiyle aşağıdaki gibi bir arz eğrisi elde edilir.

$$Q' = 0,0025 + 0,6841\ Z \qquad R^2 = 0,898$$

Böylece talep fonksiyonu P'ye göre normalleştirilir, arz fonksiyonundan P değişkeni çıkarılır ve her iki denklem de ayrı olarak normal en küçük kareler yöntemi ile tahmin edilerek tutarlı tahminler elde edilebilir.

*9.4. Tanımlama İçin Gerek ve Yeter Şartlar

Önceki kısımda tahmin edilen indirgenmiş parametrelerden yapısal parametreler hesaplanmıştı. Tanımlama problemini ele almanın alternatif bir yolu da dikkate alınan denklemin diğer denklemlerin doğrusal bir kombinasyonundan sağlanıp sağlanamayacağının araştırılmasıdır.

Örneğin 9.5'deki denklemler dikkate alınsın. Her iki eşitliğinde w ve (w-1) ağırlıkları ile ağırlıklı ortalamaları alındığında aşağıdaki denklem elde edilir.

$$q = w\,(a_1 + b_1 p + c_1 y) + (1 - w)\,(a_2 + b_2 p) + u_1'$$

$$a_1^* + b_1^* p + c_1^* y + u_1' \qquad (9.7)$$

burada $a_1^* = w a_1 + (1 - w)\, a_2$

$b_1^* = w b_1 + (1 - w)\, b_2$

$c_1^* = w c_1$ ve

$u_1' = w u_1 + (1-w)\, u_2$

Eşitlik 9.7, eşitlik 9.5'deki ilk eşitliğe benzemektedir. Talep fonksiyonunun katsayıları hesaplandığı zaman talep fonksiyonunun mu yoksa talep ve arz fonksiyonunun ağırlıklı ortalamasının mı parametrelerinin elde edildiği bilinememektedir. Bu yüzden talep fonksiyonundaki parametreler tanımlanmamıştır. Eşitlik 9.7'nin 9.5'deki arz fonksiyonuna benzemesinin tek yolu $c_1^* = 0$ olmasıyla mümkün olduğundan aynı şey arz fonksiyonu için söylenemez. Fakat c_1 sıfırdan farklı olduğundan $w = 0$ olmalıdır.

Yani, ağırlıklı ortalama talep fonksiyonuna sıfır ağırlık vermektedir. Bu yüzden arz fonksiyonu tahmin edildiğinde elde edilen tahminlerin arz fonksiyonuna ait olunduğundan emin olunmasıyla arz fonksiyonunun parametreleri tanımlanmış olur.

Yukarıdaki yöntemi iki eşitlik bulunduğu zaman kullanmak kolaydır. Fakat fazla sayıda eşitlik bulunduğu zaman daha sistematik bir kontrol yöntemi bulmak gereklidir. Gösteri amaçlı olarak üç dâhili değişken y_1, y_2, y_3 ve üç harici değişken z_1, z_2, z_3 durumu dikkate alınsın. Eğer değişken bir denklemde mevcut ise x, değil ise 0 ile işaretlenecektir. Denklem sisteminin aşağıdaki gibi olduğunu varsayalım.

Denklem	y_1	y_2	y_3	z_1	z_2	z_3
1	x	0	x	x	0	x
2	x	0	0	x	0	x
3	0	x	x	x	x	0

Herhangi bir denklem için tanımlanma kuralı aşağıdaki gibidir.

1. Tanımlanması incelenen denklemin katsayılar satırı çıkarılır.
2. İncelenen denkleme ait satırdaki sıfır olan sütunları dikkate alınır. Böylece bu denklemde bulunmayan ama modelin öteki denklemlerinde yer alan değişkenler kalır.
3. Eğer bu sütun dizisinden tümü sıfır olmayan (g-1) satır ve sütun bulunursa ki burada g dâhili değişken sayısını verir ve tüm parametre değerleri için hiçbir satır ve sütun diğer satır ve sütuna orantılı değil ise o zaman denklem tanımlanmıştır. Aksi durumda tanımlanmamıştır. Eğer böyle bir satır veya sütun var ise o satır veya sütunun çıkarılması gerekir.

Bu şart tanımlama için işaret şartı olarak adlandırılır ve gerekli ve yeterli şarttır. Yukarı da verilen örneğe bu işaret şartı uygulanabilir. Denklem sayısı yani g = 3'dür. Böylece birinci denklemde olmayan değişken sayısı 2, ikinci denklemde olmayan değişken sayısı 3 ve üçüncü denklemde olmayan değişken sayısı 2'dir. Bu nedenle *düzen şartına* (order condition) göre ilk ve üçüncü denklemler tam tanımlanmış ve ikinci denklemde aşırı tanımlanmış olacaktır. Hâlbuki *işaret şartı* (rank condition) örnekte görüleceği gibi farklı bir cevap verecektir.

Birinci denklem için işaret şartı kullanılırsa ilk satır çıkarılır ve birinci denklemde olmayan y_2 ve z_2'nin sütunları dikkate alınır ise aşağıdaki sütunlar elde edilir.

$$\begin{matrix} 0 & 0 \\ x & x \end{matrix}$$

Dikkat edilirse elemanlarının tümü sıfır olmayan tek bir satır vardır. Bu nedenle işaret şartına göre bu denklem tanımlanmamıştır. İkinci denklem için ikinci satır çıkarılır ve

bu denklemde olmayan y_2, y_3 ve z_2'ye tekabül eden sütunlar dikkate alınır. Bu durumda aşağıdaki matriks elde edilir.

$$\begin{matrix} 0 & x & 0 \\ x & x & x \end{matrix}$$

Bu durumda elemanların tümü sıfır olmayan iki satır (üç sütun) mevcut olduğundan denklem tanımlanmıştır. Ayni şekilde üçüncü denklem için üçüncü satır çıkarılır ve bu sırada sıfır olan y_1 ve z_3'e tekabül eden sütunlar dikkate alınır ise aşağıdaki matriks elde edilir.

$$\begin{matrix} x & x \\ x & x \end{matrix}$$

Yine elemanların tümü sıfır olmayan iki satır olduğundan denklem tanımlanmıştır. Görülmektedir ki işaret şartı denklemin tanımlanıp tanımlanmadığını ortaya koymaktadır. Hâlbuki düzen şartından, denklemin tam mı yoksa aşırı mı tanımlandığı öğrenilmektedir.

Özet olarak, ikinci ve üçüncü denklemlerin tahmin edilebilir olduğu sonucuna varılmaktadır. Birinci denklemin tahmin edilebilir olmamasına rağmen düzen şartı tahmin edilebilir olduğu sonucunu vererek yanılgıya neden olmaktadır. Düzen şartı karşılanmadığı halde çalışmayan, fakat düzen şartı karşılandığında işaret şartı karşılanmazsa bile parametre tahminleri veren ve simultane denklemler için kullanılan çok sayıda tahmin yöntemleri vardır. Bu nedenle işaret şartıyla kontrol yapmak gerekmektedir. Örnekteki birinci denklem için işaret şartı karşılanmamaktadır. Bunun gerçekte anlamı, eşitlik 1'deki parametreler için yapılan tahminler aslında tüm eşitliklerdeki parametrelerin doğrusal kombinasyonlarının tahminleridir. Bu nedenle özel bir yorumu yoktur. Birinci eşitliğin tahmin edilebilir olmamasının anlamı budur.

Uygulamalı Örnek 9.2

Örnek olarak yedi dâhili ve üç harici değişkenli bir makroekonomik model dikkate alınsın. Bu model aşağıdaki dâhili ve harici değişkenlere sahiptir.

Dâhili değişkenler:

C = reel tüketim N = istihdam

R = faiz oranı W = parasal ücret oranı

I = reel yatırım P = fiyat seviyesi Y = reel gelir

Harici değişkenler

G = reel hükümet harcamaları T = reel vergi gelirleri

M = Nominal para stoku

Modele ait denklemler aşağıdaki gibidirler

(1) $C = a_1 + b_1 Y - c_1 T + d_1 R + u_1$ (tüketim fonksiyonu)

(2) $I = a_2 + b_2 Y + c_2 R + u_2$ (yatırım fonksiyonu)

(3) $Y = C + I + G$ (tanım)

(4) $M = a_3 + b_3 Y + c_3 R + d_3 P + u_3$ (likidite tercih fonksiyonu)

(5) $Y = a\ 4 + b_4 N + u_4$ (üretim fonksiyonu)

(6) $N = a_4 + b_5 W + c_5 P + u_5$ (işgücü talebi)

(7) $N = a_6 + b_6 W + c6\ P + u_6$ (işgücü arzı)

Buradaki soru, bu denklemlerden hangileri tanımlanmamış, hangileri tam tanımlanmış ve hangileri aşırı tanımlanmıştır. Bu sorunun cevaplanması için aşağıdaki tablo hazırlanmıştır. Bu çizelgede "1" değişkenin denklemde var olduğunu, "0" ise olmadığını göstermektedir.

Denklem	C	I	N	P	R	Y	W	G	T	M
1	1	0	0	0	1	1	0	0	1	0
2	0	1	0	0	1	1	0	0	0	0
3	1	1	0	0	0	1	0	1	0	0
4	0	0	0	1	1	1	0	0	0	1
5	0	0	1	0	0	1	0	0	0	0
6	0	0	1	1	0	0	1	0	0	0
7	0	0	1	1	0	0	1	0	0	0

Burada dikkat edilecek husus, 3. denklemin tahmin edilecek herhangi bir parametreye sahip olmadığıdır. Bu nedenle bu denklemin tanımlanması gerekmemektedir. Bu modelde dâhili değişkenlerin sayısı eksi bir 6'ya eşit olmaktadır. Düzen şartına göre 1 ve 4. denklemlerin tam tanımlandığı, 2, 5, 6 ve 7. denklemlerin aşırı tanımlandığı sonucu ortaya çıkmaktadır.

Denklem 1 için işaret şartına bakmak için ilk sıra matriksten çıkarılır ve bu sırada mevcut olmayan I, N, P, W, G ve M değişkenlerin sütunları alınarak aşağıdaki matriks elde edilir. Aynı işlem diğer denklemler için de tekrar edilir.

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{matrix}$$

Elemanların tümü sıfır olmayan 6 sıra (6 sütun) olduğundan 1. denklem tanımlanmıştır. Aynı durum 2, 4 ve 5. denklemler için de söz konusudur. Fakat 6 ve 7. denklemler için tüm elementleri sıfır olmayan 6 sıra bulunamadığından bu denklemler düzen şartına göre aşırı tanımlanmasına rağmen işaret şartına göre tanımlanmamıştır.

*9.5. Araç Değişkenler Yöntemi

Tanımlama problemini inceledikten sonra tahmin metotları konusu tartışmak yararlo olacaktır. Tahmin metotlarından Dolaylı En Küçük Kareler yöntemini indirgenmiş denklemleri kullanarak inceledik. Fakat bu metot fazla sayıda denklem bulunduğu zaman genellikle kullanılmaz. Şimdi daha fazla uygulama imkânı olan metotları inceleyeceğiz. Bu metotlar iki temel kategoride ele alınmaktadır:

- Tek Denklem Metotları (Sınırlı Bilgi Metotları)
- Denklem Sistemi Metotları (Tam Bilgi Metotları)

Bu konu kapsamında Sistem Metotlarını çok fazla matematiksel detay içerdiği için tartışmayacağız, sadece Tek Eşitlik Metotlarını inceleyeceğiz. Tek Denklem Metotlarında her bir denklem ayrı ayrı tahmin edilir. Bir eşitliği tahmin ederken diğer eşitliklerdeki sınırlayıcılar sadece tanımlama kontrolünde kullanılır tahminde kullanılmaz. Bundan dolayı da bu metotlara sınırlı bilgi metotları denir.

Simultane denklem modellerinde parametrelerin etkin bir şekilde tahmin edilmelerinin genel yolu Araç Değişkenler Yöntemidir. Araç değişken eşitlikte hata terimi ile sıfır korelasyona sahip, ancak açıklayıcı değişkenler ile korelasyonu bulunan bir değişkendir.

Örneğin;, elimizde $Y = \beta x + u$ şeklinde bir model bulunsun. Bu modelde x ile u arasında bir korelasyon bulunduğu için modelin en küçük kareler yöntemi ile tutarlı (consistent) bir tahmini yapılamaz. Şayet u ile korelasyonu sıfır olan bir z değişkeni modele ilave edilirse etkin bir tahmin yapılabilir.

Aşağıdaki gibi bir simultane denklem modeli bulunsun:

$$y_1 = b_1 y_2 + c_1 z_1 + c_2 z_2 + u_1 \qquad (9.8)$$
$$y_2 = b_2 y_1 + c_3 z_3 + u_2$$

Bu modelde y_1, y_2 dâhili değişkenler, z_1, z_2, z_3 harici değişkenler ve u_1, u_2 hata terimleridir. Birinci denklemi tahmin edeceğimizi kabul edelim. Bu denklemde z_1, z_2 değişkenleri u_1'den bağımsızdır [cov (z_1, u_1) = 0 ve cov (z_2, u_1) = 0]. Ancak y_2 değişkeni u_1'den bağımsız değildir [cov (y_2, u_1) $\neq$ 0]. Birinci denklemde tahmin edilecek üç değişken vardır ve tahmini gerçekleştirmek için u_1'den bağımsız bir değişken bulmak zorunluluğu bulunmaktadır. İkinci denklemde y_2'nin araç değişkeni olan ve u_1 ile aralarında korelasyon bulunmayan z_3 araç değişkeni bulunmaktadır. Bu araç değişkenini de kullanarak kovaryans denklemleri yeniden aşağıdaki gibi yazılır.

$$1 / n \sum z_1 (y_1 - b_1 y_2 - c_1 z_1 - c_2 z_2) = 0$$
$$1 / n \sum z_2 (y_1 - b_1 y_2 - c_1 z_1 - c_2 z_2) = 0 \qquad (9.8')$$
$$1 / n \sum z_3 (y_1 - b_1 y_2 - c_1 z_1 - c_2 z_2) = 0$$

En küçük kareler metodunun normal denklemleri ile araç değişken metodu arasındaki tek fark son eşitliktir.

Araç değişken yönteminde sistemde araç değişkeni olarak kullanılabilecek yeter miktarda dışsal değişkenin bulunup bulunmadığı ayrı bir problemdir. Şayet eşitlik tanımlanmamış bir denklem ise yeter miktarda araç değişkeni yoktur. Tam tanımlanmış bir denklem ise yeter miktarda araç değişkeni vardır ve denklem aşırı belirlenmiş ise gereğinden fazla araç değişkeni bulunmaktadır. Aşırı belirlenmiş bir denklemde en etkili tahmini yapmak için araç değişkenlerin ağırlıklı ortalamasını kullanmak gerekir.

Yukarıdaki 9.8 numaralı modeli tahmin etmede öncelikle indirgenmiş formların bulunması için $\hat{y}_1$ ve $\hat{y}_2$'nin z_1, z_2, z_3 üzerine regresyonunun tahmin edilmesi gereklidir. Sistemde birinci denklemi tahmin etmek için $\hat{y}_2$, ikinci denklemi tahmin etmek için $\hat{y}_1$ kullanılır. Şimdi $\hat{y}_1$ve $\hat{y}_2$'yi z_1, z_2 ve z_3' ün birer doğrusal fonksiyonu olarak yazalım:

$$\hat{y}_1 = a_{11} z_1 + a_{12} z_2 + a_{13} z_3$$

$$\hat{y}_2 = a_{21} z_1 + a_{22} z_2 + a_{23} z_3$$

Burada a'lar indirgenmiş formdaki denklemlerin en küçük kareler yöntemiyle yapılmış tahmininden sağlanmıştır. 9.8 numaralı sistemin birinci denklemini tahmin etmek için y_2, z_1 ve z_2 araç olarak kullanılır. Bu z_1, z_2 ve z_3' ün kullanılmasıyla aynıdır. Çünkü;

$$\sum \hat{y}_2 u_1 = 0 \quad \Rightarrow \quad \sum (a_{21} z_1 + a_{22} z_2 + a_{23} z_3) u_1 = 0$$

$$\Rightarrow \quad a_{21} \sum z_1 u_1 + a_{22} \sum z_2 u_1 + a_{23} \sum z_3 u_1 = 0$$

Fakat ilk iki terim 9.8'deki denklemlerin ilk ikisinin etkisinden dolayı sıfırdır. $\sum \hat{y}_2 u_1 = 0 \Rightarrow \sum z_3 u_1 = 0$. Bu nedenle y_2 değişkeninin araç değişkeni olarak kullanılması, tıpkı z_3 değişkeninin araç değişkeni olarak kullanılması gibidir. Tam belirlenmiş eşitliklerde tek bir seçenek olduğu için araç değişkenleri seçme şansı yoktur.

Fakat 9.8 numaralı sistemin ikinci eşitliğinde durum farklıdır. Başlangıçta, y_1 için z_1 ve z_2 değişkenlerinin araç değişkeni olarak seçilme şansı vardır. Burada $\hat{y}_1$'in kullanılması optimum ağırlıkları verir. Normal eşitlikler aşağıdaki gibidir.

$$\sum \hat{y}_1 u_2 = 0 \quad \text{ve} \quad \sum z_3 u_2 = 0$$

$$\sum \hat{y}_1 u_2 = 0 \quad \Rightarrow \quad \sum (a_{11} z_1 + a_{12} z_2 + a_{13} z_3) u_2 = 0$$

$$\Rightarrow \quad \sum (a_{11} z_1 + a_{12} z_2) u_2 = 0$$

çünkü $\sum z_3 u_2 = 0$'dır. Bu nedenle z_1 ve z_2'nin optimum ağırlıkları a_{11} ve a_{12}'dir.

<u>R^2'nin Ölçümü</u>

Bir eşitliği araç değişken kullanarak tahmin ettiğimizde R^2 ölçüsünün modelin açıklayıcı özelliğini nasıl ifade ettiği sorulan bir sorudur. Bu soru İki Aşamalı En Küçük Kareler yönteminde de gündeme gelir. Bu konuda iki ölçü düşünülür:

1. R^2, y ve ŷ arasındaki korelasyonun karesi,
2. Rezidual kareler toplamını esas alan ölçülerdir.

$$R^2 = 1 - \left[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum (y_i - y_{ort})^2 \right]$$

Bu ders notlarında verilen sonuçlarda ikinci ölçü kullanılacaktır.

Araç değişken metodunda hesaplanan R^2 değeri En Küçük Kareler Yöntemiyle hesaplanan R^2 değerinden küçüktür. Aynı zamanda R^2 değeri negatifte olabilir ancak bu değerin negatif olması bir spesifikasyon hatasının yapıldığını yani eşitliğin tanımlanmamış olduğunu haber verir. Bu konuyu açıklayan bir örnek verilecektir.

Çoğu bilgisayar programı simultane denklem modelleri için R^2 değerini verir. Ancak bu tip modellerin tek bir R^2 ölçümü olmadığından ilgili programın R^2 değerini hesaplamak için hangi formülü kullandığını kontrol etmek gereklidir. Örneğin SAS programı bir R^2 değeri verir fakat kullanılan yöntem yukarıda açıklananlardan hiçbirine benzemez.

Uygulamalı Örnek 9.3

Çizelge 9.2'de Avustralya şarap endüstrisinin 1955-56'dan 1974-75'e kadar olan bazı özelliklerini gösteren veriler sunulmuştur. Bu endüstri için uygun talep ve arz modellerinin aşağıdaki gibi olduğu varsayılmaktadır.

$$Q_t = a_0 + a_1 P_t^w + a_2 P_t^b + a_3 Y_t + a_4 A_t + u_t \qquad \text{talep}$$

$$Q_t = b_0 + b_1 P_t^w + b_2 S_t + v_t \qquad \text{arz}$$

Burada:

Q_t : Kişi başına reel şarap tüketimi

P_t^w : TÜFE'ye uyarlanmış şarap fiyatı

P_t^b : TÜFE'ye uyarlanmış bira fiyatı

Y_t : Kişi başına reel tüketilebilir gelir

A_t : Kişi başına reel reklam harcamaları

S_t : Depolama maliyetleri indeksi

Bu sistemde Q_t ve P_t^w dâhili değişkenlerdir. Diğer değişkenlerin tümü harici değişkenlerdir. Talep fonksiyonunun tahmini için sadece bir araç değişkeni mevcuttur o da S_t'dir. Fakat arz fonksiyonunun tahmini için üç tane araç değişkeni vardır ve bunlarda P_t^b, Y_t, A_t'dir.

Talep fonksiyonunun En Küçük Kareler Tahmini aşağıdaki sonuçları vermektedir. Bütün değişkenler logaritmik formdadır ve parantez içindeki değerler t değerleridir.

$$Q = \underset{(-6,04)}{-23,621} + \underset{(4,0)}{1,158}\, P_w - \underset{(-0,45)}{0,275}\, P_b - \underset{(-1,3)}{0,603}\, A + \underset{(4,50)}{3,212}\, Y \qquad R^2 = 0,9772$$

9.2. Avustralya şarap endüstrisi ile ilgili veriler

Yıl	Q	S	P_w	P_b	A	Y
1955-1956	0,91	85,4	77,5	35,7	89,1	1056
1956-1957	1,05	88,4	80,2	37,4	83,3	1037
1957-1958	1,18	89,1	79,5	37,7	84,4	1006
1958-1959	1,27	90,5	84,9	37,1	90,1	1047
1959-1960	1,27	93,1	94,9	36,2	89,4	1091
1960-1961	1,37	97,2	92,7	35,0	89,3	1093
1961-1962	1,46	100,3	92,5	37,6	89,8	1102
1962-1963	1,59	100,3	92,7	40,1	96,7	1154
1963-1964	1,86	101,5	97,1	39,7	99,9	1234
1964-1965	1,96	104,8	93,9	38,3	103,2	1254
1965-1966	2,32	107,5	102,7	37,0	102,2	1241
1966-1967	2,86	111,8	100,0	36,1	100,0	1299
1967-1968	3,50	114,9	119,5	35,4	103,0	1287
1968-1969	3,96	117,9	119,7	35,1	104,2	1369
1969-1970	4,21	122,3	125,2	34,5	113,0	1443
1970-1971	4,54	128,2	134,1	34,5	132,5	1517
1971-1972	4,93	134,1	124,3	34,3	143,6	1562
1972-1973	5,40	145,1	119,0	34,3	176,2	1678
1973-1974	6,13	174,9	108,5	31,9	159,9	1769
1974-1975	6,29	237,2	107,9	31,0	182,1	1847

Y'ninki hariç bütün katsayılar yanlış işarete sahiptir. P_w'nun katsayısı sadece yanlış işarete sahip değil aynı zamanda önemlidir. Burada P_w'yi dâhili değişken sayılıp S'de araç değişkeni olarak kullanılır ise aşağıdaki sonuçlar elde edilir.

$$Q = \underset{(-5,09)}{-26,195} + \underset{(0,98)}{0,643}\, P_w - \underset{(-0,20)}{0,140}\, P_b - \underset{(-1,51)}{0,985}\, A + \underset{(3,28)}{4,082}\, Y \qquad R^2 = 0,9724$$

Yine P_w'nun katsayısı hala yanlış işarete sahiptir en azından istatistiksel olarak anlamlı değildir. Her durumda ulaşılan sonuç, talep edilen miktarın fiyatlara tepkili olmadığıdır ve fakat reklam harcamaları gelire karşı tepkilidir. Şarap için talebi gelir elastikiyeti yaklaşık 4'dür. Bu rakam anlamlı olarak birim elastikiyetten yüksektir.

Arz fonksiyonuna dönülür ise, P_w için P_b, A ve Y gibi dört araç değişken mevcuttur. Bunun yanında P_w'nin P_b, A, Y ve S üzerine regresyonu ile araç değişkeni elde edilebilir. En küçük kareler yöntemi ve değişik araç değişkenlerinin kullanılması ile elde edilen sonuçlar çizelge 9.3'deki gibidir. Parantez içindeki rakamlar, araç değişkeni metotları için asimptotik t değerleridir. Araç değişkeni metodu için R^2 değerleri daha önce açıklandığı gibi elde edilmiştir.

Değişik araç yöntemleriyle sağlanan tahminler arasında önemli farklılıklar olduğu görülmektedir. Özellikle P_b'nin kullanımının çok değişik sonuçlar ürettiği ve Y'nin en iyi araç değişkeni olduğu ortaya çıkmaktadır.

Çizelge 9.3. EKK ve farklı araç değişkenleri yöntemlerine göre regresyon sonuçları

		Araç Değişkenleri			
Metot	EKK	P_b	A	Y	P_w
Sabit	-15,57 (-18,36)	-10,76 (-0,28)	-17,65 (-6,6)	-16,98 (-14,56)	-16,82 (-15,57)
Pw	2,145 (8,99)	0,336 (0,02)	2,928 (3,02)	2,676 (7,30)	2,616 (7,89)
S	1,383 (8,95)	2,131 (0,36)	1,058 (2,47)	1,163 (5,72)	1,188 (6,24)
R^2	0,9632	0,8390	0,9400	0,9525	0,9548

Araç değişkeni tahminleri daha önce belirtilen yöntem ile sağlanabilir. Fakat şu çok rahatlıkla ifade edilebilir ki araç değişkeni tahmin edicileri ve bir sonraki konuda ele alınacak iki aşamalı en küçük kareler (2SLS) tahmin edicileri denklem tam tanımlanıyor ise aynı sonuçları verir. Arz fonksiyonunun tam tanımlanması için modeli değiştirerek 2SLS metodu ile araç değişkeni tahmin edicisinin elde edileceği ima edilmektedir. Bu durumda SAS gibi standart bir bilgisayar program, tahminleri ve onların standart hatalarını elde etmek için kullanılabilir. Burada takip edilen yöntem de gerçekte budur.

Örneğin, Y'yi araç değişkeni olarak kullanarak arz fonksiyonunun araç değişkeni tahmini aşağıdaki model gibi yapılabilir.

$$Q = \beta_0 + \beta_1 P_w + \beta_2 Y + u \qquad \text{talep fonksiyonu}$$

$$Q = \alpha_0 + \alpha_1 P_w + \alpha_2 S + v \qquad \text{arz fonksiyonu}$$

Şimdi arz fonksiyonu tam olarak tanımlanmıştır.

*9.6. İki Aşamalı En Küçük Kareler Yöntemi

İki Aşamalı En Küçük Kareler yöntemi (2SLS), Araç Değişkenler yönteminden farklı olmasına rağmen ikisi de benzer sonuçlar vermektedir. Tahmin edilecek denklemi aşağıdaki gibi olduğunu varsayalım.

$$y_1 = b_1 y_2 + c_1 z_1 + u_1 \qquad (9.9)$$

Sistemde z_2, z_3 ve z_4 gibi diğer dışsal değişkenler bulunduğunu varsayalım. Burada y_2'nin tahmin edilmiş değeri olarak $\hat{y}_2$'yi y_2'nin z_1, z_2, z_3 ve z_4'ün üzerine regresyonundan elde edilmiş olsun. O zaman bu durum aşağıdaki gibi ifade edilir.

$$y_2 = \hat{y}_2 + v_2$$

Burada rezidual v_2'nin y_2 ve z_1, z_2, z_3, z_4 ile korelasyonu yoktur. Bu EKK regresyonunun özelliğidir. Etkin araç değişkeni metodu için normal denklemler aşağıda 9.10'da verilen denklemler gibidir.

$$\sum \hat{y}_2 (y_1 - b_1 y_2 - c_1 z_1) = 0 \qquad (9.10)$$

$$\sum z_1 (y_1 - b_1 y_2 - c_1 z_1) = 0$$

$y_2 = y_2 + v_2$ değerini formüllerde yerine yazarsak 9.11'deki sonuçlar elde edilir.

$$\sum \hat{y}_2 (y_1 - b_1 y_2 - c_1 z_1) - b_1 \sum y_2 v_2 = 0 \qquad (9.11)$$
$$\sum z_1 (y_1 - b_1 y_2 - c_1 z_1) - b_1 \sum z_1 v_2 = 0$$

Yukarıdaki eşitliklerde $\sum z_1 v_2 = 0$ ve $\sum y_2 v_2 = 0$'dır ve z_1 ve y_2, v_2 ile korelasyona sahip değildir. O halde denklemler aşağıdaki şekle dönüşür.

$$\sum \hat{y}_2 (y_1 - b_1 y_2 - c_1 z_1) = 0 \qquad (9.12)$$
$$\sum z_1 (y_1 - b_1 y_2 - c_1 z_1) = 0$$

Elde edilen bu eşitlikler en baştaki 9.9 nolu eşitlikte y_2'nin yerine $\hat{y}_2$'nin konulmasıyla elde edilebilecek normal eşitliklerdir ve OLS ile tahmin edilebilirler. Bu metot eşitliğin sağında bulunan içsel değişkenlerin yerine indirgenmiş formlardan tahminlerinin konulması ve eşitliğin OLS ile hesaplanması şeklinde iki adımdan oluştuğu için İki Aşamalı En Küçük Kareler (2SLS) yöntemi adı verilir. Bu yöntemi oluşturan adımlar aşağıdaki gibidir.

1. Adım: OLS ile indirgenmiş form eşitliklerinin tahmini ve öngörülen y'nin elde edilmesi,
2. Adım: Sağdaki dâhili değişkenlerin yerine y'nin konulması ve eşitliğin OLS ile tahmini.

y_1 yerine $\hat{y}_1$'nın 9.9 eşitliğine koyulması ile tahminler değişmez. Normal eşitlikler 9.12'deki gibidir. Yazılacak olunursa,

$$y_1 = \hat{y}_1 + v_1$$

elde edilir. Burada v_1'in y_1,y_2 ve z_1, z_2, z_3, z_4 ile korelasyonu yoktur. Eşitliklerde $y_1 = y_1 + v_1$ denklemini yerine yazarsak aşağıdaki eşitlikleri elde ederiz.

$$\sum \hat{y}_2 (\hat{y}_1 - b_1 \hat{y}_2 - c_1 z_1) + b_1 \sum \hat{y}_2 v_2 = 0$$
$$\sum z_1 (\hat{y}_1 - b_1 \hat{y}_2 - c_1 z_1) + b_1 \sum z_1 v_1 = 0$$

Bu iki eşitliğin son terimleri sıfıra eşittir. Kalan denklemler, OLS ile hesaplanan aşağıdaki eşitlikten elde edilmiş normal denklemlerdir.

$$\hat{y}_1 - b_1 \hat{y}_2 + c_1 z_1 + w$$

2SLS yönteminin ikinci adımında bütün dâhili değişkenlerin yerine onların indirgenmiş formlardan kestirilen değerleri yerleştirilir. Sonra eşitlik OLS ile tahmin edilir.

Böyle yapmakla farklılık nereden doğar? Farklılık bağımlı değişkenlerde y_1'in yerine $\hat{y}_1$'nin yazılmasından dolayı tahmin edilen farklı standart hatalarda oluşur. Ancak, ikinci aşamadan elde edilen standart hatalar, aşağıda gösterileceği gibi ister eşitliğin sağındaki dahili değişkenleri ister 2SLS'in ikinci aşamada ŷ'leri değiştirelim yanlıştır. Standart hata tartışmasına basit bir model ile devam edeceğiz. Bütün y'ler indirgenmiş form

rezidualları ile korelasyona sahip olmayacağı için tartışmalar genel olacaktır. Tartışmanın temeli y yerine $\hat{y}$ + ύ'nin ikamesi olur ve bu genel modellere taşınabilir.

Standart Hatanın Hesaplanması

Burada, 2SLS metodunun ikinci adımında standart hatanın nasıl yanlış elde edildiğini ve doğrusunun nasıl elde edileceğini açıklanacaktır. Aşağıdaki basit modeli ele alınsın:

$$y_1 = \beta\, y_2 + u \qquad (9.13)$$

Modelde y_1 ve y_2 dâhili değişkenlerdir ve sistemde birkaç da dâhili değişken bulunmaktadır. Öncelikle y_2'nin indirgenmiş formu hesaplanır.

$$y_2 = \hat{y}_2 + v_2$$

Araç değişkeni tahmini olan β, $\sum y_2 (\hat{y}_1 - b_1 y_2) = 0$ formülünden elde edilir. veya

$$\beta_{IV} = \sum \hat{y}_2\, y_1 / \sum \hat{y}_2\, y_2$$

2SLS metodunda β'nın tahmincisi $\sum y_2 (\hat{y}_1 - b_1 y_2) = 0$ eşitliği çözülerek elde edilir, veya

$$\beta_{2SLS} = \sum \hat{y}_2\, y_1 / \sum \hat{y}_2^2$$

$\sum \hat{y}_2\, y_2 = \sum \hat{y}_2 (\hat{y}_2 + v_2) = \sum \hat{y}_2^2$ olduğundan iki tahminci aynı olduğundan β'nın farklı indisleri kaldırılarak aşağıdaki gibi yazılmıştır.

$$\beta = \sum \hat{y}_2\, y_1 / \sum \hat{y}_2^2 = \sum \hat{y}_2 (\beta y_2 + u) / \sum y_2^2$$
$$= \beta + \sum \hat{y}_2\, u / \sum \hat{y}_2^2$$

Burada $\sum \hat{y}_2\, y_2 = \sum \hat{y}_2^2$ olduğundan

$$= \beta - \beta = [(1/n) \sum \hat{y}_2\, u\,] / [(1/n) \sum \hat{y}_2^2]$$

n örnek sayısıdır. Burada u sistemde dışsal değişkenlerden bağımsız ve $\hat{y}_2$ dışsal değişkenlerin doğrusal bir fonksiyonu olduğundan aşağıdakilere sahip olunur.

$$\text{plim } (1/n) \sum y_2^2 \neq 0$$

$$\text{plim } (1/n) \sum y_2\, u = 0 \text{ olduğunu varsayılır ise,}$$

plim β' = β olur ve bu nedenle β', β'nın etkin bir tahmincisidir. β''nın asimptotik varyansı $n \to \infty$'a giderken var (β') → 0'dır. Bu nedenle n var(β')'yı dikkate almak geleneksel bir yoldur. Bu durumda aşağıdakiler elde edilir.

$$n \text{ var } (\beta') = n \text{ plim } (\beta' - \beta)^2$$
$$= n \text{ plim } [(1/n) \sum \hat{y}_2\, u]\, [(1/n) \sum \hat{y}_2\, u] / [(1/n) \sum \hat{y}_2^2]^2$$
$$= \text{plim } [(1/n) (\sum \hat{y}_2\, u) (\sum \hat{y}_2\, u)] / [(1/n) \sum \hat{y}_2^2]^2$$
$$= [\sigma_u^2 \text{ plim } (1/n) \sum \hat{y}_2^2] / [\text{plim } (1/n) \sum \hat{y}_2^2]^2$$
$$= \sigma_u^2 / \text{plim } (1/n) \sum \hat{y}_2^2 \qquad (9.14)$$

burada σ_u^2 = var (u)'dur. Pratikte var (β')'nin tahmincisi $s_u^2 / \sum \hat{y}_2^2$'dir. Burada σ_u^2'nin tahmincisi s_u^2'dir.

2SLS metodunun ikinci adımında standart hataların elde edilişinde hatalı olan nedir? Hatayı görmek için eşitlikler 9.15'deki gibi yazılabilir:

$$y_1 = \beta\hat{y}_2 + (u + \beta v_2) \qquad (9.15)$$
$$= \beta\hat{y}_2 + w$$

Yukarıdaki eşitlik OLS ile tahmin edildiğinde β'nın standart hatası:

$[s_w^2 / \sum\hat{y}_2^2]^{1/2}$'dir.

Bu eşitlik $\sigma_u^2/\sum\hat{y}_2^2$ formülü ile karşılaştırıldığında s_u^2'ya ihtiyaç olduğu ve 2SLS'nin ikinci adımında α_w^2' nin elde edildiği görülür. Payda doğrudur fakat hata varyansının tahmini yanlıştır. Bu σ_u / σ_w'den elde edilen standart hatayla düzeltilebilir.

$$s_u^2 = (1/n) \sum (y_1 - \beta' \hat{y}_2)^2$$
$$s_w^2 = (1/n) \sum (y_1 - \beta' \hat{y}_2)^2$$

İkinci yazılış OLS'nin ikinci adımında elde edilir. Çoğu bilgisayar programı doğru bir şekilde hesaplar. Problemden emin olmak için standart programlar kullanmak gerekir.

Uygulamalı Örnek 9.4

Tablo 9.4'de ticari bankaların, ABD'de 1979-1984 yıllarında firmalara sağladığı kredileri göstermektedir. Aşağıdaki talep-arz denklemleri bu veriler kullanılarak tahmin edilmişlerdir. Burada talep firmalar, arz ise bankalar tarafından yapılmaktadır.

$Q_t = \beta_0 + \beta_1 R_t + \beta_2 RD_t + \beta_3 X_t + u_t$ krediler için talep

$Q_t = \alpha_0 + \alpha_1 R_t + \alpha_2 RS_t + \alpha_3 y_t + v_t$ krediler için arz

Burada:

Q_t : toplam ticari krediler (milyar dolar)

R_t : bankalar tarafından tahsil edilen ortalama kredi oranları

RS_t : 3 aylık devlet kağıtları oranı (bankalar için bir alternatif gelir oranı)

RD_t : AAA şirketinin tahvil oranları (firmaları finanse eden alternatif fiyat)

X_t : endüstriyel üretim indeksi ve firmaların gelecekteki ekonomik faaliyetleriyle ilgili beklentileri ifade eder

y_t : toplam banka depozitleri: bir ölçü değişkenidir (milyar dolar)

Her iki eşitlik de aşırı tanımlanmıştır. Bu nedenle denklemler 2SLS metoduyla tahmin edilecektir. Burada R_t, talep fonksiyonunda negatif, arz fonksiyonunda pozitif işarete sahip olması beklenir. RS_t'nin katsayısının negatif olduğu beklenir. RD_t, X_t ve y_t'nin katsayılarının ise pozitif olması beklenir. Normal EKK ve 2SLS tahminlerinin her ikisi

de beklenen işaretlere sahiptirler. Bu tahminler Tablo 9.5'de verilmiştir. Normal EKK yöntemi ve 2SLS kullanıldığında R^2'nin artabileceği ve azalabileceği görülmektedir.

Tablo 9.4. Ticari bankaların, ABD'de 1979-1984 yıllarında firmalara sağladığı krediler

N	Q	R	RD	X	RS	Y	N	Q	R	RD	X	RS	Y
1	251,8	11,75	9,25	150,8	9,35	994,3	37	359,8	15,75	15,18	143,4	12,28	1251.5
2	255,6	11,75	9,26	151,5	9,32	1002,5	38	364,6	46,56	15,27	140,7	13,48	1258,3
3	259,8	11,75	9,37	152,0	9,48	994,0	39	372,4	16,50	14,58	142,7	12,68	1295,0
4	264,7	11,75	9,38	153,0	9,46	997,4	40	374,7	16,50	14,46	141,5	12,70	1272,1
5	268,8	11,75	9,50	150,8	9,61	1013,2	41	379,3	16,50	14,26	140,2	12,09	1286,1
6	274,6	11,65	9,29	152,4	9,06	1015,6	42	386,7	16,50	14,81	139,2	12,47	1325,8
7	276,9	11,54	9,20	152,6	9,24	1012,3	43	384,4	16,26	14,61	138,7	11,35	1307,3
8	280,5	11,91	9,23	152,8	9,52	1020,9	44	384,5	14,39	13,71	138,8	8,68	1321,7
9	288,1	12,90	9,44	151,6	10,26	1043,6	45	395,0	13,50	12,94	138,4	7,92	1335,5
10	288,3	14,39	10,13	152,4	11,70	1062,6	46	393,7	12,52	12,12	137,3	7,71	1345,2
11	287,9	15,55	10,76	152,4	11,79	1058,5	47	398,9	11,85	11,68	135,7	8,07	1358,1
12	295,0	15,30	11,31	152,1	12,64	1076,3	48	395,3	11,50	11,83	139,9	7,94	1409,7
13	295,1	15,25	11,86	152,2	13,50	1063,1	49	392,4	11,16	11,79	135,2	7,86	1385,4
14	298,5	15,63	12,36	152,7	14,35	1070,0	50	392,3	10,98	12,01	137,4	8,11	1412,6
15	301,7	18,31	12,96	152,6	15,20	1073,5	51	395,9	10,50	11,73	138,1	8,35	1419,5
16	302,0	19,77	12,04	152,1	13,20	1101,1	52	393,5	10,50	11,51	140,0	8,21	1411,0
17	298,1	16,57	10,99	148,3	8,58	1097,1	53	391,7	10,50	11,46	142,6	8,19	1413,1
18	297,8	12,63	10,58	144,0	7,07	1088,7	54	395,3	10,50	11,74	144,4	8,79	1443,8
19	301,2	11,48	11,07	141,5	8,06	1099,9	55	397,7	10,50	12,15	146,4	9,08	1438,1
20	304,7	11,12	11,64	140,4	9,13	1111,1	56	400,6	10,89	12,51	149,7	9,34	1461,4
21	308,1	12,23	12,02	141,8	10,27	1122,2	57	402,7	11,00	12,37	151,8	9,00	1448,9
22	315,6	13,79	12,31	144,1	11,62	1161,4	58	405,3	11,00	12,25	153,8	8,64	1459,0
23	323,1	16,06	11,94	143,9	13,73	1200,6	59	412,0	11,00	12,41	155,0	8,76	1499,4
24	330,6	20,35	13,21	149,4	15,49	1239,9	60	420,1	11,00	12,57	155,3	9,00	1508,9
25	330,9	20,16	12,81	151,0	15,02	1223,5	61	424,4	11,00	12,20	156,2	8,90	1504,1
26	331,3	19,43	13,35	151,7	14,79	1207,1	62	428,8	11,00	12,08	158,5	9,09	1499,3
27	331,6	18,04	13,33	151,5	13,36	1190,6	63	433,1	11,21	12,57	160,0	9,52	1494,5
28	336,2	17,15	13,88	152,1	13,69	1206,0	64	439,7	11,93	12,81	160,8	9,69	1501,5
29	340,9	19,61	14,32	151,9	16,30	1221,4	65	447,3	12,39	13,28	162,1	9,83	1541,3
30	345,5	20,03	13,75	152,7	14,73	1236,7	66	452,9	12,60	13,55	162,8	9,87	1532,9
31	350,3	20,39	14,38	152,9	14,95	1221,5	67	454,4	13,00	13,44	164,4	10,12	1535,5
32	354,2	20,50	14,89	153,9	15,51	1250,3	68	455,2	13,00	12,87	165,9	10,47	1539,0
33	366,3	20,08	15,49	153,6	14,70	1293,7	69	459,9	12,97	12,66	166,0	10,37	1549,9
34	361,7	18,45	15,40	151,6	13,54	1224,6	70	467,7	12,58	12,63	165,0	9,74	1578,9
35	365,5	16,84	14,22	149,1	10,86	1254,1	71	468,7	11,77	12,29	164,4	8,61	1578,2
36	361,4	15,75	14,23	146,3	10,85	1288,7	72	476,8	11,06	12,13	164,8	8,06	1631,2

Talep fonksiyonu için elde edilen 2SLS tahminleri normal EKK tahminleri ile kıyaslandığında çok küçük değişmeler olduğu görülür. Arz fonksiyonuna bakılır ise, sadece α_1 ve α_2 parametrelerinde farklılıklar vardır. Bu örnek çok önemlidir. Çünkü arz edilen miktar faiz oranlarındaki değişmelere EKK tahminlerine göre daha tepkilidir.

İndirgenmiş formlardan elde edilen R^2 değerleri çok yüksek ise 2SLS ve normal EKK tahminleri birbirinin benzeridirler. Bunun nedeni 2SLS tahmin yönteminde ikame edilen

y, R^2'lerin çok yüksek olması durumunda y'lere çok yakın olmasındandır. Bu, çok sayıda harici değişkenin olduğu çok geniş modellerde sıkça karşılaşılan bir durumdur.

Çizelge 9.5. Ticari kredi piyasası arz ve talep modeli için normal EKK ve 2SLS tahminleri

	EKK		2SLS*	
	Katsayı	t -değeri	Katsayı	t-değeri
		Talep fonksiyonu		
β_0	-203,70	-2,9	-210,53	-2,8
β_1	-15,99	-12,0	-20,19	-12,6
β_2	-2,29	5,4	2,34	5,2
β_3	36,07	14,2	40,76	14,4
R^2	0,7884		0,75	
				Arz Fonksiyonu
α_0	-77,41	-6,9	-87,97	-6,3
α_1	2,41	2,9	6,90	3,6
α_2	-1,89	-1,8	-7,08	-3,1
α_3	0,33	51,3	0,33	42,9
R^2	0,98		0,97	

*2SLS için t değerleri asimptotiktir.

Normalleştirme Problemi

Daha önce bahsedilen 9.8'deki eşitliğe tekrar dönecek olursak, birinci eşitlikte y_1'in ikinci eşitlikte y_2'nin katsayılarının her birinin birim olduğu dikkati çekmektedir. Bu durum birinci eşitliğin y_1 karşısında ikinci eşitliğin y_2'ye göre normalleştirildiğini göstermektedir. Burada y_1 birinci eşitliğin y_2 ise ikinci eşitliğin bağımlı değişkenidir. Bu özellikle 2SLS metodunda böyledir. Açık söylemek gerekirse, bu simultane denklem modellerinin ruhuna ters düşer çünkü y_1 ve y_2 ortak olarak tanımlanmışlardır ve hiçbir değişkeni herhangi bir eşitlikte bağımlı değişken olarak gösteremeyiz. Değişkenlerden birinin katsayısını "1" olarak kabul etmek zorundayız. Fakat tahmin metodunda normalleştirme için hangi değişkenin seçileceğinin fark etmemesi gerekir. Tahmin metotlarından Tam Bilgiyle Maksimum Olabilirlik Metodu (Full-Information Maximum Likelihood, FIML) ve Sınırlı Bilgiyle Maksimum Olabilirlik Metodu (Limited-Information Maximum Likelihood, LIML) metotları bu gerekliliğe uyarlar. Fakat daha popüler olan metotlardan örneğin 2SLS metodu buna uymaz ve açık bir şekilde simultane denklem modellerinin ruhuna uymadığı söylenir.

Pratikte, normalleştirme ekonomik teori yoluyla belirlenmiştir. Örneğin; geliri belirleyen bir makro modelde tüketimin (C) ve gelirin (y) ortak belirlendiğinin kabul edilmesine rağmen, tüketim fonksiyonu aşağıdaki gibi yazılır.

$$C = \alpha + \beta\, y + u$$

Fakat aşağıdaki gibi yazılmaz.

$$y = \alpha' + \beta'\, C + u'$$

Çünkü tüketim gelir tarafından belirlenmektedir. Mikro düzeyde gelir ile tüketim arasında bir sebep-sonuç ilişkisi vardır ve bu ilişki makro düzeye de taşınır.

Tam belirlenmiş sistemlerde normalleştirme mesele değildir. Bunu görmek için yukarıdaki 9.8 eşitlik sistemindeki birinci eşitliği ele alalım. Daha önce z_1, z_2, z_3'ün araç değişken olarak kullanıldığını görmüştük. Bu araç değişkenlerin ağırlıklı olarak kullanılmasında problem yoktur ve hangi eşitliğin normalleştirileceği mesele değildir. Diğer taraftan yukarıdaki ikinci eşitlikte z_1 ve z_2 araç değişken olarak seçilmiştir ve optimum araç değişkeni z_1 ve z_2'nin ağırlıklı ortalamasıdır. Bu ağırlıkların y_1'in indirgenmiş formundan elde edildiğini görmüştük. Diğer taraftan, y_1'e karşı normalleştirilmiş ise aşağıdaki, gibi yazılabilir.

$$y_1 = b'_2 y_2 + c'_3 z_3 + \hat{u}_2$$

Burada z_1 ve z_2'nin ağırlıkları y_2'nin indirgenmiş formundan elde edilir. Bu nedenle 2SLS ve araç değişkeni tahminleri aşırı tanımlanmış sistemlerde değişik normalleştirmeler için farklıdırlar.

*9.7. Sınırlı Bilgiyle Maksimum Olabilirlik Metodu (LIML)

LIML metodu, simultane denklem modelleri için önerilen tek eşitlik metotlarının ilkidir ve En Küçük Varyans Oranı (LVR) olarak bilinir. Anderson ve Rubin tarafından 1949 yılında önerilmiş olup 1950'li yılların sonlarında Theil tarafından tanıtılan 2SLS metoduna kadar popüler bir metot olarak gelmiştir. LIML metodu sıkıcı, fakat basit modeller için kullanışlı ve kolaydır.

Yine 9.8'deki eşitliklere tekrar dönülür ise ilk eşitlik 9.16'daki gibi yazılabilir.

$$y_1^* = y_1 - b_1 y_2 = c_1 z_1 + c_2 z_2 + u_1 \qquad (9.16)$$

Her bir b_1 için y_1^* oluşturulabilir. Bu durumda y_1^*'nın sadece z_1 ve z_2 üzerine regresyonu alındığında rezidual kareler toplamı hesaplanır (RSS_1). Daha sonra y_1^*'nın bütün harici değişkenler z_1, z_2 ve z_3 üzerine regresyonunu alıp rezidual kareler toplamına RSS_2 diyelim. Eşitlik 6.17, y_1^*'nın belirlenmesinde z_3'ün önemsiz olduğunu ifade eder. RSS'de düşüşün sağlanması için z_3'ün minimum olması gerekir. LIML veya LVR metotları, $(RSS_1 - RSS_2) / RSS_1$ veya RSS_1 / RSS_2'nin minimum olması için b_1'in seçilmesi gerektiğini söyler. Böylece b_1 belirlendikten sonra, y_1^*'nın z_1 ve z_2'ye regresyonundan c_1 ve c_2 tahminleri elde edilir.

LIML ve 2SLS metotları arasında bazı önemli ilişkiler vardır:

- 2SLS metodu ($RSS_1 - RSS_2$) farkının minimum yapılmasını gösterir, LIML metodu RSS_1 / RSS_2 oranını minimize eder.
- Eğer tahmin edilecek eşitlik tam belirlenmiş ise 2SLS ve LIML metotları aynı sonucu verir.
- LIML sonuçları normalleştirme ile değişmez.

- LIML tahminlerinin, asimptotik varyansları ve kovaryansları 2SLS tahminleriyle aynıdır. Fakat standart hatalar farklı olacaktır. Çünkü hata varyansları yapısal parametrelerin farklı tahminlerinden hesaplanmaktadır.
- LIML metodu tahminlerinde varyanslar ve kovaryanslar kullanılır. 2SLS metodu bu bilgilere bağlı değildir.

Uygulamalı Örnek 9.5

Uygulamalı örnek 9.3'de kullanılan Avustralya şarap sektöründeki örneği dikkate alındığında talep fonksiyonu tam olarak tanımlandığından, 2SLS ve LIML tahminleri benzer olacaktır ve onlar araç değişkeni tahminleri ile de benzerlik göstereceklerdir. Fakat arz fonksiyonu aşırı tanımlandığından, 2SLS ve LIML tahminleri farklı olacaktır. Arz fonksiyonunun SAS bilgisayar programı kullanılarak elde edilmiş 2SLS ve LIML tahminleri çizelge 9.7'de verilmiştir. Bütün değişkenler logaritmik formdadırlar.

Çizelge 9.7. Avustralya şarap endüstrisi arz fonksiyonu için 2SLS ve LIML tahminleri

	2SLS			LIML		
Değişkenler	Katsayı	Standart Hata	t-Değeri	Katsayı	Standart Hata	t-Değeri
Sabit	-16,820	1,080	-15,57	-16,849	1,087	-15,49
P_w	2,616	0,331	7,89	2,627	0,334	7,86
S	1,188	9,190	6,24	1,183	0,191	6,18
R^2		0,95			0,954	

Bu örnekte dikkat edilirse LIML ve 2SLS tahminleri birbirinden çok farklı değiller. Fakat bu, aşırı tanımlanmış modellerle ilgili çoğu çalışmalarda karşılaşılmayan bir durumdur.

*9.8. Simültane Denklem Modellerinde EKK'nın Kullanımı Üzerine

Simülasyon problemleri sonuçlarında parametre tahminlerinin etkinsiz olduğu durumlarda, yapısal eşitliklerin OLS ile tahmin edildiği zaman, simülasyon denklem modellerinin OLS ile tahminlerinin kullanışsız olduğu söylenemez. Bazı örneklerde, sapmanın yönü hakkında büyük örnekler için bir şeyler söylenebilir ve bu faydalı bir bilgidir. Şayet eşitlik belirsiz ise o eşitlik parametreler hakkında hiçbir şey söylenemez.

Aşağıdaki talep ve arz modellerinin ele alındığını varsayalım.

$$q_t = \beta\, p_t + u_t \qquad \text{Talep Fonksiyonu}$$

$$q_t = \alpha\, p_t + v_t \qquad \text{Arz Fonksiyonu}$$

Burada q_t ve p_t logaritmik formdadır ve β ve α talep ve arz elastikiyetlerini ifade ederler. $Var(u_t) = \sigma_u^2$, $var(v_t) = \sigma_v^2$ ve $cov(u_t, v_t)\ \sigma_{uv}$ olarak kabul edilir. β'nın OLS tahmincisi:

$$\beta' = (\sum p_t q_t) / \sum p_t^2 = \beta + (\sum p_t u_t) / \sum p_t^2$$

$$= \beta + [\,(1/n) \sum p_t u_t\,)\,] / [\,(1/n) \sum p_t^2\,]$$

burada n örnek büyüklüğüdür.

$$\beta\, p_t + u_t = \alpha\, p_t + v_t$$

veya

$$p_t = (v_t - u_t) / \beta - \alpha$$

Aşağıdakilere sahip olunduğunda

$$\text{plim } (1/n)\ (\sum p_t\, u_t) \;=\; \text{cov}\,(p_t, u_t) = (\sigma_{uv} - \sigma_u^2) / (\beta - \alpha) \;\text{ ve}$$

$$\text{plim } (1/n) \sum p_t^2 \;=\; \text{var}\,(p_t) \;=\; (\sigma_v^2 + \sigma_u^2 - 2\,\sigma_{uv}) / (\beta - \alpha)^2$$

Böylece;

$$\text{plim } \beta' \;=\; \beta + (\beta - \alpha)\ [\ (\sigma_{uv} / \sigma_u^2) / (\sigma_v^2 + \sigma_u^2 - 2\,\sigma_{uv})\ ]$$

Bu formül pek kullanışlı değildir. Ancak $\sigma_{uv} = 0$ kabul edilerek yeniden düzenlenirse aşağıdaki formül elde edilir.

$$\text{plim } \beta' = \beta + (\ \beta - \alpha\ ) - [\ \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)\ ]$$

Burada β'nın negatif, α'nın pozitif ve sapma teriminin pozitif olması beklenir. Şayet talep elastikiyetini -0.8 bulunur ise, gerçek elastikiyet bundan küçük olacaktır.

Şayet OLS ile arz fonksiyonu tahmin ediliyor ise, aynı sonucu gösterebiliriz:

$$\alpha' \;= (\sum p_t q_t) / \sum p_t^2$$

$\sigma_{uv} = 0$ için

$$\text{plim } \alpha' \;=\; \alpha + (\beta - \alpha)\ [\ \sigma_v^2 / (\sigma_v^2 + \sigma_u^2)\ ]$$

$(\beta - \alpha)$ negatif olarak beklendiği için sapma negatiftir. Örneğin; arz elastikiyeti 0,6 olarak tahmin edilmiş ise gerçek fiyat elastikiyeti 0,6'dan büyüktür.

Uygulamada q'nun p üzerine regresyonunda talep fonksiyonunun mu yoksa arz fonksiyonunun mu tahmin edildiği bilinmez. Şayet regresyon katsayısı + 0,3 olarak bulunmuş ise arz elastikiyetinin 0,3 ve ayrıca pozitif bir sayı ve talep elastikiyetinin 0,3 ve ayrıca negatif bir sayı olduğunu biliriz. Pratik olarak kullanışlı sonuç arz elastikiyetinin 0,3'den büyük olduğudur. Diğer taraftan, regresyon katsayısı - 0,9 bulunmuş ise arz elastikiyetinin - 0,9 ve ayrıca pozitif bir sayı ve talep elastikiyeti - 0,9 ve ayrıca negatif bir sayı olduğunu biliriz. Burada da pratik olarak kullanışlı sonuç talep elastikiyetinin -0,9'dan küçük veya mutlak değer olarak 0,9'dan büyüktür.

Yukarıdaki örnekle ilgili olarak aşağıdaki duruma da dikkat etmek gerekir.

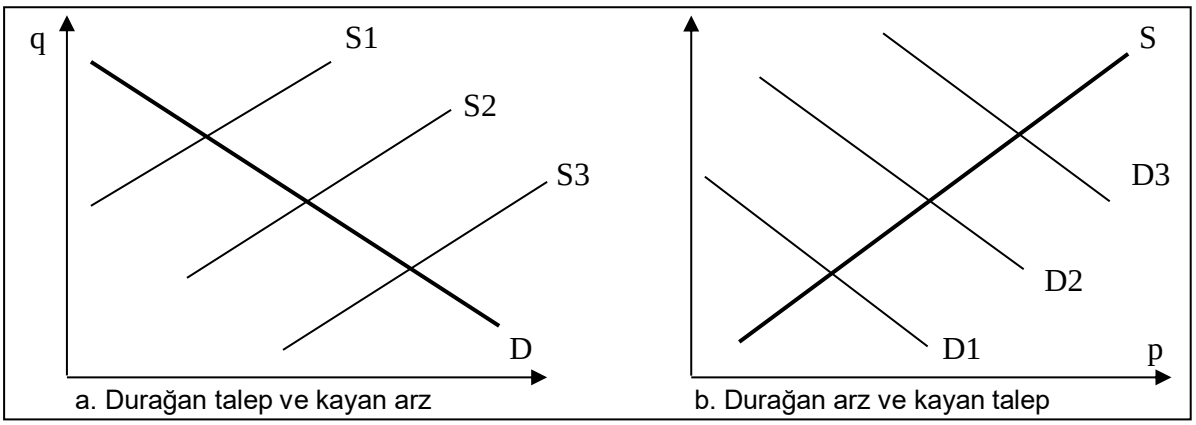
$$\text{plim } (\sum p_t\, q_t) / \sum p_t^2 \;\rightarrow \beta \text{ as } \sigma_u^2 \;\rightarrow 0$$

$$\rightarrow \alpha \text{ as } \sigma_v^2 \;\rightarrow 0$$

Burada σ_u^2 talep fonksiyondaki şansa bağlı kaymanın varyansı olduğundan, $\sigma_u^2 \rightarrow 0$ talep fonksiyonunun kaymadığı yani istikrarlı olduğunu ifade etmektedir. Bu nedenle

q'nun p üzerine regresyonu, eğer talep fonksiyonu durağan ise talep elastikiyetini ve eğer arz fonksiyonu durağan ise arz elastikiyetini tahmin edecektir.

Şekil 6.2 a'da talep eğrisinin kaymadığı ve fakat arz eğrisinin kaydığı durumu ifade gösterilmektedir. Burada pratikte gözlenen talep eğrisinin kesişme noktaları ve arz eğrileridir. Görüldüğü gibi bu kesişme noktaları talep eğrisini oluşturmaktadır. Yine şekil 6.2 b'de talep eğrisi kayarken arz eğrisi kaymamaktadır ve bu durum arz eğrisini belirlemektedir.



Şekil 6.2. Arz ve talep eğrilerinin belirlenmesi

Working'in Tanımlama Kavramı

Talep ve arz fonksiyonundan oluşan iki modelin varlığında talep fonksiyonuna ilave önemli bir değişken ekleyerek, arzın hareketliliğini azaltıp talebin fiyat elastikiyeti daha iyi tahmin edilebilir. Gelirin ilave edildiği talep fonksiyonuyla birlikte sistem aşağıdaki gibi yazılabilir.

$q_t = \beta\, p_t + \gamma\, y_t + u_t$ talep fonksiyonu

$q_t = \alpha\, p_t + v_t$ arz fonksiyonu

Working'in söylediği şey şu ki talep fonksiyonuna y_t'nin ilavesiyle β'nın daha iyi bir tahmininin elde edilebileceğidir. Fonksiyona y_t'nin ilavesiyle u_t'nin varyansı düşecektir.

Daha önce tartışılan tanımlamadan elde edilen sonuçlar kullanılırsa, arz fonksiyonunun tanımlandığı fakat talep fonksiyonunun tanımlanmadığı sonucuna varılır. Bu Working'in kastettiğinin tam karşıtıdır. Buradaki problem olan tartışılmış tanımlama kavramı, tutarlı tahminler yapma kabiliyeti ile ifade edilir. Burada Working'in endişesi en küçük kareler sapmasıyla ilgilidir. Onun iddiası σ_u^2'nun bir şekilde düşmesi ve OLS ile iyi tahminler yapılması üzerinedir. Burada β için β'_1 ve β'_2 gibi iki tahmin olduğu varsayıldığında plim $\beta'_1 = \beta$ ve plim $\beta'_2 = 0{,}999\ \beta$ olması durumunda β'_1 tutarlı bir tahminci olurken β'_2 tutarlı olmayan bir tahmindir.

Talep fonksiyonunun konu ile ilgili bütün açıklayıcı değişkenleri içinde bulundurduğunu bu yüzden σ_u^2'nun çok düşük olduğunu ve arz fonksiyonunun

açıklayıcı değişken olarak sadece fiyatı bulundurduğu kabul edilebilir. Talep fonksiyonu belirsiz olsa bile OLS ile tahmini yapılabilir. Diğer taraftan arz fonksiyonu tam belirlenmiş olsa bile σ_v^2 çok büyükse arz elastikiyetinin tahmini iyi yapılamaz. Burada α'nın etkin bir tahmini yapılabilir ancak varyansı çok yüksek olur.

Özet olarak, tanımlanmamış bir eşitliği tahmin etmekten vazgeçmek fikri doğru değildir. Aşırı tanımlanmış bir denklemin tam tanımlanmış veya belirsiz bir denklemden daha iyi tahmin edilebileceği söylenemez.

Simultane denklem modellerinde değişkenlerin harici veya dâhili diye sınıflandırılması oldukça keyfidir. Sıklıkla bazı değişkenlerin halihazır değerleri ile bu değişkenler arasında yüksek bir korelasyon olduğu zamanlarda bile dahili değişken geri kalanları harici değişken olarak atanabilir.

Tekrarlamalı (Recursive) Sistemler

Simultane denklem modellerinin tamamı OLS ile tahmin edildikleri zaman sapmalı tahminler vermezler. OLS tahminleri için geçerli önemli bir model sınıfı da Tekrarlamalı Sistemlerdir. Bu modellerde farklı eşitliklerde hatalar bağımsızdır ve dâhili değişken katsayıları üçgene benzer bir şekil gösterir. Örneğin, y_1, y_2, y_3 şeklinde 3 dâhili, z_1, z_2, z_3 şeklinde 3 harici değişkene sahip aşağıdaki modeli ele alınsın:

$$y_1 + \beta_{12} y_2 + \beta_{13} y_3 + \alpha_1 z_1 = u_1$$

$$y_2 + \beta_{23} y_3 + \alpha_2 z_2 = u_2$$

$$y_3 + \alpha_3 z_3 = u_3$$

Bu modelde u_1, u_2, u_3 bağımsızdır ve dahili değişken katsayıları şu şekildedir:

$$\begin{matrix} 1 & \beta_{12} & \beta_{13} \\ & 1 & \beta_{23} \\ & & 1 \end{matrix}$$

Bu sistemlerde her bir model OLS ile tahmin edilebilir. Yukarıdaki örnekte u_2 ile u_3 arasında korelasyon bulunduğu ve u_1'in bağımsız olduğunu varsayalım. Bu tür modellere "Block-Recursive" Modeller denir. İkinci ve üçüncü eşitlikler birlikte tahmin edilmek zorundadır, fakat birinci eşitlik OLS ile tahmin edilebilir. Burada y_2 ve y_3 sadece u_2 ve u_3' e bağımlıdır u_1'den bağımsızdırlar.

Cobb-Douglas Üretim Fonksiyonlarının Tahmini

Üretim fonksiyonlarının tahmininde çıktı, işgücü ve sermaye genellikle dâhili değişkenler olarak, işgücü ve sermaye fiyatları harici değişkenler olarak kabul edilirler. Aşağıdaki Cobb-Douglas üretim fonksiyonunun kullanıldığı varsayılsın:

$$X_i = A L_i^{\alpha} K_i^{\beta} e^{u1i} \qquad (9.17)$$

her iki tarafın logaritması alındığında aşağıdaki 9.18 eşitliğini elde ederiz.

$$Y_{1i} - \alpha\, y_{2i} - \beta\, y_{3i} = c_1 + u_{1i} \qquad (9.18)$$

Burada: $Y_{1i} = \log X_i$; $y_{2i} = \log L_i$; $y_{3i} = \log K_i$, $c_1 = \log A$'dır.

Bu durumda y_{2i} ve y_{3i} dahili değişkenler olduğu için bu eşitlikler OLS ile tahmin edilemez. Y_{2I} ve y_{3I} için eşitlikler ilave etmek zorundayız ve bunlar marjinal verimliliklerden elde edilmiştir. Zellner, Kmenta ve Dreze 9.18 nolu eşitliğin OLS ile tahmin edilebilmesini tartışmışlardır. Onların tartışmaları 9.17 nolu eşitlikte verilen outputun stokastik olması ve firmanın beklenen karlarını maksimum etmesinin gerektiğidir. Eğer $u_{1i} \sim IN\ (0, \sigma^2)$ ise bu durumda aşağıdaki sonuç elde edilir.

$$E\ (e^{u1i} = \exp\ (1/2\ \sigma^2)$$

Böylece

$$E(X_i) = A\ L_i^{\alpha}\ K_i^{\beta} \exp\ (\ 1/2\ \sigma^2\ )$$

Burada p; output fiyatı, w; işgücü fiyatı, r; sermaye fiyatı olması kaydıyla beklenen kârın hesaplanması $R_i = p_i\ E\ (\ X_i\ ) - w_i\ L_i - r_i\ K_i$ eşitliği ile yapılır. Beklenen kârın maksimum yapılması aşağıdaki şartları vermektedir.

$$\partial R_i / \partial L_i = 0 \text{ için } X_i / L_i = (w_i / \alpha p_i) \exp\ (u_{1i} - ½\ \sigma^2)$$

$$\partial R_i / \partial K_i = 0 \text{ için } X_i / K_i = (r_i / \beta p_i) \exp\ (u_{1i} - ½\ \sigma^2)$$

Logaritmaları alınıp u_{2I} ve u_{3I} hata terimleri ilave edilirse, aşağıdaki sonuçlar elde edilir. Buradaki hatalar, beklenen karların maksimum yapılmasındaki hataları vermektedir.

$$Y_{1i} - y_{2i} = c_{2i} + u_{1i} + u_{2i}$$

$$Y_{1i} - y_{3i} = c_{3i} + u_{1i} + u_{3i}$$

Burada;

$$C_{2i} = \log\ (\ w_i / \alpha p_i\ ) - ½\ \sigma^2 \qquad (9.19)$$

$$C_{3i} = \log\ (\ r_i / \beta p_i\ ) - ½\ \sigma^2$$

Şimdi 9.18 ve 9.19 nolu eşitlikler kolayca çözebilir ve y_{1i}, y_{2i} ve y_{3i} için indirgenmiş formları elde edebiliriz. Kolayca görebiliriz ki eşitlik 9.18'dan y_{1i}'yi eşitlik 9.19'a ikame eder isek u_{1i} terimleri gider. Bu nedenle y_{2i} ve y_{3i} sadece u_{2i} ve u_{3i} ile ilgilidir.

*9.9. Üç Aşamalı En Küçük Kareler Yöntemi

Üç aşamalı en küçük kareler yöntemi (3SLS) bir denklemler grubu metodudur. Yani bu denklemlerin tümü aynı anda tahmin işlemine sokulur ve tüm parametreler aynı anda tahmin edilir. Bu yöntem, iki aşamalı en küçük kareler yönteminin mantığı dikkate alınarak Theil ve Zellner tarafından geliştirilmiştir. En küçük kareler yönteminin art arda üç aşamada uygulanmasıdır. Tek denklemli modellerin tahminine göre daha fazla bilgi

kullanır. Modelin tüm yapısının yanında parametre değerlerine konan tüm sınırlamaları dikkate alır. Tek denklemli modeller tahmin edilirken sadece o modelde olan değişkenlerin ve sınırlamaların etkisi dikkate alınırken, diğer denklemlerdeki değişkenlerin bu denkleme etkisi göz ardı edilir. Halbuki bu yöntemde farklı denklemlerdeki değişkenlerin ve hata teriminin birbirleri üzerine olan kesin etkisi dikkate alınır. Üç aşamalı en küçük kareler yöntemi diğer yöntemlere göre çok daha karmaşık ve daha fazla örnek sayısına ihtiyaç duyar. Böyle bir denklemler sisteminde parametre sayısı çok olduğundan örnek büyüklüğünün parametre sayısından daha çok olacak şekilde belirlenmesi gerekir.

3SLS 2SLS'nin doğrudan bir uzantısıdır. Yani en küçük karelerin üç aşamada kullanılmasıdır. İlk iki aşama 2SLS'in aynıdır. Fakat burada sistemdeki tüm değişkenlerin indirgenmiş formlarıyla çalışılır. Üçüncü aşamada genelleştirilmiş en küçük kareler yöntemi kullanılır. Yani dönüştürülmüş denklemlere en küçük kareler uygulanır. Bu yöntemin varsayımlarını aşağıdaki gibi sıralayabiliriz.

- Bütün sistemin yapısı tam olarak bilinmelidir.
- Her bir denklemin şansa bağlı terimi ardışık olarak bağımsızdır.
- Sistemdeki denklemlerde birbirinden farklı değişkenler vardır ve bu farklı değişkenler birbirini etkiler. Eğer bütün denklemlerdeki bağımsız değişkenler aynı ise bu durumda denklemlerin ayrı ayrı tahmini aynı sonuçları verir.
- Sistemdeki denklemler aşırı belirlenmiştir. Eğer eksik tanımlama söz konusu ise sistemin tanımlanması için gerekli değişiklikler yapılır.

Üç aşamalı en küçük kareler tahminleri sapmalı fakat tutarlıdır ve iki aşamalı tahminlerden daha etkindir. Çünkü daha fazla bilgi kullanır. Yukarıdaki varsayımları temin edildiğinden kuşku duyuluyorsa, 2SLS yöntemi kullanılabilir. Bu yöntem SHAZAM programında olan SYSTEM komutu ile tahmin edilmektedir. Bu komutta denklem ve değişken sayılarını belirlemek gerekiyor.

YAZILIM UYGULAMASI

```
******************************************************************************
   SHAZAM - FOR WIN95/WIN98/WINNT PENTIUM       SITE NO. 2939A5XWU
   ** Copyright (C)  1999  by K.J. White - All Rights Reserved **

   FOR USE ONLY BY: DOC.DR. FAHRI YAVUZ
   AT: ATATURK UNIVERSITESI - ERZURUM

   FOR USE ON SINGLE COMPUTER ONLY - NO COPIES PERMITTED
******************************************************************************
```

Uygulamalı Örnek 9.1, Sayfa 157

```
******************************************************************************
|_*Burada ilk önce indirgenmiş formdaki denklemleri tahmin edelim. Bu
|_*konuyla ilgili metodoloji ders notlarından ziyade başka çalışmalardan
```

```
|_*alınmıştır. Denklemler, logaritmaların ilk farkları alınarak tahmin
|_*edilmiştir.
|_SAMPLE 1 20
|_READ (MTABLE9.1) YEAR P Q Y Z
 UNIT 88 IS NOW ASSIGNED TO: MTABLE9.1
    5 VARIABLES AND       20 OBSERVATIONS STARTING AT OBS       1
|_*Burada LOG(x) tabii logaritmayı hesaplar. Bu nedenle değişkenlerin 10
|_*temeline çevrilmesi için, bu değişkenler 10'un tabii logaritmasına
|_*bölünmesi gerekir.
|_GENR LOGP=LOG(P)/LOG(10)
|_GENR LOGQ=LOG(Q)/LOG(10)
|_GENR LOGY=LOG(Y)/LOG(10)
|_GENR LOGZ=LOG(Z)/LOG(10)
|_GENR FRSTDIFP=LOGP-LAG(LOGP)
 ..NOTE.LAG VALUE IN UNDEFINED OBSERVATIONS SET TO -99999.
|_GENR FRSTDIFQ=LOGQ-LAG(LOGQ)
 ..NOTE.LAG VALUE IN UNDEFINED OBSERVATIONS SET TO -99999.
|_GENR FRSTDIFY=LOGY-LAG(LOGY)
 ..NOTE.LAG VALUE IN UNDEFINED OBSERVATIONS SET TO -99999.
|_GENR FRSTDIFZ=LOGZ-LAG(LOGZ)
 ..NOTE.LAG VALUE IN UNDEFINED OBSERVATIONS SET TO -99999.
|_*Birinci farklar alındığından ilk gözlemlerin çıkarılması gerekmektedir.
|_*Tahmin edilen katsayılar ileride kullanılmak üzere kaydedilmiştir.

|_SAMPLE 2 20
|_OLS FRSTDIFP FRSTDIFY FRSTDIFZ / COEF=CA
        19 OBSERVATIONS     DEPENDENT VARIABLE= FRSTDIFP
 ...NOTE..SAMPLE RANGE SET TO:      2,     20
  R-SQUARE =   0.8929     R-SQUARE ADJUSTED =   0.8795
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.79514E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.28198E-01
 SUM OF SQUARED ERRORS-SSE=  0.12722E-01
 MEAN OF DEPENDENT VARIABLE = -0.18651E-02
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFY   1.0818     0.1344       8.049     0.000 0.896     0.6642    -3.3591
 FRSTDIFZ -0.83243     0.1162      -7.164     0.000-0.873    -0.5912    -1.0902
 CONSTANT -0.10164E-01 0.6518E-02  -1.559     0.138-0.363     0.0000     5.4493
|_OLS FRSTDIFQ FRSTDIFY FRSTDIFZ / COEF=CB
        19 OBSERVATIONS     DEPENDENT VARIABLE= FRSTDIFQ
 ...NOTE..SAMPLE RANGE SET TO:      2,     20
  R-SQUARE =   0.8991     R-SQUARE ADJUSTED =   0.8865
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.19854E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.14090E-01
 SUM OF SQUARED ERRORS-SSE=  0.31766E-02
 MEAN OF DEPENDENT VARIABLE =  0.92056E-03
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFY  0.16828E-03 0.6716E-01  0.2506E-02 0.998 0.001     0.0002     0.0011
 FRSTDIFZ  0.68740     0.5806E-01   11.84     0.000 0.947     0.9482    -1.8239
 CONSTANT  0.25986E-02 0.3257E-02  0.7979     0.437 0.196     0.0000     2.8229
|_* Sayfa 117'de açıklanan dolaylı en küçük kareler yöntemini kullanarak
|_*yapısal formu çözmek için indirgenmiş formu kullanalım.
|_GEN1 B1HAT=CB(2)/CA(2)
|_GEN1 B2HAT=CB(1)/CA(1)
|_GEN1 C1HAT=(-CA(1))*(B1HAT-B2HAT)
```

```
|_GEN1 C2HAT=CA(2)*(B1HAT-B2HAT)
|_GEN1 A1HAT=CB(3)-(B1HAT*CA(3))
|_GEN1 A2HAT=CB(3)-(B2HAT*CA(3))
|_* Talep fonksiyonunun tahmin edilen katsayılarını yazdıralım.
|_PRINT A1HAT B1HAT C1HAT
     A1HAT             B1HAT             C1HAT
  -0.5794211E-02       -0.8257659        0.8935059
|_* Arz fonksiyonunun tahmin edilen katsayılarını yazdıralım.
|_PRINT A2HAT B2HAT C2HAT
      A2HAT             B2HAT             C2HAT
      0.2600230E-02     0.1555512E-03     0.6875257
|_*Bu örnekte her iki denklemde tam olarak tanımlanmıştır. Bu gibi
|_*durumlarda dolaylı en küçük kareler yöntemi, araç değişken yöntemine
|_*veya iki aşamalı en küçük kareler yöntemine benzer sonuçlar vermektedir.
|_*Bu metot SHAZAM programında 2SLS komutu ile sağlanabilir. Şimdi 2SLS
|_*komutu ile yukarıdaki tahmini tekrar edelim. Genel olarak 2SLS komutunun
|_*formatı, 2SLS depvar rhsvars (exog) / options şeklindedir. Burada
|_*depvar: bağımlı değişken, rhsvars: sağ taraftaki değişkenler, exog:
|_*sistemdeki bağımsız değişkenlerdir.

|_2SLS FRSTDIFQ FRSTDIFP FRSTDIFY (FRSTDIFY FRSTDIFZ)
  TWO STAGE LEAST SQUARES - DEPENDENT VARIABLE = FRSTDIFQ
   2 EXOGENOUS VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 19 OBSERVATIONS
  R-SQUARE =   0.8864     R-SQUARE ADJUSTED =   0.8722
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.22346E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.14948E-01
 SUM OF SQUARED ERRORS-SSE=  0.35753E-02
 MEAN OF DEPENDENT VARIABLE =  0.92056E-03
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFP -0.82577     0.7400E-01  -11.16     0.000-0.941    -1.6040     1.6731
 FRSTDIFY  0.89351     0.1139       7.845     0.000 0.891     1.0656     5.6211
 CONSTANT -0.57942E-02 0.3515E-02  -1.648     0.119-0.381     0.0000    -6.2942
|_2SLS FRSTDIFQ FRSTDIFP FRSTDIFZ (FRSTDIFY FRSTDIFZ)
  TWO STAGE LEAST SQUARES - DEPENDENT VARIABLE = FRSTDIFQ
   2 EXOGENOUS VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 19 OBSERVATIONS
  R-SQUARE =   0.8991     R-SQUARE ADJUSTED =   0.8864
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.19864E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.14094E-01
 SUM OF SQUARED ERRORS-SSE=  0.31782E-02
 MEAN OF DEPENDENT VARIABLE =  0.92056E-03
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFP  0.15555E-03 0.6210E-01  0.2505E-02 0.998 0.001     0.0003    -0.0003
 FRSTDIFZ  0.68753     0.8262E-01   8.321     0.000 0.901     0.9484    -1.8243
 CONSTANT  0.26002E-02 0.3247E-02  0.8008     0.435 0.196     0.0000     2.8246
|_*Şimdi normal en küçük kareler yöntemi kullanılarak Q ve Y'ye göre
|_*normalleştirilmiş yapısal talep ve arz fonksiyonlarını tahmin edelim.
|_*Bu, dolaylı en küçük kareler sonuçlarının normal en küçük kareler ile
|_*karşılaştırmak için yapılmıştır.

|_OLS FRSTDIFQ FRSTDIFP FRSTDIFY
        19 OBSERVATIONS     DEPENDENT VARIABLE= FRSTDIFQ
 ...NOTE..SAMPLE RANGE SET TO:      2,     20
  R-SQUARE =   0.9043     R-SQUARE ADJUSTED =   0.8924
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.18821E-03
```

```
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.13719E-01
 SUM OF SQUARED ERRORS-SSE=  0.30114E-02
 MEAN OF DEPENDENT VARIABLE =  0.92056E-03
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFP -0.72313     0.5929E-01  -12.20     0.000-0.950    -1.4047     1.4651
 FRSTDIFY  0.76960     0.9658E-01   7.969     0.000 0.894     0.9178     4.8416
 CONSTANT -0.48852E-02 0.3213E-02  -1.520     0.148-0.355     0.0000    -5.3068

|_OLS FRSTDIFP FRSTDIFQ FRSTDIFY
        19 OBSERVATIONS     DEPENDENT VARIABLE= FRSTDIFP
 ...NOTE..SAMPLE RANGE SET TO:      2,     20
  R-SQUARE =   0.9562     R-SQUARE ADJUSTED =   0.9508
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.32497E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.18027E-01
 SUM OF SQUARED ERRORS-SSE=  0.51995E-02
 MEAN OF DEPENDENT VARIABLE = -0.18651E-02
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      16 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFQ  -1.2486     0.1024      -12.20     0.000-0.950    -0.6428     0.6162
 FRSTDIFY   1.0781     0.8585E-01   12.56     0.000 0.953     0.6619    -3.3477
 CONSTANT -0.69597E-02 0.4168E-02  -1.670     0.114-0.385     0.0000     3.7314
|_*Şimdi, arzın fiyata simultane tepkisi olmadığından fiyat terimini arz
|_*fonksiyonundan çıkaralım ve normal en küçük kareler yöntemini
|_*kullanalım.

|_OLS FRSTDIFQ FRSTDIFZ
        19 OBSERVATIONS     DEPENDENT VARIABLE= FRSTDIFQ
 ...NOTE..SAMPLE RANGE SET TO:      2,     20
  R-SQUARE =   0.8991     R-SQUARE ADJUSTED =   0.8932
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.18686E-03
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.13670E-01
 SUM OF SQUARED ERRORS-SSE=  0.31766E-02
 MEAN OF DEPENDENT VARIABLE =  0.92056E-03
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 FRSTDIFZ  0.68738     0.5585E-01   12.31     0.000 0.948     0.9482    -1.8239
 CONSTANT  0.25996E-02 0.3139E-02  0.8282     0.419 0.197     0.0000     2.8239
|_STOP
*******************************************************************************
```

Uygulamalı Örnek 9.3, sayfa 165

```
*******************************************************************************
|_SAMPLE 1 20
|_READ (MTABLE9.2) YEAR Q S PW PB A Y
 UNIT 88 IS NOW ASSIGNED TO: MTABLE9.2
    7 VARIABLES AND       20 OBSERVATIONS STARTING AT OBS       1
|_GENR LNQ=LOG(Q)
|_GENR LNS=LOG(S)
|_GENR LNPW=LOG(PW)
|_GENR LNPB=LOG(PB)
|_GENR LNA=LOG(A)
|_GENR LNY=LOG(Y)
|_* Normal en küçük kareler kullanarak talep fonksiyonunu tahmin edelim.
|_OLS LNQ LNPW LNPB LNA LNY / LOGLOG
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
```

```
  R-SQUARE =   0.9772     R-SQUARE ADJUSTED =   0.9711
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.12018E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.10963
 SUM OF SQUARED ERRORS-SSE=  0.18027
 MEAN OF DEPENDENT VARIABLE =  0.87271
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      15 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       1.1583     0.2898       3.996     0.001 0.718     0.2934     1.1583
 LNPB     -0.27483     0.6077     -0.4523     0.658-0.116    -0.0275    -0.2748
 LNA      -0.60299     0.4497      -1.341     0.200-0.327    -0.2278    -0.6030
 LNY        3.2121     0.7140       4.499     0.000 0.758     0.9324     3.2121
 CONSTANT  -23.651      3.913      -6.045     0.000-0.842     0.0000   -23.6512
|_*Şimdi yeni talep fonksiyonunu tahmin etmek için INST komutunu
|_*kullanalım. INST komutu, 2SLS komutu gibi aynı sonucu verir ve aynı
|_*formdadır.
|_INST LNQ LNPW LNPB LNA LNY (LNPB LNA LNY LNS)
 INSTRUMENTAL VARIABLES REGRESSION - DEPENDENT VARIABLE = LNQ
   4 INSTRUMENTAL VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 20 OBSERVATIONS
  R-SQUARE =   0.9724     R-SQUARE ADJUSTED =   0.9650
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.14547E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.12061
 SUM OF SQUARED ERRORS-SSE=  0.21820
 MEAN OF DEPENDENT VARIABLE =  0.87271
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      15 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       0.64336     0.6541      0.9835     0.341 0.246     0.1630     3.4026
 LNPB     -0.13966     0.6851     -0.2038     0.841-0.053    -0.0140    -0.5729
 LNA      -0.98508     0.6515      -1.512     0.151-0.364    -0.3721    -5.2876
 LNY        4.0821      1.244       3.280     0.005 0.646     1.1849    33.4735
 CONSTANT  -26.195      5.147      -5.089     0.000-0.796     0.0000   -30.0157
|_*Şimdi normal en küçük kareler yöntemini kullanarak arz fonksiyonunu
|_*tahmin edelim.
|_OLS LNQ LNPW LNS / LOGLOG
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9632     R-SQUARE ADJUSTED =   0.9589
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.17091E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.13073
 SUM OF SQUARED ERRORS-SSE=  0.29054
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.1450     0.2385       8.992     0.000 0.909     0.5434     2.1450
 LNS        1.3826     0.1545       8.948     0.000 0.908     0.5408     1.3826
 CONSTANT  -15.567     0.8480      -18.36     0.000-0.976     0.0000   -15.5671
|_*Şimdi dört farklı araç değişkeni senaryosunu kullanarak arz fonksiyonunu
|_*tahmin edelim.

|_OLS LNQ LNPW LNS
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9632     R-SQUARE ADJUSTED =   0.9589
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.17091E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.13073
 SUM OF SQUARED ERRORS-SSE=  0.29054
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
```

```
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.1450     0.2385       8.992     0.000 0.909     0.5434    11.3443
 LNS        1.3826     0.1545       8.948     0.000 0.908     0.5408     7.4933
 CONSTANT  -15.567     0.8480      -18.36     0.000-0.976     0.0000   -17.8376

|_INST LNQ LNPW LNS (LNPB LNS)
 INSTRUMENTAL VARIABLES REGRESSION - DEPENDENT VARIABLE = LNQ
   2 INSTRUMENTAL VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 20 OBSERVATIONS
  R-SQUARE =   0.8390     R-SQUARE ADJUSTED =   0.8200
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.74885E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.27365
 SUM OF SQUARED ERRORS-SSE=   1.2731
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW      0.33627      14.49      0.2321E-01 0.982 0.006     0.0852     1.7785
 LNS        2.1310      6.000      0.3552     0.727 0.086     0.8335    11.5497
 CONSTANT  -10.759      38.53     -0.2792     0.783-0.068     0.0000   -12.3281
|_INST LNQ LNPW LNS (LNA LNS)
 INSTRUMENTAL VARIABLES REGRESSION - DEPENDENT VARIABLE = LNQ
   2 INSTRUMENTAL VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 20 OBSERVATIONS
  R-SQUARE =   0.9399     R-SQUARE ADJUSTED =   0.9329
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.27930E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.16712
 SUM OF SQUARED ERRORS-SSE=  0.47481
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.9282     0.9683       3.024     0.008 0.591     0.7419    15.4870
 LNS        1.0584     0.4285       2.470     0.024 0.514     0.4140     5.7366
 CONSTANT  -17.649      2.673      -6.603     0.000-0.848     0.0000   -20.2236
|_INST LNQ LNPW LNS (LNY LNS)
 INSTRUMENTAL VARIABLES REGRESSION - DEPENDENT VARIABLE = LNQ
   2 INSTRUMENTAL VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 20 OBSERVATIONS
  R-SQUARE =   0.9525     R-SQUARE ADJUSTED =   0.9469
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.22076E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.14858
 SUM OF SQUARED ERRORS-SSE=  0.37529
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.6762     0.3669       7.295     0.000 0.871     0.6780    14.1538
 LNS        1.1627     0.2032       5.722     0.000 0.811     0.4548     6.3020
 CONSTANT  -16.979      1.166      -14.56     0.000-0.962     0.0000   -19.4558

|_INST LNQ LNPW LNS (LNPB LNA LNY LNS)
 INSTRUMENTAL VARIABLES REGRESSION - DEPENDENT VARIABLE = LNQ
   4 INSTRUMENTAL VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 20 OBSERVATIONS
  R-SQUARE =   0.9548     R-SQUARE ADJUSTED =   0.9495
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.21013E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.14496
 SUM OF SQUARED ERRORS-SSE=  0.35722
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.6162     0.3315       7.893     0.000 0.886     0.6628    13.8364
 LNS        1.1876     0.1902       6.243     0.000 0.834     0.4645     6.4366
```

```
 CONSTANT  -16.820      1.080      -15.58     0.000-0.967     0.0000   -19.2730
|_* Şimdi LNQ'nün LNPW ve LNA üzerine regresyonunu yapalım.
|_OLS LNQ LNPW LNA / LOGLOG
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.9431     R-SQUARE ADJUSTED =   0.9364
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.26473E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.16271
 SUM OF SQUARED ERRORS-SSE=  0.45004
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.0644     0.3128       6.600     0.000 0.848     0.5230     2.0644
 LNA        1.4175     0.2098       6.758     0.000 0.854     0.5355     1.4175
 CONSTANT  -15.296      1.055      -14.50     0.000-0.962     0.0000   -15.2961
|_*Şimdi INST komutuyla regresyonu yapalım. Burada R2'nin sıfırdan büyük
|_*olması garanti edilememektedir. Bu durumda negatif R2 değeri elde
|_*edilir.
|_INST LNQ LNPW LNA (LNS LNA)
 INSTRUMENTAL VARIABLES REGRESSION - DEPENDENT VARIABLE = LNQ
   2 INSTRUMENTAL VARIABLES
   2 POSSIBLE ENDOGENOUS VARIABLES
        20 OBSERVATIONS
  R-SQUARE =-467.3346     R-SQUARE ADJUSTED =-522.4328
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   217.79
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   14.758
 SUM OF SQUARED ERRORS-SSE=   3702.4
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       119.04      4071.      0.2924E-01 0.977 0.007    30.1580   629.5589
 LNA       -52.177      1865.     -0.2797E-01 0.978-0.007   -19.7118  -280.0700
 CONSTANT  -304.13     0.1005E+05 -0.3025E-01 0.976-0.007     0.0000  -348.4889
|_* Şimdi indirgenmiş formdaki denklemleri tahmin edelim.
|_OLS LNQ LNA LNS / LOGLOG
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNQ
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.8208     R-SQUARE ADJUSTED =   0.7997
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.83349E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.28870
 SUM OF SQUARED ERRORS-SSE=   1.4169
 MEAN OF DEPENDENT VARIABLE =  0.87271
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNA        1.3159     0.7517       1.751     0.098 0.391     0.4971     1.3159
 LNS        1.0851     0.7260       1.495     0.153 0.341     0.4244     1.0851
 CONSTANT  -10.424      1.284      -8.120     0.000-0.892     0.0000   -10.4243

|_OLS LNPW LNA LNS / LOGLOG
        20 OBSERVATIONS     DEPENDENT VARIABLE= LNPW
 ...NOTE..SAMPLE RANGE SET TO:      1,     20
  R-SQUARE =   0.4668     R-SQUARE ADJUSTED =   0.4041
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.15915E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.12615
 SUM OF SQUARED ERRORS-SSE=  0.27055
 MEAN OF DEPENDENT VARIABLE =   4.6156
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
```

```
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNA       0.44939     0.3285       1.368     0.189 0.315     0.6701     0.4494
 LNS       0.91160E-02 0.3173      0.2873E-01 0.977 0.007     0.0141     0.0091
 CONSTANT   2.4674     0.5610       4.398     0.000 0.730     0.0000     2.4674
|_DELETE / ALL
ALL VARIABLES HAVE BEEN DELETED
|_STOP
*******************************************************************************
```

Uygulamalı Örnek 9.4, sayfa 170

```
*******************************************************************************
|_SAMPLE 1 72
|_READ (MTABLE9.3) OBS Q R RD X RS Y
 UNIT 88 IS NOW ASSIGNED TO: MTABLE9.3
    7 VARIABLES AND       72 OBSERVATIONS STARTING AT OBS       1
|_OLS Q R RD X
        72 OBSERVATIONS     DEPENDENT VARIABLE= Q
 ...NOTE..SAMPLE RANGE SET TO:      1,     72
  R-SQUARE =   0.7804     R-SQUARE ADJUSTED =   0.7707
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   838.46
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   28.956
 SUM OF SQUARED ERRORS-SSE=   57015.
 MEAN OF DEPENDENT VARIABLE =   360.31
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      68 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 R         -15.992      1.336      -11.97     0.000-0.823    -0.8400    -0.6251
 RD         36.066      2.546       14.17     0.000 0.864     0.9955     1.2396
 X          2.2880     0.4228       5.412     0.000 0.549     0.3081     0.9508
 CONSTANT  -203.69      69.42      -2.934     0.005-0.335     0.0000    -0.5653
|_OLS Q R RS Y
        72 OBSERVATIONS     DEPENDENT VARIABLE= Q
 ...NOTE..SAMPLE RANGE SET TO:      1,     72
  R-SQUARE =   0.9768     R-SQUARE ADJUSTED =   0.9757
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   88.737
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   9.4200
 SUM OF SQUARED ERRORS-SSE=   6034.1
 MEAN OF DEPENDENT VARIABLE =   360.31
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      68 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 R          2.4151     0.8278       2.918     0.005 0.334     0.1269     0.0944
 RS        -1.8885      1.071      -1.764     0.082-0.209    -0.0767    -0.0564
 Y         0.33152     0.6458E-02   51.34     0.000 0.987     1.0018     1.1769
 CONSTANT  -77.415      11.23      -6.895     0.000-0.641     0.0000    -0.2149
|_2SLS Q R RD X (RD X RS Y)
  TWO STAGE LEAST SQUARES - DEPENDENT VARIABLE = Q
   4 EXOGENOUS VARIABLES   2 POSSIBLE ENDOGENOUS VARIABLES
        72 OBSERVATIONS
  R-SQUARE =   0.7485     R-SQUARE ADJUSTED =   0.7374
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   960.13
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   30.986
 SUM OF SQUARED ERRORS-SSE=   65289.
 MEAN OF DEPENDENT VARIABLE =   360.31
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      68 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 R         -20.190      1.600      -12.62     0.000-0.837    -1.0605    -0.7892
 RD         40.761      2.840       14.35     0.000 0.867     1.1251     1.4009
 X          2.3402     0.4525       5.172     0.000 0.531     0.3151     0.9725
 CONSTANT  -210.51      74.30      -2.833     0.006-0.325     0.0000    -0.5843
```

```
|_2SLS Q R RS Y (RD X RS Y)
  TWO STAGE LEAST SQUARES - DEPENDENT VARIABLE = Q
   4 EXOGENOUS VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 72 OBSERVATIONS
  R-SQUARE =   0.9667     R-SQUARE ADJUSTED =   0.9652
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   127.13
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   11.275
 SUM OF SQUARED ERRORS-SSE=   8645.0
 MEAN OF DEPENDENT VARIABLE =   360.31
 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      68 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 R          6.9051      1.898       3.639     0.001 0.404     0.3627     0.2699
 RS        -7.0819      2.268      -3.122     0.003-0.354    -0.2877    -0.2115
 Y         0.33405     0.7783E-02   42.92     0.000 0.982     1.0095     1.1858
 CONSTANT  -87.988      13.97      -6.299     0.000-0.607     0.0000    -0.2442
|_DELETE / ALL
 ALL VARIABLES HAVE BEEN DELETED
|_STOP
```

****

Uygulamalı Örnek 9.5, sayfa 174

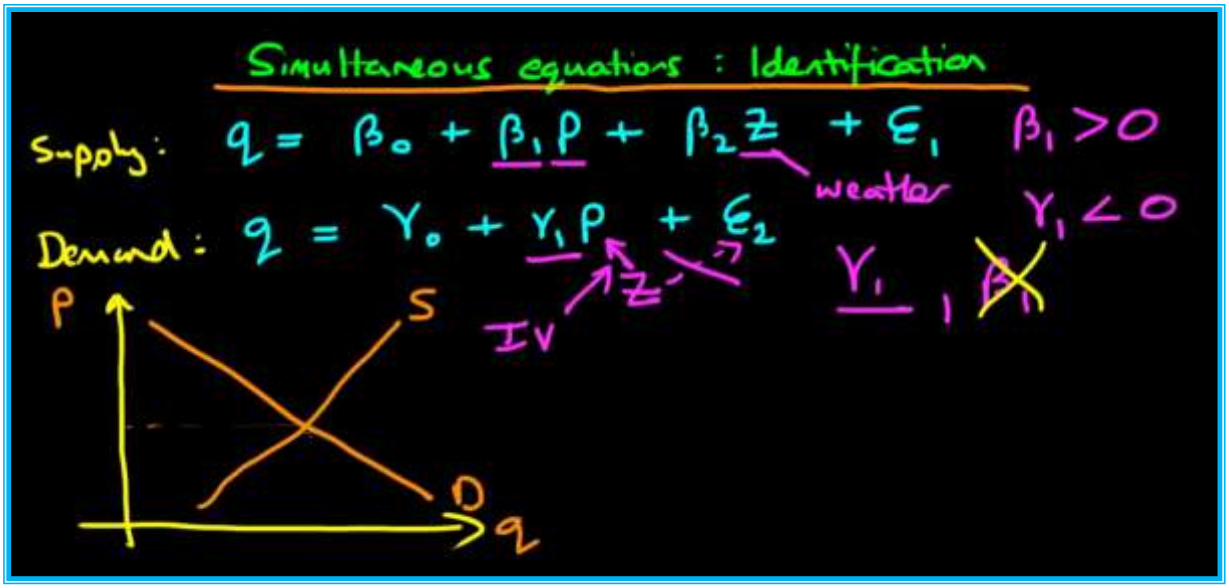
****

```
|_SAMPLE 1 20
|_READ (MTABLE9.2) YEAR Q S PW PB A Y
 UNIT 88 IS NOW ASSIGNED TO: MTABLE9.2
    7 VARIABLES AND       20 OBSERVATIONS STARTING AT OBS       1
|_GENR LNQ=LOG(Q)
|_GENR LNS=LOG(S)
|_GENR LNPW=LOG(PW)
|_GENR LNPB=LOG(PB)
|_GENR LNA=LOG(A)
|_GENR LNY=LOG(Y)
|_*Şimdi ilk olarak iki aşamalı en küçük kareler yöntemini kullanarak
|_*modeli tahmin edelim.
|_2SLS LNQ LNPW LNS (LNPB LNA LNY LNS)
  TWO STAGE LEAST SQUARES - DEPENDENT VARIABLE = LNQ
   4 EXOGENOUS VARIABLES 2 POSSIBLE ENDOGENOUS VARIABLES 20 OBSERVATIONS
  R-SQUARE =   0.9548     R-SQUARE ADJUSTED =   0.9495
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.21013E-01
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.14496
 SUM OF SQUARED ERRORS-SSE=  0.35722
 MEAN OF DEPENDENT VARIABLE =  0.87271

 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      17 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 LNPW       2.6162     0.3315       7.893     0.000 0.886     0.6628    13.8364
 LNS        1.1876     0.1902       6.243     0.000 0.834     0.4645     6.4366
 CONSTANT  -16.820      1.080      -15.58     0.000-0.967     0.0000   -19.2730
|_*Şimdi denklemi, Sınırlı Bilgiyle Maksimum Olabilirlik Metodunu (LIML)
|_*kullanarak tahmin edelim. LIML adlı bir dosyadaki SHAZAM prosedürü, bu
|_*metodu kullanarak denklem tahmini yapmaktadır. FILE PROC komutu, bu
|_*prosedürü yüklemektedir.

|_FILE PROC LIML
 UNIT 82 IS NOW ASSIGNED TO: LIML
|_* LIMITED INFORMATION MAXIMUM LIKELIHOOD
|_set noecho
```

```
|_*Gerekli matris hesaplamalarında kullanılması amacıyla sabit terimi
|_*hazırlamak için bir değerine sahip bir vektör tanımlanacaktır.
|_GENR ONE=1
|_*Sistemdeki tüm harici değişkenleri tanımlayalım.
|_EXOG:ONE LNPB LNA LNY LNS
|_*Eşitlikteki tüm dahili değişkenleri tanımlayalım.
|_ENDOG:LNQ LNPW
|_*Sağ taraftaki harici değişkenleri belirleyelim.
|_RHSEXOG:ONE LNS
|_* Sağ taraftaki dahili değişkenleri belirleyelim.
|_RHSENDOG:LNPW
|_* Sol taraftaki dahili değişkeni belirleyelim.
|_LHS:LNQ
|_* Şimdi EXEC komutunu kullanarak PROC (prosedür) ü çalıştıralım.
|_EXEC LIML
 _PROC LIML
 _         set nodoecho
 ..NOTE..CURRENT VALUE OF $ROWS=    20.000
 ..NOTE..CURRENT VALUE OF $COLS=    3.0000
     LAMBDA_
    4.680791        1.014600
 NAME         N     MEAN        ST. DEV      VARIANCE     MINIMUM      MAXIMUM
 LAMBDA_      2     2.8477      2.5924        6.7205       1.0146       4.6808
     K_
    1.014600
  The estimated coefficient vector is
   -16.84869        1.183065        2.627052
   0.2119668E-01
    1.182932       0.4003964E-01 -0.2970916
   0.4003964E-01   0.3665978E-01 -0.4624329E-01
  -0.2970916      -0.4624329E-01   0.1117562
     SE_
    1.087627       0.1914674       0.3342995
|_DELETE / ALL
 ALL VARIABLES HAVE BEEN DELETED
|_STOP
```



X. ZAMAN SERİSİ ANALİZLERİNE GİRİŞ

*10.1. Giriş

Ekonomik verileri analiz etmenin bir amacı, ilgilendiğimiz belirli değişkenlerin gelecekteki değerlerini kestirmektir. Kestirmeye yönelik yaklaşım; bir ekonometri modeli oluşturmak, mevcut verilerden parametrelerini tahmin etmek ve daha sonra bu model ile ilgilenilen değişkenlerin gelecekteki değerlerini kestirmek için kullanılmaktır. Bu yaklaşım, bu ders notlarının önceki bölümlerinde izlendi. Özellikle kısa vadeli tahmin için oldukça başarılı olduğu kanıtlanmış alternatif bir yaklaşım ise belirli bir değişkenin gelecekteki değerlerini kestirmek için yalnızca geçmiş değerlerini kullanmaktır. Bu yöntem, incelenen belirli bir değişken için değerleri oluşturan süreç hakkında sahip olabileceğimiz ekonomik bilgiyi kullanmaz. Bunun yerine, verilerin stokastik bir süreç tarafından üretildiği varsayılır ve bu stokastik süreç için bir model oluşturulur. Bu yönleriyle ekonometriden ayrılır.

Ekonomik yapı hakkındaki bilgileri kullanmayarak mevcut bilgileri ihmal ettiğimiz ve dolayısıyla verileri verimsiz kullandığımız düşünülebilir. Eğer ekonomi modelleri gerçekten doğru olsaydı ve gözlemlenen veriler tam olarak ekonomistlerin ekonomik fenomenleri açıklamak için sağladıkları modeller tarafından üretilmiş olsaydı, bu aslında doğru olurdu. Ancak gerçek durum böyle değildir. Aksine, ekonomi modelleri en iyi ihtimalle gerçeğe kaba yaklaşımlardır. Bu iddia, yalnızca tek bir değişken üzerindeki bir dizi gözlemden elde edilen bilgileri kullanan zaman serisi modellerinin, ekonometri modellerinden elde edilen tahminlerden daha üstün olan tahminler sağladığı gerçeğini bir dereceye kadar doğrulanmaktadır.

Verileri kaydettiğimiz sürece zaman çok önemli bir faktör olmuştur. Zaman serisi analizinde zaman verilerin önemli bir değişkenidir. Zaman serisi analizi dünyamızı incelememize ve içinde nasıl ilerlediğimizi öğrenmemize yardımcı olur.

Zaman serisi analizinin tanımlanmasını, neden kullanılıyor olduğunu, iyi uygulama örneklerini, tiplerini, modellerini, tekniklerini, ilgili kitaplarını ve R ilişkisini ele almak başlangıç olarak yararlı olacaktır.

*10.2. Zaman Serisi Analizi Nedir?

Zaman serisi analizi, belirli bir zaman aralığında toplanan bir dizi veri noktasını analiz etmenin özel bir yoludur. Zaman serisi analizinde analistler veri noktalarını aralıklı veya rastgele kaydetmek yerine belirli bir süre boyunca tutarlı aralıklarla veri noktalarını kaydeder. Ancak bu analizler yalnızca zaman içinde veri toplama eylemi değildir.

Zaman serisi verilerini diğer verilerden ayıran şey, analizin değişkenlerin zaman içinde nasıl değiştiğini gösterebilmesidir. Başka bir deyişle, zaman çok önemli bir değişkendir

çünkü verilerin veri noktaları ve nihai sonuçları boyunca nasıl ayarlandığını gösterir. Ek bir bilgi kaynağı ve veriler arasında bir dizi bağımlılık sırası sağlar.

Zaman serisi analizi, tutarlılık ve güvenilirliği sağlamak için tipik olarak çok sayıda veri noktası gerektirir. Kapsamlı bir veri seti, temsili bir örneklem büyüklüğüne sahip olmanızı ve bu analizin gürültülü verileri ortadan kaldırabilmesini sağlar. Ayrıca, keşfedilen herhangi bir eğilimin veya modelin aykırı değerler olmamasını ve mevsimsel değişkenliği açıklayabilmesini sağlar. Ek olarak, zaman serisi verileri geçmiş verilere dayalı olarak gelecekteki verileri kestirmek için model tahmini amacıyla kullanılabilir.

*10.3. Zaman Serisi Analizleri Neden Kullanıyor?

Zaman serisi analizi, zaman içindeki eğilimlerin veya sistemik modellerin altında yatan nedenleri anlamaya yardımcı olur. İş çevrelerinden kullanıcılar veri görselleştirmelerini kullanarak sezonluk eğilimleri görebilir ve bu eğilimlerin neden oluştuğunu daha derinlemesine inceleyebilir. Modern analiz platformlarıyla bu görselleştirmeler, çizgi grafiklerin çok ötesine geçebilir.

Verileri tutarlı aralıklarla analiz edildiğinde, gelecekteki olayların olasılığını kestirmek için zaman serisi modellerinin tahmininde kullanılabilirler. Zaman serisi kestirmesi, tahmin analitiğinin bir parçasıdır. Veri değişkenlerinin daha iyi anlaşılmasını sağlayan ve daha iyi kestirme yapılmasına yardımcı olan mevsimsellik veya döngüsel davranış gibi verilerdeki olası değişiklikleri gösterebilir.

Örneğin, devlet okullarından biri risk altındaki öğrencilerini belirlemek ve zaman içindeki ilerlemeyi izlemek için beş yıllık öğrenci başarı verilerini analiz etti. Günümüz teknolojisi, her gün çok büyük miktarlarda veri toplamamıza imkân sağlıyor ve kapsamlı analiz için yeterince tutarlı veri toplamak her zamankinden daha kolay.

Zaman serisi analizi durağan olmayan veriler, yani zaman içinde sürekli dalgalanan veya zamandan etkilenen oluşumlar için kullanılır. Para birimi ve satışlar her zaman değiştiği için finans, perakende ve ekonomi gibi sektörler sıklıkla zaman serisi analizini kullanır. Hisse senedi piyasası analizi, özellikle otomatik ticaret algoritmaları ile eylem halindeki zaman serisi analizinin mükemmel bir örneğidir. Aynı şekilde zaman serisi analizi, hava değişikliklerini tahmin etmek için de idealdir ve meteorologların yarının hava raporundan gelecek yıllardaki iklim değişikliğine kadar her şeyi kestirmesine yardımcı olur. Eylem halindeki zaman serisi analizi örnekleri şunları içerir:

- Hava durumu verileri
- Yağış ölçümleri
- Sıcaklık okumaları
- Nabız izleme (EKG)
- Beyin izleme (EEG)

- Üç aylık satışlar
- Stok fiyatları
- Otomatik hisse senedi ticareti
- Sektör tahminleri
- Faiz oranları

*10.4. Zaman Serisi Analizi Türleri

Zaman serisi analizi birçok veri kategorisi veya varyasyonu içerdiğinden, analistlerin bazen karmaşık modeller oluşturması gerekir. Ancak analistler tüm varyasyonları hesaba katamazlar ve belirli bir modeli her örneğe genelleyemezler. Çok karmaşık veya çok fazla şey yapmaya çalışan modeller uyum eksikliğine yol açabilir. Uyum eksikliği veya aşırı uydurma modeller, bu modellerin rastgele hata ile gerçek ilişkiler arasında ayrım yapmamasına yol açarak analizleri çarpık ve yanlış sonuçlara götürür.

Zaman serisi analizi modelleri şunları içerir:

- Sınıflandırma: Verileri ait kategorileri tanımlar ve kategorilere yerleştirir.
- Eğri uyumlama: Veri içindeki değişkenlerin ilişkilerini çalışmak için verileri bir eğri boyunca çizer.
- Tanımlayıcı analiz: Trendler, döngüler veya mevsimsel değişimler gibi zaman serisi verilerindeki kalıpları tanımlar.
- Açıklayıcı analiz: Verileri ve içindeki neden ve sonuç ilişkilerini anlamaya çalışır.
- Keşifsel analiz: Zaman serisi verilerinin temel özelliklerini genellikle görsel bir formatta vurgular.
- Kestirme: Gelecekteki verileri kestirir. Bu tür tarihsel eğilimlere dayanmaktadır. Gelecekteki çizim noktaları boyunca meydana gelebilecek senaryoları kestirerek geçmiş verileri gelecekteki veriler için bir model olarak kullanır.
- Müdahale analizi: Bir olayın verileri nasıl değiştirebileceğini çalışır.
- Segmentasyon: Kaynak bilginin altında yatan özellikleri göstermek için verileri segmentlere ayırır.

Veri sınıflandırması

Ayrıca, zaman serisi verileri iki ana kategoride sınıflandırılabilir:

- Stok zaman serisi verileri, bilgilerin olduğu gibi statik bir anlık görüntüsü gibi zamanın belirli bir noktasında özniteliklerin ölçülmesi anlamına gelir.
- Akış zaman serisi verileri, genellikle toplam bütünün bir parçası olan ve sonuçların bir kısmını oluşturan özelliklerin belirli bir süre boyunca etkinliğini ölçmek anlamına gelir.

Veri varyasyonları

Zaman serisi verilerinde, veriler boyunca ara sıra değişiklikler meydana gelebilir:

- İşlevsel analiz, kayda değer olayları belirlemek için verilerdeki kalıpları ve ilişkileri ortaya çıkarabilir.
- Trend analizi, belirli bir yönde tutarlı hareketin belirlenmesi anlamına gelir. İki tür eğilim vardır: altta yatan nedeni bulabileceğimiz deterministik ve rastgele ve açıklanamayan stokastik.
- Mevsimsel değişim, bir yıl boyunca belirli ve düzenli aralıklarla meydana gelen olayları ifade eder. Seri bağımlılık zaman içinde birbirine yakın olan veri noktalarının ilişkili olma eğiliminde olduğunda ortaya çıkar.

Zaman serisi analizi ve kestirme modelleri, iş hayatıyla ilgili soruları yanıtlamakla ilgili veri türlerini tanımlamalıdır. Analistler analiz etmek istedikleri ilgili verileri seçtikten sonra hangi tür analizlerin ve tekniklerin en uygun olduğunu seçerler.

Zaman Serisi Analizi İçin Önemli Hususlar

Zaman serisi verileri zaman içinde toplanan veriler olsa da, o zaman verilerinin nasıl ve ne zaman kaydedildiğini açıklayan farklı veri türleri vardır. Örneğin:

- Zaman serisi verileri, tutarlı zaman aralıklarında kaydedilen verilerdir.
- Kesit verileri, aynı anda kaydedilen birkaç değişkenden oluşur.
- Havuzlanmış veriler, hem zaman serisi verilerinin hem de yatay kesit verilerinin bir kombinasyonudur.

*10.5. Zaman Serisi Analiz Modelleri ve Teknikleri

Pek çok tür ve model olduğu gibi, verileri incelemek için de çeşitli yöntemler vardır. İşte en yaygın üç tanesi.

- Box-Jenkins ARIMA modelleri: Bu tek değişkenli modeller zamana bağlı sıcaklık gibi tek bir zamana bağlı değişkeni daha iyi anlamak ve değişkenlerin gelecekteki veri noktalarını kestirmek için kullanılır. Bu modeller, verilerin durağan olduğu varsayımıyla çalışır. Analistler, geçmiş veri noktalarındaki mümkün olduğu kadar çok değişkenliği ve mevsimselliği hesaba katmalı ve kullanmalıdır. Neyse ki ARIMA modeli, model içinde hareketli ortalamaları, mevsimsel fark operatörlerini ve otoregresif terimleri hesaba katan terimler içerir.
- Box-Jenkins Çok Değişkenli Modeller: Çok değişkenli modeller, sıcaklık ve nem gibi birden fazla zamana bağlı değişkeni zaman içinde analiz etmek için kullanılır.
- Holt-Winters Yöntemi: Holt-Winters yöntemi üstel bir düzeltme tekniğidir. Veri noktalarının mevsimselliği içermesi koşuluyla sonuçları tahmin etmek için tasarlanmıştır.

*10.6. Zaman Serisi Analizi Örnek Kitapları

Zaman serisi analizi, teknolojinin erişimi kolaylaştırmasına rağmen yeni bir çalışma değildir. Konunun temel teorilerini ve uygulamalarını öğreten önerilen metinlerin çoğu, birkaç on yıldır ekonometri literatüründe yer almaktadır. Ayrıca yöntemin kendisi bu teorilerde bile daha eskidir. Zaman serisi analizlerini binlerce yıldır gezegen hareketleri ve navigasyonun eski araştırmalarına kadar birçok alanda insanoğlu kullanagelmektedir.

Dolayısıyla bu alanda binlerce kitap var ve bazıları çok eski ve modası geçmiş olduğundan zaman serisi analiziyle ilgili en iyi kitapların bir listesini vermek yararlı olur. Bunlar kitaplar ders kitapları ve başvuru kılavuzlarının bir karışımıdır ve yeni başlayanlardan uzmanlara kadar farklı seviyelere hitap eder. Bu kitaplarda teori, örnekler, vaka çalışmaları, uygulamalar ve daha fazlasını bulabilirisiniz.

1. "Introductory Time Series with R (Use R!)" Paul S.P. Cowpertwait and Andrew V. Metcalfe
2. "The Analysis of Time Series: An Introduction" Chris Chatfield
3. "Time Series Analysis" James Douglas Hamilton
4. "Forecasting: Principles and Practice" Rob J. Hyndman and George Athanasopoulos
5. "Introduction to Time Series Analysis and Forecasting" Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci
6. "Practical Time Series Forecasting with R: A Hands-On Guide" Galit Shmueli and Kenneth C. Lichtendahl Jr
7. "Time Series Analysis: Forecasting and Control" George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

Zaman serisi analizi ve R

Açık kaynak programlama dili ve R ortamı, çizim gibi genel zaman serisi analiz işlevlerini yalnızca birkaç tuşa basarak tamamlayabilir. Daha karmaşık işlevler mevsimsel değerleri veya düzensizlikleri bulmayı da içerir.

Zaman serisi analizi teknik ve sağlam bir konudur ve burada ifade edilenler sadece yüzeyseldir. Zaman serisi analizi verilerinizi anlamak için çok önemlidir. İleriye ve geriye bakma, yıllardan günlere kadar detaya inme ve farklı zaman dilimlerindeki veri eğilimlerini görme yeteneği, en kapsamlı analizler için çok önemlidir.

KAYNAKLAR

Anonim, 2023. Time Series Analysis: Definition, Types, Techniques, and When It's Used, https://www.tableau.com/learn/articles/time-series-analysis

Green, W. H., 2012. Econometric Analysis, Seventh Edition, Prentice Hall, Columbus.

Gujarati, D.N., 1995. Basic Econometrics, Third Edition, McGraw Hill, New York.

Johnston, J., 1984. Econometric Methods, Third Edition, McGraw-Hill, New York.

Judge, G. G., Hill, R.C., Griffits, W. E., Lutkepohl, H., Lee, T.C., 1982. Introduction to the Theory and Practice of Econometrics, Second Edition, John Wiley, NewYork.

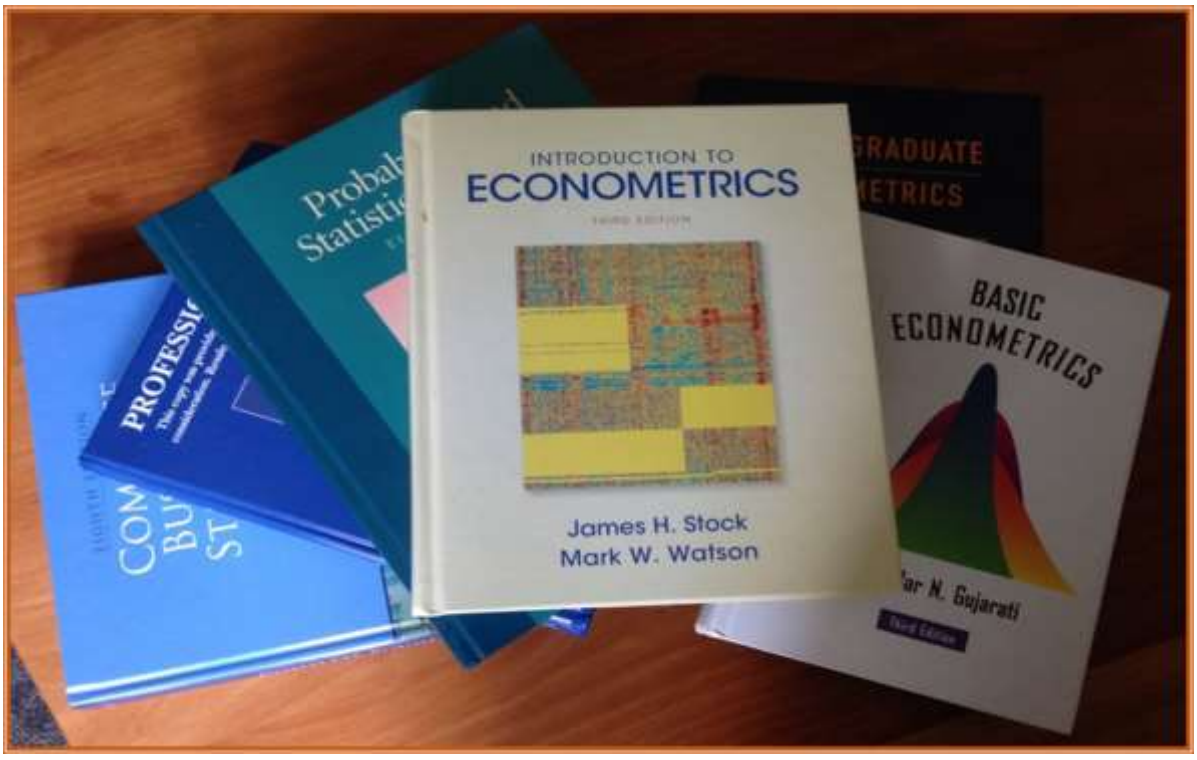
Kennedy, P., A., 1985. Guide to Econometrics, Second Edition, Blackwell, Oxford.

Koutsoyiannis, A., 1989. Ekonometri Kuramı: Ekonometri Yöntemlerinin Tanımına Giriş, Çevirenler: Ü. Şenesen, G. Günlük Şenesen, Feryal Matbaacılık, Ankara.

Maddala, G. S., 1992. Introduction to Econometrics, Second Edition, Macmillan, New York.

Varian, H. V., 1984. Microeconomic Analysis, Second Edition, W. W. Norton Company Inc., New York, London.

White, K.J., Wong, S.D., Whistler, D., Haun, S.A., 1997. Shazam Econometrics Computer Program, User's Reference Manual, Version 8.0, Irwin / McGraw-Hill Inc., Canada.



EK 1. TABLOLAR

χ^2 Dağılımı

n/g.a.	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,995
1	$0,0^4393$	$0,0^3316$	$0,0^3982$	$0,0^2393$	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,60
3	0,0717	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,30	12,80
4	0,207	0,297	0,484	0,711	1,06	1,92	2,36	5,39	7,78	9,49	11,10	13,30	14,90
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,10	12,80	15,10	16,70
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,60	12,60	14,40	16,80	18,50
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,00	14,10	16,00	18,50	20,30
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,20	13,40	15,50	17,50	20,10	22,00
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,40	14,70	16,90	19,00	21,70	23,60
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,50	16,00	18,30	20,50	23,20	25,20
11	2,60	3,05	3,82	4,57	5,58	7,58	10,30	13,70	17,30	19,70	21,90	24,70	26,80
12	3,07	3,57	4,40	5,23	6,30	8,44	11,30	14,80	18,50	21,00	23,30	26,20	28,30
13	3,57	4,11	5,01	5,89	7,04	9,30	12,30	16,00	19,80	22,40	24,70	27,70	29,80
14	4,07	4,66	5,63	6,57	7,79	10,20	13,30	17,10	21,10	23,70	26,10	29,10	31,30
15	4,60	5,23	6,26	7,26	8,55	11,00	14,30	18,20	22,30	25,00	27,50	30,60	32,80
16	5,14	5,81	6,91	7,96	9,31	11,90	15,30	19,40	23,50	26,30	28,80	32,00	34,30
17	5,70	6,41	7,56	8,67	10,10	12,80	16,30	20,50	24,80	27,60	30,20	33,40	35,70
18	6,26	7,01	8,23	9,39	10,90	13,70	17,30	21,60	26,00	28,90	31,50	34,80	37,20
19	6,84	7,63	8,91	10,10	11,70	14,60	18,30	22,70	27,20	30,10	32,90	36,20	38,60
20	7,43	8,26	9,59	10,90	12,40	15,50	19,30	23,80	28,40	31,40	34,20	37,60	40,00
21	8,03	8,90	10,30	11,60	13,20	16,30	20,30	24,90	29,60	32,70	35,50	38,90	41,40
22	8,64	9,54	11,00	12,30	14,00	17,20	21,30	26,00	30,80	33,90	36,80	40,30	42,80
23	9,26	10,20	11,70	13,10	14,80	18,10	22,30	27,10	32,00	35,20	38,10	41,60	44,20
24	9,89	10,90	12,40	13,80	15,70	19,00	23,30	28,20	33,20	36,40	39,40	43,00	45,60
25	10,50	11,50	13,10	14,60	16,50	19,90	24,30	29,30	34,40	37,70	40,60	44,30	46,90
26	11,20	12,20	13,80	15,40	17,30	20,80	25,30	30,40	35,60	38,90	41,90	45,60	48,30
27	11,80	12,90	14,60	16,20	18,10	21,70	26,30	31,50	36,70	40,10	43,20	47,00	49,60
28	12,50	13,60	15,30	16,90	18,90	22,70	27,30	32,60	37,90	41,30	44,50	48,30	51,00
29	13,10	14,30	16,00	17,70	19,80	23,60	28,30	33,70	39,10	42,60	45,70	49,60	52,30
30	13,80	15,0	16,80	18,50	20,60	24,50	29,30	34,80	40,30	43,80	47,00	50,90	53,70

t Dağılımı

n/g.a.	0,75	0,90	0,95	0,975	0,99	0,995	0,9995
1	1,000	3,078	6,314	12,706	31,821	63,657	636,62
2	0,816	1,886	2,920	4,303	6,965	9,925	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	6,859
6	0,718	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	5,405
8	0,706	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	1,296	1,671	2,00	2,390	2,660	3,460
120	0,677	1,289	1,658	1,98	2,358	2,617	3,373
∞	0,674	1,282	1,645	1,96	2,326	2,576	3,291

F Dağılımı
(0.95 Güven Aralığı - 0.05 Önem Seviyesi)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,35	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

F Dağılımı
(0.99 Güven Aralığı - 0.01 Önem Seviyesi)

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,9	26,7	26,6	26,5	26,4	26,3	26,2	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,2	14,0	13,9	13,8	13,7	13,7	13,6	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,01
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,4	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,60	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,70	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,0	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	3,92	2,84	2,75	2,66	2,57
19	8,19	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,30	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,53	2,45	2,36	2,27	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

Durbin - Watson Testi için Kritik Değerler
(%5 Önem Seviyesi)

	k = 3		k = 6		k = 9		k = 12		k = 15		k = 21	
n	**A**	**Ü**	**A**	**Ü**	**A**	**Ü**	**A**	**Ü**	**A**	**Ü**	**A**	**Ü**
10	0,70	1,64	0,24	2,82								
15	0,95	1,54	0,56	2,22	0,25	2,98						
20	1,10	1,54	0,79	1,99	0,50	2,52	0,26	3,06	0,10	3,54		
25	1,21	1,55	0,95	1,89	0,70	2,28	0,47	2,70	0,27	3,12	0,04	3,79
30	1,28	1,57	1,07	1,83	0,85	2,14	0,64	2,48	0,45	2,82	0,16	3,47
35	1,34	1,58	1,16	1,80	0,97	2,05	0,78	2,33	0,60	2,62	0,30	3,19
40	1,39	1,60	1,23	1,79	1,06	2,00	0,90	2,23	0,73	2,47	0,43	2,97
45	1,43	1,62	1,29	1,78	1,14	1,96	0,99	2,16	0,84	2,37	0,55	2,81
50	1,46	1,63	1,34	1,77	1,20	1,93	1,06	2,10	0,93	2,29	0,66	2,68
55	1,49	1,64	1,37	1,77	1,25	1,91	1,13	2,06	1,00	2,23	0,75	2,57
60	1,51	1,65	1,41	1,77	1,30	1,89	1,19	2,03	1,07	2,18	0,84	2,49
65	1,54	1,66	1,44	1,77	1,34	1,88	1,23	2,01	1,12	2,14	0,91	2,42
70	1,55	1,67	1,46	1,77	1,37	1,87	1,27	1,99	1,17	1,11	0,97	2,36
75	1,57	1,68	1,49	1,77	1,40	1,87	1,31	1,97	1,22	2,08	1,03	2,32
80	1,59	1,69	1,51	1,77	1,43	1,86	1,34	1,96	1,25	2,06	1,08	2,28
85	1,60	1,70	1,53	1,77	1,45	1,86	1,37	1,95	1,29	2,04	1,12	2,24
90	1,61	1,70	1,54	1,78	1,47	1,85	1,40	1,94	1,32	2,03	1,16	2,21
95	1,62	1,71	1,56	1,78	1,49	1,85	1,42	1,93	1,35	2,01	1,20	2,19
100	1,63	1,72	1,57	1,78	1,51	1,85	1,44	1,92	1,37	2,00	1,23	2,16
150	1,71	1,76	1,67	1,80	1,62	1,85	1,58	1,90	1,54	1,94	1,44	2,04
200	1,75	1,79	1,72	1,82	1,69	1,85	1,65	1,89	1,62	1,92	1,55	1,99