Ekonometri 1
Ders Notları

A. TALHA YALTA

TÜRKİYE BİLİMLER AKADEMİSİ
AÇIK DERS MALZEMELERİ PROJESİ

SÜRÜM 2.0
EKİM 2011

Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011

İçindekiler

1 İstatistiksel Kavramların Gözden Geçirilmesi 8
1.1 Anlamlı Basamaklar ve Yuvarlama Kuralları 8
1.2 Olasılık Konusu ve Olasılık Dağılımları 11
1.2.1 Olasılık ve Olasılık Yoğunluk İşlevi 11
1.2.2 Olasılık Dağılımlarının Beklemleri 15
1.2.3 Bazı Kuramsal Olasılık Dağılımları 20
1.3 İstatistiksel Çıkarsama 25
1.3.1 Tahmin Sorunu 25
1.3.2 Önsav Sınaması 28

2 Ekonometri Nedir? 32
2.1 Ekonometri Nedir? 32
2.1.1 Ekonometrinin Konusu 32
2.1.2 Ekonometrinin Yöntembilimi 35
2.1.3 Uygulama: Keynesçi Tüketim Kuramı 35

3 Bağlanım Çözümlemesi 43
3.1 Temel Kavramlar 43
3.1.1 Bağlanım Teriminin Anlamı 43
3.1.2 Ekonometrik Çözümlemede Kullanılan Verilerin Niteliği . . 45
3.2 Varsayımsal Bir Örnek 48
3.2.1 Koşullu Olasılık ve Koşullu Ortalama 48
3.2.2 Anakütle Bağlanım İşlevi 49
3.2.3 Örneklem Bağlanım İşlevi 52

4 İki Değişkenli Bağlanım Modeli - Tahmin Sorunu 56
4.1 Sıradan En Küçük Kareler Yöntemi 56
4.1.1 SEK Tahmincilerinin Türetilmesi 58
4.1.2 SEK Tahmincilerinin Arzulanan Özellikleri 61
4.1.3 SEK Yönteminin Ardındaki Varsayımlar 64
4.2 SEK Yönteminin Güvenilirliği 71

4.2.1 SEK Tahmincilerinin Ölçünlü Hataları 71
4.2.2 Belirleme Katsayısı r^2 73
4.2.3 Monte Carlo Yöntemi 75
4.3 Sayısal Bir Örnek 76

5 Normallik Varsayımı ve Ençok Olabilirlik Yöntemi 79
5.1 Normallik Varsayımı ve İlişkin Dağılımlar 79
5.1.1 Hata Teriminin Olasılık Dağılımı 79
5.1.2 Normal Dağılıma İlişkin Dağılımlar 81
5.2 Ençok Olabilirlik Yöntemi 84
5.2.1 Ençok Olabilirlik Yaklaşımı 84
5.2.2 İkiterimli Dağılım Örneği 85
5.2.3 İkiterimli Dağılım EO Tahmincisi 86
5.3 Açıklayıcı Örnekler 88
5.3.1 Poisson Dağılımı EO Tahmincisi 88
5.3.2 Üstel Dağılım EO Tahmincisi 89
5.3.3 Normal Dağılım EO Tahmincisi 89

6 İki Değişkenli Bağlanım Modeli - Çıkarsama Sorunu 93
6.1 Aralık Tahmini 93
6.1.1 Bazı Temel Noktalar 93
6.1.2 SEK Tahmincilerinin Güven Aralıkları 94
6.2 Önsav Sınaması 97
6.2.1 Güven Aralığı Yaklaşımı 97
6.2.2 Anlamlılık Sınaması Yaklaşımı 99
6.2.3 Anlamlılık Konusu 100
6.3 Çıkarsamaya İlişkin Konular 103
6.3.1 Varyans Çözümlemesi 103
6.3.2 Kestirim Sorunu 104
6.3.3 Bağlanım Bulgularının Değerlendirilmesi 106

7 İki Değişkenli Bağlanım Modelinin Uzantıları 110
7.1 Sıfır Noktasından Geçen Bağlanım 110
7.2 Hesaplamaya İlişkin Konular 114
7.2.1 Ölçekleme ve Ölçü Birimleri 114
7.2.2 Sayısal Hesaplama Sorunları 115
7.3 Bağlanım Modellerinin İşlev Biçimleri 118
7.3.1 Log-Doğrusal Model 118
7.3.2 Yarı-logaritmasal Modeller 120
7.3.3 Evrik ve Log-Evrik Modeller 124

8 Çoklu Bağlanım Çözümlemesi - Tahmin Sorunu 129
8.1 Üç Değişkenli Model 129
8.1.1 Gösterim ve Varsayımlar 129
8.1.2 Kısmi Bağlanım Katsayılarının Tahmini 131
8.2 Çoklu Bağlanımda Yakışmanın İyiliği 135
8.2.1 Çoklu Belirleme ve İlinti Katsayıları 135
8.2.2 Kısmi İlinti Katsayıları 138
8.2.3 Çoklu Bağlanım Açıklayıcı Örnek 139
8.3 Çokterimli Bağlanım Modelleri 143

9 Çoklu Bağlanım Çözümlemesi - Çıkarsama Sorunu 147
9.1 T Sınamaları 147
9.1.1 Çoklu Bağlanımda Önsav Sınaması 147
9.1.2 Tek Bir Katsayının Sınanması 148
9.1.3 İki Katsayının Eşitliğinin Sınanması 149
9.2 F Sınamaları 151
9.2.1 Bağlanımın Bütününün Anlamlılık Sınaması 151
9.2.2 Bir Açıklayıcı Değişkenin Marjinal Katkısı 153
9.2.3 Sınırlamalı Enküçük Kareler Yöntemi 156
9.3 Diğer Sınama ve Konular 160
9.3.1 Chow Sınaması 160
9.3.2 MWD Sınaması 164
9.3.3 Diğer Bazı Sınama ve Konular 166

10 Kukla Değişkenlerle Bağlanım 169
10.1 Nitel Değişkenlerle Bağlanım 169
10.1.1 VARÇÖZ Modelleri 170
10.1.2 KOVÇÖZ Modelleri 172
10.2 Kukla Değişken Kullanım Şekilleri 173
10.2.1 Chow Sınamasının Kukla Almaşığı 173
10.2.2 Karşılıklı Etkileşim 175
10.2.3 Parça-Yollu Doğrusal Bağlanım 176
10.3 Kukla Değişkenlere İlişkin Konular 178
10.3.1 Mevsimsel Çözümlemeler 178
10.3.2 Yarı-Logaritmasal İşlevler 180
10.3.3 İleri Çalışma Konuları 182

Önsöz

Bu ekonometri ders notları uzun ve titiz bir çalışmanın ürünüdür. Aynı zamanda, uzun bir süredir içinde yer aldığım açık kaynak hareketinin önemine olan inancımın göstergesi ve bu oluşuma verdiğim desteğin bir parçasıdır. Ders notlarımı ekonometri öğrenmeyi ve öğretmeyi arzulayan herkesin açık ve özgür kullanımına mutlulukla sunuyorum. Yararlanacak kişiler için; var olan malzemenin kapsamı, sayfa düzeni ve kullandığı terminoloji ile ilgili birkaç bilginin açıklayıcı olacağını düşünüyorum.

Notların İçeriği

- Ders notları TOBB Ekonomi ve Teknoloji Üniversitesi'nde 2007 yılından bu yana vermiş olduğum Ekonometri 1 ve Ekonometri 2 derslerinden ortaya çıkmıştır.
- Notlar, genel olarak, önceki bir baskısı Ümit Şenesen ve Gülay Günlük Şenesen tarafından Türkçe'ye de çevrilmiş olan Gujarati ve Porter'ın Basic Econometrics ders kitabı konu sırasını izlemektedir.
- Tüm görsel öğeler tarafımdan Türkçe'ye kazandırılmış olan gretl (GNU Regression, Econometrics and Time-series Library) ekonometri yazılımı kullanılarak oluşturulmuştur.
- Notlarda yer alan çözümleme ve örneklerin tamama yakını Türkiye'yi konu almakta, Türkiye verilerini kullanmaktadır.
- Bu özgün veri setleri ders notlarını tamamlayıcıdır ve gretl gdt ve csv dosyası olarak iki ayrı biçimde ekte verilmiştir.

Sayfa Düzeni

- Tüm konu anlatımları yatay düzende ve sunum biçiminde hazırlanmıştır. Bunun nedeni, öğrenmeyi özendiren çekici bir yaklaşım benimsemek ve notların bilgisayar ekranında okunabilmesini kolaylaştırmaktır.

- Benimsemiş olduğum yöntemin çizim, çizelge, ve tahmin çıktıları gibi görsel öğelere dayalı uygulamalı bir bilim olan ekonometriyi öğretmede elverişli olduğunu düşünüyorum.
- A4 düzenine getirildiğinde, her bir konu ortalama 15 - 20 sayfa tutmaktadır. Bu şekilde hazırlanmış olan bir “kitap” sürümü de ilgilenenler için ayrıca sunulmaktadır.
- Konu anlatımlarının yanı sıra, ikişer takım sınav soru ve yanıtları da açık ders malzemeleri içinde yer almaktadır. Bu ek belgeler de A4 sayfa boyutundadır.

Kullanılan Terminoloji

- Türkçe terimler konusunda çeşitli akademisyenlerin değerli katkıları bulunmakla birlikte, yerleşmiş ve kendi içerisinde tutarlı bir ekonometrik terminolojinin eksikliği bir gerçektir.
- Ders notlarında kullanılan Türkçe konusunda büyük titizlik gösterilmiş ve çeşitli ekonometri kaynakları taranarak daha önce farklı yazarlarca önerilmiş karşılıklara dayalı, anlam ve dilbilgisi yönünden doğru bir terimler seti hazırlanmıştır. Bu konuda yerli ve yabancı dilbilimci ve ekonometricilerden de sıkça yardım alınmıştır.
- Çeşitli ekonometrik terimlerin İngilizce karşılıklarının metin içerisinde düzenli olarak verilmesi, notlarının bir özelliğidir.
- İki sözcükten oluşan ancak tek bir kavrama karşılık gelen ve terim özelliği gösteren sözcüklerin bitişik yazılması ise bilinçli bir seçimdir. (*Örnek:* Bandwidth = Kuşakgenişliği)

Terminolojide Yararlanılan Kaynaklar
Ders notlarında kullanılan terminolojide yararlanılan başlıca kaynaklar şunlardır:

- Akalın H. vd., *TDK Ekonometri Sözlüğü*, `http://www.emu.edu.tr/mbalcilar/eets/Ana\_Sayfa.html`
- Ceyhan İ. vd., *İstatistik Terimleri Sözlüğü*, Türk Dil Kurumu, 1983.
- Güriş S. ve E. Çağlayan, *Ekonometrik Terimler Sözlüğü*, Derin Yayınevi, 2007.
- Kutlar A., *Uygulamalı Ekonometri*, 2. b., Nobel Yayın Dağıtım, 2005.

- Şenesen Ü. ve G. G. Şenesen, *Temel Ekonometri*, 4. b., Literatür Yayıncılık, 2006.
- Tarı R., *Ekonometri*, 4. b., Kocaeli Üniversitesi Yayınları, 2006.

Terim Seçimine Örnek

- Kullanmakta olduğum terimler konusunda ısrarcı değilim. Öte yandan, belli bir terim için şu sözcük kullanılmalıdır denilecek olursa bunu nedeninin gösterilebilmesi gerek diye düşünüyorum.
- Örnek olarak, “asymptote” terimi için Türkçe kaynaklarda “kavuşmaz,” “sonuşmaz,” ve ”yanaşık” gibi karşılıkların kullanılmış olduğu görülmektedir. Diğer yandan, -iş -ış eki Türkçe’de yalnızca fiillerin sonuna geldiği için “sonuşmaz” sözcüğü dilbilgisi yönünden yanlıştır.
- Terimin kavramsal içeriğine dikkat ederek ve Türk Dili ve Edebiyatı Bölümü’nden hocalarıma danışarak “kavuşmaz” terimini yeğledim ve tüm akademisyen arkadaşlarıma da bir öneri olarak sundum.
- Buna benzer örnekleri çoğaltmak mümkündür.

Olası Yanlışlar Konusunda

Büyük titizlikle hazırladığım notlarımı zaman içerisinde çok kez gözden geçirme fırsatım olduğu için mutluyum. Ayrıca, bu ders malzemeleri TÜBA Açık Ders Malzemeleri Projesi kapsamında anonim ekonometriciler tarafından da incelenmiştir. En ufak bir yazım yanlışı bile olmaması gereken bu malzemelerde bir hata görürseniz, düzeltmem için lütfen benimle bağlantıya geçiniz.

A. Talha Yalta, Ekim 2011
http://yalta.etu.edu.tr

Bölüm 1

İstatistiksel Kavramların Gözden Geçirilmesi

1.1 Anlamlı Basamaklar ve Yuvarlama Kuralları

Anlamlı Basamaklar
Ondalık bir sayının *"anlamlı basamakları"* (significant digits), o sayının kesinlik ve doğruluğuna katkıda bulunan tüm basamaklarını gösterir.

- Veri ve ölçümleri elde etmek için çeşitli süreç ve işlemler kullanılabilmektedir.
- Eğer eldeki ölçüme ait bazı rakamlar, o ölçümü elde etmek için kullanılan sürecin doğruluk sınırı dışındaysa, bunları kullanmanın anlamı yoktur.
- Örnek olarak, kol saatimize bakıp "saat 10:18:37:3" demek anlamlı değildir. Saat 10:18'dir.

Anlamlı Basamakları Belirleme Kuralları

1. Sıfır olmayan tüm basamaklar anlamlıdır. *Örnek:* 123456 sayısının anlamlı basamak sayısı altıdır.
2. İki sıfır-dışı basamak arasındaki tüm sıfırlar anlamlıdır. *Örnek:* $103{,}406$ sayısının anlamlı basamak sayısı altıdır.
3. Baştaki sıfırlar anlamsızdır. *Örnek:* 000012 ve $0{,}012$ için anlamlı basamak sayısı ikidir.

4. Ondalık ayraç içeren sayılarda sondaki sıfırlar anlamlıdır. *Örnek:* $1{,}20300$ için anlamlılık düzeyi altı basamaktır.

5. Tam sayılarda sondaki sıfırlar anlamlı ya da anlamsız olabilir. *Örnek:* $(10\bar{0}00)$, $(10\underline{0}00)$, (1230000) ve $(100{,})$ sayıları için anlamlılık düzeyi üçtür. Sonuncu örnekte ondalık ayraçının anlamlılık düzeyini vurgulamak için kullanılmış olduğuna dikkat ediniz.

Bilimsel Gösterim

- *"Bilimsel gösterim"* (scientific notation), baştaki ve sondaki anlamlı olmayan sıfırları kullanmayarak anlamlı basamak sayısındaki olası bir karışıklığı önlemeyi hedefler.
- Kısaca bilimsel gösterimde tüm basamaklar anlamlıdır.
- *"Üstel gösterim"* (exponential notation) adı da verilen bilimsel gösterimde tüm sayılar $a \times 10^b$ biçiminde yazılır.
- Burada b bir tam sayıdır. a ise $1 \leq |a| < 10$ olan bir *"oranlı sayı"* (rational number) biçimindedir. *Örnek:* $0{,}00123$ bilimsel gösterimi $1{,}23 \times 10^{-3}$'tür. *Örnek:* $0{,}0012300$ bilimsel gösterimi $1{,}2300 \times 10^{-3}$'tür. *Örnek:* 1230000 eğer dört basamağa kadar anlamlı ise $1{,}230 \times 10^6$ diye gösterilir. *Örnek:* Üç basamağa kadar anlamlıysa da $1{,}23 \times 10^6$ olur.
- *Dikkat:* Bilimsel gösterimde, baştaki oranlı sayının her zaman 1 ile 10 arasında olduğuna dikkat ediniz.

Yuvarlama Kuralları

"Yuvarlama" (rounding) kavramı anlamlı basamak kavramı ile yakından ilişkilidir. Çeşitli hesaplamalarda sıradan yuvarlama yerine *"istatistikçi yuvarlaması"* (statistician's rounding) yöntemini kullanmak, sonuçların yukarı *"yanlı"* (biased) olmasını önlemede gereklidir:

1. Tutulacak son basamak seçilir. Bir sonra gelen basamak eğer < 5 ise tutulacak basamak değişmez. *Örnek:* $1{,}2345$ sayısı üç basamağa yuvarlanırsa $1{,}23$ olur. *Örnek:* 1230000 iki basamağa yuvarlanırsa 1200000 olur.

2. Bir sonraki basamak > 5 ise tutulacak basamak bir artırılır. *Örnek:* $0{,}126$ sayısı iki basamağa yuvarlanırsa $0{,}13$ olur.

3. Bir sonra gelen basamak $= 5$ ise; tutulacak basamak tek sayıysa bir artırılır, çift sayıysa değiştirilmez. *Örnek:* 13500 sayısı iki basamağa yuvarlanırsa 14000 olur. *Örnek:* $0{,}125$ sayısı iki basamağa yuvarlanırsa $0{,}12$ olur.

Anlamlı Basamaklar ve Aritmetik

Anlamlı basamaklar ile ilgili olarak, veri ve ölçümler arası aritmetik işlemlerinde aşağıdaki kurallar uygulanır:

1. Öncelikle, örnek olarak $0{,}12$ gibi bir değerin gerçekte $0{,}115$ ile $0{,}125$ arasında olduğu unutulmamalıdır.

2. Toplama ve çıkarma işlemlerinde sonuç, girdiler içinde en az ondalık basamak içeren sayı ile aynı ondalık basamak sayısında olacak şekilde yuvarlanmalıdır. *Örnek:* $0{,}12 + 0{,}1277$ yanıtı $0{,}2477$ değil $0{,}25$ olmalıdır.

3. Çarpma ve bölme işlemlerinde sonuç, girdiler içindeki en az anlamlı basamak içeren sayı ile aynı anlamlılık düzeyinde olmalıdır. *Örnek:* $0{,}12 \times 1234$ yanıtı $148{,}08$ değil 150 olmalıdır.

4. Ancak ara işlemlerde izleyici basamakları elde tutmak gereklidir. Böylece yuvarlama hataları azaltılmış olur.

1.2 Olasılık Konusu ve Olasılık Dağılımları

1.2.1 Olasılık ve Olasılık Yoğunluk İşlevi

Örneklem Uzayı ve Örneklem Noktası

"Rastsal" (random) bir deneyin olabilecek tüm sonuçlarına *"örneklem uzayı"* (sample space), bu örneklem uzayının her bir üyesine de *"örneklem noktası"* (sample point) denir.

- *Örnek:* İki madeni para ile yazı-tura atma deneyinin 4 örneklem noktalı bir örneklem uzayı vardır:

$$Y = \{\text{YY, YT, TY, TT}\}$$

Rastsal Olay

Rastsal bir deneye ait örneklem uzayının olası her bir alt kümesine *"rastsal olay"* (random event) denir.

- *Örnek:* Bir yazı ve bir tura gelmesi olayı: $\{\text{YT, TY}\}$

Karşılıklı Dışlamalı Olay

Bir olayın gerçekleşmesi diğer bir olayın oluşmasını önlüyorsa, bu iki olay *"karşılıklı dışlamalı"* (mutually exclusive) olaylardır.

- *Örnek:* $\{\text{YY, YT, TY}\}$ ve $\{\text{TT}\}$ karşılıklı dışlamalıdır.

Rastsal Değişken

Değerleri rastsal bir deney sonucu belirlenen değişkene *"rastsal değişken"* (random variable) ya da kısaca *"rd"* (rv) denir.

- Rastsal değişkenler genellikle X, Y, Z gibi büyük harflerle ve aldıkları değerler de x, y, z gibi küçük harflerle gösterilir.
- Rastsal bir değişken ya *"kesikli"* (discrete) ya da *"sürekli"* (continuous) olur.
- Kesikli bir rd ancak sonlu sayıda farklı değerler alabilir.*Örnek:* Zar.
- Sürekli bir rd ise belli bir aralıkta her sayısal değeri alabilir.*Örnek:* Rastsal olarak seçilmiş bir kişinin boyu.

Olasılık
A, örneklem uzayındaki bir olay olsun. Rastsal deney sürekli yinelendiğinde, A olayının gerçekleşme sıklık oranına A olayına ait *"olasılık"* (probability) denir, $P(A)$ ya da $Prob(A)$ ile gösterilir.

- $P(A)$ aynı zamanda *"göreli sıklık"* (relative frequency) olarak da adlandırılır.

$P(A)$ gerçek değerli bir *"işlev"* (function) olup, şu özellikleri taşır:

1. Her A için $0 \leq P(A) \leq 1$'dir. ($1 = \%100$)
2. $A, B, C, \ldots$ örneklem uzayını oluşturuyorsa şu geçerlidir:
$$P(A + B + C + \ldots) = 1$$
3. A, B ve C karşılıklı dışlamalı olaylar ise şu geçerlidir:
$$P(A + B + C) = P(A) + P(B) + P(C)$$

Örnek: Altı yüzlü bir zarı atma deneyi düşünelim: Bu deneyde örneklem uzayı= $\{1, 2, 3, 4, 5, 6\}$ biçimindedir ve $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$'dır. Ayrıca, $P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$ olur.

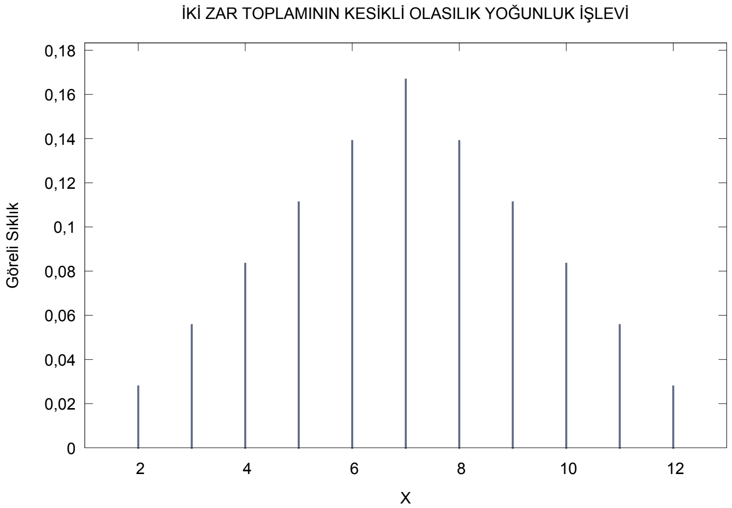
Kesikli Bir Değişkenin Olasılık Yoğunluk İşlevi
X değişkeni $x_1, x_2, x_3, \ldots$ gibi ayrık değerler alan bir rd olsun.

$$\begin{aligned} f(x) &= P(X = x_i) && i = 1, 2, \ldots, n \text{ için} \\ &= 0 && X \neq x_i \text{ için} \end{aligned}$$

işlevine X'e ait *"kesikli olasılık yoğunluk işlevi"* (discrete probability density function) denir.

- *Örnek:* İki zar atıldığında zarların toplam değerini gösteren kesikli rastsal değişken X, 11 farklı değer alabilir:

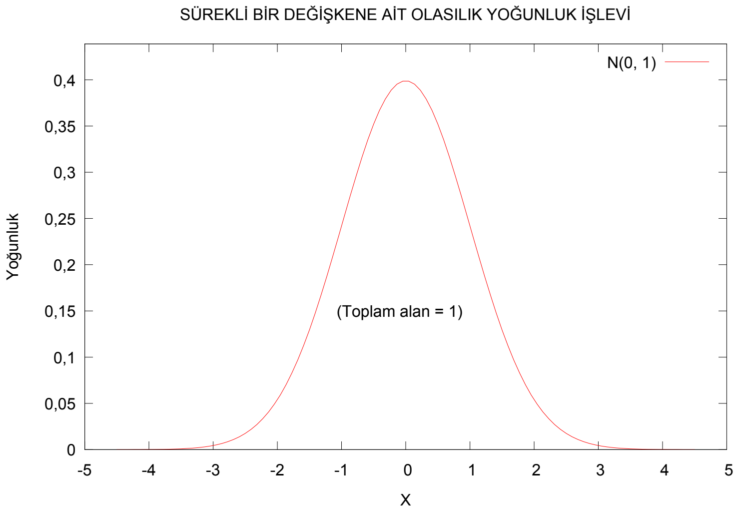
$$x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}\ f(x) = \{\tfrac{1}{36}, \tfrac{2}{36}, \tfrac{3}{36}, \tfrac{4}{36}, \tfrac{5}{36}, \tfrac{6}{36}, \tfrac{5}{36}, \tfrac{4}{36}, \tfrac{3}{36}, \tfrac{2}{36}, \tfrac{1}{36}\}$$



Sürekli Bir Değişkenin Olasılık Yoğunluk İşlevi
X sürekli bir rd olsun.

$$f(x) \geq 0,$$
$$\int_{-\infty}^{\infty} f(x)dx = 1,$$
$$\int_a^b f(x)dx = P(a \leq x \leq b)$$

Eğer yukarıdaki koşullar sağlanırsa, $f(x)$'e X'in *"sürekli olasılık yoğunluk işlevi"* (continuous probability density function) denir.



Birleşik Olasılık Yoğunluk İşlevi
X ve Y iki kesikli rd olsun.

$$
\begin{aligned}
f(x,y) &= P(X = x_i \wedge Y = y_j), \\
&= 0 \qquad\qquad\qquad\qquad X \neq x_i \wedge Y \neq y_j \text{ için}
\end{aligned}
$$

işlevi, *"kesikli birleşik olasılık yoğunluk işlevi"* (discrete joint probability density function) adını alır.

- Birleşik OYİ, X'in x_i değerini ve Y'nin de y_j değerini aynı anda almasının birleşik olasılığını gösterir.
- Aşağıdaki çizelgede X ve Y kesikli değişkenlerine ait bir birleşik OYİ gösterilmektedir:

		X		
		1	**2**	**3**
Y	**0**	0,2	0,3	0,1
	1	0,1	0,1	0,2

- Buna göre $X = 2$ değerini aldığında $Y = 0$ olma olasılığı $f(2,0) = 0{,}3$ ya da diğer bir deyişle %30'dur.
- Tüm olasılıklar toplamının 1 olduğuna dikkat ediniz.

Marjinal Olasılık Yoğunluk İşlevi

$f(x,y)$ birleşik OYİ'sine ilişkin olarak $f(x)$ ve $f(y)$ işlevlerine *"marjinal olasılık yoğunluk işlevi"* (marginal probability density function) adı verilir:

$$
\begin{aligned}
f(x) &= \textstyle\sum_y f(x,y) \quad && X\text{'in marjinal OYİ'si} \\
f(y) &= \textstyle\sum_x f(x,y) \quad && Y\text{'nin marjinal OYİ'si}
\end{aligned}
$$

- Önceki örnekteki verileri ele alalım. X'in marjinal OYİ'si:

$$
\begin{array}{rclclcl}
f(x=1) & = & \sum_y f(x=1,y) & = & 0{,}2+0{,}1 & = & 0{,}3 \\
f(x=2) & = & \sum_y f(x=2,y) & = & 0{,}3+0{,}1 & = & 0{,}4 \\
f(x=3) & = & \sum_y f(x=3,y) & = & 0{,}1+0{,}2 & = & 0{,}3 \\
 & & & & & + & \\
\hline
 & & & & & & 1{,}0
\end{array}
$$

- Aynı şekilde Y'nin marjinal OYİ'si de aşağıdaki gibidir:

$$
\begin{array}{rclclcl}
f(y=0) & = & \sum_x f(y=0,x) & = & 0{,}2+0{,}3+0{,}1 & = & 0{,}6 \\
f(y=1) & = & \sum_x f(y=1,x) & = & 0{,}1+0{,}1+0{,}2 & = & 0{,}4 \\
 & & & & & + & \\
\hline
 & & & & & & 1{,}0
\end{array}
$$

İstatistiksel Bağımsızlık

X ve Y rastsal değişkenlerinin ancak ve ancak

$$f(x, y) = f(x) \cdot f(y)$$

çarpımı olarak yazılabilmeleri durumunda bunlara *"istatistiksel bağımsız"* (statistically independent) değişkenler denir.

- Örnek olarak bir torbada üzerlerinde 1, 2, 3 yazılı üç top olduğunu düşünelim. Torbadan iki top (X ve Y) yerine koyularak çekilirse, X ve Y'nin birleşik OYİ'si şöyle olur:

$$\begin{array}{cc|ccc} & & & X & \\ & & 1 & 2 & 3 \\ \hline & 1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ Y & 2 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ & 3 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{array}$$

- Burada $f(x = 1, y = 1) = \frac{1}{9}$'dur.
- $f(x = 1) = \sum_y f(x = 1, y) \qquad = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{1}{3}$
- $f(y = 1) = \sum_x f(x, y = 1) \qquad = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{1}{3}$

- Bu örnekte $f(x, y) = f(x) \cdot f(y)$ olduğuna göre, bu iki değişken istatistiksel olarak bağımsızdır diyebiliriz.

1.2.2 Olasılık Dağılımlarının Beklemleri

- Matematikte, bir noktalar kümesinin nasıl bir şekil gösterdiğini anlatan sayısal ölçüye *"beklem"* (moment) denir.
- Dolayısıyla, bir olasılık dağılımı o dağılıma ait bir dizi beklem ile özetlenebilir.
- Beklemler, *"merkezi beklem"* (central moment) ve *"ham beklem"* (raw moment) olarak ikiye ayrılır.
- En yaygın kullanılan iki beklem ise *"ortalama"* (mean) (μ) ve *"varyans"* (variance) (σ^2) olarak karşımıza çıkar.
- Ortalama, aynı zamanda *"beklenen değer"* (expected value) olarak da adlandırılır.

Beklenen Değer

Kesikli bir rd olan X'e ait ortalama ya da beklenen değer $E(X)$ şöyle tanımlanır:

$$E(X) = \sum_x xf(x)$$

- Örnek olarak, iki zarın toplamını gösteren kesikli rd X'in olasılık dağılımını ele alalım:

$$E(X) = \sum_x x\, f(x) = 2\tfrac{1}{36} + 3\tfrac{2}{36} + 4\tfrac{3}{36} + \cdots + 11\tfrac{2}{36} + 12\tfrac{1}{36} = 7$$

- Demek ki iki zar atıldığında gözlenecek sayıların beklenen değeri 7'dir.

Beklenen değer kavramına ilişkin bazı özellikler şunlardır:

1. Sabit bir sayının beklenen değeri kendisidir. *Örnek:* Eğer $b = 2$ ise $E(b) = 2$'dir.
2. Eğer a ve b birer sabitse, $E(aX + b) = aE(X) + b$'dir.
3. Eğer X ve Y bağımsız rd ise, $E(XY) = E(X)E(Y)$'dir.
4. X, $f(X)$ olasılık yoğunluk işlevli bir rd ve $g(X)$ de X'in herhangi bir işleviyse, şu kural geçerlidir:

$$\begin{aligned} E[g(X)] &= \textstyle\sum_x \; g(X)f(x) && X \text{ kesikli ise,} \\ &= \textstyle\int_{-\infty}^{\infty} g(X)f(x)dx && X \text{ sürekli ise.} \end{aligned}$$

   Buna göre eğer $g(X) = X^2$ ise:

$$\begin{aligned} E(X^2) &= \textstyle\sum_x \; x^2 f(X) && X \text{ kesikli ise,} \\ &= \textstyle\int_{-\infty}^{\infty} x^2 f(X)dx && X \text{ sürekli ise.} \end{aligned}$$

- Örnek olarak, aşağıdaki OYİ'yi ele alalım:

$$\begin{array}{rcl} x & = & \{\text{-2}, \quad 1, \quad 2\} \\ f(x) & = & \{\tfrac{5}{8}, \quad \tfrac{1}{8}, \quad \tfrac{2}{8}\} \end{array}$$

- Buna göre X'in beklenen değeri şudur:

$$\begin{aligned} E(X) = \textstyle\sum_x xf(x) &= -2\tfrac{5}{8} + 1\tfrac{1}{8} + 2\tfrac{2}{8} \\ &= -\tfrac{5}{8} \end{aligned}$$

- Ayrıca X^2'nin beklenen değeri ise şudur:

$$\begin{aligned} E(X^2) = \sum_x x^2 f(x) &= 4\tfrac{5}{8} + 1\tfrac{1}{8} + 4\tfrac{2}{8} \\ &= \tfrac{29}{8} \end{aligned}$$

Varyans (Değişirlik)
X bir rd ve $E(X) = \mu$ ise, X değerlerinin beklenen değerleri etrafındaki yayılımı *"varyans"* (variance) ile ölçülür:

$$\begin{aligned} \text{var}(X) = \sigma_X^2 &= \textstyle\sum_x \; (X-\mu)^2 f(x) && X \text{ kesikli ise,} \\ &= \textstyle\int_{-\infty}^{\infty} (X-\mu)^2 f(x)dx && X \text{ sürekli ise.} \end{aligned}$$

- σ_X^2'nin artı değerli kare kökü σ_X, X'e ait *"ölçünlü sapma"* (standard deviation) olarak adlandırılır.
- Varyans ve ölçünlü sapma, her bir rastsal x değerinin X'in ortalaması etrafında ne genişlikte bir alana yayıldığının göstergesidir.

Varyans kavramına ilişkin bazı özellikler şunlardır:

1. Sabit bir sayının varyansı sıfırdır.
2. Eğer a ve b birer sabitse, $\text{var}(aX + b) = a^2\text{var}(X)$'dir.
3. Eğer X ve Y bağımsız birer rd ise şu yazılabilir:

$$\begin{aligned} \text{var}(X+Y) &= \text{var}(X) + \text{var}(Y) \\ \text{var}(X-Y) &= \text{var}(X) + \text{var}(Y) \end{aligned}$$

4. Eğer X ve Y bağımsız birer rd ve a, b, c de birer sabit ise, aşağıdaki kural geçerlidir:

$$\text{var}(aX + bY + c) = a^2\text{var}(X) + b^2\text{var}(Y)$$

- Hesaplama kolaylığı bakımından varyans formülü şöyle de yazılabilir:

$$\begin{aligned} \text{var}(X) = \sigma_X^2 &= (1/n)\textstyle\sum((X_i - E(X))^2) \\ &= (1/n)\textstyle\sum(X_i^2 - 2X_iE(X) + E(X)^2) \\ &= \textstyle\sum(X_i^2)/n - \sum 2X_iE(X)/n + \sum E(X)^2/n \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ \\ &= E(X^2) - E(X)^2 \end{aligned}$$

- Buna göre önceki örnekteki rastsal değişkenin varyansı şudur:

$$\text{var}(X) = \frac{29}{8} - \left(-\frac{5}{8}\right)^2 = \frac{207}{64}$$

Kovaryans (Eşdeğişirlik)
X ve Y rd'lerinin ortalamaları sırasıyla $E(X)$ ve $E(Y)$ olsun. Bu iki değişkenin birlikte değişirlikleri *"kovaryans"* (covariance) ile ölçülür:

$$\begin{aligned}\text{cov}(X,Y) &= \textstyle\sum_y \sum_x XY\, f(x,y) - E(X)E(Y) && \text{kesikliyse,}\\ &= \textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY\, f(x,y)\, dxdy - E(X)E(Y) && \text{sürekliyse.}\end{aligned}$$

- Kovaryans formülü şöyle de gösterilebilir: $\text{cov}(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$
- Görüldüğü gibi bir değişkenin varyansı aynı zamanda kendisiyle olan kovaryansıdır.

Kovaryans kavramına ilişkin birkaç önemli özellik şunlardır:

1. Eğer X ve Y bağımsız rd'ler ise kovaryansları 0 olur:

$$\begin{aligned}\text{cov}(X,Y) &= E(XY) - E(X)E(Y)\\ &= E(X)E(Y) - E(X)E(Y) = 0\end{aligned}$$

2. Eğer a, b, c, d birer sabitse şu kural geçerlidir:

$$\text{cov}(a + bX, c + dY) = bd\,\text{cov}(X,Y)$$

3. Bağımsız olmayan X ve Y rd'lerinin bileşimlerinin varyanslarını hesaplarken kovaryans bilgisi de gereklidir:

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\text{cov}(X,Y)$$

İlinti Katsayısı
"İlinti katsayısı" (correlation coefficient) iki rd arasındaki doğrusal ilişkinin bir ölçüsüdür ve $[-1, 1]$ değerleri arasında yer alır:

$$\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\text{cov}(X,Y)}{\sigma_x\sigma_y}.$$

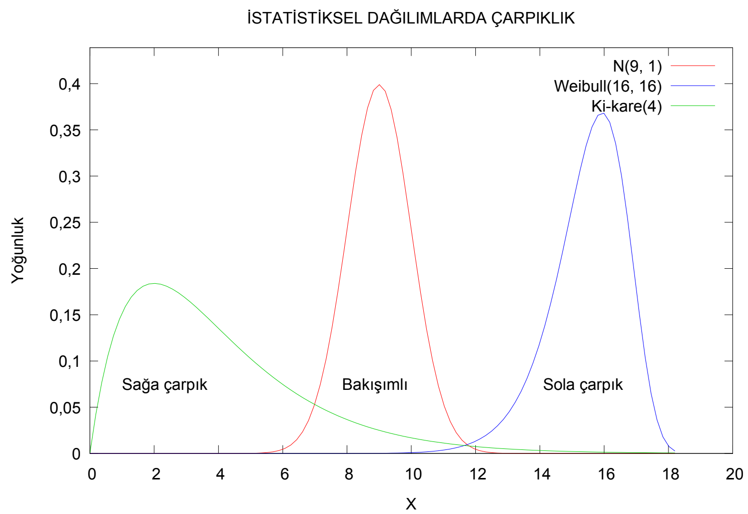
- Yukarıdaki formülden şu görülebilir: $\text{cov}(X,Y) = \rho\sigma_x\sigma_y$

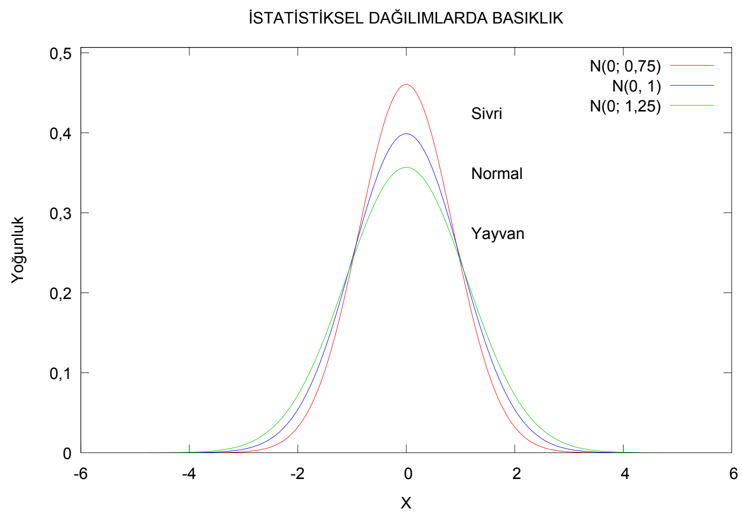
Diğer Merkezi Beklemler

- Genel olarak, $f(x)$ tek değişkenli OYİ'sinin kendi ortalaması dolayındaki merkezi beklemleri şöyle tanımlanır:

Beklem	Tanım	Açıklama
1	$E(X-\mu)$	0
2	$E(X-\mu)^2$	varyans
3	$E(X-\mu)^3$	çarpıklık
4	$E(X-\mu)^4$	basıklık
$\vdots$	$\vdots$	$\vdots$
n	$E(X-\mu)^n$	n. derece

- *"Çarpıklık"* (skewness), bakışımdan uzaklığı ölçer.
- *"Basıklık"* (kurtosis), yayvanlığı incelemek için kullanılır.
- Bir rastsal değişkenin normal dağılıma uyup uymadığını anlamak için çarpıklık ve basıklık değerlerine bakılabilir.





1.2.3 Bazı Kuramsal Olasılık Dağılımları

Normal Dağılım
Ortalaması ve varyansı sırasıyla μ ve σ^2 olan *"normal dağılım"* (normal distribution) aşağıdaki OYİ ile gösterilir:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \qquad -\infty \leq x \leq \infty$$

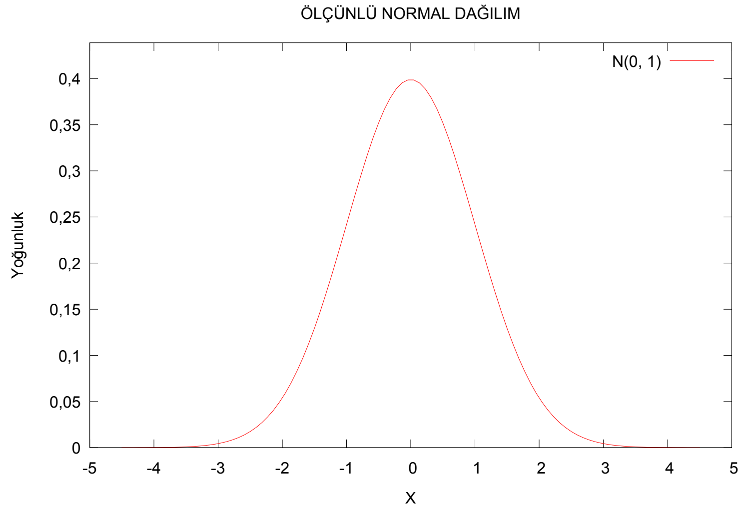
- Normal dağılan bir rd, $X \sim N(\mu, \sigma^2)$ şeklinde gösterilir.
- Normal eğri altında kalan alanın yaklaşık yüzde 68'i $\mu \pm \sigma$ değerleri, yüzde 95 kadarı $\mu \pm 2\sigma$ değerleri ve yüzde 99,7 kadarı da $\mu \pm 3\sigma$ değerleri arasında yer alır.

Ölçünlü Normal Dağılım
"Ölçünlü normal dağılım" (standard normal distribution) için $\mu = 0$, $\sigma^2 = 1$'dir ve $X \sim N(0, 1)$ diye gösterilir. OYİ'si şudur:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Z^2\right), \qquad Z = \frac{x-\mu}{\sigma}$$

- Formülde görülen $\exp$ işlemcisi, e üzeri anlamına gelir.
- μ ve σ^2 değerleri verili ve normal dağılan X rd'si, $Z = \frac{x-\mu}{\sigma}$ formülü ile ölçünlü normal değişken Z'ye dönüştürülür.

- *Örnek:* $X \sim N(8,4)$ olsun. X'in $[6,12]$ arası değerler alma olasılığı için $Z_1 = \frac{6-8}{2} = -1$ ve $Z_2 = \frac{12-8}{2} = 2$'dir. Çizelgeden $P(0 \leq Z \leq 2) = 0{,}4772$ olduğunu görürüz. Bakışım nedeniyle $P(-1 \leq Z \leq 0) = 0{,}3413$ bulunur. Demek ki istenilen olasılık $0{,}3413 + 0{,}4772 = 0{,}8185$'tir.



Normal dağılıma ilişkin bazı özellikler şunlardır:

1. Normal dağılımın 3. ve 4. merkezi beklemleri şöyledir:

   3. merkezi beklem: $E(X-\mu)^3 = 0$
   4. merkezi beklem: $E(X-\mu)^4 = 3\sigma^4$

   Buna göre, ölçünlü normal dağılımın basıklığı 3'tür. Ayrıca çarpıklığı 0 olduğu için *"bakışımlı"* (symmetric) olur.

2. Normal dağılan bir rd'nin tek sayılı tüm beklemleri sıfırdır.

3. Normal rd'lerin doğrusal bileşimleri de normal dağılır. *Örnek:* $X_1 \sim N(\mu_1, \sigma_1^2)$ ve $X_2 \sim N(\mu_2, \sigma_2^2)$ iki bağımsız rd olsun. Eğer $Y = aX_1 + bX_2$ ise,

$$Y \sim N\left[(a\mu_1 + b\mu_2), (a^2\sigma_1^2 + b^2\sigma_2^2)\right] \text{ olur.}$$

- Normal dağılıma ilişkin önemli bir nokta da *"Merkezi limit kanıtsavı"* (central limit theorem) ya da kısaca *"MLK"* (CLT) konusudur.

- Merkezi limit kanıtsavı günümüz olasılık kuramının yapı taşlarından biridir.

- MLK'yi kısaca açıklamak için, bağımsız ve benzer şekilde dağılan (ortalama $= \mu$, varyans $= \sigma^2$) n sayıda $X_1, \ldots, X_n$ rastsal değişken varsayalım.
- Kanıtsava göre bu rd'ler, n sonsuza giderken ortalaması μ ve varyansı da σ^2/n olan normal dağılıma yakınsarlar.
- Başlangıçtaki OYİ ne olursa olsun bu sonuç geçerlidir.

χ^2 (Ki-Kare) Dağılımı

$Z_1, Z_2, Z_3, \ldots, Z_k$, k sayıda ölçünlü normal değişken olsun. Bu durumda

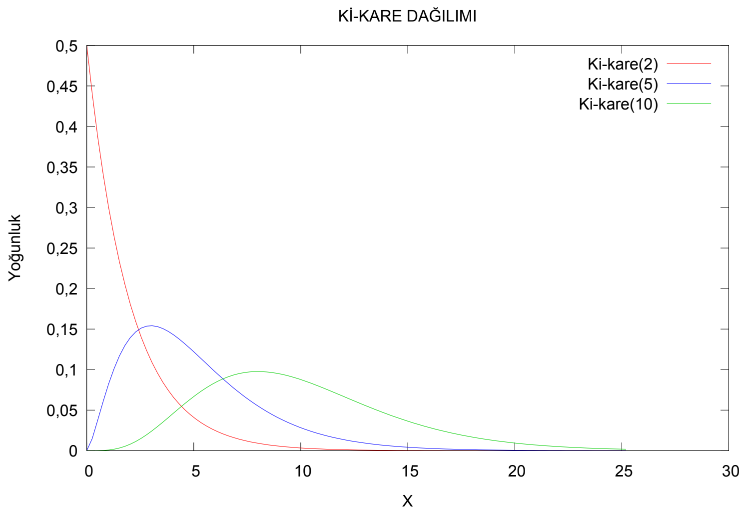
$$\chi^2 = \sum_{i=1}^{k} Z_i^2$$

rastsal değişkeni, χ^2 şeklinde gösterilen *"ki-kare"* (chi-square) dağılımına uyar.

- Buradaki k değeri, ki-kare değişkenine ait *"serbestlik derecesi"* (degrees of freedom) ya da kısaca *"sd"* (df) olarak tanımlanır.

Ki-kare dağılımına ilişkin bazı özellikler şunlardır:

1. Ki-kare, *"sağa çarpık"* (right-skewed) bir dağılımdır ancak serbestlik derecesi arttıkça bakışıma yaklaşır.
2. k sd'li bir χ^2 dağılımının ortalaması k, varyansı ise $2k$'dir.
3. Eğer Z_1 ve Z_2 iki bağımsız dağılan ki-kare değişkeniyse, $Z_1 + Z_2$ toplamı da sd $= k_1 + k_2$ olan bir χ^2 değişkeni olur.



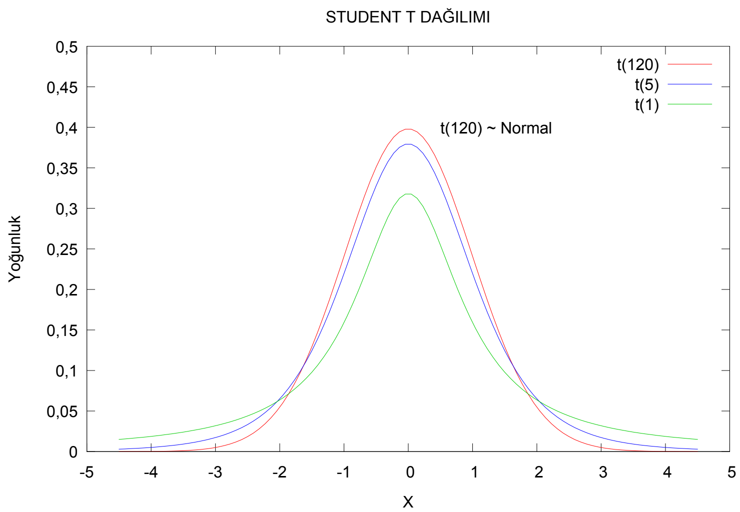
Student T Dağılımı

Z_1 bir ölçünlü normal değişken ve Z_2 de Z_1'den bağımsız bir ki-kare değişkeni olsun. Bu durumda:

$$t = \frac{Z_1}{\sqrt{Z_2/k}}$$

değişkeni, k sd ile *"Student t"* (Student's t) dağılımına uyar.

- Neredeyse tüm çalışmalarını "Student" takma adı ile yazmış olan istatistikçi William Sealy Gosset (1876-1937) tarafından bulunmuştur.
- t dağılımı da normal dağılım gibi bakışımlı ancak daha basıktır. Sd'si yükseldikçe normal dağılıma yakınsar.
- Ortalaması 0, varyansı ise $k > 2$ için $k/(k-2)$'dir.



Fisher-Snedecor F Dağılımı

Z_1 ve Z_2, k_1 ve k_2 sd'li bağımsız iki ki-kare değişkeni olsun. Bu durumda:

$$F = \frac{Z_1/k_1}{Z_2/k_2},$$

k_1 ve k_2 sd'li bir *"F dağılımı"* (F distribution) biçiminde dağılır.

F dağılımına ilişkin bazı özellikler ise şunlardır:

1. Ki-kare dağılımı gibi F dağılımı da sağa çarpıktır ama k_1 ve k_2 büyüdükçe F dağılımı da normale yakınsar.

2. $k_2 > 2$ için F dağılımının ortalaması şöyledir:

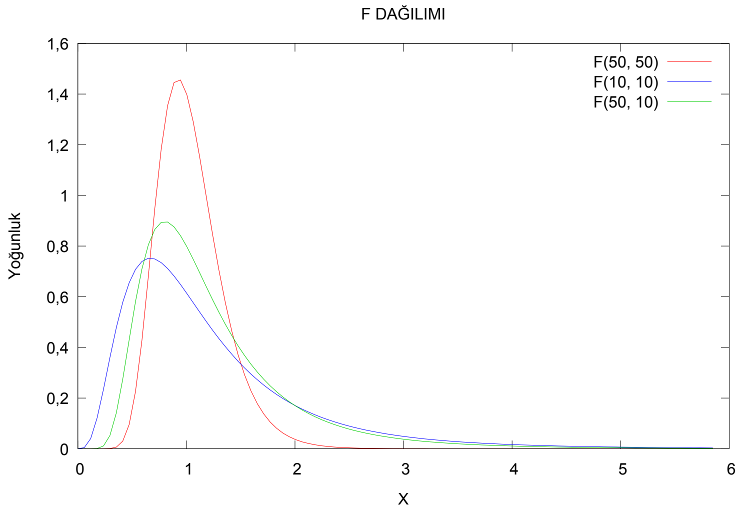
$$\mu = \frac{k_2}{(k_2-2)}$$

3. $k_2 > 4$ için F dağılımının varyansı şöyledir:

$$\sigma^2 = \frac{2k_2^2(k_1+k_2-2)}{k_1(k_1-2)^2(k_2-4)}$$

4. F ile t dağılımları arasında şu ilişki vardır: $t_k^2 = F_{1,k}$

5. Eğer payda sd'si k_2 yeterince büyükse F ve ki-kare dağılımları arasında şu ilişki vardır: $k_1 F_{k_1,k_2} \sim \chi^2_{k_1}$



1.3 İstatistiksel Çıkarsama

1.3.1 Tahmin Sorunu

- İstatistikte bilinmeyenleri tahmin etmenin genel yolu, bilinen bir olasılık dağılımından çekilen n boyutundaki rastsal örneklem verilerini kullanmaktır.
- X, OYİ'si $f(x;\theta)$ olan bir rastsal değişken olsun.
- Burada θ, dağılıma ait herhangi bir anakütle katsayısıdır.
- Rastsal bir örneklem çekilip şöyle bir örneklem değerleri işlevi geliştirilebilir: $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$
- Bize θ'nın bir tahminini veren $\hat{\theta}$'ya *"istatistik"* (statistic) ya da *"tahminci"* (estimator) denir ve *"teta şapka"* (theta hat) diye okunur.
- *"Tahmin"* (estimation) denilen bu süreç iki bölüme ayrılır:

  "Nokta tahmini" (point estimation) *"Aralık tahmini"* (interval estimation)

Nokta Tahmini ve Aralık Tahmini

- Nokta tahmini, θ'nın tahminini tek bir değer olarak verir.
- *Örnek:* Eğer $\hat{\theta} = 20$ ise bu θ'nın nokta tahminidir.
- *"En küçük kareler"* (least squares) ve *"ençok olabilirlik"* (maximum likelihood) yöntemleri en yaygın kullanılan iki nokta tahmincisidir.
- Aralık tahmini ise öncelikle θ için $\hat{\theta}_1 = f(x_1, x_2, \ldots, x_n)$ ve $\hat{\theta}_2 = f(x_1, x_2, \ldots, x_n)$ gibi iki tahminci tanımlar.
- Daha sonra, gerçek θ değerinin belli bir güvenle (olasılıkla) bulunduğu $[\hat{\theta}_1, \hat{\theta}_2]$ aralığı tahmin edilir.
- *Örnek:* θ'nın %95 güven aralığı şu olabilir: $19 \leq \theta \leq 21$
- Böyle bir aralığın θ'yı içerdiği kesin olarak bilinemez. Belirlenen aralığın θ'yı içerme olasılığı ya 0'dır ya da 1'dir.
- Öyleyse, bu aralığın yorumu şudur: Eğer böyle 100 aralık hesaplanırsa, bunlardan 95'i aslında değeri bilinemeyen gerçek θ'yı içermelidir.

Arzulanan İstatistiksel Özellikler

- En küçük kareler ve ençok olabilirlik gibi tahmincilerde *"arzulanan"* (desired) bir takım istatistiksel özellikler vardır.
- Bunları iki kümede inceleyebiliriz:

"küçük örneklem özellikleri" (small sample properties)
"kavuşmazsal özellikler" (asymptotic properties)

- Küçük örneklem özellikleri, tahmincinin sınırlı sayıda gözlemden oluşan örneklemlerde taşıdığı özelliklerdir.
- Tahmincinin kavuşmazsal ya da büyük örneklem özellikleri ise örneklem büyüklüğü sonsuza yaklaştıkça gözlenir.

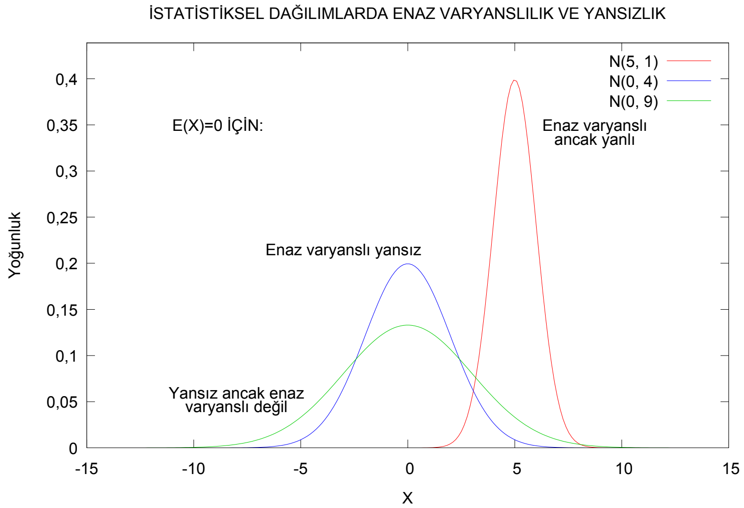
Yansızlık
Eğer $\hat{\theta}$ gibi bir tahmincinin beklenen değeri gerçek θ'ya eşitse, bu tahminciye θ'nın *"yansız"* (unbiased) tahmincisi denir:

$$E(\hat{\theta}) = \theta \qquad \text{ya da} \qquad E(\hat{\theta}) - \theta = 0$$

- Kuramsal olarak yansızlık, aynı büyüklükte farklı farklı örneklemler çekilip de katsayı tahmini yapılabilirse, bu tahminlerin ortalamasının giderek anakütledeki gerçek değere yaklaşacağı anlamına gelir.
- Bu durumda yansızlık bir *"tekrarlı örnekleme"* (repeated sampling) özelliğidir.

Enaz Varyanslı Tahminci
$\hat{\theta}_1$'in varyansı; θ'ya ilişkin $\hat{\theta}_2, \hat{\theta}_3, \ldots$ gibi diğer tahmincilerin varyansından küçük ya da ona eşit olsun. Bu durumda, $\hat{\theta}_1$'ya *"enaz varyanslı tahminci"* (minimum variance estimator) denir.

Enaz Varyanslı Yansız Tahminci
$\hat{\theta}_1$ ve $\hat{\theta}_2$, θ'nın iki yansız tahmincisi olsun. Eğer $\hat{\theta}_1$'nın varyansı $\hat{\theta}_2$'nın varyansından küçük ya da ona eşitse $\hat{\theta}_1$ tahmincisine *"enaz varyanslı yansız"* (minimum variance unbiased) ya da *"en iyi yansız"* (best unbiased) ya da *"etkin"* (efficient) tahminci denir.



Kavuşmazsal Yansızlık

n gözlemli bir örneklem için $\hat{\theta}_n$ tahmincisinin *"kavuşmazsal yansız"* (asymptotically unbiased) bir tahminci olabilmesi için θ'nın şu koşulu sağlaması gereklidir:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

- Diğer bir deyişle, örneklem büyüklüğü artarken eğer $\hat{\theta}$'nın beklenen ya da ortalama değeri gerçek θ'ya yakınsıyorsa, $\hat{\theta}$ tahmincisi kavuşmazsal yansızdır.

Tutarlılık

Örneklem büyüklüğü n artarken $\hat{\theta}$ tahmincisi θ'ya yakınsıyorsa, $\hat{\theta}$'ya *"tutarlı"* (consistent) tahminci denir.

- Diğer bir deyişle, tutarlı tahmincilerde n büyürken $\hat{\theta}$'nın beklenen değeri gerçek θ'ya yaklaşır ve aynı zamanda varyansı da küçülür.
- *Dikkat:* Yansızlık ve tutarlılık özellikleri kavramsal olarak çok farklıdır. Tutarlılık yalnızca kavuşmazsal bir özelliktir.
- Tutarlılığın yeterli koşulu örneklem sonsuza yaklaşırken hem yanlılığın hem de varyansın sıfıra doğru gitmesidir.
- $\hat{\theta}$ tahmincisinin kavuşmazsal dağılımının varyansına, $\hat{\theta}$'ya ait *"kavuşmazsal varyans"* (asymptotic variance) denir.

Kavuşmazsal Etkinlik
Eğer $\hat{\theta}$ tutarlıysa ve $\hat{\theta}$'nın kavuşmazsal varyansı diğer tüm tahmincilerin kavuşmazsal varyanslarından küçükse, $\hat{\theta}$'ya *"kavuşmazsal etkin"* (asymptotically efficient) tahminci denir.

Kavuşmazsal Normallik
Örneklem büyürken eğer $\hat{\theta}$ tahmincisinin örneklem dağılımı da normal dağılıma yakınsıyorsa, bu tahmincinin *"kavuşmazsal normal"* (asymptotically normal) dağıldığı söylenir.

- Kavuşmazsal normallik özelliği, merkezi limit kanıtsavının bir sonucudur.

Doğrusallık
$\hat{\theta}$ tahmincisi eğer örneklem gözlemlerinin doğrusal bir işlevi ise, buna θ'nın *"doğrusal"* (linear) tahmincisi denir. Örnek olarak:

$$\hat{\theta} = (ax_1 + bx_2 + cx_3 + \dots) \qquad \{a, b, c, \dots\} \in R$$

tahmincisi θ'nın doğrusal bir tahmincisidir.

En iyi Doğrusal Yansız Tahminci
$\hat{\theta}$ eğer θ'nın farklı doğrusal tahmincileri arasında yansız ve enaz varyanslı tahminciyse, $\hat{\theta}$'ya *"en iyi doğrusal yansız tahminci"* (best linear unbiased estimator), kısaca *"EDYT"* (BLUE) denir.

1.3.2 Önsav Sınaması

Önsav sınaması konusu aşağıdaki gibi özetlenebilir:

- X, OYİ'si $f(x;\theta)$ bilinen bir rastsal değişken olsun.
- Burada θ, dağılımın herhangi bir anakütle katsayısıdır.
- Genellikle gerçek θ bilinemez ancak tahmin edilebilir.
- n büyüklüğünde bir rastsal örneklem çekilerek $\hat{\theta}$ tahmincisi bulunmuş olsun.
- Önsav sınaması yöntemi kullanılarak, anakütle katsayısı θ'nın varsayılan bir θ^* değeriyle uyumluluğu sınanabilir.
- Bunun için, eldeki $\hat{\theta}$ tahmini ve bu tahminin olasılık dağılımı ile ilgili bilgi ya da varsayımlardan yararlanılır.

Sıfır Önsavı ve Almaşık Önsav

- Anakütle katsayısı θ'nın seçili bir θ^* değerine eşit olup olmadığı sınanmak isteniyor olsun.
- Bu durumda, $\theta = \theta^*$ savına *"sıfır önsavı"* (null hypothesis) adı verilir ve $H_0 : \theta = \theta^*$ ile gösterilir.
- Bu sıfır önsavı, $H_1 : \theta \neq \theta^*$ ile gösterilen *"almaşık önsav"* (alternative hypothesis) savına karşı sınanır.

I. ve II. Tür Hatalar

- Sınama sonuçları değerlendirilirken dikkatli olunmalıdır.
- Sınama sonucu bir olasılık değeri olacağı için hatalı bir karara varılması olasıdır.
- Eğer H_0 aslında doğruyken reddedilirse, buna *"I. tür hata"* (type I error) denir.
- Eğer H_0 aslında yanlışken reddedilmezse, buna da *"II. tür hata"* (type II error) denir.

Çizelge: I. ve II. Tür Hatalar

	Gerçek Durum	
Karar	H_0 Doğru	H_0 Yanlış
H_0 Reddedilir	I. tür hata	Hata yok
H_0 Reddedilmez	Hata yok	II. tür hata

Anlamlılık Düzeyi

- Yazında I. tür hata olasılığı α ile gösterilir ve *"anlamlılık düzeyi"* (significance level) adıyla anılır.
- Önsav sınamasına klasik yaklaşım I. tür hatanın II. türe göre daha ciddi olduğudur.
- Dolayısıyla, uygulamada α 0,01 ya da 0,05 gibi düşük bir düzeyde tutularak I. tür hata yapma olasılığı azaltılır.
- $(1 - \alpha)$ değeri I. tür hatayı yapmama olasılığını gösterdiği için buna *"güven katsayısı"* (confidence coefficient) denir.
- Örnek olarak, eğer anlamlılık düzeyi $\alpha = 0{,}05$ olarak seçilmişse, güven katsayısı $(1 - \alpha) = 0{,}95$ ya da %95 olur.

Anlamlılık Sınaması ve Güven Aralığı

- Önsav sınamasına iki farklı yaklaşım vardır:

  "güven aralığı" (confidence interval)
  "anlamlılık sınaması" (test of significance)

- Güven aralığı yaklaşımında, anakütle katsayısı θ için tahmin edilen $\hat{\theta}$'ya dayanan bir $\%100(1-\alpha)$ aralığı kurulur ve bunun $\theta = \theta^*$ değerini içerip içermediğine bakılır.
- Eğer bulunan güven aralığı θ^*'ı içeriyorsa sıfır önsavı reddedilmez, içermiyorsa reddedilir.
- Anlamlılık sınaması yaklaşımında ise $\theta = \theta^*$ varsayımına ilişkin bir sınama istatistiği hesaplanır ve bu istatistiği elde etme olasılığının ne olduğuna bakılır.
- Eğer bu olasılık seçilen α değerinden küçükse sıfır önsavı reddedilir, büyükse reddedilmez.
- Belli bir uygulamada bu iki yaklaşım aynı sonucu verir.

Önsav Sınaması Özet

İstatistiksel bir önsavın sınanmasının adımları kısaca şöyledir:

1. Bir sınama istatistiği alınır. *Örnek:* $\bar{X}$
2. Sınama istatistiğinin olasılık dağılımı belirlenir. *Örnek:* $\bar{X} \sim N(\mu, \sigma^2/2)$
3. Sıfır önsavı ve almaşık önsav belirtilir. *Örnek:* $H_0 : \mu = 75, \qquad H_1 : \mu \neq 75$
4. Anlamlılık düzeyi α seçilir. *Örnek:* $\alpha = 0{,}05$
5. Sınama istatistiğinin olasılık dağılımından bir $\%100(1-\alpha)$ güven aralığı kurulur ya da sıfır önsavına ilişkin istatistik hesaplanarak bunu elde etmenin olasılığına bakılır.
6. Elde edilen sonuçlara göre sıfır önsavı reddedilir ya da reddedilmez. Karar verilirken her 100 deneyde 100α kez yanlış sonuç bulma riski olduğu unutulmaz.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Appendix A* “A Review of Some Statistical Concepts” okunacak.

Önümüzdeki Ders
Ekonometri Nedir?

Bölüm 2

Ekonometri Nedir?

2.1 Ekonometri Nedir?

2.1.1 Ekonometrinin Konusu

Ekonometri
Sözcük anlamı ile ekonometri, *ekonomik ölçüm* demektir. Matematiksel araç ve istatistiksel hesaplama yöntemlerinin ekonomi kuramı ile birleştirildiği uygulamalı bir bilim dalıdır.

Neden Ayrı Bir Bilim Dalı?
Ekonometri; kuramsal iktisat, matematiksel iktisat ve iktisadi istatistikten ayrı bir bilim dalıdır çünkü:

- İktisat kuramı çoğunlukla nitel söylem ya da önsavlar öne sürer. İki değişken arasındaki ilişkinin yönüne odaklanırken ilişkinin sayısal ölçümünü vermez. Ekonometri ise iktisat kuramına *"görgül"* (empirical) bir içerik katar.

- Matematiksel iktisatın görevi iktisat kuramını matematiksel kalıplar içerisine sokmaktır. Diğer yandan matematiksel eşitliklerin ekonometrik eşitlikler haline sokulması ve doğrulanması ile ekonometri ilgilenir.

- İktisadi istatistiğin ilgi alanı iktisadi verileri derlemek ve işlemektir. Toplanan bu verilerin kullanılarak iktisadi kuramların doğrulanması ve sayısal kestirimler yapılması işini ise ekonometri üstlenir.

Ekonometrinin İlgilendiği Konular
Ekonometri temel olarak üç konu ile ilgilenir:

1. Ekonomik ilişkilerin incelenerek değerlendirilmesi,
2. Ekonomik davranışla ilgili çeşitli önsavların sınanması,
3. Ekonomik büyüklüklere yönelik *"yordama"* (forecast) yapılması.

Ekonomik İlişkilerin İncelenmesi Konusu

Verilere dayanarak ekonomik ilişkilerin incelenmesi ve değerlendirilmesi ile ilgili olarak aşağıdaki örnekler verilebilir:

- Kamu ya da özel sektör tarafından piyasada var olan çeşitli mal ve hizmetlere olan arz ve talebin ölçülmesi.
- Özel bir firma tarafından yapılan reklam harcamalarının satış ve kazanç üzerindeki etkisinin incelenmesi.
- Bir şirketin hisse senetlerinin değeri ile o şirketin ekonomik performans göstergeleri arasında ilişki kurulması.
- Devletin uyguladığı çeşitli maliye ve para politikalarının toplam gelir, işsizlik, faiz oranları, ithalat ve ihracat gibi önemli değişkenler üzerindeki etkilerinin çözümlenmesi.
- Bir belediyenin o bölgede iş yapmakta olan bir şirketin konut, işgücü gibi değişkenler ve okul, elektrik gibi kamu hizmetlerine olan talebi nasıl etkilediğini araştırması.

Ekonomik Önsavların Sınanması Konusu

Ekonomik davranışla ilgili çeşitli önsavların sınanmasına ilişkin olarak aşağıdaki örnekler verilebilir:

- Ticari bir şirketin yeni başlattığı reklam kampanyasının satışları artırıp artırmadığına karar verilmesi.
- Bir mal ya da hizmete olan talebin fiyat ya da gelir yönünden esnek olup olmadığının incelenmesi.
- Verili bir üretim ölçeğine göre getirinin artan düzeyde mi yoksa azalan düzeyde mi olduğunun saptanması.
- Kamuya ait çeşitli ekonomi politikalarının kısa ya da uzun dönemde etkili olup olmadığının sınanması.
- Yeni bir yasal düzenlemenin toplum ve ekonomi üzerinde kayda değer bir etkisinin olup olmadığına karar verilmesi.

Ekonomik Değişkenlerin Yordanması Konusu

Ekonomik değişkenlerin yordanmasına ilişkin olarak aşağıdaki örnekler verilebilir:

- Ticari firmaların satış, kar, maliyet ve envanter gibi bilgilerini düzenli olarak yordamaları.
- Gelecekteki enerji talebinin tahmin edilerek gereksinim duyulan yatırımlara bugünden karar verilmesi.
- Borsa endekslerinin ve çeşitli firmaların hisse senedi fiyatlarının izleyeceği yönün kestirilmesi.
- Devlet ve kamu kuruluşlarının kısa ve uzun vadede işsizlik, enflasyon, vergi gelirleri, kamu harcamaları ve bütçe açığı gibi konularda geleceğe yönelik tahminde bulunmaları.
- Belediyelerin düzenli olarak nüfus, istihdam, kirlilik, yol, polis ve okul ihtiyacı gibi değişkenleri yordamaları.

Ekonometrinin Uygulama Alanları

Ekonometrik yöntemler günümüzde ekonomi dışında birçok alanda yaygın şekilde kullanılmaktadır. Bunlara örnek olarak aşağıdakiler gösterilebilir:

- Siyaset bilimi
- Tarih
- Sosyoloji
- Psikoloji
- Biyoloji
- Tıp
- Finans
- Eğitim
- Pazarlama

2.1.2 Ekonometrinin Yöntembilimi

Ekonometrik konuların incelenmesi genellikle rastsal örneklem verileri üzerinde *"çözümleme"* (analysis) yapılmasına dayanır. Bu nedenle tüm ekonometrik çalışmalarda bir hata etmeni de bulunur.

- Ölçülmüş olan ilişkilerin kesin olmaması,
- Önsav sınama sonuçlarının hatalı olabilmesi,
- Tahminlerin tutturulamaması

gibi riskler karşısında ekonometriciler genellikle değişkenler arasında birden fazla ilişkiyi incelerler. Daha sonra, çeşitli sınamalar kullanılarak gerçek davranışın hangi ilişki tarafından en iyi şekilde açıklandığına karar verilir.

Ekonometri Yöntembiliminin Ana Çizgileri

Bir iktisadi sorunun çözümlenmesinde ekonometrinin izlediği *"yöntembilim"* (methodology) şu şekilde özetlenebilir:

1. Kuramın ya da önsavın ortaya konulması
2. Kuramın matematiksel modelinin kurulması
3. Kuramın ekonometrik modelinin belirtimi
4. Verilerin elde edilmesi
5. Ekonometrik modelin katsayı tahmini
6. Çeşitli önsav sınamalarının yapılması
7. Çıkarsama ve yordama
8. Modelin denetim ya da politika amaçlı kullanılması

2.1.3 Uygulama: Keynesçi Tüketim Kuramı

Ekonometrinin 8 adımla özetlemiş olduğumuz yöntembilimini gösterebilmek için, Keynes'in ünlü tüketim kuramını ele alalım.

Adım 1: Kuramın ortaya konulması

Öncelikle incelenecek konu üzerinde düşünülmeli ve bu konuda iktisat kuramının neler söylediği dikkatlice gözden geçirilmelidir.

- Keynes der ki:

  > "Temel psikolojik yasa ... insanlar gelirleri arttıkça, kural olarak ve ortalama olarak, tüketimlerini artırma eğilimindedirler. Yalnız bu artış gelirlerindeki artış kadar olmaz."

- Kısaca Keynes, gelirdeki 1 birimlik artışa karşılık tüketimde görülen değişikliği ölçen marjinal tüketim eğilimi MTüE'nin 0'dan büyük ve 1'den küçük bir değer alacağını söylemiştir.

Adım 2: Matematiksel modelin belirtilmesi
İktisadi ilişki matematiksel olarak anlatılabilmelidir.

- Keynes, tüketim ile gelir arasında aynı yönlü bir ilişki öne sürmüşse de bu ilişkinin işlev biçimini açıkça belirtmemiştir.

- Bunun için şu tek denklemli matematiksel model önerilebilir:

$$C = \beta_1 + \beta_2 Y, \qquad 0 < \beta_2 < 1$$

- Burada
  C (tüketim) *"bağımlı değişken"* (dependent variable),
  Y (gelir) *"açıklayıcı değişken"* (explanatory variable),
  β_1 *"sabit terim"* (constant term),
  β_2 ise *"eğim değiştirgesi"* (slope parameter)
  olarak tanımlanır.

Adım 3: Ekonometrik modelin belirtilmesi
İktisadi değişkenler arasındaki ilişkiler kesin olmadığı için rastlantısallığı dikkate alan bir ekonometrik model belirtilmelidir.

- Gelirin yanı sıra aile büyüklüğü, yaş, eğitim gibi etmenlerin de tüketim harcamalarını etkileyebildiğini biliyoruz.

- Kesin olmayan ilişkileri dikkate alabilmek için, ekonometrik model aşağıdaki gibi tanımlanır:

$$C = \beta_1 + \beta_2 Y + \epsilon$$

- Buradaki "ϵ" terimi tüketimi etkileyen ama gözlenemeyen, ölçülemeyen ya da basitlik ilkesi nedeniyle açıkça dikkate alınmayan diğer tüm etmenleri temsil eder.

- *"Hata terimi"* (error term) olarak da bilinen ϵ, iyi tanımlı olasılıksal özellikleri olan bir rastsal değişkendir.

Adım 4: Verilerin Elde Edilmesi
Ekonometrik modelin tahmin edilebilmesi için ilgili verilerin elde bulunması gereklidir.

- 1987–2006 yılları arasında Türkiye'deki toplam tüketim ve gayrisafi yurtiçi hasıla (1987 fiyatları, milyon TL) şöyledir:

Çizelge: Türkiye'de Tüketim ve GSYH (1987–2006)

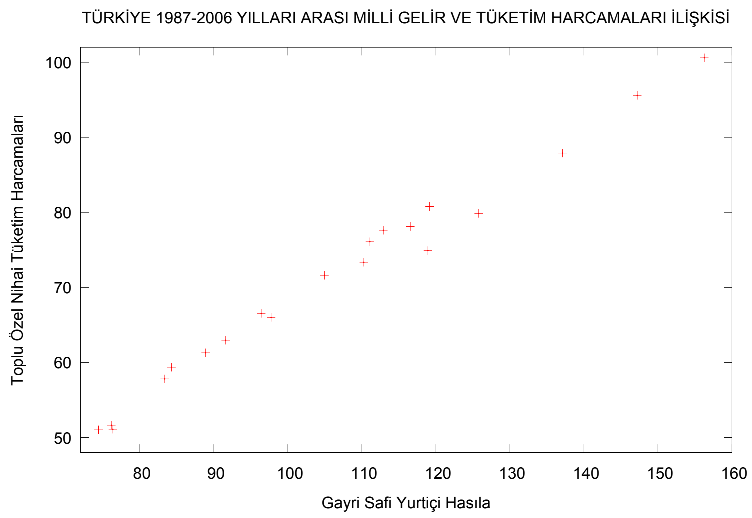
Yıl	C	Y	Yıl	C	Y
1987	51.019	74.416	1997	77.620	112.892
1988	51.638	76.143	1998	78.113	116.541
1989	51.105	76.364	1999	76.077	111.083
1990	57.803	83.371	2000	80.774	119.147
1991	59.366	84.271	2001	73.356	110.267
1992	61.282	88.893	2002	74.894	118.923
1993	66.545	96.391	2003	79.862	125.778
1994	62.962	91.600	2004	87.897	137.110
1995	66.011	97.729	2005	95.594	147.200
1996	71.614	104.940	2006	100.584	156.249

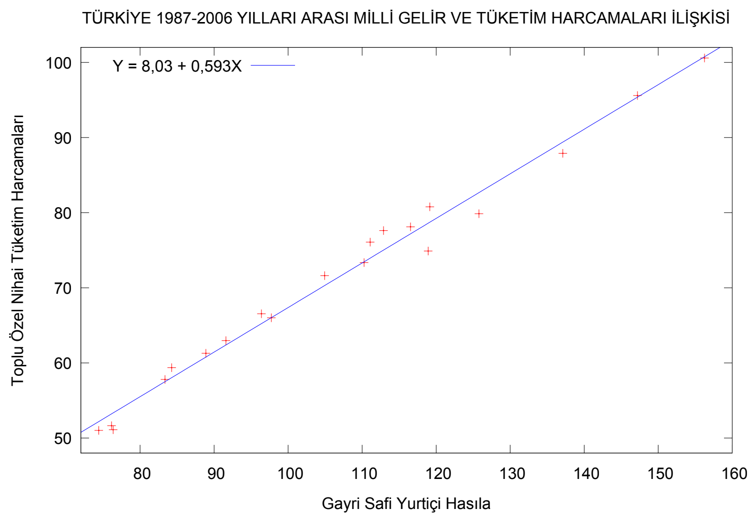
Adım 5: Modelin tahmini
Verilerin sağlanmasının ardından, başta belirtilen ekonometrik model eldeki verilere yakıştırılır.

- *"Bağlanım"* (regression) çözümlemesi denilen yöntemin uygulanması ile bulunan β_1 ve β_2 tahminleri şöyledir:

$$\hat{C} = 8{,}03 + 0{,}59Y$$

- C teriminin üzerindeki ($\hat{}$) *"şapka"* (hat) işareti, bunun bir tahmin olduğunu göstermektedir.
- 1987–2006 arası dönem için otonom tüketimi gösteren β_1, 8,03 milyon TL olarak tahmin edilmiştir.
- β_2 katsayısı ise yaklaşık 0,59 bulunmuştur.
- Buna göre, örneklem döneminde *"gerçek"* (real) gelirdeki 1 milyon TL büyüklüğündeki bir artış tüketim harcamalarında ortalama yaklaşık 590.000 TL'lik bir artışa yol açmaktadır.





Adım 6: Önsav sınamaları
İktisat kuramlarının verilerden elde edilen kanıtlara dayanarak doğrulanması ya da yanlışlanması *"istatistiksel çıkarsama"* (statistical inference) ile olur.

- Keynes, tüketimdeki marjinal artışın sıfırla bir arasında olmasını bekliyordu.
- Model tahminine göre tüketim doğrusunun eğimi 0,59'dur.
- Ancak eldeki bulgunun Keynes'in kuramını doğruladığı sonucuna varmadan önce, bunun bir rastlantı eseri olup olmadığına belli bir güvenle karar vermek gereklidir.

- Bu noktada 0,59 değerinin istatistiksel olarak 1'den küçük olup olmadığını bulmak için çeşitli istatistiksel çıkarsama yöntemleri uygulanır.

Adım 7: Yordama
Eldeki model eğer kuramı ve önsavı doğruluyorsa, bağımsız değişken ya da bağımlı değişkenin gelecekteki değerlerini tahmin etmede de kullanılabilir.

- Örnek olarak, 2007 yılında GSYH'nin 165 milyon TL olacağı beklenirse tüketim harcamaları tahmini de şöyle bulunur:

$$\begin{aligned} \hat{C} &= 8{,}03 + 0{,}59 \times 165 \\ &= 105{,}38 \end{aligned}$$

- Demek ki tüketimin gelecek değerinin yordaması 105,38 milyon liradır.

Adım 8: Politika amaçlı kullanım
Yararlılığı kanıtlanan model son olarak denetim ve politika amaçlarıyla da kullanılabilir.

- Tüketim harcamaları 100 milyon TL düzeyinde tutulursa yüzde 5'lik bir enflasyonun sağlanabileceğini varsayalım. Buna göre hedeflenecek gelir şudur:

$$\begin{aligned} 100 &= 8{,}03 + 0{,}59 \times Y^* \\ Y^* &\cong 156 \end{aligned}$$

- Demek ki uygun bir ekonomi politikasıyla devlet C *"denetim değişkeni"* (control variable) aracılığıyla *"hedef değişken"* (target variable) Y için istenen düzeyi elde edebilir.

Bilgisayar ve Bilgisayar Yazılımlarının Rolü

- Bilgisayar ve bilgisayar yazılımları olmadan ekonometrinin temelini oluşturan gelişmiş bağlanım ve veri çözümleme yöntemlerini uygulamayı düşünmek bile olanaksızdır.
- 1940'larda ortaya çıkan bilgisayarlar, iktisatçılar tarafından ilk kez 1950'lerde kullanılmaya başlamıştır.
- İlk ekonometrik yazılımlar ise kamu kurumları ve üniversite ana bilgisayarlarında kullanılmak üzere 1960'larda ortaya çıkmıştır.

- Başta sayıca iki elin parmaklarını geçmeyen ekonometrik yazılımlar, 1980'lerde kişisel bilgisayarların yaygınlaşması ile birlikte hızla çoğalarak yüzler düzeyine ulaşmıştır.
- Bilgisayarlardaki büyük *"başarım"* (performance) artışları ekonometri biliminde de çığırlar açmıştır. Giderek daha da karmaşıklaşan ve yüksek maliyetler gerektiren ekonometri yazılımları 1990'lardan sonra sayıca azalmaya başlamıştır.
- Günümüzde, ekonometrik çalışmalarda yaygın olarak kullanılan yazılımlardan bazıları şunlardır:

Çizelge: Ekonometrik Çözümlemeye Yönelik Bazı Yazılımlar

Yazılım	Türü	Yazılım	Türü
Eviews	Ekonometri yazılımı	R	Programlama dili
GAUSS	Programlama dili	RATS	Ekonometri yazılımı
Gretl	Ekonometri yazılımı	SAS	Çok amaçlı paket
LimDep	Ekonometri yazılımı	SPSS	Çok amaçlı paket
MATLAB	Çok amaçlı paket	Stata	Ekonometri yazılımı
OxMetrics	Ekonometri yazılımı	TSP	Ekonometri yazılımı

- Bir iktisat öğrencisi yukarıdaki programlardan en az birini kullanmayı öğrenmelidir.
- Bu doğrultuda ticari yazılım olarak Stata ve Eviews, açık kaynaklı yazılım olarak ise R ve Gretl önerilir.

Bu ders notlarında Gretl temel alınmış, tüm görsel öğeler Gretl kullanılarak hazırlanmış ve örneklerde kullanılan özgün veriler de yine Gretl dosyası olarak sunulmuştur.
Bunun başlıca beş nedeni vardır:

- Gretl, modern ve ileri ekonometri yöntem ve araçlarını içeren kapsamlı bir yazılımdır.
- Öğrenmesi ve kullanması zevkli ve son derece kolaydır.
- Özgür ve açık kaynaklı yazılım olduğu için ücretsizdir ve herkesin kullanımına açıktır.
- Türkçe dilini destekleyen ilk ve tek ekonometri yazılımıdır.
- Hızlı bir şekilde gelişmekte ve giderek yaygınlaşmaktadır.

Son olarak, Gretl ya da başka bir yazılımı iyi öğrendikten sonra diğerlerini de öğrenmenin görece kolay olduğu unutulmamalıdır.

Ekonometri Şiiri

- "Ekonometri Nedir?" tartışmasını Gazi Üniversitesi öğretim üyesi Prof. Dr. Nihat Bozdağ'ın güzel şiiri ile noktalayalım:

EKONOMETRİ

Hep elden geçirilir birer birer
Titizlikle ayıklanır veriler
Bilimsel çalışma en büyük hüner
Bilimlere rehber Ekonometri

Sosyal araştırmalara baz verir
Bilimsel gelişmelere hız verir
Araştırmacıya büyük haz verir
Gönüllerde eser Ekonometri

Lafla gerçek bulgulara varılmaz
Yöntemsiz çalışma bilim sayılmaz
Ekonometrisiz hiç çalışılmaz
Verileri süzer Ekonometri

Teknolojik gelişimli kulvarda
İktisaden problemli yollarda
Bol sorunlu ikibinli yıllarda
Sorunları çözer Ekonometri

- "http://www.nihatbozdag.net/" adresinden Prof. Dr. Nihat Bozdağ'ın kişisel Internet sayfasına ulaşabilirsiniz.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *"Introduction"* bölümü (sayfa 1–14) okunacak.

Önümüzdeki Ders
Bağlanım Çözümlemesi

Bölüm 3

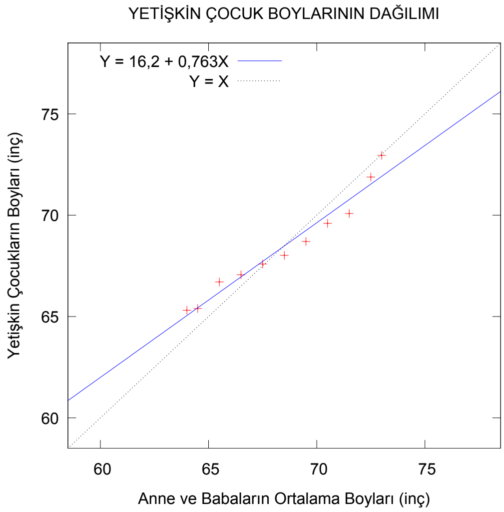
Bağlanım Çözümlemesi

3.1 Temel Kavramlar

3.1.1 Bağlanım Teriminin Anlamı

Bağlanım Teriminin Anlamı

- İngilizce "regression" teriminin sözcük anlamı, istatistikteki *"sıradanlığa doğru çekilme"* (regression toward mediocrity) olgusundan gelmektedir.
- Bu terim ilk kez İngiliz antropolog, meteorolojist, kaşif, mucit ve istatistikçi Sir Francis Galton (1822 - 1911) tarafından kullanılmıştır.
- Galton ünlü bir yazısında belli bir boydaki anne-babaların yetişkin çocuklarının ortalama boylarının genel nüfustaki ortalama boya çekilme eğiliminde olduğunu bulmuştur.
- Günümüzde kullanılan anlamıyla "regression" bağımlı bir değişkeni, tahmin ya da çıkarım amacıyla farklı bağımsız değişkenler ile ilişkilendiren istatistiksel bir yöntemdir.
- Bu terimin uygun ve doğru Türkçe karşılığı ise *"bağlanım"* sözcüğüdür (Bkz. TDK İstatistik Terimleri Sözlüğü).



Bağlanım terimi istatistikte bir çözümleme yöntemini anlatır:

Bağlanım Çözümlemesi
Bağlanım çözümlemesi, bir bağımlı değişkenin başka açıklayıcı değişkenlerle olan ilişkisini, birincinin ortalama değerini ikinci(ler)in bilinen ya da sabit değerleri cinsinden tahmin etme ya da kestirme amacıyla inceleyen bir istatistiksel yöntemdir.

- Diğer bir deyişle, bağlanım yöntemi, bağımlı değişkendeki değişiklikleri açıklayıcı değişken denilen çeşitli etmenleri denetim altında tutarak inceler.
- Bağlanım çözümlemesindeki ilgi odağı kesin ilişkiler değil istatistiksel ilişkilerdir.
- Kullanılan değişkenler genellikle rastsal ya da *"olasılıksal"* (stochastic) ya da olasılık dağılımı olan değişkenlerdir.

Bağlanım ile İlgili Temel Terimler

- Bağlanım çözümlemesinde kullanılan sol ve sağ yan değişkenleri yazında farklı adlar ile karşımıza çıkabilirler:

SOL YAN (Y)		SAĞ YAN (X)	
Türkçe	İngilizce	Türkçe	İngilizce
"Açıklanan değişken"	(Explained variable)	*"Açıklayıcı değişken"*	(Explanatory variable)
"Bağımlı değişken"	(Dependent variable)	*"Bağımsız değişken"*	(Independent variable)
"Bağlanan"	(Regressand)	*"Bağlayan"*	(Regressor)
"Kestirilen"	(Predictand)	*"Kestiren"*	(Predictor)
"Tepki değişkeni"	(Response variable)	*"Denetim değişkeni"*	(Control variable)
"İçsel değişken"	(Endogenous variable)	*"Dışsal değişken"*	(Exogenous variable)

Bağlanım ve Nedensellik

İstatistiksel bir ilişki kendi başına bir nedensellik anlamı taşımaz. M. G. Kendal ve A. Stuart'ın sözleriyle:

> "İstatistiksel bir ilişki ne denli güçlü ve ne denli anlamlı olursa olsun, asla nedensel bir ilişki kuramaz. Bizim nedensellik düşüncelerimiz istatistiğin dışından, eninde sonunda şu ya da bu kuramdan gelmelidir."

Bağlanım ve İlinti

- *"İlinti"* (correlation) çözümlemesi, iki değişken arasındaki doğrusal ilişkinin gücünü inceler.
- Bağlanım çözümlemesi ve ilinti çözümlemesi yakından ilişkili olsa da bu iki yöntem arasında önemli kavramsal farklar vardır.
- İlinti çözümlemesinde herhangi iki değişken *"bakışımlı"* (symmetric) olarak ele alınabilir.
- Diğer bir deyişle bağımlı ve açıklayıcı değişkenlerden söz edilmez.
- Bağlanım çözümlemesinde ise değişkenlerin ele alınışı tek yönlüdür. Bağımlı değişkenin olasılıksal olduğu, açıklayıcı değişken(ler)in ise değişmeyen değerler aldığı varsayılır.

3.1.2 Ekonometrik Çözümlemede Kullanılan Verilerin Niteliği

Veri Seti Türleri

"Görgül" (empirical) çözümlemelerde üç tür veri seti kullanılır:

1. *"Zaman serisi"* (time series) veri setleri
2. *"Yatay-kesit"* (cross-sectional) veri setleri
3. *"Karma"* (pooled) veri setleri

Zaman Serileri

Zaman serisi, bir değişkenin farklı zamanlarda gözlenen bir değerler setidir.

Zaman serilerine örnek olarak aşağıdakiler gösterilebilir:

- Hisse senedi fiyatları (günlük / dakikalık)
- Para arzı (haftalık)

- Tüketici Fiyat Endeksi (aylık)
- Gayri Safi Milli Hasıla (üç aylık)
- Hükümet bütçesi (yıllık)
- Genel seçim sonuçları (dört yıllık)

Yatay-Kesit Verileri
Yatay-kesitsel veriler, zaman içinde belli bir noktada derlenerek oluşturulan veri setleridir.

Yatay-kesit verilerine örnek olarak şunlar gösterilebilir:

- TÜİK tarafından belli aralıklarla düzenlenen tüketici harcamaları anketi
- Çeşitli kurumlarca yürütülen kamuoyu araştırmaları
- Hisse senedi fiyatlarının belli bir gün sonundaki değerleri

Karma Veriler
Karma veriler, hem zaman serisi hem de yatay-kesit öğeleri içeren verilerdir.

- Karma verilere örnek olarak çeşitli illere ait gelir, işsizlik, iç göç gibi istatistikleri içeren bir veri seti gösterilebilir.

"Panel" (panel) verileri denen özel bir karma veri tipi vardır:

Panel Verileri
Birden fazla değişkenin zaman içerisinde izlenilmesi ile ortaya çıkan veri seti türüdür.

- Panel verilerine örnek olarak ABD Michigan Üniversitesi tarafından düzenlenen Panel Study of Income Dynamics (PSID) veri tabanı gösterilebilir.

Aldıkları değerler bakımından ise veriler ikiye ayrılırlar:

Nicel Veriler
Gelir, fiyatlar, para arzı, faiz oranları ...

Nitel Veriler
Erkek / kadın, evli / bekar, üniversite mezunu / değil, ...

Verilerin Doğruluk Derecesi

Ekonomik araştırmalarda kullanılan veriler çoğu zaman nitelik yönünden çok iyi düzeyde olamayabilmektedirler:

- Çoğu toplum bilim verileri deneysel olmadığı için gözlem hataları içermektedir.
- Deneysel verilerde bile ölçüm hataları olabilmektedir.
- Anketle toplanan verilerde yanıt alamama sorunu ya da *"seçim yanlılığı"* (selection bias) doğabilmektedir.
- Kullanılan örnekleme yöntemi *"örnekleme yanlılığı"* (sampling bias) sorununa yol açabilmektedir.
- *"Toplulaştırmalı"* (aggregated) iktisadi veriler hane halkı gibi mikro birimler için fazla açıklayıcı olamayabilmektedir.

Sonuç olarak; ekonometrik yöntemlerin başarısı kullanılan verilerin kaynak, nitelik ve doğruluk derecesine bağlıdır.

3.2 Varsayımsal Bir Örnek

3.2.1 Koşullu Olasılık ve Koşullu Ortalama

Varsayımsal Bir Örnek

- Bağlanım çözümlemesine başlangıç olarak ikili bağlanım modelini inceleyeceğiz.
- İki değişkenli durum çoğu uygulama için yetersiz olsa da temel bilgileri olabildiğince yalın gösterebilmek açısından önemlidir.
- İkili bağlanıma varsayımsal bir örnek olarak toplam nüfusu 60 aileden oluşan bir ülke düşünelim.
- Bu ailelerin vergiden sonraki harcanabilir haftalık gelirleri X ve haftalık tüketim harcamaları Y arasındaki ilişkiyi tahmin etmek istiyor olalım.
- Bunun için öncelikle bu 60 aileyi gelirleri yaklaşık aynı olan 10 farklı öbeğe ayıralım.
- Örneğimiz ile ilgili varsayımsal veriler aşağıdadır:

Çizelge: Haftalık Aile Geliri X ile Haftalık Tüketim Harcamaları Y, \$

$Y \downarrow, X \rightarrow$	**80**	**100**	**120**	**140**	**160**	**180**	**200**	**220**	**240**	**260**
	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Toplam	325	462	445	707	678	750	685	1043	966	1211

- Buradaki her bir sütun, farklı gelir düzeylerine (X) karşılık gelen tüketim harcamaları (Y) dağılımını göstermektedir.

Koşullu Olasılık ve Koşullu Ortalama

- Örnekteki $X = 80$ değerine karşılık gelen 5 ayrı Y değeri bulunmaktadır: 55, 60, 65, 70 ve 75.
- Yukarıdaki tüketim harcamalarının her birinin gerçekleşme olasılığı ise $\frac{1}{5}$'tir.
- Bu durumda, $X = 80$ olduğunda Y'nin de 55 olma *"koşullu olasılığı"* (conditional probability) $P(Y = 55|X = 80) = \frac{1}{5}$'tir.

- *"Koşullu ortalama"* (conditional mean) ya da *"koşullu beklenen değer"* (conditional expected value) ise Y'nin her bir koşullu olasılık dağılımı için beklenen değerini gösterir.
- Koşullu ortalamayı bulmak için ilgili Y değerleri ve bunlara karşılık gelen koşullu olasılıklar çarpılıp toplanır.
- Örnek olarak, $X = 80$ iken Y'nin koşullu ortalaması $55(\frac{1}{5})+60(\frac{1}{5})+65(\frac{1}{5})+70(\frac{1}{5})+75(\frac{1}{5}) = 65$ olur.

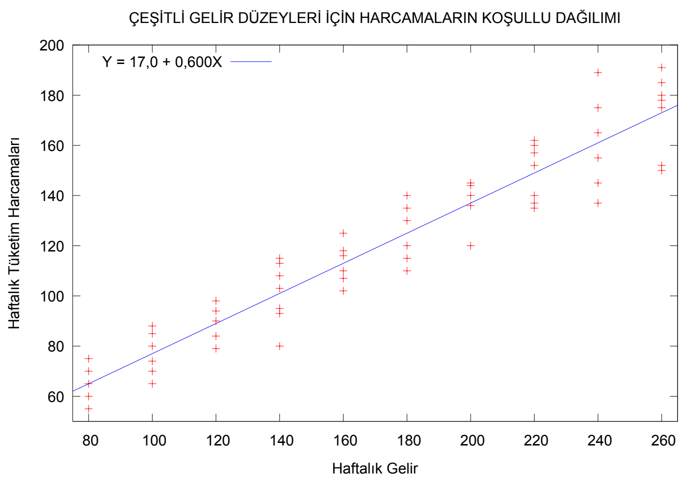
Çizelge: $P(Y|X_i)$ Koşullu Olasılık ve Koşullu Ortalamaları

$Y \downarrow, X \rightarrow$	**80**	**100**	**120**	**140**	**160**	**180**	**200**	**220**	**240**	**260**
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	–	1/6	–	1/7	1/6	1/6	–	1/7	1/6	1/7
	–	–	–	1/7	–	–	–	1/7	–	1/7
Ortalama	65	77	89	101	113	125	137	149	161	173

3.2.2 Anakütle Bağlanım İşlevi

Anakütle Bağlanım İşlevi

Verilerimizi *"serpilim çizimi"* (scatter plot) üzerinde inceleyelim:



Çizimde görülen artı eğimli doğrunun gösterdiği matematiksel işlev *"anakütle bağlanım işlevi"* (population regression function) ya da kısaca *"ABİ"* (PRF) olarak adlandırılır:

Anakütle Bağlanım İşlevi

Anakütle bağlanım işlevi, açıklayıcı değişken(ler)in sabit değerlerine karşılık gelen bağımlı değişkenin koşullu ortalamaları ya da koşullu beklenen değerlerinin geometrik yerini gösterir.

- Her koşullu ortalama X'in bir işlevidir: $E(Y|X_i) = f(X_i)$.
- Anakütle bağlanım işlevi denilen $f(X_i)$, X'teki değişmeye karşılık Y'nin dağılımının ortalama tepkisini vermektedir.
- Kısaca Y'nin ortalama değeri ile ilgileniyoruz ama özellikle de Y'nin ortalamasının X'lere bağlı olarak nasıl değiştiğini bulmaya çalışıyoruz.
- $f(X_i)$'nin işlev biçiminin ne olduğu sorusu önemlidir.
- Gerçek yaşamda tüm anakütle incelemeye açık olmadığı için burada iktisat kuramından yararlanılmalıdır.
- Örnek olarak, bir ekonomist tüketim harcamalarının gelirle doğrusal bir ilişki içinde olduğunu söylüyor olsun.
- Bu durumda varsayılabilecek doğrusal işlev de şu olur:

$$E(Y|X_i) = f(X_i) = \beta_1 + \beta_2 X_i$$

"Doğrusal" (linear) işlev, *"değişkenlerde doğrusallık"* (linearity in the variables) ve *"değiştirgelerde doğrusallık"* (linearity in the parameters) olmak üzere iki farklı anlama gelebilir:

Değişkenlerde Doğrusallık
Doğal ve basitçe bağlanım işlevinin düz bir doğruyu gösterdiği durumdur.
Doğrusal: $Y_i = \beta_1 + \beta_2 X_i$
Doğrusal-dışı: $Y_i = \beta_1 + \beta_2 X_i^2$

Değiştirgelerde Doğrusallık
$E(Y|X_i)$'nin β değiştirgelerinin doğrusal bir işlevi olduğu durumdur.
Doğrusal: $Y_i = \beta_1 + \beta_2 X_i^2$
Doğrusal-dışı: $Y_i = \beta_1 + \sqrt{\beta_2} X_i$

Rastsal Hata Terimi

- Örneğimizde görüldüğü gibi gelir artarken tüketim harcamaları da genel olarak artmaktadır.
- Diğer yandan, tekil bir ailenin harcamasının geliri daha düşük olan bir aileden fazla olması da zorunlu değildir:

Çizelge: Haftalık Aile Geliri X ile Haftalık Tüketim Harcamaları Y, \$

$Y \downarrow, X \rightarrow$	**80**	**100**	**120**	**140**	**160**	**180**	**200**	**220**	**240**	**260**
	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Toplam	325	462	445	707	678	750	685	1043	966	1211

- Tekil bir ailenin harcamasının aynı gelir düzeyindeki bütün ailelerin harcamalarının ortalaması, diğer bir deyişle koşullu beklenen değeri dolayında dağıldığını biliyoruz.
- Buna göre, bireysel Y_i'nin kendi beklenen değerinden gösterdiği *"sapma"* (deviation) şöyle gösterilebilir:

$$\begin{aligned} & u_i = Y_i - E(Y|X_i) \\ \text{ya da} \quad & Y_i = E(Y|X_i) + u_i \\ \text{ya da} \quad & Y_i = \beta_1 + \beta_2 X_i + u_i \end{aligned}$$

- Buradaki u_i *"bozukluk"* (disturbance) terimi, artı ya da eksi değerler alabilen ama gözlenemeyen *"rastsal hata terimi"* (random error term) diye adlandırılır.
- $Y_i = E(Y|X_i) + u_i$ eşitliğinin her iki yanının beklenen değeri alınırsa şu bulunur:

$$\begin{aligned} Y_i &= E(Y|X_i) + u_i \\ E(Y_i|X_i) &= E[E(Y|X_i)] + E(u_i|X_i) \\ E(Y_i|X_i) &= E(Y|X_i) + E(u_i|X_i) \\ 0 &= E(u_i|X_i) \end{aligned}$$

- $E(Y_i|X_i)$ ile $E(Y|X_i)$ aynı şey olduğu için, $E(u_i|X_i) = 0$ olur.

- Bu durumda, u_i'lerin koşullu ortalamasının sıfır olduğu varsayımına dayanılarak, bağlanım doğrusunun Y'nin koşullu ortalamasından geçtiği sonucuna ulaşılabilir.

Modele katılmayan ama Y'yi etkileyen tüm değişkenlerin yerine geçen hata terimi u_i'nin modele açıkça koyulması gereğinin nedenlerinden bazıları şunlardır:

1. Kuramın belirsizliği ya da eksikliği
2. Yeterli ya da geçerli verilerin bulunamaması
3. İlişkili ancak ortak etkisi küçük olan değişkenler
4. İnsan davranışlarının doğasında olan rastsallık
5. Güçsüz *"yaklaşık değişkenler"* (proxy variables)
6. Basitlik ilkesi
7. Bilinemeyen işlev biçimi

3.2.3 Örneklem Bağlanım İşlevi

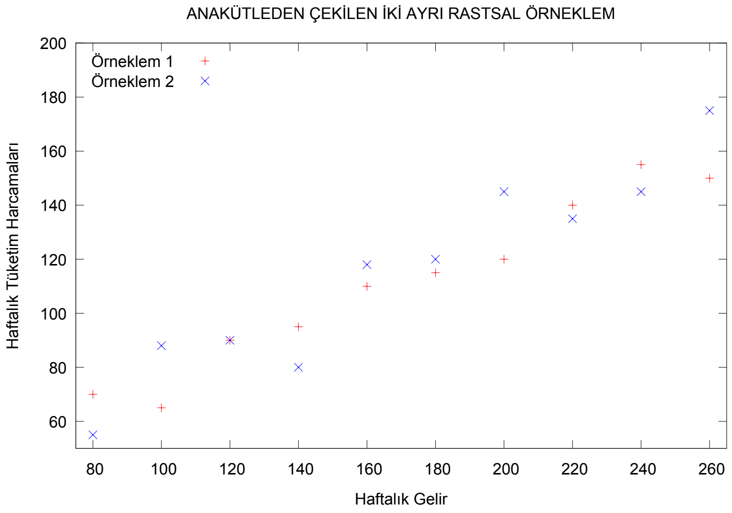
Örneklem Bağlanım İşlevi

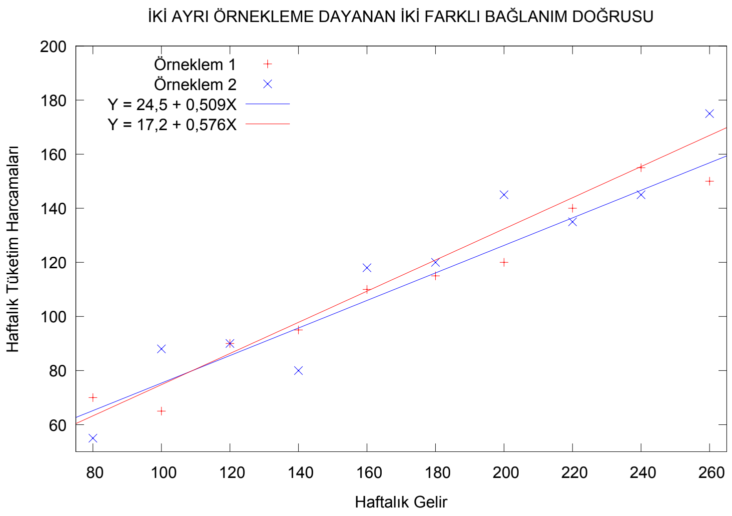
- Gerçek yaşamda anakütle verilerine ulaşabilme olasılığı düşüktür.
- Çoğu uygulamada elimizde yalnızca anakütleden alınmış örneklem verileri bulunmaktadır.
- Öyleyse yanıtlamamız gereken önemli soru, örneklem verilerini kullanarak anakütle bağlanım işlevi ABİ'yi tahmin edip edemeyeceğimiz sorusudur.
- Rastsal bir örneklem kullanarak bulunan bağlanım işlevine *"örneklem bağlanım işlevi"* (sample regression function) ya da kısaca *"ÖBİ"* (SRF) denir.
- Bu işlevi anlatan doğruya ise *"örneklem bağlanım doğrusu"* (sample regression line) adı verilir.

Anakütleden her biri 10 gözlem büyüklüğünde iki farklı rastsal örneklem çekelim:

Çizelge: Anakütleden Çekilmiş İki Rastsal Örneklem

X	Y	X	Y
80	70	80	55
100	65	100	88
120	90	120	90
140	95	140	80
160	110	160	118
180	115	180	120
200	120	200	145
220	140	220	135
240	155	240	145
260	150	260	175





- Anlaşılıyor ki rastsallık nedeniyle örneklem verilerini kullanarak anakütle bağlanım işlevini tam doğru biçimde tahmin etmek olanaksızdır.
- Elimizdeki iki değişik örneklem bağlanım doğrusundan hangisinin gerçek anakütle bağlanım doğrusunu daha iyi temsil ettiği kesin değildir.
- Genel olarak, n farklı örneklem için n sayıda farklı ÖBİ bulunabilir diyebiliriz.

Örneklem Bağlanım İşlevinin Bulunması

- Açıklamış olduğumuz tahmin sorunu yüzünden örneklem bağlanım işlevi aşağıdaki gibi gösterilir:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

- Burada:
  $\hat{\beta}_1$ "β_1 *şapka*" (β_1 hat) diye okunan β_1'in tahmincisini,
  $\hat{\beta}_2$ β_2'nin tahmincisini,
  $\hat{u}_i$ u_i'nin tahmincisini göstermektedir.
- Anakütle bağlanım işlevini başta $Y_i = 17 + 0{,}6X_i + u_i$ olarak hesaplamış olduğumuzu anımsayalım.
- Bulduğumuz birinci örneklem bağlanım işlevi şudur:

$$Y_i = 24{,}5 + 0{,}509X_i + \hat{u}_i$$

- Bulduğumuz ikinci örneklem bağlanım işlevi ise şudur:

$$Y_i = 17{,}2 + 0{,}576X_i + \hat{u}_i$$

- Örneklem bağlanım işlevlerinin her ikisi de β_1 ***"değiştirge"*** (parameter) değerini yüksek tahmin ederken, β_2 değiştirge değerini düşük tahmin etmiştir.
- O zaman buradaki önemli soru, ABİ bilinemese bile $\hat{\beta}_1$'nın gerçek β_1'e ve $\hat{\beta}_2$'nın da gerçek β_2'ye olabildiğince yakın olduğu bir ÖBİ'nin nasıl oluşturulabileceği sorusudur.

Gujarati'nin sözleriyle:

> "Burada vurguladığımız, ABİ'yi olabildiğince doğru yansıtan ÖBİ'nin nasıl kurulacağını söyleyen süreçler geliştirebileceğimizdir. ABİ'yi asla gerçekten belirleyemesek bile, bunun yapılabileceğini düşünmek heyecan vericidir."

- Bu heyecanlı tartışmayı burada şimdilik sonlandırıyoruz.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 1* “The Nature of Regression Analysis” ve *Bölüm 2* “Two-Variable Regression Analysis: Some Basic Ideas” okunacak.

Önümüzdeki Ders
İki Değişkenli Bağlanım Modeli: Tahmin Sorunu

Bölüm 4

İki Değişkenli Bağlanım Modeli - Tahmin Sorunu

4.1 Sıradan En Küçük Kareler Yöntemi

Sıradan En Küçük Kareler Yöntemi

- Bağlanım çözümlemesinde amaç, örneklem bağlanım işlevi (ÖBİ) temel alınarak anakütle bağlanım işlevinin (ABİ) olabildiğince doğru biçimde tahmin edilmesidir.
- Bunun için kullanılan en yaygın yol *"sıradan en küçük kareler"* (ordinary least squares), kısaca *"SEK"* (OLS) yöntemidir.
- SEK yönteminin 1794 yılında Alman matematikçi Carl Fredrich Gauss tarafından bulunduğu kabul edilir.
- SEK yöntemini anlamak için iki değişkenli ABİ'yi anımsayalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- ABİ gözlenemediğinden ÖBİ kullanılarak tahmin edilir:

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

- ÖBİ'nin kendisini bulmak için ise *"kalıntılar"* (residuals), diğer bir deyişle hata terimi kullanılır:

$$\begin{aligned}\hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\end{aligned}$$

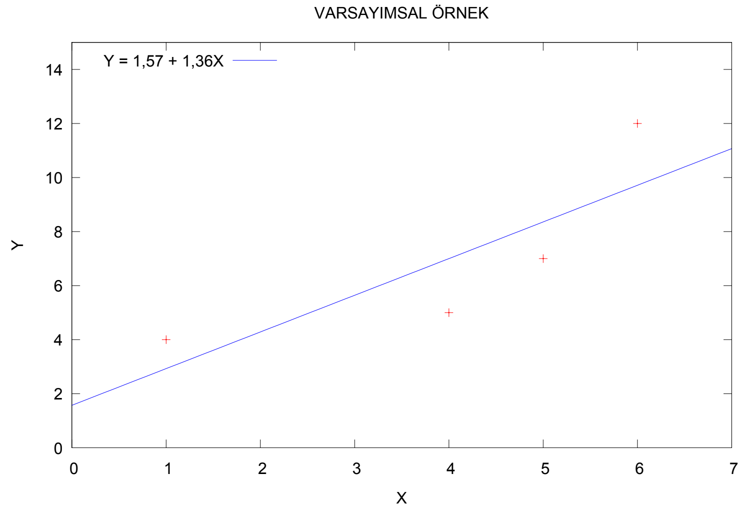
- Elimizde n tane X ve Y varken, ÖBİ'yi gözlenen Y'lere olabildiğince yakın biçimde belirlemek istiyoruz.
- Bunun için şu ölçüt benimsenebilir:

$$\min\left(\sum \hat{u}_i\right) = \min\left(\sum (Y_i - \hat{Y}_i)\right)$$

- Ancak bu durumda artı ve eksi değerli hatalar büyük ölçüde birbirlerini etkisiz hale getirecektir.
- Ayrıca burada ÖBİ'ye ne kadar yakın ya da uzak olursa olsun tüm kalıntılar eşit önem taşımaktadır.
- Öyleyse, ÖBİ'yi kalıntılar toplamı en küçük olacak şekilde seçmek iyi bir ölçüt değildir.
- Herhangi bir veri seti için farklı $\hat{\beta}_1$ ve $\hat{\beta}_2$ değerleri farklı $\hat{u}_i$ ve dolayısıyla da farklı $\sum \hat{u}_i{}^2$ toplamları verir.
- Ancak hatalar toplamı $\sum \hat{u}_i$ her zaman sıfır çıkar.
- Örnek olarak, varsayımsal bir veri seti için aşağıdaki iki ÖBİ'yi ele alalım:

$$\begin{aligned}\hat{Y}_{1i} &= 1{,}572 + 1{,}357X_i \\ \hat{Y}_{2i} &= 3{,}000 + 1{,}000X_i\end{aligned}$$

	Y_i	X_i	$\hat{Y}_{1i}$	$\hat{u}_{1i}$	$\hat{u}^2_{1i}$	$\hat{Y}_{2i}$	$\hat{u}_{2i}$	$\hat{u}^2_{2i}$
	4	1	2,929	1,071	1,147	4	0	0
	5	4	7,000	-2,000	4,000	7	-2	4
	7	5	8,357	-1,357	1,841	8	-1	1
	12	6	9,714	2,286	5,226	9	3	9
Toplam	28	16		0	12,214		0	14



- Artı ve eksi değerler alabilen kalıntıların toplamının küçük çıkma sorunundan kurtulmak için en küçük kareler ölçütü kullanılır:

En Küçük Kareler Ölçütü

$$\begin{aligned}\min\left(\sum \hat{u}_i{}^2\right) &= \min\left(\sum(Y_i - \hat{Y}_i)^2\right)\\ &= \min\left(\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2\right)\end{aligned}$$

- Yukarıdaki gösterimin $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincilerine dayanan bir matematiksel işlev olduğuna dikkat ediniz.

4.1.1 SEK Tahmincilerinin Türetilmesi

Normal Denklemler

- SEK, kalıntı kareleri toplamını *"enazlamak"* (minimize) için, ÖBİ değiştirgelerini hesaplamada basit bir *"eniyileme"* (optimization) yönteminden yararlanır.
- $\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$ teriminin $\hat{\beta}_1$ ve $\hat{\beta}_2$'ya göre kısmi türevlerini alalım:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

- Burada n örneklem büyüklüğüdür.
- Yukarıdaki denklemler *"normal denklemler"* (normal equations) olarak adlandırılırlar.
- $\hat{\beta}_1$ ve $\hat{\beta}_2$ değiştirgeleri, normal denklemlerin eşanlı olarak çözülmesi ile bulunur:

$$\begin{aligned}\hat{\beta}_2 &= \tfrac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \\ &= \tfrac{\sum x_i y_i}{\sum x_i^2}\end{aligned}$$

$$\begin{aligned}\hat{\beta}_1 &= \tfrac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - (\sum X_i)^2} \\ &= \bar{Y} - \hat{\beta}_2\bar{X}\end{aligned}$$

- $\bar{X}$ ve $\bar{Y}$ terimleri X ile Y'nin örneklem ortalamalarıdır.
- Küçük harfler ise *"ortalamadan sapma"* (deviation from the mean) olarak kullanılmıştır:

$$\begin{aligned}x_i &= (X_i - \bar{X}) \\ y_i &= (Y_i - \bar{Y})\end{aligned}$$

SEK Bağlanım Doğrusunun Özellikleri

İkili bağlanım SEK tahmincileri $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın şu özelliklerine dikkat edelim:

- Bunlar birer nokta tahmincisidirler.
- Gözlemlenebilen örneklem değerleri (X_i ve Y_i) cinsinden gösterilir ve dolayısıyla kolayca hesaplanabilirler.
- Örneklem verileri kullanılarak $\hat{\beta}_1$ ve $\hat{\beta}_2$ hesaplandıktan sonra, örneklem bağlanım doğrusu da kolayca çizilebilir.

SEK yöntemi ile bulunan örneklem bağlanım doğrusu aşağıda verilen özellikleri taşır:

1. Örneklem bağlanım doğrusu, X ve Y'nin örneklem ortalamalarından geçer. ($\bar{Y}_i = \hat{\beta}_1 + \hat{\beta}_2\bar{X}_i$)

2. $\hat{u}_i$ kalıntılarının ortalaması sıfırdır. ($\bar{\hat{u}}_i = 0$)

3. $\hat{u}_i$ kalıntıları tahmin edilen Y_i'lerle ilişkisizdir. ($\sum \hat{u}_i \hat{Y}_i = 0$)

4. $\hat{u}_i$ kalıntıları X_i'lerle ilişkisizdir. ($\sum \hat{u}_i X_i = 0$)

5. Tahmin edilen $\hat{Y}_i$'ların ortalaması, gözlemlenen Y_i değerlerinin ortalamasına eşittir. Bu ÖBİ'den görülebilir:

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i \\ &= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X}) \end{aligned}$$

Son satırın her iki yanı örneklem üzerinden toplanıp n'ye bölünürse, $\bar{\hat{Y}} = \bar{Y}$ olarak bulunabilir.

ÖBİ'nin Sapma Biçiminde Gösterimi

- ÖBİ'nin *"sapma biçimi"* (deviation form) gösterimini bulmak için $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$ işlevinin her iki yanını toplayalım:

$$\begin{aligned} \sum Y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \qquad (\textstyle\sum \hat{u}_i = 0 \text{ olduğu için}) \end{aligned}$$

- Daha sonra bu denklemin her iki yanını n'ye bölelim:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

- Yukarıdaki eşitlik, örneklem bağlanımı doğrusunun X ve Y'nin örneklem ortalamalarından geçtiğini göstermektedir.

- Son olarak yukarıdaki eşitliği ilk eşitlikten çıkaralım:

$$\begin{aligned} Y_i - \bar{Y} &= \hat{\beta}_2 (X_i - \bar{X}) + \hat{u}_i \\ y_i &= \hat{\beta}_2 x_i + \hat{u}_i \end{aligned}$$

- Sapma gösteriminde $\hat{\beta}_1$'nın bulunmadığına dikkat ediniz.

4.1.2 SEK Tahmincilerinin Arzulanan Özellikleri

Gauss - Markov Kanıtsavı

- Klasik Doğrusal Bağlanım Modeli (KDBM) varsayımları geçerli iken, en küçük kareler yöntemi ile elde edilen tahminler arzulanan bazı özellikler taşırlar.
- Gauss - Markov kanıtsavına göre $\hat{\beta}$ SEK tahmincilerine *"En iyi Doğrusal Yansız Tahminci"* (Best Linear Unbiased Estimator), kısaca *"EDYT"* (BLUE) adı verilir.

EDYT olan $\hat{\beta}$ şu üç arzulanan özelliği taşır:

1. Doğrusaldır. Diğer bir deyişle bağlanım modelindeki Y bağımlı değişkeninin doğrusal bir işlevidir.
2. Yansızdır. Beklenen değeri $E(\hat{\beta})$, anakütleye ait gerçek β değerine eşittir.
3. Tüm doğrusal ve yansız tahminciler içinde enaz varyanslı olandır. Kısaca en iyi ya da *"etkin"* (efficient) tahmincidir.

Gauss - Markov kanıtsavı hem kuramsal olarak hem de uygulamada önemlidir.

SEK Tahmincilerinin Doğrusallık Özelliği

- SEK tahmincilerinin *"doğrusallık"* (linearity) arzulanan özelliğini gösterebilmek için $\hat{\beta}_2$ formülünü şöyle yazalım:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i - \bar{Y} \sum x_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

- Bu basitçe şu şekilde de gösterilebilir:

$$\frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Y_i, \qquad k_i = \frac{x_i}{(\sum x_i^2)}$$

- x_i değerleri olasılıksal olmadığına göre k_i'ler de gerçekte Y_i'lerin önüne gelen birer *"ağırlık"* (weight) katsayısıdırlar.
- $\hat{\beta}_2$ bu durumda Y_i'lerin doğrusal bir işlevidir. Basitçe $\hat{\beta}_2$'nın Y_i'lerin bir ağırlıklı ortalaması olduğu da söylenebilir.
- $\hat{\beta}_1$'nın doğrusal olduğu da benzer biçimde kanıtlanabilir.

SEK Tahmincilerinin Yansızlık Özelliği

SEK tahmincilerinin *"yansızlık"* (unbiasedness) arzulanan özelliğini gösterebilmek için ağırlık terimi k'nin şu beş özelliği önemlidir:

1. X_i'ler olasılıksal olmadığından k_i'ler de olasılıksal değildir.
2. $\sum k_i = 0$'dır. ($\sum x_i = 0$ olduğu için)
3. $\sum k_i^2 = \sum x_i^2 / \sum (x_i^2)^2 = 1/\sum x_i^2$ olur.
4. $\sum k_i x_i = \sum x_i^2 / \sum x_i^2 = 1$'dir.
5. $\sum k_i x_i = \sum k_i X_i$ olur. ($\sum k_i x_i = \sum k_i (X_i - \bar{X}) = \sum k_i X_i - \bar{X} \sum k_i$ olduğu için)

Dikkat: Tüm bu özellikler k_i'nin tanımından türetilebilmektedir.

- $\hat{\beta}_2$'nın yansız olduğunu kanıtlamak için $Y_i = \beta_1 + \beta_2 X_i + u_i$ biçimindeki ABİ'yi $\hat{\beta}_2$ formülünde yerine koyalım:

$$\begin{aligned}\hat{\beta}_2 &= \sum k_i Y_i \\ &= \sum k_i(\beta_1 + \beta_2 X_i + u_i) \\ &= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i \\ &= \beta_2 + \sum k_i u_i\end{aligned}$$

- Yukarıdaki son adımda k_i'nin az önce sözü edilen ikinci, dördüncü ve beşinci özelliklerinden yararlanılmıştır.

- β_2 ve k_i'nin olasılıksal olmadığını ve $E(u_i) = 0$ varsayımını anımsayalım ve her iki yanın beklenen değerini alalım:

$$\begin{aligned}E(\hat{\beta}_2) &= E(\beta_2) + \sum k_i E(u_i) \\ &= \beta_2\end{aligned}$$

- $E(\hat{\beta}_2) = \beta_2$ olduğuna göre $\hat{\beta}_2$ yansız bir tahmincidir.

SEK Tahmincilerinin Enaz Varyanslılık Özelliği

- SEK tahmincilerinin *"enaz varyans"* (minimum variance) arzulanan özelliğini gösterebilmek için ise β_2'nin en küçük kareler tahmincisinden yola çıkalım:

$$\hat{\beta}_2 = \sum k_i Y_i$$

- Şimdi β_2 için başka bir doğrusal tahminci tanımlayalım:

$$\tilde{\beta}_2 = \sum w_i Y_i$$

- Buradaki ($\tilde{}$) işareti *"dalga"* (tilde) diye okunur.
- w_i'ler de birer ağırlıktır ama $w_i = k_i$ olmak zorunda değildir:
- $\tilde{\beta}_2$'nın yansız olabilmesi için gerekli koşullara bir bakalım:

$$\begin{aligned} E(\tilde{\beta}_2) &= \sum w_i E(Y_i) \\ &= \sum w_i(\beta_1 + \beta_2 X_i) \\ &= \beta_1 \sum w_i + \beta_2 \sum w_i X_i \end{aligned}$$

- Buna göre, $\tilde{\beta}_2$'nın yansız olabilmesi için şunlar gereklidir:

$$\sum w_i = 0, \qquad \sum w_i x_i = \sum w_i X_i = 1$$

- var$(\hat{\beta}_2) \leq$ var$(\tilde{\beta}_2)$ savını kanıtlamak istiyoruz. Bunun için şimdi $\tilde{\beta}_2$'nın varyansını ele alalım:

$$\begin{aligned} \text{var}(\tilde{\beta}_2) &= \text{var}(\textstyle\sum w_i Y_i) \\ &= \textstyle\sum w_i^2 \text{var}(Y_i) \qquad [\text{Dikkat: var}(Y_i) = \text{var}(u_i) = \sigma^2] \\ &= \sigma^2 \textstyle\sum w_i^2 \qquad [\text{Dikkat: cov}(Y_i, Y_j) = 0,\ (i \neq j)] \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2 \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 + 2\sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right) \left(\frac{x_i}{\sum x_i^2} \right) \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\frac{1}{\sum x_i^2} \right) \end{aligned}$$

- Son satırda bulmuş olduğumuz şey şudur:

$$\text{var}(\tilde{\beta}_2) = \sigma^2 \sum \left(w_i - \tfrac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\tfrac{1}{\sum x_i^2} \right)$$

- Yukarıda en sağdaki terim w_i'den bağımsızdır.
- Öyleyse var$(\tilde{\beta}_2)$'yı enazlayabilmek ilk terime bağlıdır ve ilk terimi sıfırlayan w_i değeri de şudur:

$$w_i = \tfrac{x_i}{\sum x_i^2} = k_i$$

- Bu durumda aşağıdaki eşitlik geçerlidir:

$$\text{var}(\tilde{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \text{var}(\hat{\beta}_2)$$

- Demek ki w_i ağırlıkları k_i ağırlıklarına eşit olduğunda $\tilde{\beta}_2$'nın varyansı enazlanarak $\hat{\beta}_2$'nın varyansına eşitlenmektedir.
- Sonuç olarak, en küçük kareler tahmincisi $\hat{\beta}_2$ tüm yansız ve doğrusal tahminciler içinde enaz varyanslı tahmincidir.

4.1.3 SEK Yönteminin Ardındaki Varsayımlar

- Ekonometrik çözümlemenin amacı yalnızca β_1 ve β_2 gibi değiştirgeleri tahmin etmek değildir. Bu değerlere ilişkin çıkarsamalar yapmak da istenir.
- Örnek olarak, $\hat{Y}_i$'ların gerçek $E(Y|X_i)$ değerlerine ne kadar yakın olduklarını bilmek önemlidir.
- Anakütle bağlanım işlevini anımsayalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Görülüyor ki Y_i hem X_i'ye hem de u_i'ye bağlıdır.
- Öyleyse Y_i, β_1 ve β_2'ye ilişkin istatistiksel çıkarım yapmak için X_i ve u_i'nin nasıl oluşturulduğunu bilmek gereklidir.
- Bu noktada Gaussçu *"Klasik Doğrusal Bağlanım Modeli"* (Gaussian Classical Linear Regression Model), kısaca *"KDBM"* (CLRM) 10 temel varsayım yapar.

Varsayım 1
Bağlanım modeli değiştirgelerde doğrusaldır:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Ancak değişkenlerde doğrusallık zorunlu değildir.
- Değiştirgelerde doğrusallık varsayımı KDBM'nin başlangıç noktasıdır.

Varsayım 2
X değerleri tekrarlı örneklemelerde değişmez.

- Bu varsayım X'in olasılıksal olmadığını söyler.
- Buna göre X ve Y değerlerinin rastsal $\{X,Y\}$ çiftleri şeklinde elde edilmemiş olduğu kabul edilir.
- Diğer bir deyişle, gelir düzeyi başta örneğin 80 olarak belirlendikten sonra rastsal bir aile seçildiğini varsayıyoruz.
- Buna göre elimizdeki çözümleme açıklayıcı X değişkenine göre bir koşullu bağlanım çözümlemesidir.
- X ve Y değerlerinin birlikte örneklenebilmesi, bazı ek koşulların sağlanması ile geçerli olur. Bu duruma ise *"Neo-Klasik Doğrusal Bağlanım Modeli"* (NKDBM) denir.

Varsayım 3
u_i hata teriminin ortalaması sıfırdır:

$$E(u_i|X_i) = 0$$

- Buna göre, modelde açıkça yer almayan ve dolayısıyla u_i içine katılmış olan etmenlerin Y'yi kurallı bir şekilde etkilemediği varsayılmaktadır.
- Artı değerli u_i'ler eksi değerli u_i'leri götürmeli ve böylece bunların Y üzerindeki ortalama etkileri sıfır olmalıdır.

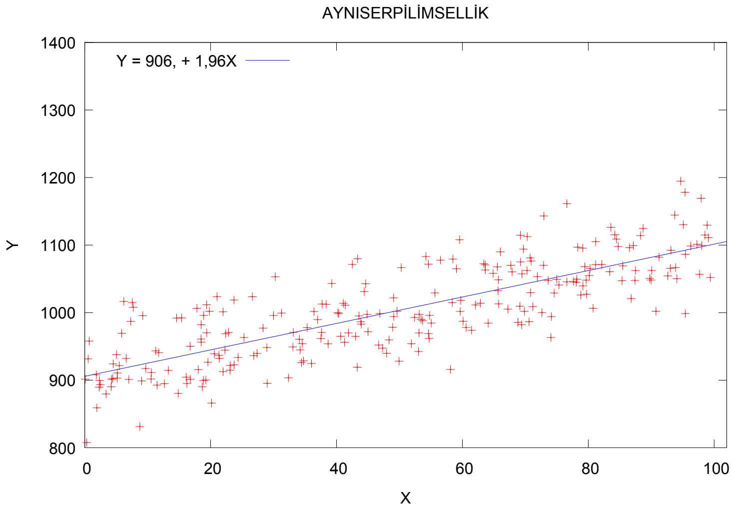
Varsayım 4
u_i hata teriminin varyansı tüm gözlemler için sabittir:

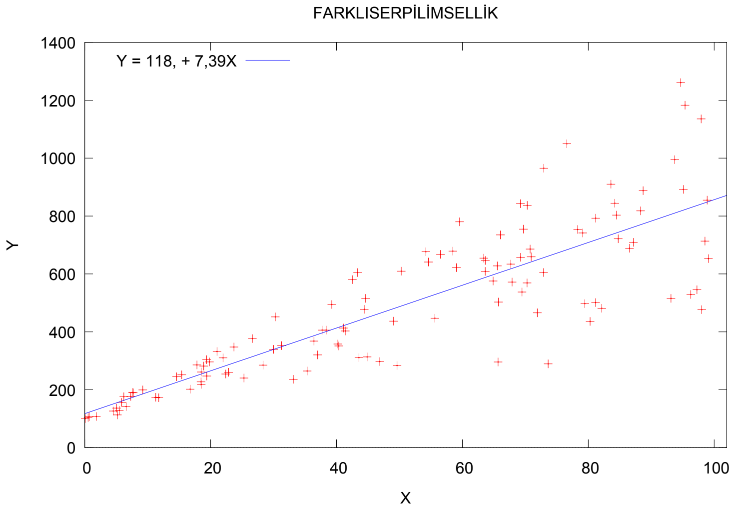
$$\text{var}(u_i|X_i) = \sigma^2$$

- *"Aynıserpilimsellik"* (homoscedasticity) varsayımına göre farklı X değerlerine karşılık gelen tüm Y'ler eşit önemdedir.
- Tersi durum ise *"farklıserpilimsellik"* (heteroscedasticity) durumudur:

$$\text{var}(u_i|X_1) \neq \text{var}(u_2|X_2) \neq \cdots \neq \text{var}(u_n|X_n).$$

- Farklıserpilimsellik durumunda çeşitli X değerlerine karşılık gelen Y değerlerinin güvenilirlikleri aynı olmaz.
- Bu yüzden kendi ortalaması etrafında farklı sıklıkta yayılan Y'leri farklı ağırlıklar vererek değerlendirmek gereklidir.



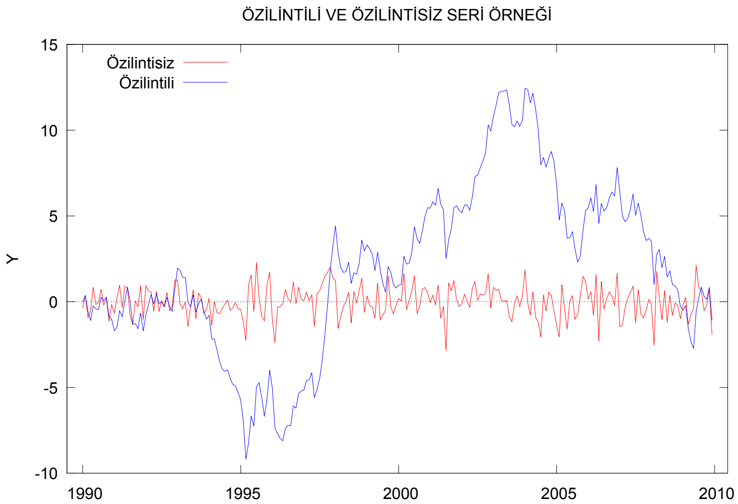


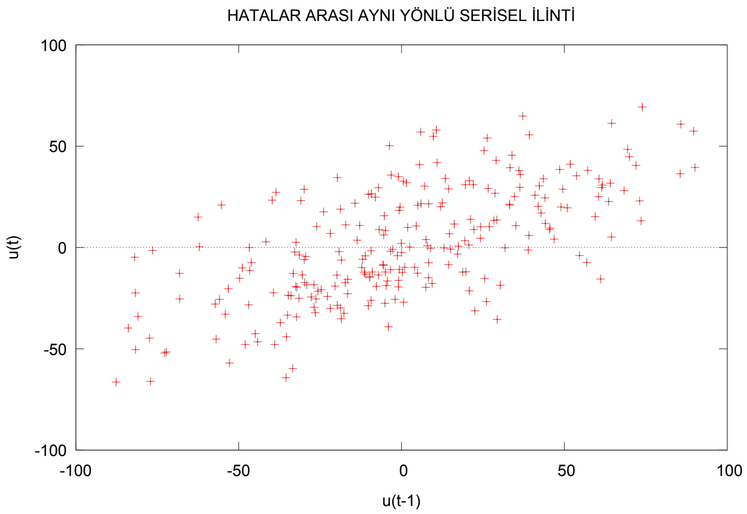
Varsayım 5

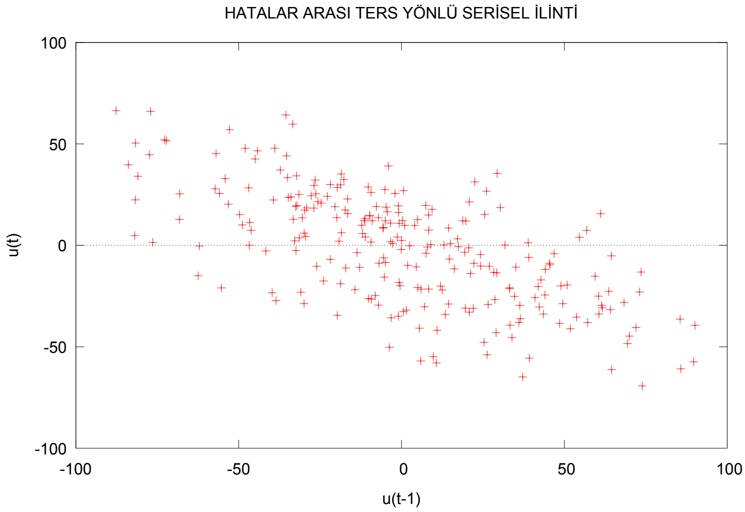
Hatalar arasında *"özilinti"* (autocorrelation) yoktur.

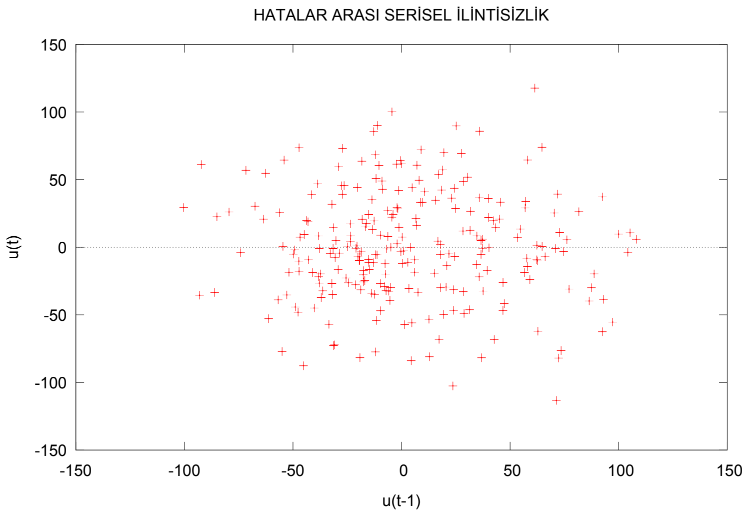
- Eğer *"bozukluklar"* (disturbances) birbirlerini kurallı biçimde izlerlerse özilinti ortaya çıkar.
- ABİ'yi $Y_t = \beta_1 + \beta_2 X_t + u_t$ olarak kabul edelim ve u_t ile u_{t-1} de aynı yönde ilişkili olsun.
- Bu durumda, Y_t yalnızca X_t'ye değil u_t'ye de bağlı olur ve bu nedenle u_t'yi bir ölçüde u_{t-1} belirler.

- Bu sorunla karşılaşmamak için hatalar arasında *"serisel ilinti"* (serial correlation) olmadığı varsayılır.









Varsayım 6
Hata terimi u_i ile X_i'nin kovaryansı sıfırdır:

$$\text{cov}(u_i, X_i) = 0$$

- Eğer X ve u ilişkiliyse, ikisinin de Y üzerindeki tekil etkilerini bulmak olanaksızlaşır.
- Ayrıca, eğer X ile u aynı yönde ilişkiliyse, X arttıkça u da artarak farklıserpilimsellik sorununa yol açar.

- Eğer 2. varsayım (X'in rastsal olmaması) ve 3. varsayım ($E(u_i|X_i) = 0$) geçerliyse, 6. varsayım da kendiliğinden gerçekleşmiş olur.

Varsayım 7
Gözlem sayısı n, tahmin edilecek anakütle katsayısından fazla olmalıdır.

- İki bilinmeyeni (β_1 ve β_2) bulmak için en az iki noktaya gereksinim vardır.
- Bu koşul çözümlemenin matematiksel olarak yapılabilmesi için gereklidir.
- Diğer yandan n serbestlik derecesi açısından önemlidir. Bu nedenle sağlıklı sonuçlar için örneklemin yeterince büyük olmasının ayrıca gerekli olduğu unutulmamalıdır.

Varsayım 8
Belli bir örneklemdeki X değerlerinin hepsi aynı olamaz:

$$\text{var}(X) \neq 0$$

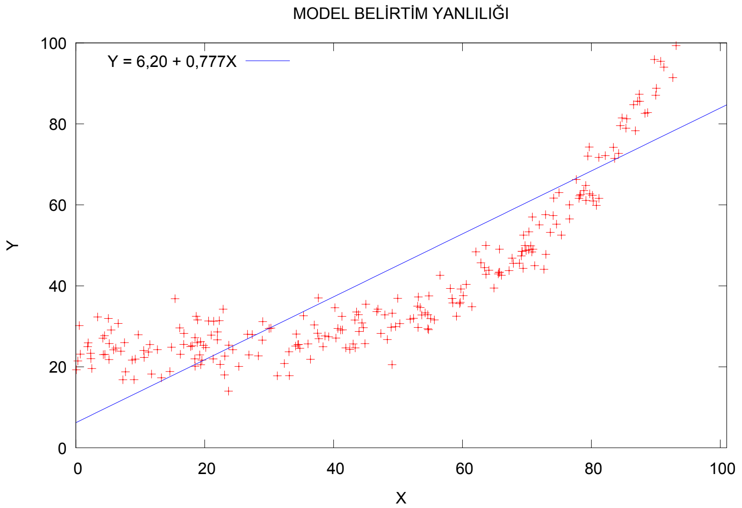
- Eğer bütün X değerleri aynı olursa:

$$\begin{array}{ll} X_i = \bar{X}, & \\ x_i = X_i - \bar{X} & \text{olduğundan} \\ \hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} & \text{formülünün paydası sıfır çıkar.} \end{array}$$

- Kısaca değişkenler değişmelidir.

Varsayım 9
Bağlanım modeli doğru biçimde belirtilmiş olmalıdır.

- Bağlanım çözümlemesi sonuçlarının güvenilirliği, seçilen modele bağlıdır.
- Özellikle de bir iktisadi olguyu açıklayan birden fazla kuram bulunuyor ise ekonometrici çok dikkatli olmalıdır.
- Her durumda modelin işlev biçiminin ne olduğu, değişken ve değiştirgelerde doğrusal olup olmadığı konuları iyice sorgulanmalıdır.
- Bağlanım modeli yanlış olduğu zaman *"model belirtim hatası"* (model specification error) ortaya çıkar.



Varsayım 10
"Tam çoklueşdoğrusallık" (exact multicollinearity) yoktur.

- Tam çoklueşdoğrusallık durumunda bağlanım katsayıları belirsiz ve bu katsayıların ölçünlü hataları da sonsuz olur.

KDBM Varsayımlarının Gerçekçiliği
Ünlü ekonomist Milton Friedman'ın "varsayımların yersizliği" tezine göre gerçek dışılık bir üstünlüktür:

> "Önemli olabilmek için . . . bir önsav, varsayımlarında betimsel olarak gerçek dışı olmalıdır."

- Ekonometrideki KDBM'nin, fiyat kuramındaki tam rekabet modelinin karşılığı olduğu söylenebilir.
- Diğer bir deyişle öne sürmüş olduğumuz bu 10 varsayım gerçekleri tümüyle yansıtmak için değil, konuyu yavaş yavaş geliştirebilmeyi kolaylaştırmak amacıyla önemlidir.
- Bu varsayımların gerçekleşmemesi durumunda doğacak sonuçları ise ilerideki bölümlerde inceleyeceğiz.

4.2 SEK Yönteminin Güvenilirliği

4.2.1 SEK Tahmincilerinin Ölçünlü Hataları

- Sıradan en küçük kareler tahmincilerinin örneklem verilerinin birer işlevi olduğunu anımsayalım:

$$\hat{\beta}_2 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \frac{\sum x_iy_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \bar{Y} - \hat{\beta}_2\bar{X}$$

- Veriler örneklemden örnekleme değişeceği için tahminler de buna bağlı olarak değişecektir.
- Öyleyse $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincilerinin güvenilirliği için bir ölçüte gereksinim vardır.
- İstatistikte rastsal bir değişkenin doğruluk derecesi *"ölçünlü hata"* (standard error), kısaca *"öh"* (se) ile ölçülür:

Ölçünlü Hata
Ölçünlü hata, bir tahminciye ait örneklem dağılımının kendi ortalamasından ortalama olarak ne kadar saptığını gösterir. Örneklem dağılımı varyansının artı değerli kare köküdür.

- Başta sözü edilmiş olan Gaussçu varsayımlar geçerli iken SEK tahmincilerinin ölçünlü hataları aşağıdaki gibidir:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$$\text{öh}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$$

$$\text{öh}(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \sigma$$

- Burada
  var değişirlik ya da varyansı,
  öh ölçünlü hatayı,
  σ^2 ise bağlanımın sabit varyansını
  göstermektedir.
- u_i'nin sabit varyansını veren σ^2 şöyle tahmin edilir:

$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n - 2}$$

- Buradaki $\hat{\sigma}^2$, bilinmeyen σ^2'nin SEK tahmincisidir.
- $\sum \hat{u_i}^2$ terimine *"kalıntı kareleri toplamı"* (residual sum of squares), kısaca *"KKT"* (RSS) denir ve şöyle bulunur:

$$\sum \hat{u_i}^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$$

- $n - 2$ değeri ise iki değişkenli çözümleme için geçerli serbestlik derecesidir.

Serbestlik Derecesi
"Serbestlik derecesi" (degree of freedom), örneklemdeki toplam gözlem sayısı (n) eksi bunlar üzerine konulmuş olan bağımsız ve doğrusal sınırlama sayısıdır.

- Örnek olarak, KKT'nin hesaplanabilmesi için önce $\hat{\beta}_1$ ve $\hat{\beta}_2$ değerlerinin bulunmuş olması gereklidir:

$$\sum \hat{u_i}^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

- Dolayısıyla bu iki tahminci KKT üzerine iki sınırlama getirir.
- Bu durumda, KKT'yi ve dolayısıyla da ölçünlü hatayı doğru hesaplayabilmek için aslında elde n değil $n - 2$ sayıda bağımsız gözlem vardır.

SEK tahmincileri $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın varyans formüllerini anımsayalım:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad \text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2$$

$\hat{\beta}_1$ ile $\hat{\beta}_2$ tahmincilerinin varyanslarının ve dolayısıyla bunların ölçünlü hatalarının şu özellikleri önemlidir:

1. Örneklem büyüklüğü n arttıkça $\sum x_i^2$ toplamındaki terim sayısı da artar. Böylece n büyüdükçe $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın doğruluk dereceleri de artar.
2. $\hat{\beta}_1$ ve $\hat{\beta}_2$, verili bir örneklemde birbirleri ile ilişkili olabilirler. Bu bağımlılık aralarındaki kovaryans ile ölçülür:

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X}\ \text{var}(\hat{\beta}_2)$$

3. Eğer $\bar{X}$ artı değerli ise kovaryans da eksi değerli olur. Bu durumda eğer β_2 katsayısı olduğundan büyük tahmin edilir ise β_1 de olduğundan küçük tahmin edilmiş olur.

4.2.2 Belirleme Katsayısı r^2

- Eldeki gözlemler çoğunlukla bağlanım doğrusu üzerinde yer almazlar.
- Artı ya da eksi işaretli $\hat{u}_i$ hataları ile karşılaşıldığına göre örneklem bağlanım doğrusunun eldeki verilerle ne ölçüde örtüştüğünü gösteren bir ölçüte gereksinim vardır:

Belirleme Katsayısı
"Belirleme katsayısı" (coefficient of determination) ya da r^2 (çoklu bağlanımda R^2), örneklem bağlanım işlevinin verilere ne kadar iyi yakıştığını gösteren özet bir ölçüttür.

Belirleme katsayısını hesaplamak için, $y_i = \hat{y}_i + \hat{u}_i$ eşitliğinin iki yanının karesi alınır ve örneklem boyunca toplanır:

$$\begin{aligned}
\sum y_i^2 &= \sum \hat{y}_i{}^2 + \sum \hat{u}_i{}^2 + 2\sum \hat{y}_i \hat{u}_i \\
&= \sum \hat{y}_i{}^2 + \sum \hat{u}_i{}^2 \\
&= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i{}^2 \\
\text{TKT} &= \text{BKT} + \text{KKT}
\end{aligned}$$

Burada
TKT *"Toplam Kareleri Toplamı"* (Total Sum of Squares),
BKT *"Bağlanım Kareleri Toplamı"* (Regression Sum of Squares),
KKT *"Kalıntı Kareleri Toplamı"* (Residual Sum of Squares)
anlamına gelmektedir.

- Yukarıdaki $\sum \hat{y}_i \hat{u}_i$ teriminin SEK bağlanım doğrusunun 3. özelliğinden dolayı sıfıra eşit olduğuna dikkat ediniz.

$$\begin{array}{ccccc} \sum y_i^2 & = & \hat{\beta}_2^2 \sum x_i^2 & + & \sum \hat{u}_i{}^2 \\ \text{TKT} & = & \text{BKT} & + & \text{KKT} \end{array}$$

Yukarıdaki eşitliğin her iki yanını TKT'ye bölelim:

$$1 = \tfrac{\text{BKT}}{\text{TKT}} + \tfrac{\text{KKT}}{\text{TKT}}$$

Buna göre r^2 aşağıdaki gibi tanımlanır:

Belirleme Katsayısı

$$r^2 = \frac{\sum \hat{y}_i{}^2}{\sum y_i^2} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{BKT}}{\text{TKT}} = 1 - \frac{\text{KKT}}{\text{TKT}}$$

r^2'nin iki temel özelliğinden söz edilebilir:

1. r^2 eksi değer almayan bir büyüklüktür.
2. Sınırları $0 \leq r^2 \leq 1$'dir.

Buna göre:

- Eğer $r^2 = 1$ olursa bu kusursuz bir yakışma demektir. Bu durumda rastsal hata yoktur ve tüm gözlemler bire bir bağlanım doğrusu üzerinde yer almaktadır.
- Sıfıra eşit bir r^2 ise bağımlı değişkenle açıklayıcı değişken arasında hiçbir ilişkinin olmadığı ($\hat{\beta}_2 = 0$) anlamına gelir.

r^2 ile yakın ilişkili ama kavramsal olarak çok uzak bir büyüklük *"ilinti katsayısı"* (coefficient of correlation), kısaca r'dir:

İlinti Katsayısı

$$r = \pm\sqrt{r^2}$$

- r değeri, bağımlı ve açıklayıcı değişkenler arasındaki doğrusal bağımlılığın bir ölçüsüdür.
- -1 ve $+1$ arasında yer alır: $-1 \leq r \leq 1$.
- Bakışımlıdır: $r_{XY} = r_{YX}$.
- Sıfır noktasından ve ölçekten bağımsızdır.
- Herhangi bir neden-sonuç ilişkisi içermez.
- İki değişken arasında sıfır ilinti ($r = 0$) mutlaka bağımsızlık göstermez çünkü r yalnızca doğrusal ilişkiyi ölçer.

4.2.3 Monte Carlo Yöntemi

- KDBM varsayımları altında SEK tahmincilerinin EDYT (En iyi Doğrusal Yansız Tahminci) olmalarını sağlayan bazı arzulanan özellikler taşıdıklarını anımsayalım.
- EDYT özelliklerinin geçerliliği, bir *"benzetim"* (simulation) yöntemi olan Monte Carlo deneyleri ile doğrulanabilir.
- Bu yöntem, anakütle katsayılarını tahmin eden süreçlerin istatistiksel özelliklerini incelemede sıkça kullanılmaktadır.
- Monte Carlo aynı zamanda istatistiksel çıkarsamanın temeli sayılan *"tekrarlı örnekleme"* (repeated sampling) kavramının anlaşılması için de yararlı bir araçtır.

Bir Monte Carlo deneyi aşağıdaki gibi yapılır:

1. Anakütle katsayıları seçilir. Örnek: $\beta_1 = 20$ ve $\beta_2 = 0{,}6$.
2. Bir örneklem büyüklüğü seçilir. Örnek: $n = 25$.
3. Her gözlem için bir X değeri belirlenir.
4. Bir rastsal sayı oluşturucu kullanılarak u_i kalıntıları üretilir.
5. β_1, β_2, X_i'ler ve u_i'ler kullanılarak Y_i değerleri bulunur.
6. Bu şekilde üretilen Y_i değerleri X_i'ler ile bağlanıma sokulur ve $\hat{\beta}_1$ ve $\hat{\beta}_2$ SEK tahmincileri hesaplanır.
7. İşlem tekrarlanır (örneğin 1000 kez) ve rastsallıktan dolayı her seferde değişen tahminlerin ortalamaları ($\bar{\hat{\beta}}_1, \bar{\hat{\beta}}_2$) alınır.
8. Eğer $\bar{\hat{\beta}}_1$ ve $\bar{\hat{\beta}}_2$ değerleri β_1 ve β_2'ye aşağı yukarı eşit ise, deney SEK tahmincilerinin yansızlığını, diğer bir deyişle $E(\hat{\beta}_1) = \beta_1$ ve $E(\hat{\beta}_2) = \beta_2$ olduğunu saptamış sayılır.

4.3 Sayısal Bir Örnek

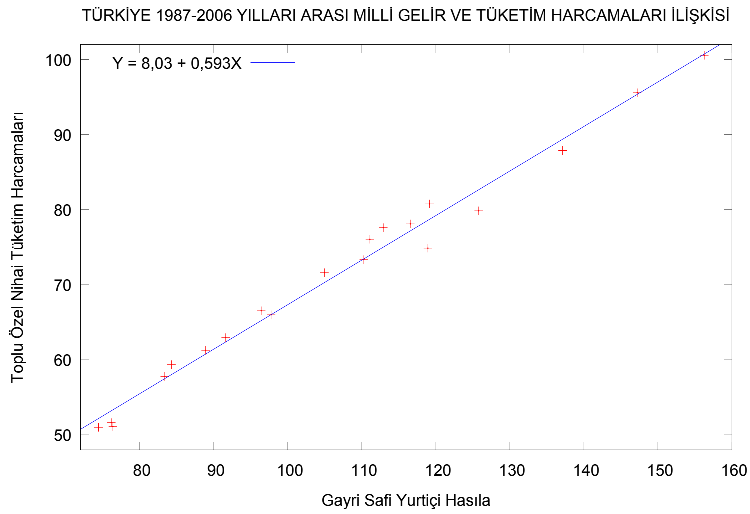
Sayısal Bir Örnek

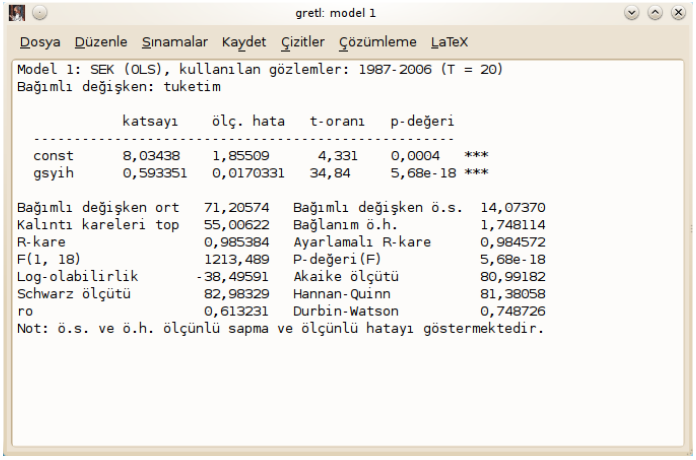
- Ele almış olduğumuz bazı kavramları sayısal bir örnek yardımı ile gözden geçirelim. Türkiye'de 1987–2006 arası toplam tüketim harcamaları ve GSYH verileri şöyledir:

Çizelge: Türkiye'de Tüketim ve GSYH (1987–2006)

Yıl	C	Y	Yıl	C	Y
1987	51.019	74.416	1997	77.620	112.892
1988	51.638	76.143	1998	78.113	116.541
1989	51.105	76.364	1999	76.077	111.083
1990	57.803	83.371	2000	80.774	119.147
1991	59.366	84.271	2001	73.356	110.267
1992	61.282	88.893	2002	74.894	118.923
1993	66.545	96.391	2003	79.862	125.778
1994	62.962	91.600	2004	87.897	137.110
1995	66.011	97.729	2005	95.594	147.200
1996	71.614	104.940	2006	100.584	156.249

- Toplu özel nihai tüketim harcamalarını (Y), gayri safi yurtiçi hasıla (X) ile ilişkilendirmek istiyor olalım.





- Gretl çıktısına göre marjinal tüketim eğilimi (MTE) 0,59'dur.
- Buna göre gelir 1 lira arttığında tüketimin de 59 kuruş artması beklenmektedir.
- Sabit terim, toplam gelir sıfır olduğunda toplam tüketimin yaklaşık 8 milyon lira olacağını göstermektedir.
- Sıfır gelirin gözlem aralığı dışında kalan ve gerçek hayatta olanaksız bir değer olmasından dolayı, sabit terimin böylesi bir mekanik yorumu iktisadi anlam içermemektedir.
- Gretl $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{u}_i$ için ölçünlü hataları sırasıyla 1,85509 ve 0,0170331 ve 1,748114 olarak hesaplamıştır.
- Yukarıdaki değerlerin karesi alınarak var$(\hat{\beta}_1) = 3{,}44136$ ve var$(\hat{\beta}_2) = 0{,}000290126$ ve $\hat{\sigma}^2 = 3{,}05590$ varyansları da kolayca bulunabilir.
- $r^2 = 0{,}985$ değeri ise bağlanım modelinin verilere gerçekçi kabul edilemeyecek kadar iyi yakıştığını göstermektedir.

Önümüzdeki Dersin Konusu ve Ödev

Ödev

Kitaptan *Bölüm 3* “Two-Variable Regression Model: The Problem of Estimation” okunacak.

Önümüzdeki Ders

Normallik Varsayımı ve Ençok Olabilirlik Yöntemi

Bölüm 5

Normallik Varsayımı ve Ençok Olabilirlik Yöntemi

5.1 Normallik Varsayımı ve İlişkin Dağılımlar

5.1.1 Hata Teriminin Olasılık Dağılımı

Ekonometrik çözümlemede amaç yalnızca ÖBİ'yi hesaplamak değil, aynı zamanda ABİ'ye ilişkin çıkarsama ve çeşitli önsav sınamaları da yapabilmektir. Bu doğrultuda u_i hatalarının olasılık dağılımının bilinmesi ya da belirlenmesi iki nedenden dolayı önemlidir:

1. SEK bir katsayı hesaplama yöntemidir. ÖBİ'den ABİ'ye yönelik çıkarsama yapmada tek başına işe yaramaz.

2. $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincileri Y'nin doğrusal işlevi ve Y'nin kendisi de u_i'lerin bir doğrusal işlevidir. Öyleyse, $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın örneklem dağılımları u_i'lerin olasılık dağılımına ilişkin varsayımlara dayanmaktadır.

- Daha önce ele alınan klasik doğrusal bağlanım modeli (KDBM), u_i hata teriminin olasılık dağılımı ile ilgili herhangi bir varsayımda bulunmaz.

- Almaşık olarak, *"Klasik Normal Doğrusal Bağlanım Modeli"* (Classical Normal Linear Regression Model) ya da kısaca *"KNDBM"* (CNLRM) ise u_i'lerin normal dağıldığını varsayar.

Normallik Varsayımı

KNDBM, her bir u_i'nin aşağıdaki değerlerle normal dağıldığı varsayımını getirir:

Ortalama:	$E(u_i) = 0$
Varyans:	$E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$
Kovaryans:	$\text{cov}(u_i, u_j) = 0, \qquad i \neq j$

- Bu varsayımlar kısaca $u_i \sim N(0, \sigma^2)$ şeklinde de gösterilir.
- Buradaki ($\sim$), *"dağılımlı"* (distributed) anlamına gelir. N ise normal dağılımı göstermektedir.
- İki rastsal değişkenin kovaryansının sıfır olması bu iki değişkenin bağımsız olduğunu gösterir.
- Bu nedenle $u_i \sim N(0, \sigma^2)$ yerine $u_i \sim \text{NBD}(0, \sigma^2)$ de yazılabilir.
- Burada NBD *"normal ve bağımsız dağılımlı"* (normally and independently distributed, kısaca NID) demektir.

Normallik varsayımının nedenleri şunlardır:

1. Merkezi limit kanıtsavına göre bağımsız ve özdeş dağılımlı (BÖD) rastsal değişkenlerin toplam dağılımı, değişken sayısı sonsuza yaklaştıkça normale yakınsar.
2. Merkezi limit kanıtsavına göre değişken sayısı çok fazla olmasa ya da değişkenler tam bağımsız dağılmasalar bile toplamları normal dağılabilir.
3. Normal dağılan değişkenlerin doğrusal işlevleri de normal dağılır. (*Örnek:* $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{\sigma}^2$)
4. Normal dağılım yalnızca iki katsayı (ortalama ve varyans) içeren basit, istatistiksel özellikleri iyi bilinen bir dağılımdır.

Normallik Varsayımı Altında SEK Tahmincileri

$\hat{\beta}_1$ ve $\hat{\beta}_2$ SEK tahmincileri ile $\hat{\sigma}^2$ varyans tahmini, normallik varsayımı altında şu istatistiksel özellikleri taşırlar:

1. Yansızdırlar. Diğer bir deyişle beklenen değerleri gerçek değerlerine eşittir. *Örnek:* $E(\hat{\beta}_1) = \beta_1$.
2. $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincileri, doğrusal ve doğrusal-dışı tüm yansız tahminciler içinde enaz varyanslıdır.
3. Tutarlıdırlar. Örneklem sonsuza doğru büyürken gerçek değerlerine yakınsarlar.

4. $\hat{\beta}_1$ şöyle dağılır: $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$.

5. $\hat{\beta}_2$ şöyle dağılır: $\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$.

6. $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincileri $\hat{\sigma}^2$'dan bağımsız olarak dağılırlar.

7. $(n-2)\hat{\sigma}^2/\sigma^2$ değeri ise $n-2$ sd ile χ^2 dağılımlıdır. Bu bilgi, σ^2'ye ilişkin çıkarsamalarda $\hat{\sigma}^2$'den yararlanabilmek içindir.

- Normallik varsayımı yardımı ile $\hat{\beta}_1$ (normal), $\hat{\beta}_2$ (normal) ve $\hat{\sigma}^2$ (ki-kare ile ilgili) örneklem dağılımı bilgilerine ulaşıyoruz.
- Bu durum güven aralıklarını belirlemek ve önsav sınaması yapabilmek için önemlidir.
- *Dikkat:* Y_i'ler u_i'lerin bir işlevi olduğuna göre, $u_i \sim N(0, \sigma^2)$ varsayımı altında $Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$ olur.

5.1.2 Normal Dağılıma İlişkin Dağılımlar

- F, t ve ki-kare (χ^2) olasılık dağılımları temelde normal dağılımla ilişkilidirler.
- Bu dağılımların normal dağılımla ilişkilerini özetleyen yedi kanıtsav bulunmaktadır.
- Bu kanıtsavların uygulamadaki önemi büyüktür. Yararları ileride daha iyi anlaşılacaktır.

Kanıtsav 1
$Z_1, Z_2, \ldots, Z_n$ değişkenleri $Z_i \sim N(\mu_i, \sigma_i^2)$ olan normal ve bağımsız dağılımlı rastsal değişkenler olsun. Bu durumda $Z = \sum k_i Z_i$ toplamı da aşağıda verilen ortalama ve varyans ile normal dağılır.

$$E(Z) = \sum k_i \mu_i$$
$$\text{var}(Z) = \sum k_i^2 \sigma_i^2$$

- Kısaca normal rastsal değişkenlerin doğrusal bileşimleri de normal dağılır.
- *Örnek:* $Z_1 \sim N(10, 2)$ ve $Z_2 \sim N(8; 1{,}5)$ varsayalım ve Z rd'si de $Z = 0{,}8Z_1 + 0{,}2Z_2$ olsun. Buna göre Z

$$0{,}8(10) + 0{,}2(8) = 9{,}6 \text{ ortalama ve}$$
$$0{,}64(2) + 0{,}04(1{,}5) = 1{,}34 \text{ varyans ile}$$

$Z \sim N(9{,}6; 1{,}34)$ olur.

Kanıtsav 2
$Z_1, Z_2, \dots, Z_n$ rastsal değişkenleri normal dağılımlı olsun ancak bağımsız olmasın. Bu durumda $Z = \sum k_i Z_i$ toplamı da aşağıda verilen ortalama ve varyans ile normal dağılır.

$$E(Z) = \sum k_i \mu_i$$
$$\text{var}(Z) = [\sum k_i^2 \sigma_i^2 + 2 \sum k_i k_j \text{cov}(Z_i, Z_j), i \neq j]$$

- *Örnek:* $Z_1 \sim N(6, 2)$, $Z_2 \sim N(7, 3)$, $\text{cov}(Z_1, Z_2) = 0{,}8$ olsun. Z rd'si de $Z = 0{,}6Z_1 + 0{,}4Z_2$ olsun. Buna göre Z

$$0{,}6(6) + 0{,}4(7) = 6{,}4 \text{ ortalama ve}$$
$$0{,}36(2) + 0{,}16(3) + 2(0{,}6)(0{,}4)(0{,}8) = 1{,}58 \text{ varyans ile}$$

$Z \sim N(6{,}4; 1{,}58)$ olur.

Kanıtsav 3
$Z_1, Z_2, \dots, Z_n$ değişkenleri $Z_i \sim N(0, 1)$ ölçünlü normal ve aynı zamanda bağımsız rastsal değişkenler olsun. Bu durumda $\sum Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ toplamı da n serbestlik derecesi ile ki-kare dağılımına uyar.

- Kısaca ölçünlü normal dağılımlı bağımsız rd'lerin kareleri toplamı, toplam terim sayısına eşit serbestlik derecesi ile ki-kare dağılımına uyar.
- Bir yazım kolaylığı olarak sd'si k olan χ^2 dağılımı χ_k^2 diye gösterilebilir.
- *Örnek:* $Z_1, Z_2, Z_3 \sim NBD(0, 1)$ olsun. Öyleyse şu geçerlidir:

$$Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

Kanıtsav 4
$Z_1, Z_2, \dots, Z_n$ değişkenleri her birinin sd'si k_i olmak üzere χ^2 dağılımlı ve bağımsız rastsal değişkenler olsun. Bu durumda bunların toplamı olan $\sum Z_i = Z_1 + Z_2 + \cdots + Z_n$ rastsal değişkeni de sd'si $k = \sum k_i$ olan ki-kare dağılımına uyar.

- *Örnek:* Z_1 ve Z_2, sd'leri sırasıyla 7 ve 9 olan iki bağımsız ki-kare değişkeni olsun. Buna göre $Z_1 + Z_2$ de serbestlik derecesi $7 + 9 = 16$ olan bir χ_{16}^2 değişkenidir.

Kanıtsav 5
$Z_1 \sim N(0,1)$ ölçünlü normal ve Z_2 de Z_1'den bağımsız ve k sd ile χ^2 dağılımına uyan bir rastsal değişken olsun. Bu durumda $t = Z_1/(\sqrt{Z_2/k})$ şeklinde tanımlanan değişken de k sd ile Student t dağılımına uyar.

- Serbestlik derecesi k sonsuza doğru yaklaştıkça Student t dağılımı ölçünlü normal dağılıma yakınsar.
- Bir yazım kolaylığı olarak k sd'li t dağılımı t_k diye gösterilir.

Kanıtsav 6
Z_1 ve Z_2, serbestlik dereceleri sırasıyla k_1 ve k_2 olan ve bağımsız dağılımlı birer ki-kare değişkeni olsun. Bu durumda $F = (Z_1/k_1)/(Z_2/k_2)$ olarak tanımlanan F değişkeni de k_1 pay ve k_2 payda serbestlik derecesi ile F dağılımına uyar.

- Diğer bir deyişle, F değişkeni kendi sd'lerine bölünmüş iki bağımsız χ^2 değişkeni arasındaki oranı gösterir.
- Bir yazım kolaylığı olarak, sd'leri k_1 ve k_2 olan F dağılımlı değişken F_{k_1,k_2} diye gösterilir.

Kanıtsav 7
Serbestlik derecesi k olan t rastsal değişkeninin karesi, pay sd'si $k_1 = 1$ ve payda sd'si $k_2 = k$ ile F dağılımlıdır.

- *Örnek:* $F_{1,4} = (t_4)^2$ 'dir.
- *Örnek:* $(t_{25})^2 = F_{1,25}$ olur.

5.2 Ençok Olabilirlik Yöntemi

5.2.1 Ençok Olabilirlik Yaklaşımı

- İstatistikte tüm anakütleler kendilerine karşılık gelen bir olasılık dağılımı ile tanımlanırlar.
- Sıradan en küçük kareler yöntemi ise özünde olasılık dağılımları ile ilgili herhangi bir varsayım içermez.
- Bu nedenle çıkarsama yapmada SEK tek başına bir işe yaramaz.
- SEK'i genel bir tahmin süreci olarak değil de örneklem bağlanım işlevlerinin katsayılarını bulmada kullanılan bir hesaplama yöntemi olarak görmeliyiz.
- SEK yönteminden daha güçlü kuramsal özellikler gösteren bir diğer nokta tahmincisi ise *"ençok olabilirlik"* (maximum likelihood), kısaca *"EO"* (ML) yöntemidir.
- Ençok olabilirlik yönteminin ardında yatan temel ilke şu beklentidir:

> "Rastsal bir olayın gerçekleşmesi, o olayın gerçekleşme olasılığı en yüksek olay olmasındandır."

- Bu yöntem 1920'li yıllarda İngiliz istatistikçi Sir Ronald A. Fisher (1890-1962) tarafından bulunmuştur.
- Ki-kare sınaması, Bayesçi yöntemler ve çeşitli ölçüt modelleri gibi birçok istatistiksel çıkarım yöntemi temelde EO yaklaşımına dayanır.
- EO yöntemini anlayabilmek için elimizde rastsal olarak belirlenmiş bir örneklem ve dağılım katsayıları bilinen farklı anakütle adayları olduğunu varsayalım.
- Bu örneklemin farklı anakütlelerden gelme olasılığı farklı ve bazı ana kütlelerden gelme olasılığı diğerlerine göre daha yüksektir.
- Elimizdeki örneklem eğer bu anakütlelerden birinden alınmışsa, alınma olasılığı ençok olan anakütleden alınmış olduğunu tahmin etmek akılcı bir yaklaşımdır.

Ençok olabilirlik yöntemi kısaca şöyledir:

1. Anakütlenin olasılık dağılımı belirlenir ya da bu yönde bir varsayım yapılır.
2. Eldeki örneklem verilerinin gelmiş olma olasılığının ençok olduğu anakütlenin hangi katsayılara sahip olduğu bulunur.

5.2.2 İkiterimli Dağılım Örneği

Ençok olabilirlik yöntemini daha iyi anlayabilmek için şu basit örneği ele alalım:

- Elimizde, içinde siyah ya da beyaz toplam on top bulunan değişik torbalar olsun.
- Torbadaki siyah top sayısı $i = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ olmak üzere 11 farklı torba olasıdır.
- Bu torbalardan birinden *"yerine koyarak"* (with replacement) dört top seçtiğimizi ve S-B-S-B sırasıyla 2 siyah top geldiğini varsayalım.
- Bu sonucun hangi torbadan gelmiş olabileceğini ençok olabilirlik yaklaşımı ile tahmin edelim.
- Elimizdeki soru *"ikiterimli"* (binomial, Bi) dağılım konusudur.
- İkiterimli dağılıma göre, örnek olarak, 8'i siyah olmak üzere içinde 10 top olan bir torbadan yerine koyarak çekilen 4 toptan 2'sinin (belli bir sıra ile) siyah gelme olasılığı şudur:

$$\text{Bi}(2|4, \frac{8}{10}) = \left(\frac{8}{10}\right)^2 \left(1 - \frac{8}{10}\right)^{4-2} = 0{,}0256$$

- Torbadaki siyah top oranı p olsun. Örneğimizde p için 11 farklı değer söz konusudur:

$$p = \{\tfrac{0}{10}, \tfrac{1}{10}, \dots, \tfrac{10}{10}\}$$

- Bu 11 farklı torba için, 4 toptan 2'sinin siyah gelmesi durumunun gerçekleşme olasılığı şöyle gösterilebilir:

$$\text{Bi}(2|4, p) = p^2(1 - p)^{4-2}$$

Çizelge: Siyah Top Sayısına Göre Olasılığın Aldığı Değerler

Siyah Top Sayısı	Olasılık				
0	$\text{Bi}(2\|4, \frac{0}{10})$	=	$(0)^2(1)^{4-2}$	=	0
1	$\text{Bi}(2\|4, \frac{1}{10})$	=	$(0{,}1)^2(0{,}9)^{4-2}$	=	0,0081
2	$\text{Bi}(2\|4, \frac{2}{10})$	=	$(0{,}2)^2(0{,}8)^{4-2}$	=	0,0256
3	$\text{Bi}(2\|4, \frac{3}{10})$	=	$(0{,}3)^2(0{,}7)^{4-2}$	=	0,0441
4	$\text{Bi}(2\|4, \frac{4}{10})$	=	$(0{,}4)^2(0{,}6)^{4-2}$	=	0,0576
5	$\textit{Bi}(2\|4, \frac{5}{10})$	=	$(0{,}5)^2(0{,}5)^{4-2}$	=	*0,0625*
6	$\text{Bi}(2\|4, \frac{6}{10})$	=	$(0{,}6)^2(0{,}4)^{4-2}$	=	0,0576
7	$\text{Bi}(2\|4, \frac{7}{10})$	=	$(0{,}7)^2(0{,}3)^{4-2}$	=	0,0441
8	$\text{Bi}(2\|4, \frac{8}{10})$	=	$(0{,}8)^2(0{,}2)^{4-2}$	=	0,0256
9	$\text{Bi}(2\|4, \frac{9}{10})$	=	$(0{,}9)^2(0{,}1)^{4-2}$	=	0,0081
10	$\text{Bi}(2\|4, \frac{10}{10})$	=	$(1)^2(0)^{4-2}$	=	0

- Çizelgeye bakarak eldeki örneklemin ençok olasılıkla siyah top sayısı 5 olan torbadan alınmış olduğunu tahmin ederiz.

5.2.3 İkiterimli Dağılım EO Tahmincisi

Tanımlamış olduğumuz ikiterimli dağılım sorusunu şimdi bir de *"çözümlemesel"* (analytical) olarak ele alalım:

- Elimizde içinde kaç siyah ve beyaz top olduğu bilinmeyen bir torba olsun.
- Torbadaki siyah top oranına $0 \leq p \leq 1$ diyelim.
- İlk örnekte 4 toptan oluşan bir örneklem alınmıştı. Şimdi ise örneklem büyüklüğü n, çıkan siyah top sayısı da k olsun.
- Farklı n ve k sonuçları veren toplam N sayıda bağımsız çekiliş yapalım.
- Ençok olabilirlik yöntemini kullanarak anakütle katsayısı p'yi tahmin etmek istiyor olalım.
- Eldeki sorunun ikiterimli dağılımı ilgilendirdiğini biliyoruz.
- İstatistikte ikiterimli dağılım, "başarı" olasılığı p olan n bağımsız deneyde başarılı olan k'lerin dağılımını gösteren bir kesikli olasılık dağılımıdır.
- Olasılık yoğunluk işlevi şudur:

$$\mathrm{Bi}(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5.1}$$

- Yukarıda verilen OYİ sırasız çekilişler içindir. Matematiksel kolaylık açısından sonuçların belirli bir sırayı izlemesi gerektiğini varsayalım.
- Bu durumda kesikli OYİ şu olur:

$$p^k (1-p)^{n-k} \tag{5.2}$$

- Elimizdeki olasılık yoğunluk işlevi şuydu:

$$\mathrm{Bi}(k|n,p) = p^k (1-p)^{n-k} \tag{5.3}$$

- Toplam N sayıdaki çekiliş için birleşik yoğunluk işlevi:

$$\mathrm{Bi}(k_i|n_i,p) = \mathrm{Bi}(k_1|n_1,p)\,\mathrm{Bi}(k_2|n_2,p)\ldots\mathrm{Bi}(k_N|n_N,p) \tag{5.4}$$

- Her bir n_i ve k_i için, (5.3)'ü (5.4)'te yerine koyalım:

$$\text{Bi}(k_i|n_i,p) = p^{\sum k_i}(1-p)^{\sum n_i - k_i} \tag{5.5}$$

- $n_1, n_2 \ldots n_N$ ve $k_1, k_2 \ldots k_N$ değerleri veriliyken anakütle katsayısı p eğer bilinmiyorsa, yukarıda gösterilen işleve *"olabilirlik işlevi"* (likelihood function) adı verilir:

$$\text{Oİ}(p) = p^{\sum k_i}(1-p)^{\sum n_i - k_i} \tag{5.6}$$

- Adından da anlaşılacağı gibi EO tahmini, verili n_i ve k_i'leri gözleme olasılığını ençoklamaya dayanır.
- Öyleyse, hedefimiz olabilirlik işlevinin *"ençoksal"* (maximal) değerini bulmak olmalıdır.
- Bu da doğrudan bir türev hesabıdır.
- Bir işlev kendi logaritması ile *"tekdüze"* (monotonous) ilişkilidir. Bu nedenle olasılık işlevi yerine *"log-olasılık"* (log-likelihood) işlevini ençoklamak hesap kolaylığı sağlar:

$$\ln \text{Oİ}(p) = \sum_{i=1}^{N} k_i \ln(p) + \sum_{i=1}^{N}(n_i - k_i)\ln(1-p) \tag{5.7}$$

- (5.7) eşitliğinin p'ye göre kısmi türevini alıp sıfıra eşitleyelim:

$$\frac{\partial \ln \text{Oİ}}{\partial p} = \sum_{i=1}^{N} k_i \frac{1}{p} + \sum_{i=1}^{N}(n_i - k_i)\frac{1}{1-p}(-1) = 0 \tag{5.8}$$

- Sadeleştirmelerden sonra, EO tahmincisi $\tilde{p}$ şöyle bulunur:

$$\tilde{p} = \frac{\sum_{i=1}^{N} k_i}{\sum_{i=1}^{N} n_i} \tag{5.9}$$

- p üzerindeki ($\sim$) *"dalga"* (tilde) imi, bunun bir EO tahmincisi olduğunu göstermek için kullanılmıştır.
- Görülüyor ki EO yöntemi anakütledeki siyah top oranı k değerini, N çekiliş sonunda bulunan siyah top sayısının çekilen toplam top sayısına oranı olarak tahmin etmektedir.

5.3 Açıklayıcı Örnekler

5.3.1 Poisson Dağılımı EO Tahmincisi

EO yöntemine bir örnek olarak Poisson dağılımını ele alalım. Bu kesikli dağılım, verili bir süre ya da uzaysal alan içerisinde bir olayın belirli bir sayıda tekrarlanma olasılığını anlatır. Örnek olarak, 5 dakikalık süre içinde belli bir noktadan geçen araç sayısı Poisson dağılımına uyan bir rastsal değişkendir. Poisson dağılımının genel gösterimi şöyledir:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{5.10}$$

- Burada x olayın gerçekleşme sayısını, λ ise belirli bir süredeki beklenen gerçekleşme sayısını göstermektedir.
- Artı değerli bir gerçel sayı olan λ'ya ait EO tahmincisini bulmak istiyoruz.
- Poisson dağılımının matematiksel gösterimini alalım ve n sayıda gözlem için olabilirlik işlevini aşağıdaki gibi yazalım.

$$\begin{aligned} \text{Oİ}(\lambda) &= f(x_1)f(x_2)\dots f(x_n) \\ &= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\dots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\ &= \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!} \end{aligned} \tag{5.11}$$

- İki yanlı logaritmasını alarak log-olabilirlik işlevini bulalım.

$$\ln \text{Oİ} = -n\lambda + \sum x_i \ln\lambda - \ln\prod x_i! \tag{5.12}$$

- Log-olabilirlik işlevinin türevini alıp sıfıra eşitleyelim:

$$\frac{\partial \ln \text{Oİ}}{\partial\lambda} = -n + \sum x_i\frac{1}{\lambda} = 0 \tag{5.13}$$

- Yukarıdaki eşitliği λ'ya göre çözersek şunu buluruz:

$$\tilde{\lambda} = \frac{\sum x_i}{n} \tag{5.14}$$

- Demek ki EO yöntemi λ değiştirgesinin tahmincisini x'in örneklem ortalaması olarak bulmaktadır.

5.3.2 Üstel Dağılım EO Tahmincisi

İkinci bir örnek olarak *"üstel"* (exponential) dağılıma bakalım. Bu sürekli dağılım, olayların belli bir sabit hızda yinelendiği Poisson türü bir süreçteki gerçekleşmeler arası süreyi anlatır. Örnek olarak, belli bir noktadan rastsal olarak geçen araçlar arasında geçen süre üstel dağılıma uyan bir rastsal değişkendir. Üstel dağılımın olasılık yoğunluk işlevi ise şöyledir:

$$\begin{aligned} f(x) &= \left(\tfrac{1}{\theta}\right) e^{-x/\theta} && X > 0 \text{ için,} \\ &= 0 && X \leq 0 \text{ için.} \end{aligned} \tag{5.15}$$

- Burada x süreyi, θ ise dağılıma ait *"hız"* (rate) değiştirgesini göstermektedir.
- Şimdi de θ'nın EO tahmincisini türetmek istiyoruz.
- Baştaki örneklerde yaptığımız gibi, önce n gözlem için olabilirlik işlevini bulalım.

$$\text{Oİ} = \frac{1}{\theta^n} \exp\left(\sum \frac{-x_i}{\theta}\right) \tag{5.16}$$

- Log-olabilirlik işlevini yazalım, türevini alıp sıfıra eşitleyelim:

$$\ln \text{Oİ} = -n \ln \theta - \sum \frac{x_i}{\theta} \tag{5.17}$$

$$\frac{\partial \ln \text{Oİ}}{\partial \theta} = -\frac{n}{\theta} + \sum \frac{x_i}{\theta^2} = 0 \tag{5.18}$$

$$\tilde{\theta} = \frac{\sum x_i}{n} \tag{5.19}$$

- Demek ki örneklem büyüklüğü n iken θ değiştirgesinin EO tahmincisi $\tilde{\theta} = \sum x_i/n$ olarak bulunmaktadır.

5.3.3 Normal Dağılım EO Tahmincisi

Son olarak, şimdi de $Y_i = \beta_1 + \beta_2 X_i + u_i$ iki değişkenli bağlanım modelini EO yöntemi ile tahmin edelim.

- Bunun için önce hata teriminin sıfır ortalama ile normal ve bağımsızca dağıldığını ($u_i \sim NBD(0, \sigma^2)$) varsayalım.

- $X \sim N(\mu, \sigma^2)$ rastsal değişken X'in olasılık yoğunluk işlevi aşağıdaki gibidir:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \tag{5.20}$$

- Yukarıdaki $\exp$ işlemcisi e üzeri anlamına gelmektedir.

- Hataların $u_i \sim N(0, \sigma^2)$ olduğunu varsaydığımıza göre $Y_i \sim N(\beta_1{+}\beta_2 X_i, \sigma^2)$'dir. Diğer bir deyişle, Y_i değerleri $\beta_1 + \beta_2 X_i$ ortalama ve σ^2 varyans ile normal dağılırlar.

- Buna göre tek bir Y'nin olasılık yoğunluk işlevi şudur:

$$f(Y|\beta_1 + \beta_2 X, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(Y-\beta_1-\beta_2 X)^2}{\sigma^2}\right\} \tag{5.21}$$

- Birbirinden bağımsız $i \in \{1, 2, \ldots, n\}$ sayıda Y_i'nin ortak olasılık yoğunluk işlevi ise n tekil OYİ'nin çarpımıdır:

$$f(Y_1|\beta_1 + \beta_2 X_1, \sigma^2)\ f(Y_2|\beta_1 + \beta_2 X_2, \sigma^2) \ldots f(Y_n|\beta_1 + \beta_2 X_n, \sigma^2) \tag{5.22}$$

- Elimizdeki n gözlemdeki her bir Y_i ve X_i için (5.21)'i (5.22)'de yerine koyarsak, olabilirlik işlevini buluruz:

$$\text{Oİ}(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n\sqrt{2\pi}^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i-\beta_1-\beta_2 X_i)^2}{\sigma^2}\right\} \tag{5.23}$$

- Bu denklemin doğal logaritmasını alırsak da log-olabilirlik işlevini elde ederiz:

$$\ln \text{Oİ} = -n\ln(\sigma) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i-\beta_1-\beta_2 X_i)^2}{\sigma^2} \tag{5.24}$$

- $\ln$ Oİ değerini ençoklamak en sondaki terimi enazlamak demektir. Bu da en küçük kareler yaklaşımı ile aynı şeydir.

- Log olabilirlik işlevinin β_1, β_2 ve σ^2'ye göre kısmi türevleri alınırsa şu eşitlikler bulunur:

$$\frac{\partial \ln \text{Oİ}}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_i)(-1) \tag{5.25}$$

$$\frac{\partial \ln \text{Oİ}}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_i)(-X_i) \tag{5.26}$$

$$\frac{\partial \ln \text{Oİ}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_i)^2 \tag{5.27}$$

(... devam)

- (5.25), (5.26) ve (5.27) sıfıra eşitlenip birlikte çözüldükten sonra ise aşağıdaki EO tahmincileri elde edilir:

$$\tilde{\beta}_1 = \bar{Y} - \tilde{\beta}_2 \bar{X} \tag{5.28}$$

$$\tilde{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{5.29}$$

$$\tilde{\sigma}^2 = \frac{\sum \hat{u}_i{}^2}{n} \tag{5.30}$$

- Görüldüğü gibi u_i'lerin normal dağıldığı varsayımı altında β bağlanım katsayılarının EO ve SEK tahmincileri aynıdır.
- Diğer yandan σ^2 tahminleri farklıdır:

SEK tahmincisi

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i{}^2}{n-2} \tag{5.31}$$

EO tahmincisi

$$\tilde{\sigma}^2 = \frac{\sum \hat{u}_i{}^2}{n} \tag{5.32}$$

- Buna göre σ^2'nin SEK tahmincisi yansızken EO tahmincisi aşağı doğru yanlıdır.
- Ancak kavuşmazsal olarak, diğer bir deyişle n sonsuza yaklaştıkça, EO tahmincisi de yansızlaşır.
- Öyleyse, σ^2'nin EO tahmincisi *"kavuşmazsal yansızlık"* (asymptotic unbiasedness) özelliğini taşımaktadır.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 4* “Classical Normal Linear Regression Model (CNLRM)” okunacak.

Önümüzdeki Ders
İki Değişkenli Bağlanım: Aralık Tahmini ve Önsav Sınaması

Bölüm 6

İki Değişkenli Bağlanım Modeli - Çıkarsama Sorunu

6.1 Aralık Tahmini

6.1.1 Bazı Temel Noktalar

- Yansız SEK tahmincilerinin ürettiği tahminlerin anakütle değerlerine eşit olması beklenir.
- Ancak, örneklemlerin rastsallığı nedeniyle sonuçların gerçek değerlerden farklı çıkabileceği de bir gerçektir.
- Hata teriminin normalliği varsayımı altında $\hat{\beta}_1$, $\hat{\beta}_2$, ve $\hat{\sigma}^2$ tahmincilerinin dağılımları ile ilgili şu bilgileri anımsayalım:

$$\begin{gathered}\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \\ \hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2}) \\ Z = (n-2)\tfrac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}\end{gathered}$$

- Rastsallık etmeni nedeniyle tahminlerin gerçek değerlerine ne kadar yakın olduğunu bilmek isteriz.
- Öyleyse, yalnızca nokta tahminine güvenmek yerine onun iki yanında öyle bir aralık oluşturalım ki anakütlenin gerçek katsayısını belli bir olasılıkla içersin:

$$P(\hat{\beta} - \delta \leq \beta \leq \hat{\beta} + \delta) = 1 - \alpha$$

- Buradaki
  $0 < \alpha < 1$'e *"anlamlılık düzeyi"* (significance level),
  $1 - \alpha$'ya *"güven katsayısı"* (confidence coefficient),
  $\hat{\beta} - \delta$'ya *"alt güven sınırı"* (lower confidence limit),
  $\hat{\beta} + \delta$'ya ise *"üst güven sınırı"* (upper confidence limit)
  adı verilir.

Aralık tahminine ilişkin bazı önemli noktalar şunlardır:

- Tanımlanan aralık rastsal bir aralıktır ve bir örneklemden diğerine değişecektir.
- Eğer $\alpha = 0{,}05$ ise, tanımlanan rastsal aralığın gerçek β değerini içerme olasılığı 0,95 ya da %95'tir.
- Belli bir örneklem alınarak bulunan sabit aralığın gerçek β'yı içerme olasılığının ise $(1 - \alpha)$ olduğu söylenmez.
- Çünkü, böyle bir durumda β ya bu aralığın içindedir ya da dışındadır. Diğer bir deyişle olasılık ya 1'dir ya da 0'dır.

6.1.2 SEK Tahmincilerinin Güven Aralıkları

β_1 ve β_2 İçin Güven Aralığı

- Hata teriminin normalliği varsayımı altında $\hat{\beta}_1$ ve $\hat{\beta}_2$ SEK tahmincilerinin normal dağılımlı olduğunu biliyoruz.
- Öyleyse, bir ölçünlü normal değişken olan Z'yi aşağıdaki gibi tanımlayabiliriz:

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\text{öh}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma}$$

- Demek ki anakütlenin gerçek varyansı σ^2 biliniyorsa, β_2'yi incelemek için normal dağılımdan yararlanılabilir.
- Ancak, σ^2 genellikle bilinemediği için uygulamada yansız tahmincisi $\hat{\sigma}^2$ kullanılır.
- σ^2 bilinmediği zaman bunun yerine yansız tahminci $\hat{\sigma}^2$ aşağıda gösterilen şekilde kullanılır:

$$\begin{aligned} Z_1 &= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \\ Z_2 &= (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \\ t &= \frac{Z_1}{\sqrt{Z_2/(n-2)}} \\ &= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \end{aligned}$$

- Demek ki, normal dağılan Z_1'in ki-kare dağılan Z_2'nin kendi serbestlik derecesine bölümünün kareköküne bölünmesi ile elde edilen t rastsal değişkeni, $n-2$ sd ile t dağılımlıdır.
- Bu işlem β_1 için öh$(\hat{\beta}_1) = \sqrt{\sum X_i^2/n\sum x_i^2}\sigma$ olması dışında benzerdir.
- Normal dağılım yerine t dağılımı kullanıldığı zaman β_1 için güven aralığı aşağıdaki gibi kurulur:

$$\begin{aligned} P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) &= 1-\alpha \\ P\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\text{öh}(\hat{\beta}_1)} \leq t_{\alpha/2}\right] &= 1-\alpha \end{aligned}$$

- Buradaki $t_{\alpha/2}$ değeri, $\alpha/2$ anlamlılık düzeyinde ve $(n-2)$ serbestlik derecesi için t dağılımından bulunan t değeridir.
- Bu $t_{\alpha/2}$ değerine $\alpha/2$ anlamlılık düzeyindeki *"kritik t değeri"* (critical t value) adı verilir.
- Normal dağıldığı bilinen β_2'nin güven aralığı da benzer şekilde bulunur.
- β_1 ve β_2'nin $\%100(1-\alpha)$ güven aralıkları kısaca aşağıdaki gibi de gösterilebilir:

$$\begin{aligned} &\hat{\beta}_1 \pm t_{\alpha/2}\text{öh}(\hat{\beta}_1) \\ &\hat{\beta}_2 \pm t_{\alpha/2}\text{öh}(\hat{\beta}_2) \end{aligned}$$

- Her iki durumda da güven aralığının genişliği tahmincinin ölçünlü hatası ile doğru orantılıdır.
- Zaman zaman β_1 ve β_2 için bir *"birleşik güven aralığı"* (joint confidence interval) kurmak gerekli olabilir. Bu durum daha sonraki konularda ele alınacaktır.

σ^2 İçin Güven Aralığı

- Normallik varsayımı altında $(n-2)\frac{\hat{\sigma}^2}{\sigma^2}$ şeklinde tanımlanan değişkenin $n-2$ sd ile ki-kare dağılımlı olduğunu biliyoruz.
- Bu bilgiden yararlanarak σ^2'nin güven aralığını bulabiliriz:

$$\begin{aligned} P(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}) &= 1-\alpha \\ P\left[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq (n-2)\frac{\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right] &= 1-\alpha \end{aligned}$$

- Bu güven aralıklarının yorumu şudur: Farklı örneklemler kullanarak σ^2 ve β'lar için $\%100(1-\alpha)$ güven sınırları bulur ve gerçek değerlerin bu sınırlar içinde olduğunu söylersek, her 100 seferde $100(1-\alpha)$ kez haklı çıkarız.

6.2 Önsav Sınaması

6.2.1 Güven Aralığı Yaklaşımı

Önsav sınaması konusu ile ilgili bazı önemli noktalar şunlardır:

- Önsav sınaması, verili bir gözlem ya da bulgunun belli bir önsav ile uyuşup uyuşmadığı sorusu ile ilgilenir.
- Buradaki uyuşmak sözcüğü, önsavdaki değere bu önsavı reddetmemeyi sağlamaya yetecek derecede yakın olmak anlamındadır.
- İleri sürülen önsava H_0 ya da *"sıfır önsavı"* (null hypothesis) denir ve H_1 ile gösterilen *"almaşık önsav"* (alternative hypothesis) karşısında sınanır.
- Almaşık önsav *"basit"* (simple) ya da *"bileşik"* (composite) olabilir. Eğer belli bir değer öne sürülüyor ise önsav basittir.
- Örnek olarak
  $H_1 : \beta_1 = 3$ basit,
  $H_1 : \beta_1 \geq 3$ bileşik,
  $H_1 : \beta_1 \neq 3$ ise yine bir bileşik önsavdır.
- Önsav sınamasına birbirini karşılıklı tamamlayıcı iki farklı yaklaşım vardır.
- Bu yaklaşımlar *"güven aralığı"* (confidence interval) ve *"anlamlılık sınaması"* (test of significance) yaklaşımlarıdır.
- Güven aralığı yaklaşımı için karar kuralı aşağıdaki gibidir:

Güven Aralığı Karar Kuralı
Sınanacak katsayı için $\%100(1 - \alpha)$ güven aralığı belirlenir. Eğer katsayı bu güven aralığının içinde ise H_0 reddedilmez. Katsayı eğer güven aralığının dışında kalıyorsa H_0 reddedilir.

- Örnek olarak, serbestlik derecesi 11 ve ölçünlü hatası 0,1 olan ve $\hat{\beta}_2 = 0{,}5$ olarak tahmin edilen katsayı için şunu ileri sürdüğümüzü düşünelim:

$$H_0 : \beta_2 = 0{,}8$$
$$H_1 : \beta_2 \neq 0{,}8$$

- Almaşık önsava göre β_2 0,8'den küçük ya da büyük olabilir. Dolayısı ile bu *"çift kuyruklu"* (two tailed) bir sınamadır.

- Gözlemlenen $\hat{\beta}_2$'nın H_0 ile uyumlu olup olmadığını bulmak için β_2'ye ait %95 güven aralığını oluşturalım:

$$0{,}28 \leq \beta_2 \leq 0{,}72$$

- $0{,}8$ değeri, %95 güven aralığının dışında kalmaktadır.
- Buna göre gerçek β_2'nin 0,8 olduğu önsavını %95 güvenle reddederiz.

Tek Kuyruklu Güven Aralığı

- Zaman zaman almaşık önsavın iki yanlı yerine tek yanlı olduğu yönünde önsel bilgi ya da kuramsal beklentilerimiz olabilir.
- Bu durumda güven aralığı *"tek yanlı"* (one sided) ya da *"tek kuyruklu"* (one tailed) olarak aşağıdaki gibi belirlenir:

$$\beta \geq \hat{\beta} - t_\alpha \text{öh}(\hat{\beta}) \qquad \text{ya da} \qquad \beta \leq \hat{\beta} + t_\alpha \text{öh}(\hat{\beta})$$

- Güven aralığının tek-kuyruklu mu yoksa çift-kuyruklu mu oluşturulacağı almaşık önsavın belirleniş biçimine bağlıdır.
- Tek kuyruklu sınamaya örnek olarak, serbestlik derecesi 11 ve ölçünlü hatası 0,1 olan $\hat{\beta}_2 = 0{,}5$ için β_2'nin 0,8'den küçük olduğu kanısında olduğumuzu varsayalım.
- Bu durumda sıfır önsavı ve almaşık önsav şöyle seçilir:

$$H_0 : \beta_2 \geq 0{,}8 \; H_1 : \beta_2 < 0{,}8$$

- Burada dağılımının sol kuyruğunu göz önüne almaya gerek olmadığı için $1 - \alpha$ güven aralığı $(-\infty,\ \hat{\beta}_2 + t_\alpha \text{öh}(\hat{\beta}_2)]$ olur.
- Tek kuyruklu %95 güven aralığı aşağıdaki gibi bulunur:

$$-\infty \leq \beta_2 \leq 0{,}6796$$

- 0,8 değeri %95 tek yanlı güven aralığının dışında olduğuna göre gerçek β_2'nin 0,8'den büyük ya da 0,8'e eşit olduğu sıfır önsavını %95 güvenle reddedebiliriz.

6.2.2 Anlamlılık Sınaması Yaklaşımı

- Anlamlılık sınaması yaklaşımı güven aralığı yaklaşımını tamamlayıcı ve ona benzer bir süreçtir.
- Normallik varsayımı altında

$$t = \frac{\hat{\beta} - \beta}{\text{öh}(\hat{\beta})}$$

  değişkeninin $(n-2)$ sd ile t dağılımına uyduğunu biliyoruz.

- Eğer sıfır önsavı altında sınanmak üzere belli bir β^* değeri seçilmiş ise, yukarıdaki t değeri örneklemden kolayca hesaplanabilir ve bir sınama istatistiği görevi görebilir.
- Anlamlılık sınaması yaklaşımındaki sınama istatistiği t dağılımlı olduğuna göre şu güven aralığını yazabiliriz:

$$P\left[-t_{\alpha/2} \leq \frac{\hat{\beta} - \beta^*}{\text{öh}(\hat{\beta})} \leq t_{\alpha/2}\right] = 1 - \alpha \tag{6.1}$$

$$|\hat{\beta} - \beta^*| \leq t_{\alpha/2}\text{öh}(\hat{\beta}) \tag{6.2}$$

- β^* burada H_0 altındaki β'dır. $t_{\alpha/2}$ ise $(\alpha/2)$ anlamlılık düzeyinde ve $(n-2)$ sd ile t çizelgesinden okunan kritik değerdir.
- Anlamlılık sınaması yaklaşımına bir örnek olarak σ^2'yi ele alalım:

$$\chi^2 = (n-2)\tfrac{\hat{\sigma}^2}{\sigma^{2*}}$$

- Yukarıda gösterilen değişkenin $(n-2)$ serbestlik derecesi ile ki-kare dağılımına uyduğunu biliyoruz.
- $n = 13$ ve $\hat{\sigma}^2 = 40$ verili olsun.
- $H_0 : \sigma^{2*} = 50$ önsavını sınamak için önce aşağıdaki ki-kare değeri hesaplanır.

$$\chi^2 = (13-2)\tfrac{40}{50} = 8{,}8$$

- 11 serbestlik derecesi ile ve anlamlılık düzeyi $\alpha = 0{,}05$ için $\chi^2_{0{,}975} = 3{,}82$ ve $\chi^2_{0{,}025} = 21{,}92$'dir.
- Hesaplanan χ^2 değeri yukarıdaki iki değer arasında kaldığı için sıfır önsavı reddedilmez.

t sınaması karar kuralları aşağıdaki gibi özetlenebilir:

Çizelge: t Anlamlılık Sınaması Karar Kuralları

Önsav Türü	Sıfır önsavı	Almaşık önsav	H_0 ret kuralı
Çift Kuyruk	$\beta = \beta^*$	$\beta \neq \beta^*$	$\|t\| > t_{\alpha/2,\mathrm{sd}}$
Sağ Kuyruk	$\beta \leq \beta^*$	$\beta > \beta^*$	$t > t_{\alpha,\mathrm{sd}}$
Sol Kuyruk	$\beta \geq \beta^*$	$\beta < \beta^*$	$t < -t_{\alpha,\mathrm{sd}}$

- *Dikkat:* İki değişkenli model için sd $= (n-2)$'dir.

6.2.3 Anlamlılık Konusu

Bir Önsavı Reddetmemenin Anlamı

- Bir anlamlılık sınamasına dayanarak sıfır önsavının desteklenmesi demek, aslında, örneklem verilerine dayanarak bu önsavı reddedecek bir neden olmadığı anlamına gelir.
- Örnek olarak gerçek $\beta = 0{,}5$ olduğunu varsayalım.
- Verilere dayanarak burada $H_0 : \beta = 0{,}4$ ve $H_0 : \beta = 0{,}5$ gibi farklı önsavlar ileri sürmek olasıdır.
- Ancak bu önsavlardan hangisinin doğru olduğu bilinemez.
- Bu nedenle, tıpkı bir mahkemenin "suçsuzdur" yerine "beraat etmiştir" demesi gibi "kabul ederiz" yerine "reddedemeyiz" sonucuna varmalıyız.

$\beta_2 = 0$ Sıfır Önsavı ve 2t Yöntemi

- Görgül çalışmalarda $H_0 : \beta_2 = 0$ önsavı sıklıkla sınanır.
- Burada amaç Y'nin açıklayıcı değişken X ile ilişkisi olup olmadığına karar vermektir.
- $H_0 : \beta_2 = 0$ sıfır önsavını sınamada *"2t başparmak kuralı"* (2t rule of thumb) kullanılabilir:

2t Yöntemi

Serbestlik derecesi 30 ya da daha fazla ise, anlamlılık düzeyi $\alpha = 0{,}05$ iken bulunan $t = \hat{\beta}_2/\text{öh}(\hat{\beta}_2)$, mutlak değer olarak eğer 2'den büyükse, $\beta_2 = 0$ sıfır önsavı reddedilir.

- Bunun nedeni, sd> 30 olduğunda, t dağılımındaki alanın yüzde 95'ten büyük bölümünün $(-2, 2)$ değerleri arasında yer almasıdır. Bu durum t çizelgesinden de görülebilir.
- $H_0 : \beta_2 = 0$'a karşı $\beta_2 < 0$ ya da $\beta_2 > 0$ tek yanlı sınamaları için ise kullanılacak değer 2 değil 1,7'dir.

Anlamlılık Düzeyinin Seçimi

- Uygulamada anlamlılık düzeyi α çoğu zaman $\%1$, $\%5$ ya da en çok $\%10$ olarak seçilmektedir.
- Aslında bu değerlerin yerine başka herhangi bir değer de aynı işi görebilir.
- H_0'ı kabul ya da ret kararı verilirken iki tür hata yapılabilir:

I. Tür Hata:	Aslında doğru olan H_0'ı reddetmek.
II. Tür Hata:	Aslında yanlış olan H_0'ı reddetmemek.

- Örneklem büyüklüğü veriliyken, I. tür hata yapma olasılığı azaltılmak istenirse II. tür hata yapma olasılığı artar. Eğer II azaltılırsa bu sefer de I artar.
- Anlamlılık düzeyi seçimindeki klasik yaklaşım, uygulamada I. tür hatanın II. türe göre daha ciddi olduğudur.
- Dolayısıyla, $\alpha = 0{,}01$ ya da $\alpha = 0{,}05$ seçilerek I. tür hata yapma olasılığı olabildiğince düşük tutulur.

Anlamlılığın Kesin Düzeyi

- Önsav sınamasındaki zayıf noktanın α'nın seçimindeki gelişigüzellik olduğunu biliyoruz. *"p-değeri"* (p-value) kavramı, α değerini seçme sorununu ortadan kaldırır:

P değeri

P-değeri ya da *"olasılık değeri"* (probability value), anlamlılığın gözlenen kesin düzeyi ya da I. tür hata yapma olasılığının kesin düzeyinin ölçüsüdür.

- Diğer bir deyişle p değeri, sıfır önsavının reddedilebileceği en düşük anlamlılık düzeyini verir.
- Belli bir örneklem veriliyken $|t|$ büyüdükçe p değeri azalır ve sıfır önsavı da gittikçe artan bir güvenle reddedilebilir.
- Güncel ekonometri yazılımları çeşitli sınama istatistiklerine ilişkin p-değerlerini de hesaplayıp verebilmektedir.

İstatistikte Anlamlılık ve Uygulamada Anlamlılık

İstatistiksel anlamlılık uygulamada anlamlığı gerektirmez. Buna ilişkin olarak Türkiye milli gelir-tüketim örneğimizi anımsayalım:

- Örneklemden elde ettiğimiz $\hat{\beta}_2$ değeri 0,59 idi.
- $\hat{\beta}_2$ için %95 güven aralığı (0,56, 0,63) olarak hesaplanır. Buna göre $\beta_2 = 0{,}64$ sıfır önsavını reddedebiliriz.
- Öte yandan, $\hat{\beta}_2$'yı 0,56 ya da 0,63 almak arasındaki farkın uygulamada önemli olup olmadığı da dikkate alınmalıdır.
- Bu sorunun yanıtı modelden modele değişir.
- Örnek olarak, burada $\hat{\beta}_2$ marjinal tüketim eğilimi MTüE'dir. İktisat kuramına göre yatırım çarpanı ise $1/(1 - \text{MTüE})$'dir.
- Buna göre eğer MTüE $= 0{,}56$ ise çarpan 2,27 olurken MTüE $= 0{,}63$ ise de çarpan 2,70 olacaktır.
- Görüldüğü gibi, bu örnekteki fark hem istatistiksel olarak hem de uygulama açısından önemlidir.

6.3 Çıkarsamaya İlişkin Konular

6.3.1 Varyans Çözümlemesi

- *"Varyans çözümlemesi"* (analysis of variance) ya da kısaca *"VARÇÖZ"* (ANOVA), istatistiksel çıkarsama sorununa tamamlayıcı ve aydınlatıcı bir yaklaşım sunar.

- Aşağıdaki özdeşliği anımsayalım:

$$\begin{aligned} \sum y_i^2 &= \sum \hat{y_i}^2 + \sum \hat{u_i}^2 \\ \sum y_i^2 &= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u_i}^2 \\ \text{TKT} &= \text{BKT} + \text{KKT} \end{aligned}$$

- VARÇÖZ yaklaşımının temelinde TKT'nin bu iki parçasının incelenmesi yatar.

- BKT 1 sd ile ve KKT de iki değişkenli model için $(n-2)$ sd ile ki-kare dağılımlıdır.

- O halde, toplamların kendi sd'lerine bölünmesi ile bulunan *"ortalama kareleri toplamı"* (mean sum of squares) ya da kısaca *"OKT"* (MSS) değerlerini kullanarak şunu yazabiliriz:

$$\begin{aligned} F &= \frac{\text{BKT'nin OKT'si}}{\text{KKT'nin OKT'si}} \\ &= \frac{(\hat{\beta}_2^2 \sum x_i^2)/1}{\sum \hat{u_i}^2/(n-2)} \end{aligned}$$

- Yukarıdaki değişken, hata teriminin normalliği varsayımı altında pay 1 ve payda $(n-2)$ sd ile F dağılımına uyar.

- Tanımladığımız F oranından nasıl yararlanabileceğimizi görmek için aşağıdaki eşitliklere bakalım:

$$\begin{aligned} E\left((\hat{\beta}_2^2 \sum x_i^2)/1\right) &= \ldots = \sigma^2 + \beta_2^2 \sum x_i^2 \\ E\left(\sum \hat{u_i}^2/(n-2)\right) &= E(\hat{\sigma}^2) = \sigma^2 \end{aligned}$$

- β_2 ve σ^2 gerçek anakütle katsayılarıdır.
- Eğer β_2 sıfır ise eşitliklerin her ikisi de aynı çıkar.
- Demek ki, F oranı bize $H_0 : \beta_2 = 0$ sıfır önsavını sınamada kullanılabilecek bir sınama istatistiği vermektedir.
- *Dikkat:* Bu durum iki değişkenli bağlanım için geçerlidir. F oranının çoklu bağlanımdaki yorumu farklıdır.

Varyans çözümlemesine örnek olarak, Türkiye gelir-tüketim örneğimize dönelim.

- F değeri 1213,49 olarak hesaplanmaktadır.
- Anlamlılık düzeyi %5 iken, 1 ve 18 sd için kritik F değeri 4,41 olarak verilidir.
- Elimizdeki F istatistiği kritik değerden büyük olduğu için, $\beta_2 = 0$ önsavını reddederek Türkiye'de gelirin, özel tüketim harcamaları üzerinde etkili olduğunu söyleyebiliriz.
- Bu noktada, k sd ile t dağılımına uyan değişkenin karesinin de 1 ve k sd ile F dağılımına uyduğunu da anımsayalım.
- $H_0 : \beta_2 = 0$ altında tahmin edilen t değeri 34,84'tür.
- Yuvarlama hatalarını bir yana bırakırsak $t^2 = (34{,}84)^2 = F$ eşitliğinin geçerli olduğunu görüyoruz.
- Bu nedenle, iki değişkenli bağlanım için F sınamasına aslında gerek yoktur.
- Şimdilik F ve t sınamalarının $\beta_2 = 0$ sınamasının iki farklı ve birbirini tamamlayıcı yolu olduğunu söyleyebiliriz.
- F sınamasının önemini ve farklı uygulamalarını çoklu bağlanım konusu içerisinde ele alacağız.

6.3.2 Kestirim Sorunu

Ortalama Kestirimi

- Örneklem katsayıları yanında tekil $\hat{Y}_i$ değerleri için de aralık tahmini ve önsav sınaması yapılabilir.

- Örnek olarak, aşağıdaki örneklem bağlanımına bakalım:

$$\hat{Y}_i = 25 + 2X_i$$

- Katsayı tahminlerine dayanarak $E(Y|X_0 = 100)$ kestirimini yapmak istediğimizi varsayalım.
- Bu *"ortalama kestirimi"* (mean prediction) şöyle bulunur:

$$\begin{aligned}\hat{Y}_0 &= \hat{\beta}_1 + \hat{\beta}_2 X_0 \\ &= 25 + 2(100) = 225\end{aligned}$$

- $\hat{Y}_0$ burada $E(Y|X_0)$ tahmincisidir.
- $\hat{Y}_0$'nın, bir tahminci olmasından dolayı, kendi gerçek değerinden farklı çıkması söz konusudur.
- $\hat{Y}_0$ tahmincisinin aşağıda gösterilen ortalama ve varyans ile normal dağılımlı olduğu kanıtlanabilir:

$$E(\hat{Y}_0) = \beta_1 + \beta_2 X_0 \qquad \text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

- Bilinmeyen σ^2 yerine yansız tahminci $\hat{\sigma}^2$ koyulduğunda ise bulunan değişken $(n-2)$ sd ile t dağılımına uyacaktır:

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\text{öh}(\hat{Y}_0)}$$

- Öyleyse, t dağılımını kullanarak $E(Y_0|X_0)$ güven aralığını bulabilir ve bunu önsav sınaması yapmada kullanabiliriz.
- $E(Y_0|X_0)$ güven aralığının tüm X'ler için hesaplanması ile anakütle bağlanım işlevine ilişkin bir *"güven kuşağı"* (confidence band) elde edilebilir.
- Bu güven kuşağı $X_0 = \bar{X}$ olduğunda en dar noktadadır. X_0 değeri $\bar{X}$'den uzaklaştıkça kemer de genişler.
- Dolayısıyla, örneklem ortalaması $\bar{X}$'den uzaklaşıldıkça örneklem bağlanımının kestirim yeteneği de azalacaktır.

6.3.3 Bağlanım Bulgularının Değerlendirilmesi

Verilere yakıştırılan model sonuçları yorumlanırken aşağıdaki üç ölçüt göz önüne alınmalıdır:

1. Tahmin edilen katsayıların işaretlerinin kuramsal ya da önsel bilgilere dayalı beklentilerle uyumluluğu,

2. Kuramsal ilişkinin istatistiksel olarak anlamlı olup olmadığı,

3. Bağlanım modelinin güvenilirliği ve kuramsal ilişkiyi açıklayabilme derecesi.

Türkiye gelir-tüketim örneği için gretl bağlanım çıktısı şöyledir:

```
gretl: model 1
Dosya  Düzenle  Sınamalar  Kaydet  Çizitler  Çözümleme  LaTeX

Model 1: SEK (OLS), kullanılan gözlemler: 1987-2006 (T = 20)
Bağımlı değişken: tuketim

             katsayı    ölç. hata   t-oranı   p-değeri
  ------------------------------------------------------
  const      8,03438    1,85509      4,331    0,0004   ***
  gsyih      0,593351   0,0170331   34,84     5,68e-18 ***

Bağımlı değişken ort   71,20574   Bağımlı değişken ö.s.  14,07370
Kalıntı kareleri top   55,00622   Bağlanım ö.h.          1,748114
R-kare                 0,985384   Ayarlamalı R-kare      0,984572
F(1, 18)               1213,489   P-değeri(F)            5,68e-18
Log-olabilirlik       -38,49591   Akaike ölçütü          80,99182
Schwarz ölçütü         82,98329   Hannan-Quinn           81,38058
ro                     0,613231   Durbin-Watson          0,748726
Not: ö.s. ve ö.h. ölçünlü sapma ve ölçünlü hatayı göstermektedir.
```

Bağlanım bulgularını incelediğimizde şunları görürüz:

- Keynesci tüketim kuramı çerçevesinde β_1 otonom tüketimi, β_2 ise marjinal tüketim eğilimi MTüE'yi göstermektedir.

- Sıfır gelir gerçek hayatta gözlenen bir durum olmadığı için β_1'e otonom tüketim anlamı yüklemekten kaçınılmalıdır.

- β_2 ise önsel beklentilere uygun şekilde 1'den küçük ve 0,59 olarak tahmin edilmiştir. Buna göre, Türkiye'de milli gelir 1 TL arttığında tüketim de 59 kuruş artmaktadır.

- t istatistikleri ilgili anakütle değerinin sıfır olduğu varsayımı altında bulunmuştur. β_2 için $34{,}84 = 0{,}59335/0{,}017033$'tür.

- $\hat{\beta}_2$'ya ait p-değeri de 18 sd ile 34,84 ya da daha yüksek bir t değeri bulma olasılığını $5{,}68 \times e^{-18}$ olarak vermektedir. Demek ki MTüE'nin sıfırdan farklı olduğunu söyleyebiliriz.

- Yaklaşık 0,98 büyüklüğündeki r^2 değeri, özel tüketim harcamalarındaki değişimin %98 oranında milli gelirdeki değişim ile açıklanabildiğini söylemektedir.
- Bağlanım sonuçlarının güvenilir olduğuna karar verebilmek için modelimizin KNDBM varsayımlarını sağladığını da onaylamak zorundayız.
- Şu an tüm KNDBM'nin varsayımlarını denetleyemesek de u_i hata teriminin normalliği varsayımına bakabiliriz.
- Yazında çeşitli normallik sınamaları bulunmaktadır. Biz bunlardan ki-kare *"yakışmanın iyiliği"* (goodness of fit) ve Jarque-Bera normallik sınamalarını ele alacağız.
- Bu sınamaların ikisi de $\hat{u}_i$ kalıntılarını ve ki-kare olasılık dağılımını temel almaktadır.

χ^2 Yakışmanın İyiliği Sınaması

χ^2 yakışmanın iyiliği sınamasının adımları şöyledir:

1. Bağlanım işlevi bulunur ve $\hat{u}_i$ kalıntıları elde edilir.
2. $\hat{u}_i$'nın örneklem ölçünlü sapması hesaplanır.
3. Örneklem büyüklüğüne göre bir *"kap"* (bin) sayısı belirlenir. Kalıntılar büyüklük sırasına sokulur ve sıfırdan kaç ölçünlü sapma uzaklıkta olduklarına göre bu kaplara bölüştürülür.
4. Gözlenen sıklıklar (G_i) ile normal dağılım için beklenen sıklıklar (B_i) arasındaki farkların kareleri alınır, beklenen sıklıklara bölünür ve bunların toplamı hesaplanır:
$$\chi^2 = \sum_{i=1}^{k} \frac{(G_i - B_i)^2}{B_i}$$
k = kap sayısı iken, yukarıdaki değişken $(k - 3)$ (normal dağılıma karşı sınadığımız için) sd ile χ^2 dağılımına uyar.
5. Eğer p değeri yüksekse H_0 : normallik önsavı reddedilmez.

Jarque-Bera Normallik Sınaması

JB sınaması bir *"kavuşmazsal"* (asymptotic) ya da büyük örneklem sınamasıdır. Şu şekilde yapılır:

1. Öncelikle SEK kalıntılarının *"çarpıklık"* (skewness) ve *"basıklık"* (kurtosis) ölçüleri bulunur.

2. Daha sonra aşağıdaki istatistik hesaplanır:

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

   Burada S çarpıklığı, K ise basıklığı göstermektedir. Jarque ve Bera, 1987 tarihli bir çalışmalarında, kalıntıların normal dağıldığı varsayımı altında JB istatistiğinin büyük örneklemde 2 sd ile χ^2 dağılımlı olduğunu göstermişlerdir.

3. Eğer hesaplanan sınama istatistiğine ait p değeri yüksekse H_0 : normallik önsavı reddedilmez.

- Ki-kare sınamasının çekici yanı, *"yığınsal dağılım işlevi"* (cumulative distribution function) hesaplanabilen her türlü dağılım için yakışmayı sınamak için kullanılabilmesidir.
- Sakıncası ise kap sayısının nesnel bir ölçütü olmadığı için hesaplanan χ^2 değerinin farklılık gösterebilmesidir.
- SEK yönteminin istatistiksel özelliklerinden dolayı, büyük örneklemlerde normallik sınaması çoğu zaman gerekmez.
- Bağlanım ile ilgili olarak normallik sınaması daha çok bir küçük örneklem konusudur.
- Diğer yandan, JB kavuşmazsal bir sınama olduğu için küçük örneklemlerde ki-kare dağılımından sapmaktadır.
- Örnek olarak, $n = 70$ gibi çok da küçük sayılamayacak örneklemlerde bile bulunan JB p değeri yanıltıcı olabilir.
- Bu yüzden Jarque-Bera yerine Doornik-Hansen sınaması yazın *"literature"* içinde yeğlenmektedir.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 5* "Two-Variable Regression: Interval Estimation and Hypothesis Testing" okunacak.

Önümüzdeki Ders
İki Değişkenli Doğrusal Bağlanım Modelinin Uzantıları

Bölüm 7

İki Değişkenli Bağlanım Modelinin Uzantıları

7.1 Sıfır Noktasından Geçen Bağlanım

Kuram bazen modelde sabit terimin bulunmamasını öngörür:

$$Y_i = \beta_2 X_i + u_i$$

Sıfır noktasından geçen bağlanım modelinin uygun olduğu bazı durumlar şunlardır:

- *"sermaye varlığı fiyatlama modeli"* (capital asset pricing model) ya da kısaca *"SVFM"* (CAPM),
- Milton Friedman'ın *"kalıcı gelir önsavı"* (permanent income hypothesis),
- *"Maliyet çözümlemesi kuramı"* (cost analysis theory),
- Enflasyon oranının para arzındaki değişim ile orantılı olduğunu ileri süren para kuramı çeşitlemeleri.

Sıfır noktasından geçen bağlanım için ÖBİ aşağıdaki gibidir:

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i$$

Yukarıdaki modele ait $\hat{\beta}_2$ SEK tahmincisi şu şekilde bulunur:

Alışılmış Model

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$
$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$
$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-2}$$

$\beta_1 = 0$ **Modeli**

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$$
$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$$
$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-1}$$

- Yukarıdaki büyük ve küçük harf kullanımına dikkat ediniz.
- Kısaca, $\beta_1 = 0$ modeli formüllerinde ortalamalardan sapma yerine X ve Y'lerin asıl değerlerini kullanıyoruz.

Sabit terimsiz modelin iki özelliğinin bilinmesinde yarar vardır:

1. Bu modellerde $\sum \hat{u}_i$ kalıntı toplamı her zaman sıfır olmak zorunda değildir.
2. Bu modellerde kalıntı kareleri toplamı, toplam kareleri toplamından küçük olmak zorunda değildir. Bu nedenle, alışıldık modeller için hesaplanan belirleme katsayısı r^2 sıfır noktasından geçen bağlanımlarda zaman zaman eksi değerler alabilir ve kullanılması uygun değildir. Sabit terimsiz modellerde *"ham"* (raw) r^2 kullanılabilir:

   Ham r^2

$$\text{ham } r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

Ham r^2 de 0 ve 1 arasındadır ama diğer r^2 ile karşılaştırılamaz.

- Önsel dayanaklar çok güçlü olmadığı sürece sabit terimin modele eklenmesinde yarar vardır.
- Eğer modele sabit terim eklenir ve bu terim istatistiksel olarak anlamlı bulunmazsa, zaten elde sıfır noktasından geçen bir bağlanım modeli var demektir.
- Diğer yandan, gerçekte modelde sabit terim varken sabit terimsiz model yakıştırılmaya çalışılırsa *"model belirtim hatası"* (model specification error) yapılmış olur.

Sıfır Noktasından Geçen Bağlanım Açıklayıcı Örnek

- Sıfır noktasından geçen bağlanıma bir örnek olarak, Güz 2007 döneminde TOBB ETÜ ekonometri öğrencilerinin arasınav ve dönem sonu sınav notu sıralamalarını alalım:

$$Y_i = \alpha + \beta X_i + u_i$$

- Burada $\begin{array}{ll} Y & \text{öğrencinin dönem sonu sınavında kaçıncı olduğunu,} \\ X & \text{öğrencinin arasınavda kaçıncı olduğunu} \end{array}$ göstermektedir.

- Tekil öğrencilere ilişkin motivasyon değişikliği ya da özel durumlar gibi rastsal kabul edilebilecek etmenler dışında sıralamanın değişmeyeceğini varsaymak yanlış olmaz.

- Bu durumda önsel beklentimiz $\alpha = 0$ ve $\beta = 1$ olmasıdır.

- Bu modeli sıfır noktasından geçen bağlanım olarak hesaplarsak aşağıdaki bulguları elde ederiz:

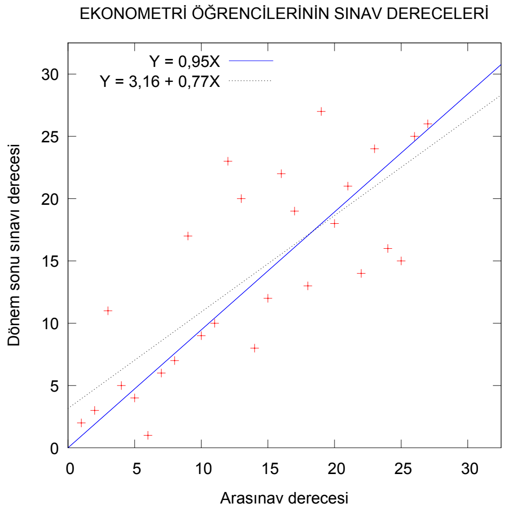
$$\begin{array}{rll} \hat{Y}_i = & 0{,}9466 & X_i \\ \text{öh} & (0{,}0632) & \\ t & (14{,}9721) & \text{ham } r^2 = 0{,}8961 \end{array}$$

- Sabit terimsiz bağlanımın uygun olup olmadığını sınamak için alışılmış bağlanıma da bakalım:

$$\begin{array}{rllll} \hat{Y}_i = & 3{,}1624 & + & 0{,}7741 & X_i \\ \text{öh} & (2{,}0284) & & (0{,}1266) & \\ t & (1{,}5591) & & (6{,}1142) & r^2 = 0{,}5992 \end{array}$$

- İlk bağlanımda $\hat{\beta}$, 1'e oldukça yakındır. İkinci bağlanımda sabit terimin sıfır olduğu sıfır önsavı reddedilmez.

- Eğer baştaki varsayımımız doğru ise, r^2'den dolayı, rastsal etmenlerin başarıda %10 etkili olduğunu söyleyebiliriz.


EKONOMETRİ ÖĞRENCİLERİNİN SINAV DERECELERİ
Y = 0,95X
Y = 3,16 + 0,77X
Dönem sonu sınavı derecesi
Arasınav derecesi
0
5
10
15
20
25
30

7.2 Hesaplamaya İlişkin Konular

7.2.1 Ölçekleme ve Ölçü Birimleri

Bağlanım çözümlemesinde dikkat edilmesi gereken bir nokta da *"verileri ölçekleme"* (data scaling) konusudur. Verilerin ölçeklenmesi ile ilgili iki önemli soru şudur:

1. X ve Y değişkenlerinin ölçü birimleri bağlanım bulgularını etkiler mi?
2. Bağlanım çözümlemesi için ölçü biriminin seçilmesinde izlenilmesi gereken bir yol var mıdır?

Türkiye'ye ait aşağıda verilen 1987 fiyatları ile gayrisafi sabit sermaye oluşumu ve gayrisafi yurtiçi hasıla verilerine bakalım:

Çizelge: Türkiye'de Sabit Sermaye Oluşumu ve GSYH (1987–2000)

Yıl	GSSSO (milyon TL)	GSSSO (milyon TL)	GSYH (milyar TL)	GSYH (milyar TL)
1987	18.491	74.416	0.018491	0.074416
1988	18.299	76.143	0.018299	0.076143
1989	18.701	76.364	0.018701	0.076364
1990	21.670	83.371	0.021670	0.083371
1991	21.764	84.271	0.021764	0.084271
1992	23.147	88.893	0.023147	0.088893
1993	29.247	96.391	0.029247	0.096391
1994	24.577	91.600	0.024577	0.091600
1995	26.823	97.729	0.026823	0.097729
1996	30.598	104.940	0.030598	0.104940
1997	35.137	112.892	0.035137	0.112892
1998	33.768	116.541	0.033768	0.116541
1999	28.473	111.083	0.028473	0.111083
2000	33.281	119.147	0.033281	0.119147

Ortaya atmış olduğumuz iki soruyu yanıtlayabilmek için aşağıda verilen bağlanım bulgularını inceleyelim:

Hem GSSSO, hem GSYH milyon TL:

$$\widehat{\text{GSSSO}}_t = \underset{(2,73489)}{-8,76891} + \underset{(0,0283565)}{0,364933}\ \text{GSYH}_t \qquad r^2 = 0,9324$$

Hem GSSSO, hem GSYH milyar TL:

$$\widehat{\text{GSSSO}}_t = \underset{(0,00273489)}{-0,00876891} + \underset{(0,0283565)}{0,364933}\ \text{GSYH}_t \qquad r^2 = 0,9324$$

GSSSO milyon dolar, GSYH milyar TL:

$$\widehat{\text{GSSSO}}_t = \underset{(2,73489)}{-8,76891} + \underset{(28,3565)}{364,933}\ \text{GSYH}_t \qquad r^2 = 0,9324$$

GSSSO milyar dolar, GSYH milyon TL:

$$\widehat{\text{GSSSO}}_t = \underset{(0,00273489)}{-0,00876891} + \underset{(0,0000283565)}{0,000364933}\ \text{GSYH}_t \qquad r^2 = 0,9324$$

Not: Ölçünlü hatalar parantez içerisinde verilmiştir.

- Bağlanım bulgularının dördü de GSYH'deki bir milyon liralık bir değişimin GSSSO'de ortalama 0,364933 milyon liralık bir değişime yol açtığını göstermektedir.
- Öyleyse, SEK tahmincilerinin bilinen özellikleri farklı ölçü birimlerinin kullanılmasından etkilenmemektedir.
- Öte yandan, bağlanım hesapları bilgisayar kullanılarak yapıldığı için, verilerin uygun biçimde ölçeklendirilmesi uygulamada zaman zaman önemli olabilir.

7.2.2 Sayısal Hesaplama Sorunları

- Ekonometri, birçok karmaşık matematiksel ve istatistiksel yöntem içeren bir bilim dalıdır.
- Ancak çoğu araştırmacı çeşitli tekniklerin yalnızca birkaç fare tıklaması ile uygulanabileceği izlenimini taşımaktadır.
- Bilgisayar yazılımlarının her zaman sayısal olarak tutarlı olduğunu varsaymak hatalı bir yaklaşımdır.
- Günümüz bilimsel yazılımlarının çoğu tüm hesaplamalarda 64bit *"kayan nokta"* (floating point) aritmetik kullanmaktadır.
- Bu altyapı gerçel sayı sistemini tümüyle karşılayamayarak dört tür hataya yol açabilmektedir:

"Yuvarlama hataları" (rounding errors)
"İptal etme hataları" (cancellation errors)
"Budama hataları" (truncation errors)
"Çözümyolu hataları" (algorithm errors)

Yuvarlama Hataları

- Yuvarlama hatası, bazı sayıların bilgisayarların kullandığı ikili düzende tam olarak gösterilememesinden kaynaklanır.
- Örnek olarak 0,1 ondalık sayısının ikili düzende gösterimi $0{,}0\overline{0011}$'dir. Bu sayı yeniden ondalık sisteme çevrildiğinde 0,09999999403953 olur.
- Bu nedenle, cebirsel olarak birbirine eşdeğer olan $(p = q)$ ve $(p - q = 0)$ gibi iki denklem bilgisayarda uygulandığında farklı sonuçlar verebilmektedir.

İptal Etme Hataları

- İptal etme hatası, yuvarlama hatasının özel bir durumudur. Gözlemlerde fazla sayıda sabit öncül basamak olduğunda ortaya çıkar.
- Bu özelliği gösteren veri setlerine *"katı"* (stiff) veri seti denir.
- Örnek olarak 1,000,000,001 sayısından 1,000,000,000 çıkarılınca geriye yalnızca en sağdaki tek basamak kalır.
- Baştaki sayının büyüklüğünden dolayı, bu son basamak yuvarlama hatalarına fazla duyarlıdır.

Budama Hataları

- Budama hatası, *"yinelemesel"* (iterative) işlemlerde görülen ve yazılımdan zorunlu olarak kaynaklanan bir hata türüdür.
- Örnek olarak $\exp(x)$ işlevi $x = 1$ noktasında aşağıdaki gibi genişletilir:

$$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!} = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = e$$

- Görüldüğü gibi, *"oransız sayı"* (irrational number) e'nin hesaplanabilmesi sonsuz sayıda toplama gerektirmektedir.
- Ancak bilgisayar hesaplaması sınırlı sayıda işlem içerebilir ve sonuçta bir budama hatası ortaya çıkar.

Çözümyolu Hataları

- Çözümyolu hatası, bir problemin çoğu zaman birden fazla şekilde çözülebileceği gerçeğinden kaynaklanır.
- Sonuçta bazı çözümler diğerlerinden daha iyidir.
- Örnek olarak, doğrusal SEK modelini hesaplamak için kullanılabilecek yöntemlerden bazıları şunlardır:

"Gaussçu eleme" (Gaussian elimination) *"Tekil değer ayrıştırması"* (singular value decomposition) *"Cholesky çarpanlaması"* (Cholesky factorization) *"QR çarpanlaması"* (QR factorization)

- Bunlar içinde QR yöntemi, çoklueşdoğrusal veriler dışında diğerlerine göre daha güvenilir sonuçlar vermektedir.

Sayısal Hesaplama Sorunları Özet

Özetle, sayısal hesaplama sorunlarına ilişkin dikkat edilmesi gereken noktalar şunlardır:

- Bilgisayar matematiğinin kağıt-kalem matematiğinden tümüyle farklı olduğu unutulmamalıdır.
- Sayısal hataları azaltmanın kolay yolu, çözümleme öncesi verileri uygun şekilde ölçeklemektir.
- Tüm verileri öntanımlı olarak $[0, 1)$ ya da $[0,10)$ aralıklarına göre ölçeklemek doğru bir yaklaşımdır.
- Çok büyük ve çok küçük sayıları birlikte kullanmanın hatalı sonuçlara davetiye çıkarmak olduğu unutulmamalıdır.
- Ayrıca araştırmacı çalışmasında yalnızca veri kaynaklarını belirtmekle yetinmemeli, verilerin nasıl ölçüldüğünü ve ölçeklendiğini de mutlaka açıklamalıdır.

7.3 Bağlanım Modellerinin İşlev Biçimleri

- *"Doğrusallık"* (linearity) kavramının değişkenlerde doğrusallık ve değiştirgelerde doğrusallık olmak üzere iki ayrı şekilde tanımlandığını anımsayalım.
- KDBM için değiştirgelerde doğrusallık zorunlu olsa da değişkenlerde doğrusallık zorunlu değildir.
- Öyleyse, değişkenlerde doğrusal-dışı ama değiştirgelerde doğrusal olan ya da uygun dönüştürmelerle doğrusal yapılabilen modelleri KDBM ile tahmin etmek olanaklıdır.
- Bu bağlamda ele alacağımız model biçimleri şunlardır:

"Log-doğrusal model" (log-linear model)
"Yarı-logaritmasal model" (semi-logarithmic model)
"Evrik model" (reciprocal model)
"Log-evrik model" (log-reciprocal model)

7.3.1 Log-Doğrusal Model

- *"Üstel"* (exponential) bağlanım modeli diye adlandırılan aşağıdaki modeli ele alalım:

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$$

- Yukarıdaki gösterim aşağıdaki şekilde doğrusallaştırılabilir:

$$\begin{aligned}\ln Y_i &= \ln \beta_1 + \beta_2 \ln X_i + u_i \\ &= \alpha + \beta_2 \ln X_i + u_i\end{aligned}$$

- Bu model, α ve β_2 anakütle katsayılarında doğrusaldır ve SEK yöntemiyle aşağıdaki gibi tahmin edilebilir:

$$Y_i^* = \alpha + \beta_2 X_i^* + u_i$$

- Burada $Y_i^* = \ln Y_i$ ve $X_i^* = \ln X_i$'dir.

- Her iki yanının logaritması alınarak doğrusallaştırılmış modellere *"log-doğrusal"* (log-linear), *"log-log"* (log-log) ya da *"çifte-log"* (double-log) modeller adı verilir.
- Log-doğrusal modeldeki $\hat{\alpha}$ ve $\hat{\beta}_2$ SEK tahmincileri, başta gördüğümüz doğrusal modellerde olduğu gibi EDYT'dirler.
- Ancak $\hat{\beta}_1 = \text{antilog}(\hat{\alpha})$ biçiminde tahmin edildiği için $\hat{\alpha}$, $\hat{\beta}_1$'nın yanlı bir tahmincisidir.
- Birçok uygulamada sabit terim ikinci derecede önemli olduğundan, $\hat{\beta}_1$'in yanlı olmasına aldırılmayabilir.

Log-doğrusal modelin yaygınlığına yol açan çekici özelliği, β_2 eğim katsayısının Y'nin X'e göre esnekliğini vermesidir:

Doğrusal Model

$Y_i = \alpha + \beta_2 X_i + u_i$

Eğim (birim değişim):

$\frac{\mathrm{d}Y_i}{\mathrm{d}X_i} = \beta_2$

Esneklik (yüzde değişim):

$\frac{\mathrm{d}Y_i/Y_i}{\mathrm{d}X_i/X_i} = \frac{\mathrm{d}Y_i}{\mathrm{d}X_i}\frac{X_i}{Y_i} = \beta_2 \frac{X_i}{Y_i}$

Log-doğrusal Model

$Y_i = \exp(\alpha + \beta_2 \ln X_i + u_i)$

Eğim (birim değişim):

$\frac{\mathrm{d}Y_i}{\mathrm{d}X_i} = \exp(\alpha + \beta_2 \ln X_i + u_i)\beta_2 \frac{1}{X_i}$
$= \beta_2 \frac{Y_i}{X_i}$

Esneklik (yüzde değişim):

$\frac{\mathrm{d}Y_i/Y_i}{\mathrm{d}X_i/X_i} = \frac{\mathrm{d}Y_i}{\mathrm{d}X_i}\frac{X_i}{Y_i} = \left(\beta_2 \frac{Y_i}{X_i}\right)\frac{X_i}{Y_i} = \beta_2$

- Bu özelliğinden dolayı log-doğrusal model *"sabit esneklik"* (constant elasticity) modeli diye de adlandırılır.
- Örnek olarak kahve talebi modeline bakalım.
- Veriler üzerinde log-log doğrusallaştırması yapıldıktan sonra hesaplanan bağlanım şu sonuçları vermektedir:

$$\begin{array}{lcll} \widehat{\ln Y}_i = & 0{,}7774 & -\ 0{,}2530\ \ln X_i & \\ \text{öh} & (0{,}0152) & (0{,}0494) & r^2 = 0{,}7448 \\ t & (51{,}1447) & (-5{,}1214) & F_{1,9} = 26{,}23 \end{array}$$

- Fiyat esnekliği katsayısı $-0{,}25$ olarak bulunmuştur.
- Buna göre kahve fiyatında yüzde 1 artış olması durumunda kahve tüketiminin ortalama yüzde 0,25 azalması beklenir.
- Öyleyse kahve talebinin kendi fiyatına göre esnek olmadığı söylenebilir.
- Zaman zaman doğrusal ve log-doğrusal model arasında bir seçim yapmak gerekli olabilir.
- Bağımlı değişkenler aynı olmadığı için, böyle bir durumda iki r^2 değerini doğrudan karşılaştırma yoluna gidilemez.
- Katsayı tahminlerini karşılaştırma konusunda ise $\beta_2(\bar{X}/\bar{Y})$ tanımından yararlanılarak doğrusal model için bir ortalama esneklik hesaplanabilir.
- Kahve talebi örneğinde, log-log modelden elde edilen β_2 esneklik katsayısı $-0{,}25$ iken, doğrusal modelin ortalama esnekliği de benzer biçimde $-0{,}22$ olarak bulunur.
- *Dikkat:* $\beta_2(\bar{X}/\bar{Y})$ kullanılarak bulunan ortalama esneklik farklı $\bar{X}$ ve $\bar{Y}$ değerlerine bağlıdır. Log-doğrusal modelin esneklik katsayısı β_2 ise her fiyat düzeyinde aynıdır.

7.3.2 Yarı-logaritmasal Modeller

Log-Doğ Modeli

- Ekonomistler sık sık para arzı, istihdam, GSYH gibi değişkenlerin büyüme oranlarının tahmini ile ilgilenirler.
- Bileşik faiz formülünü anımsayalım:

$$Y_t = Y_0(1+r)^t$$

- Burada r, Y'nin zaman içindeki (bileşik) büyüme hızıdır. Yukarıdaki denklemin logaritmasını alalım:

$$\ln Y_t = \ln Y_0 + t \ln(1 + r)$$

- $\beta_1 = \ln Y_0$ ve $\beta_2 = \ln(1+r)$ tanımlamalarını yapıp hata terimini de ekledikten sonra modeli şöyle yazabiliriz:

$$\ln Y_t = \beta_1 + \beta_2 t + u_t$$

- Yukarıda gösterilen modele *"log-doğ"* (log-lin) modeli denir.

Bu noktada, sık sık karşılaştığımız *"mutlak değişim"* (absolute change), *"göreli değişim"* (relative change) ve *"yüzde değişim"* (percentage change) terimleri arasındaki farka dikkat edelim:

Mutlak değişim

$$\Delta X$$

Göreli değişim

$$\Delta X/X$$

Yüzde değişim

$$100 \times \Delta X/X$$

Eğer X'deki değişim küçükse, aşağıda gösterilen *"yaklaştırma"* (approximation) uygulamada sıklıkla kullanılır:

$$\Delta \ln X \approx \Delta X/X \text{ (göreli değişim)}$$

- Log-doğ modeline geri dönelim:

$$\ln Y_t = \beta_1 + \beta_2 t + u_t$$

- Bu modelde β_2 katsayısı, açıklayıcı değişken t'deki mutlak bir değişmeye karşılık Y'deki göreli değişimi ölçmektedir:

$$\beta_2 = \frac{\Delta \ln Y}{\Delta t}$$

- Diğer bir deyişle, β_2 katsayısı Y_t değişkenindeki büyüme hızını ($\beta_2 > 1$) ya da küçülme hızını ($\beta_2 < 1$) vermektedir.
- Bu nedenle, log-doğ modeline aynı zamanda *"sabit büyüme"* (fixed growth) modeli de denir.

Reel GSSSO örneğine dönersek, log-doğ modeline dayanan bağlanım bulgularının aşağıdaki gibi olduğunu görürüz:

$$\begin{array}{lllll} \widehat{\text{GSSSO}}_t = & 2{,}8516 & + \; 0{,}0509 \; t & \\ \text{öh} & (0{,}0517) & (0{,}0061) & \\ t & (55{,}1830) & (8{,}3932) & r^2 = 0{,}8544 \end{array}$$

- Buna göre, 1987-2000 döneminde Türkiye'de gayri safi sabit sermaye oluşumu yılda ortalama yüzde 5,09'dur.
- Ayrıca, $\ln Y_0 = 2{,}8516$'nın anti-logaritmasını alırsak bulacağımız 17,3155 değeri de 1987 yılı için GSSSO'nun yaklaşık 17,3 milyon TL olarak tahmin edildiğini gösterir.

Doğrusal Eğilim Modeli

- Araştırmacılar kimi zaman log-doğ modeli yerine aşağıdaki modeli tahmin ederler:

$$Y_t = \beta_1 + \beta_2 t + u_t$$

- $\ln Y_t$ yerine Y_t'nin zamana göre bağlanımının hesaplandığı bu modele *"doğrusal eğilim"* (linear trend) modeli denir.
- Buradaki t, *"eğilim"* (trend) değişkeni diye adlandırılır.
- Eğer β_2 eğim katsayısı artı çıkarsa Y_t'de zaman içinde bir artış eğilimi, eksi çıkarsa da bir düşüş eğilimi var demektir.

Log-doğ ve doğrusal eğilim modellerine ilişkin iki noktayı özellikle belirtmekte yarar vardır:

1. İki modelin bağımlı değişkenleri farklı olduğu için bu modellerin r^2 değerlerini karşılaştırmak doğru değildir.
2. Bağımlı değişkenin zaman içinde değişiminin bu şekilde incelenmesi ancak zaman serisinin *"durağan"* (stationary) olması durumunda uygundur.

Durağanlık kavramı ileride zaman serileri ekonometrisi konusu altında incelenecektir.

GSSSO örneğimize geri dönelim ve şimdi de doğrusal eğilim modelini tahmin edelim:

$$\begin{array}{llll} \widehat{\text{GSSSO}}_t = & 16{,}3621 & + \ 1{,}2848 & t \\ \text{öh} & (1{,}4170) & (0{,}1664) & \\ t & (11{,}5466) & (7{,}7202) & r^2 = 0{,}8324 \end{array}$$

- Buna göre, 1987-2000 döneminde Türkiye'de reel GSSSO yılda yaklaşık 1,3 milyon TL olarak gerçekleşmiştir.
- Demek ki bu dönemde reel GSSSO'da artış eğilimi vardır.

Doğ-Log Modeli

- Eğer X'deki yüzde değişime karşılık Y'deki mutlak değişim ile ilgileniyorsak, buna uygun bir modeli şöyle yazabiliriz:

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

- Yukarıdaki modele *"doğ-log"* (lin-log) modeli denir.
- Bu modelde β_2 katsayısını kullanarak şunu gösterebiliriz:

$$\beta_2 = \frac{\Delta Y}{\Delta X / X} \quad \Rightarrow \quad \Delta Y = \beta_2 (\frac{\Delta X}{X})$$

- Böylece X'deki 0,01 (yüzde 1) oranındaki göreli değişmeye karşı Y'de $\beta_2 \times 0{,}01$ boyutunda mutlak değişme olmaktadır.
- Dolayısıyla, doğ-log modelini yorumlarken eğim katsayısı β_2'yi önce 0,01 ile çarparız.

Örnek olarak 1987-2006 yıllarında Türkiye'deki GSYH ve M2 para arzı verilerini kullanarak doğ-log modelini tahmin edelim:

$$\begin{array}{llll} \widehat{\text{GSYH}}_t = & -26{,}7905 & + \ 41{,}9796 & \ln \text{M2}_t \\ \text{öh} & (13{,}6546) & (4{,}2488) & \\ t & (-1{,}9620) & (9{,}8805) & r^2 = 0{,}8443 \end{array}$$

- 41,98 büyüklüğündeki eğim katsayısının anlamı, örneklem döneminde para arzındaki yüzde 1'lik bir artışın GSYH'de ortalama 0,4198 milyon liralık artışa yol açmış olduğudur.

7.3.3 Evrik ve Log-Evrik Modeller

Evrik Model

- Aşağıda gösterilen türden modellere evrik model denir:

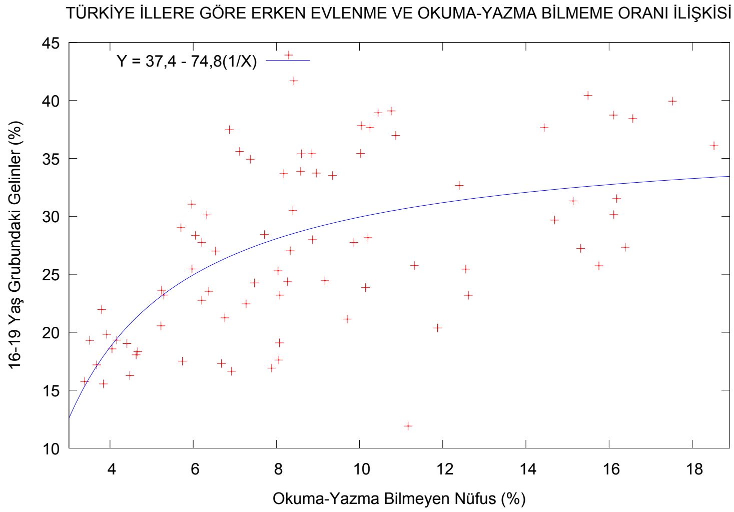
$$Y_i = \beta_1 + \beta_2 \frac{1}{X_i} + u_i$$

- Yukarıdaki model, X değişkeni modele evrik girdiğinden, X'te doğrusal değildir ama β_1 ve β_2'de doğrusaldır.
- Modelin önemli özelliği, X sonsuza yaklaşırken Y'nin de β_1 *"kavuşmazsal"* (asymptotic) değerine yakınsamasıdır.
- Dolayısıyla, evrik modellerde açıklayıcı değişken artarken bağımlı değişkenin yaklaştığı bir limit değeri bulunur.
- Bu tür modellere örnek olarak Phillips eğrisi ya da üretimin ortalama sabit gider ile olan ilişkisi verilebilir.

Bir evrik model uygulaması olarak 2009 yılında Türkiye'de illere göre 16-19 yaş grubundaki gelinlerin oranı (Y) ile okuma yazma bilmeyenlerin toplam nüfusa oranı (X) verilerine bakalım:

$$\begin{array}{lllll} \widehat{Y}_i = & 37{,}4131 & - & 74{,}7805 & 1/X_i \\ \text{öh} & (1{,}7221) & & (11{,}7015) & \qquad r^2 = 0{,}3408 \\ t & (21{,}7253) & & (-6{,}3907) & \qquad F_{1,79} = 40{,}8407 \end{array}$$

- Buna göre erken evliliklerde tavan oran yaklaşık %36,7'dir.
- Şöyle ki $X = \%100$ ve $1/X = 0{,}01$ olunca 16-19 yaşında evlenen bayanların oranı da % $(37{,}4131 - 0{,}7478)$ olur.
- *Dikkat:* Gelir gibi diğer önemli etmenleri de göz önüne alan bir modelde bu kavuşmazsal oran daha düşük çıkacaktır.



Log-Evrik Model

- Evrik modelin bir türü olan *"log-evrik"* (log-reciprocal) model aşağıdaki biçimi alır:

$$\ln Y_i = \beta_1 - \beta_2 \frac{1}{X_i} + u_i$$

- Türev hesabı kullanılarak burada Y'nin X'e göre eğimi $\mathrm{d}/\mathrm{d}X(\ln Y_i) = \beta_2(1/X_i^2)$ olarak bulunur.
- Model çizim üzerinde incelendiğinde de X artarken Y'deki artışın önce dışbükey ve daha sonra da içbükey görünüm sergilediği anlaşılır.
- Öyleyse böyle bir model sermaye sabitken üretimin önce artarak arttığı ve sonra da azalarak arttığı üretim-işgücü ilişkisini çözümlemede kullanılabilir.

İşlev Biçiminin Seçimi

Ele almış olduğumuz çeşitli model işlev biçimlerine ilişkin eğim ve esneklik bilgileri aşağıdaki çizelgede verilmiştir.

Çizelge: Çeşitli İşlev Biçimlerinin Eğim ve Esneklikleri

Model	İşlev Biçimi	Eğim ($\frac{dY}{dX}$)	Esneklik ($\frac{dY}{dX}\frac{X}{Y}$)
Doğrusal	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2 \frac{X}{Y}$
Log-Log	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2 \left(\frac{Y}{X}\right)$	β_2
Log-Doğ	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 (Y)$	$\beta_2 (X)$
Doğ-Log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2 \left(\frac{1}{X}\right)$	$\beta_2 \left(\frac{1}{Y}\right)$
Evrik	$Y = \beta_1 + \beta_2 \left(\frac{1}{X}\right)$	$-\beta_2 \left(\frac{1}{X^2}\right)$	$-\beta_2 \left(\frac{1}{XY}\right)$
Log-Evrik	$\ln Y = \beta_1 - \beta_2 \left(\frac{1}{X}\right)$	$\beta_2 \left(\frac{Y}{X^2}\right)$	$\beta_2 \left(\frac{1}{X}\right)$

Görgül çalışmalarda model seçiminin deneyim gerektirdiği açıktır. Yardımcı olabilecek birkaç nokta şunlardır:

1. Bazı durumlarda iktisat kuramı belli bir işlev biçimini gösterebilir ya da öngörebilir.
2. Tahmin edilen katsayıların önsel beklentileri karşıladığı doğrulanmalıdır.
3. Almaşık modelleri karşılaştırmak için eğim ve esneklik katsayılarını hesaplamak yardımcı olabilir.
4. Veri setine iki farklı model yakıştırıldığında, eğer bağımlı değişkenler aynı ise r^2 değerleri karşılaştırılabilir.
5. Ancak iki modeli r^2 temelinde karşılaştırmak her zaman uygun değildir. Bunun bir nedeni, eklenen her açıklayıcı değişkenin r^2'yi yükseltecek olmasıdır.

İşlev biçiminin seçimine ilişkin olarak, aşağıdaki hata terimsiz bağlanım modelini ele alalım:

$$Y_i = \beta_1 X_i^{\beta_2}$$

Bu modeli tahmin amacıyla üç farklı şekilde yazabiliriz:

$Y_i = \beta_1 X_i^{\beta_2} u_i$

$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$

$Y_i = \beta_1 X_i^{\beta_2} + u_i$

İki yanlı logaritmalarını alırsak da şunları elde ederiz:

$\ln Y_i = \alpha + \beta_2 \ln X_i + \ln u_i$

$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i$

$\ln Y_i = \ln(\beta_1 X_i^{\beta_2} + u_i)$

- Yukarıda görülen $\alpha = \ln \beta_1$'dir.
- İlk iki model değiştirgelerde doğrusalken, üçüncü modelin özünde doğrusal-dışı olduğuna dikkat ediniz.
- SEK'in EDYT özelliğinin hatalarda sıfır ortalama ve sabit varyans aradığını anımsayalım.
- Ayrıca önsav sınaması için u_i'lerin normal dağılımlı olduğu, kısaca $u_i \sim N(0, \sigma^2)$ varsayılmaktadır.
- Buna göre, örneğimizdeki ikinci modeli kullanmak istersek $\ln u_i \sim N(0, \sigma^2)$ varsaymamız gereklidir.
- Ancak eğer $\ln u_i \sim N(0, \sigma^2)$ ise, ilk modeldeki u_i de $e^{\sigma^2/2}$ ortalama, $e^{\sigma^2}(e^{\sigma^2} - 1)$ varyansla log-normal dağılımlı olur.
- Üçüncü model ise değiştirgelerde doğrusal-dışı olduğu için ancak yinelemesel bir yöntem ile çözülebilir.
- Sonuç olarak, modeli bağlanım için dönüştürürken hata terimine özel bir dikkat göstermek gereklidir.
- Hatalı doğrusallaştırma, arzulanan istatistiksel özellikleri taşımayan bir modele yol açabilir.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 6* “Extensions of the Two-Variable Regression Model” okunacak.

Önümüzdeki Ders
Çoklu Bağlanım Çözümlemesi: Tahmin Sorunu

Bölüm 8

Çoklu Bağlanım Çözümlemesi - Tahmin Sorunu

8.1 Üç Değişkenli Model

8.1.1 Gösterim ve Varsayımlar

- Önceki bölümlerde bağımlı değişken Y'nin yalnızca bir açıklayıcı değişken X tarafından etkilendiği varsayılmıştı.
- Ancak iktisat kuramı bu denli basit değildir.
- *Örnek:* Bir mala olan talep yalnızca o malın fiyatına değil; ikame ya da tamamlayıcı malların fiyatına, gelir düzeyine, nüfusa ve diğer değişkenlere de bağlı olabilir.
- *Örnek:* Tüketim harcamaları yalnızca gelir ile değil; kişinin yaşı, eğitim düzeyi, cinsiyeti, toplam serveti ve benzer değişkenler ile de ilişkili olabilir.
- Modele başka değişkenler eklemek bizi çoklu bağlanım çözümlemesine götürür.
- En basit çoklu bağlanım modeli, bir bağımlı ve iki açıklayıcı değişkenden oluşan üç değişkenli bağlanımdır:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

- Burada Y bağımlı değişken, X_2 ve X_3 açıklayıcı değişkenler, u olasılıksal hata terimi, i gözlem no'sudur.

- β_1, modelde bulunmayan tüm değişkenlerin Y üzerindeki ortalama etkisini gösteren sabit terimdir.
- β_2 ve β_3'e de *"kısmi bağlanım katsayısı"* (partial regression coefficient) adı verilir.

Üç değişkenli modeldeki kısmi bağlanım katsayılarının anlamı şudur:

- β_2, X_3 sabit tutulurken X_2'deki bir birimlik değişmeye karşı Y'nin beklenen değeri $E(Y|X_2, X_3)$'teki değişmeyi ölçer.
- Bir başka deyişle β_2, X_3 sabitken $E(Y|X_2, X_3)$'ün X_2'ye göre eğimini verir.
- Diğer bir deyişle β_2, X_2'deki bir birimlik değişmenin Y üzerindeki X_3'ten ayrı, net etkisini gösterir.
- β_3'ün yorumu da benzer şekildedir.

Üç Değişkenli Model Varsayımları

Daha önce KDBM çerçevesinde yapılmış olan varsayımlar, k değişkenli çoklu bağlanım modeli için de geçerlidir:

1. Çoklu bağlanım modeli değiştirgelerde doğrusaldır.
2. Açıklayıcı değişkenler tekrarlı örneklemlerde değişmez.
3. Açıklayıcı değişkenlerde yeterli değişkenlik bulunur.
4. Hata teriminin ortalaması sıfırdır: $E(u_i|X_{2i}, X_{3i}, \ldots, X_{ki}) = 0$
5. Hata teriminin varyansı sabittir: $\text{var}(u_i) = \sigma^2$
6. u_i ve X'ler birbirlerinden bağımsız dağılmaktadır:
$$\text{cov}(u_i, X_{2i}) = \text{cov}(u_i, X_{3i}) = \ldots = \text{cov}(u_i, X_{ki}) = 0$$
7. *"Serisel ilinti"* (serial correlation) bulunmamaktadır:
$$\text{cov}(u_i, u_j) = 0 \; (i \neq j)$$
8. *"Model belirtim hatası"* (model specification error) yoktur.
9. X_2 ile X_3 arasında *"tam eşdoğrusallık"* (exact collinearity) bulunmamaktadır.

Eşdoğrusallık Kavramı

- X_2 ile X_3 arasında tam doğrusal ilişki olmadığı yönündeki SEK varsayımını anımsayalım.
- *"eşdoğrusal-dışılık"* (non-collinearity) varsayımına göre, aşağıdaki gibi tanımlanan iki değişken doğrusal bağımlıdır:

$$X_{2i} = aX_{3i} \qquad \text{ya da} \qquad X_{2i} - aX_{3i} = 0, \qquad a \in \mathrm{R}$$

- Dolayısıyla, X_2 ve X_3 eğer aynı modelde yer alırlarsa tam eşdoğrusal ilişki ortaya çıkar.
- Tam eşdoğrusallık çoklu bağlanımda önemli bir konudur çünkü bu durumda açıklayıcı değişkenlerin bağımlı değişken üzerindeki tekil etkilerini bulmanın yolu yoktur.
- Tam eşdoğrusallık olması durumunda kısaca elde iki değil bir bağımsız değişken var demektir.
- *Örnek:* $4X_{2i} = X_{3i}$ olsun. Bu durumda üç değişkenli model ikili modele indirgenir:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3(4X_{2i}) + u_i \\ &= \beta_1 + \quad (\beta_2 + 4\beta_3)X_{2i} \quad + u_i \\ &= \beta_1 + \qquad \alpha X_{2i} \qquad + u_i \end{aligned}$$

- Diğer yandan, eğer $X_{3i} = X_{2i}^2$ ise iki değişken arasındaki ilişki doğrusal değildir. Bu durumda da eşdoğrusal-dışılık varsayımı çiğnenmiş olmaz.

8.1.2 Kısmi Bağlanım Katsayılarının Tahmini

SEK Tahmincileri

- Üç değişkenli modelin SEK tahmincilerini bulmak için önce örneklem bağlanım işlevini aşağıdaki gibi yazalım:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

- SEK yöntemi, anakütle tahmincilerini kalıntı kareleri toplamı ($\sum \hat{u_i}^2$) en küçük olacak biçimde hesaplar:

$$\min \sum \hat{u_i}^2 = \min \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$$

- Yukarıdaki eşitliği enazlayacak en doğrudan süreç eşitliğin $\hat{\beta}$'lara göre türevini almak, bunları sıfıra eşitlemek ve daha sonra eşanlı olarak çözmektir.

Üç değişkenli model için SEK yöntemi şu tahmincileri verir:

$$\begin{aligned}
\hat{\beta}_1 &= \hat{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \\
\hat{\beta}_2 &= \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \\
\hat{\beta}_3 &= \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}
\end{aligned}$$

- $\hat{\beta}_2$ ve $\hat{\beta}_3$ tahmincileri bakışımlıdır ve paydaları aynıdır.
- Demek ki X_2 ile X_3'ün yerleri değiştirilirse $\hat{\beta}_2$ ile $\hat{\beta}_3$'ün de yeri değişir ama bu bağlanım sonuçlarını etkilemez.

Varyans ve Ölçünlü Hatalar

SEK tahmincilerinin varyansları ise aşağıdaki gibi bulunur:

$$\begin{aligned}
\text{var}(\hat{\beta}_1) &= (\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i} x_{3i}}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2})\sigma^2 \\
\text{var}(\hat{\beta}_2) &= \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\sigma^2 = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \\
\text{var}(\hat{\beta}_3) &= \frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\sigma^2 = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}
\end{aligned}$$

- var($\hat{\beta}_2$) ve var($\hat{\beta}_3$) formüllerinde yer alan r_{23}, X_2 ve X_3 arasındaki örneklem ilinti katsayısı r'dir.
- Ölçünlü hatalar ise varyansların artı değerli kareköklerididir:

$$\text{öh}(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})}.$$

- Varyans ve ölçünlü hata formüllerindeki σ^2'nin anakütle hata terimi u_i'nin sabit varyansı olduğunu biliyoruz.
- Bu anakütle katsayısının yansız tahmincisi şöyledir:

$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-3}$$

- σ^2'nin bu tahmincisi ile iki değişkenli modeldeki tahmincisi $(\sum \hat{u_i}^2/n-2)$ benzerdir. Aralarındaki tek fark üç değişkenli model için serbestlik derecesinin artık $(n-3)$ olmasıdır.
- Kalıntılar bulunduktan sonra $\hat{\sigma}^2$ kolayca hesaplanabilir.
- Kalıntı kareleri toplamı ise şu eşitlik ile kolayca bulunabilir:

$$\sum \hat{u_i}^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

SEK Tahmincilerinin Özellikleri

Üç değişkenli model için SEK tahmincilerinin özellikleri iki değişkenli model ile aynıdır:

1. Üç değişkenli bağlanım doğrusu (düzlemi) $\bar{Y}$, $\bar{X}_2$, $\bar{X}_3$ ortalamalarından geçer: $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X}_2 + \hat{\beta}_3\bar{X}_3$
2. $\hat{Y}_i$'nın ortalaması gözlenen Y_i ortalamasına eşittir: $\bar{\hat{Y}}_i = \bar{Y}_i$
3. Kalıntılar toplamı sıfıra eşittir: $\sum \hat{u}_i = n\bar{\hat{u}}_i = \bar{\hat{u}}_i = 0$
4. Kalıntılar X_{2i} ve X_{3i} ile ilişkisizdir: $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$
5. $\hat{u}_i$ kalıntıları $\hat{Y}_i$ ile de ilişkisizdir: $\sum \hat{u}_i \hat{Y}_i = 0$
6. Varyans formüllerinden görüldüğü gibi, X_2 ile X_3 arasındaki ilinti katsayısı r_{23} artarken $\hat{\beta}_2$ ve $\hat{\beta}_3$'nın varyansları yükselir.
7. Gözlem sayısı n artarken $\hat{\beta}_2$ ve $\hat{\beta}_3$'nın varyansları da azalır.
8. $\hat{\beta}_2$ ve $\hat{\beta}_3$ tahmincileri, en iyi doğrusal yansız tahminci ya da kısaca EDYT'dirler.

ÖBİ'nin Sapmalar Biçimi Gösterimi

Çok değişkenli modelde ÖBİ'nin sapmalar biçiminde gösterimi aşağıda gösterilen şekilde elde edilir:

1. Üçlü bağlanım modelini ele alalım:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

2. Bağlanım yüzeyi $\bar{Y}$, $\bar{X}_2$, $\bar{X}_3$ ortalamalarından geçtiği için:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$$

3. İkinci denklemi birinciden çıkartırsak şunu buluruz:

$$\begin{array}{rcccccc} & \hat{Y}_i = & \hat{\beta}_1 & + & \hat{\beta}_2 X_{2i} & + & \hat{\beta}_3 X_{3i} \\ - & \bar{Y} = & \hat{\beta}_1 & + & \hat{\beta}_2 \bar{X}_2 & + & \hat{\beta}_3 \bar{X}_3 \\ \hline \end{array}$$

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_1 - \hat{\beta}_1 + \hat{\beta}_2 (X_{2i} - \bar{X}_2) + \hat{\beta}_3 (X_{3i} - \bar{X}_3)$$

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

Ençok Olabilirlik Tahmincileri

- İki değişkenli modelde olduğu gibi çoklu modeller için de bağlanım katsayılarının SEK ve EO tahmincileri aynıdır.
- Ancak üçlü modelde σ^2'nin SEK tahmincisi $\sum \hat{u}_i{}^2/(n-3)$ iken EO tahmincisi modelde kaç değişken olursa olsun $\sum \hat{u}_i{}^2/n$ olarak bulunur.
- Diğer bir deyişle SEK tahmincisi serbestlik derecesini hesaba katarken yanlı EO tahmincisi bunu dikkate almaz.
- Eğer n çok büyükse kuşkusuz EO ve SEK tahmincileri birbirlerine yaklaşırlar.

8.2 Çoklu Bağlanımda Yakışmanın İyiliği

8.2.1 Çoklu Belirleme ve İlinti Katsayıları

- İki değişkenli durum için geliştirmiş olduğumuz r^2, ikiden çok değişkenli bağlanım modellerine de genişletilebilir.
- Çoklu modelde bu istatistiğe R^2 ya da *"çoklu belirleme katsayısı"* (multiple coefficient of determination) denir.
- R^2, bağımlı değişken Y'deki değişimin $X_2, X_3, \ldots, X_n$ ile topluca açıklanabilme oranını gösterir.

  Çoklu Belirleme Katsayısı

$$R^2 = \frac{\text{BKT}}{\text{TKT}} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \cdots + \hat{\beta}_n \sum y_i x_{ni}}{\sum y_i^2}$$

- R^2 de r^2 gibi 0 ile 1 arasındadır.
- R^2 1'e ne kadar yakınsa modelin verilere yakışması da o kadar iyidir. Eğer $R^2 = 1$ ise, yakıştırılan bağlanım Y'deki değişimin tamamını açıklıyor demektir.
- İki değişken arasındaki doğrusal ilişkinin derecesini ölçen r'nin çoklu bağlanımdaki karşılığı da *"çoklu ilinti katsayısı"* (coefficient of multiple correlation) olup, R ile gösterilir:

Çoklu İlinti Katsayısı

$$R = \pm\sqrt{R^2}$$

- R değeri, bağımlı değişken Y ile tüm açıklayıcı değişkenler arasındaki ortak ilişkinin derecesini ölçer.
- Diğer taraftan uygulamada R'nin önemi azdır. Bağlanım çözümlemesi çerçevesinde asıl anlamlı büyüklük R^2'dir.
- R^2'nin önemli bir özelliği, modelde bulunan açıklayıcı değişken sayısının azalmayan bir işlevi olmasıdır.
- Diğer bir deyişle açıklayıcı değişken sayısı arttıkça R^2 hemen hemen her zaman artar, asla azalmaz.

- Bunu görebilmek için belirleme katsayısının tanımını anımsayalım:

$$R^2 = 1 - \frac{\text{KKT}}{\text{TKT}} = 1 - \frac{\sum \hat{u}_i{}^2}{\sum y_i^2}$$

- Burada TKT, X'lerin sayısından bağımsızdır. KKT ise açıklayıcı değişken sayısı arttıkça azalma eğilimine girer.
- Bu nedenle, bağımlı değişkeni aynı olan ama farklı sayıda açıklayıcı değişken içeren iki ayrı bağlanım modeline ait R^2 değerleri karşılaştırırken dikkatli olunmalıdır.

R^2 Değerlerinin Karşılaştırılması

- İki R^2 değerini karşılaştırırken modelde var olan açıklayıcı değişken sayısını da dikkate alma gereksinimi *"ayarlamalı"* (adjusted) belirleme katsayısı $\bar{R}^2$ tanımına yol açmıştır:

  Ayarlamalı Belirleme Katsayısı

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i{}^2/(n-k)}{\sum y_i^2/(n-1)} \qquad \text{ya da} \qquad \bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

- Burada k, sabit terimle birlikte modeldeki katsayı sayısıdır. s_Y^2 ise Y'nin örneklem varyansıdır.
- Ayarlamalı sözcüğü, giren kareler toplamının serbestlik derecesine göre ayarlanmış olduğu anlamına gelir.
- *Dikkat:* Üç değişkenli bağlanım için $\sum \hat{u}_i{}^2$ sd'sinin $(n-3)$ olduğunu anımsayınız.
- $\bar{R}^2$'nin R^2 ile ilişkisini aşağıdaki eşitlikle gösterebiliriz:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

- Buradan da görülüyor ki $k > 1$ olduğunda $\bar{R}^2 < R^2$'dir.
- Diğer bir deyişle, X'lerin sayısı arttıkça ayarlamalı R^2 *"ayarlamasız"* (unadjusted) R^2'ye göre daha az artar.

- Ayrıca $\bar{R}^2$'nin eksi değerler de alabildiği görülmektedir. Eğer $\bar{R}^2$ eksi bulunursa uygulamada sıfır kabul edilir.

- Tüm modern ekonometri yazılımları alışıldık R^2'nin yanısıra ayarlamalı R^2 istatistiğini de verir.

İki farklı modeli ayarlamalı ya da ayarlamasız R^2 temelinde karşılaştırabilmek için iki noktaya daha dikkat edilmelidir:

1. Örneklem büyüklüğü n her iki model için aynı olmalıdır. *Dikkat:* Modele gözlem eklendiğinde ya da çıkartıldığında, hesaplanan R^2'nin de değişeceğini unutmayınız.

2. Bağımlı değişken Y de her iki model için aynı olmalıdır. *Dikkat:* R^2 değerinin, X açıklayıcı değişkenlerinin Y'deki değişimi açıklama oranını gösterdiğini anımsayınız. Eğer Y'ler farklıysa, hesaplanan R^2'ler de farklı şeylerin değişim oranını göstereceği için karşılaştırılamaz.

- Bağımlı değişkenleri aynı olmayan iki model düşünelim:

$$\begin{array}{rcccccc} \hat{Y}_i & = & \beta_1 & + & \beta_2 X_{1i} & + & \beta_3 X_{2i} \\ \widehat{\ln Y_i} & = & \alpha_1 & + & \alpha_2 \ln X_{1i} & + & \alpha_3 \ln X_{2i} \end{array}$$

- Burada R^2 değerlerini karşılaştırmak için 2 yol izlenebilir:

1. Yol
İkinci modelden tahmin edilen $\widehat{\ln Y_i}$'ların anti-logaritmaları alınır. Bulunan değerler ile Y_i arasında hesaplanan r^2 değeri birinci modeldeki R^2 ile karşılaştırılabilir.

2. Yol
Birinci modelden tahmin edilen $\hat{Y}_i$'ların logaritmaları alınır. Bulunan değerler ile $\ln Y_i$ arasında hesaplanan r^2 değeri ikinci modeldeki R^2 ile karşılaştırılabilir.

- Bağımlı değişkenleri farklı modelleri karşılaştırmak için, iki değişken arasındaki ilinti formülünün karesine dayanan şu r^2 formülü kullanılabilir:

$$r^2 = \frac{\sum (y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i{}^2)}$$

- Son olarak, R^2'nin yakışmanın iyiliğini ölçmede kullanılan istatistiklerden yalnızca biri olduğu unutulmamalıdır.

- Model seçimi için başka ölçütler de bulunmaktadır:

"Akaike bilgi ölçütü" (Akaike information criterion)
"Schwarz Bayesçi ölçüt" (Schwarz Bayesian criterion)
"Hannan-Quinn ölçütü" (Hannan-Quinn criterion)

- Araştırmacının asıl ilgisi, açıklayıcı değişkenlerin bağımlı değişken ile olan mantıksal ya da kuramsal ilişkilerine ve bunların istatistiksel anlamlılıklarına yönelik olmalıdır.

8.2.2 Kısmi İlinti Katsayıları

- İki değişken arasındaki doğrudan ilişkinin bir ölçüsü olarak tanımlanan ilinti katsayısı r kavramını anımsayalım.
- Üç değişkenli model için böyle üç ayrı *"basit ilinti katsayısı"* (simple correlation coefficient) değerinden söz edilebilir:

Basit İlinti Katsayıları

Y ile X_2 arasındaki ilinti katsayısı: r_{12}
Y ile X_3 arasındaki ilinti katsayısı: r_{13}
X_2 ile X_3 arasındaki ilinti katsayısı: r_{23}

- Bunlara aynı zamanda *"sıfırıncı dereceden ilinti katsayısı"* (correlation coefficient of zero order) da denmektedir.
- Eğer bir X_3 değişkeni hem Y hem de X_2 ile ilişkiliyse, bu durumda Y ve X_2 arasındaki basit ilinti r_{12} yanıltıcıdır.
- İki değişken arasında, üçüncü bir değişkenin etkisinden bağımsız olarak bulunan *"kısmi ilinti katsayısı"* (partial correlation coefficient) ise şöyle tanımlanır:

Kısmi İlinti Katsayıları

X_3 sabitken Y ile X_2 arasındaki kısmi ilinti: $r_{12.3}$
X_2 sabitken Y ile X_3 arasındaki kısmi ilinti: $r_{13.2}$
Y sabitken X_2 ile X_3 arasındaki kısmi ilinti: $r_{23.1}$

- Bunlara *"birinci dereceden"* (first order) ilinti katsayıları denir. Buradaki derece ikincil alt imlerin sayısıdır.
- Buna göre, X_3 ve ikinci bir X_4 sabit tutulurken bulunan $r_{12.34}$ değerine de ikinci dereceden bir ilinti katsayısı denir.

- Birinci dereceden kısmi ilinti katsayılarını hesaplamak için aşağıdaki eşitlikler kullanılabilir:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$
$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$
$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

Çok değişkenli modellerde basit ilinti katsayılarını yorumlarken şu noktalara dikkat etmek gereklidir:

- $r_{12} = 0$ olsa bile, aynı anda r_{13} ya da r_{23} de sıfır olmazsa $r_{12.3} = 0$ olmaz.
- $r_{12.3}$ ile r_{12} aynı işareti taşımak zorunda değildir.
- $r_{13} = r_{23} = 0$ olması $r_{12} = 0$ anlamına gelmez.
- İkili bağlanımdaki $0 \leq r^2 \leq 1$ tanımını anımsayalım. Kısmi ilinti katsayıları kareleri için de geçerli olan bu durumdan yararlanılarak, üç sıfırıncı dereceden ilinti katsayısı arasındaki ilişki şöyle gösterilebilir:

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$$

- Yukarıdaki eşitsizlikten de anlaşılabileceği gibi, Y ile X_2'nin ve X_2 ile de X_3'ün ilintisiz olması Y ile X_3'ün ilintisiz olacağı anlamına gelmemektedir.

8.2.3 Çoklu Bağlanım Açıklayıcı Örnek

- Çoklu bağlanıma örnek olarak 2005-2009 aylık verilerini alalım ve Türkiye için bir *“beklentilerle-genişletmeli Phillips eğrisi”* (expectations-augmented Phillips curve) modeli belirtelim:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t$$

- Burada
  Y_t TÜFE değerini (2005 Ocak=100),
  X_{2t} işsiz sayısını (bin kişi, mevsimsel ayarlamalı),
  X_{3t} ise beklenen TÜFE değerini
  göstermektedir.
- İktisat kuramına göre β_2 eksi, β_3 ise artı değerli olmalıdır.
- Aslında kurama göre $\beta_3 = 1$ beklentisi vardır.

SEK yöntemi ile elde edilen bağlanım bulguları şöyledir:

$$\begin{array}{lllll} \ln \hat{Y}_t = & -0{,}1879 & - \quad 0{,}0364 \quad \ln X_{2t} + & 1{,}1012 \quad \ln X_{3t} & \\ \text{öh} & (0{,}1072) & (0{,}0166) & (0{,}0120) & \\ t & (-1{,}7535) & (-2{,}1960) & (91{,}8156) & R^2 = 0{,}9963 \end{array}$$

- $\hat{\beta}_2$ ve $\hat{\beta}_3$ önsel beklentilerle uyumlu işaret taşımaktadır.
- $\hat{\beta}_1$'ya göre, X_2 ve X_3 dışındaki diğer tüm etmenler TÜFE üzerinde ortalama $e^{-0{,}1879} \approx 0{,}83$ etkiye yol açmaktadır.
- $\hat{\beta}_2$ kısmi bağlanım katsayısı ise X_3 sabit tutulduğunda işsizlikteki %1'lik bir artışa karşılık TÜFE'nin de yaklaşık %0,036 düşeceği anlamına gelir.
- Bulunan bu düşük değer, Türkiye'de enflasyon ve işsizlik arasındaki ilişkinin zayıf olduğu önsel bilgisi ile uyumludur.
- R^2 değeri, enflasyon oranındaki değişimin %99'unun bu iki açıklayıcı değişkenle açıklanabildiğini öne sürer. Bu kadar yüksek bir R^2 bağlanıma kuşkuyla yaklaşmayı gerektirir.

Model Belirtim Yanlılığı Sorunu

- Klasik doğrusal bağlanım modeli varsayımlarına göre bağlanım modeli doğru kurulmuş olmalıdır.
- Eğer çözümlemede kullanılacak bağlanım modeli yanlış kurulursa *"model belirtim yanlılığı"* (model specification bias) ortaya çıkar.
- Bu varsayımın önemini vurgulayabilmek için elimizdeki Phillips eğrisi modeli yardımcı olabilir.

(... devam)

- Az önce ele almış olduğumuz aşağıdaki üçlü bağlanım modelinin "doğru" model olduğunu varsayalım:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_{1t}$$

- Elimizdeki Türkiye verilerini şu iki değişkenli modele yakıştırmakta diretiyor olalım:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + u_{2t}$$

- Y_t burada t dönemindeki TÜFE değerini, X_{2t} ise toplam işsiz sayısını göstermektedir.
- Birinci model "doğru" olduğuna göre ikinci model bir model belirtim hatası içermektedir.
- Buradaki hata, X_{3t} beklenen TÜFE değişkenini modelden dışlamış olmaktır.
- Birinci modeldeki $\hat{\beta}_2$'nın gerçek β_2'nin yansız bir tahmincisi olduğunu biliyoruz.
- Diğer yandan ikinci modeldeki $\hat{\alpha}_2$ değiştirgesi β_2'nin yansız tahmincisi değildir.
- α_2'nin aslında X_3'ün X_2'ye göre bağlanımından ortaya çıkan eğim değiştirgesi α_3 ile ilişkili olduğu gösterilebilir:

$$\alpha_2 = \beta_2 + \beta_3\alpha_3 + \text{hata terimi}$$

- Buna göre $E(\alpha_2)$ beklenen değeri β_2 değil de $\beta_2 + \beta_3\alpha_3$ olarak karşımıza çıkmaktadır.
- Sonuç olarak, ilk modeldeki β_2 değiştirgesi X_2'nin Y üzerindeki doğrudan ya da tekil etkisini ölçmektedir.
- Hatalı modeldeki α_2 değiştirgesi ise X_2'nin Y üzerindeki hem doğrudan hem de X_3 üzerinden dolaylı etkisini verir.

Hatalı modelin SEK tahmini aşağıdaki bulguları vermektedir:

$$\begin{array}{llll} \ln \hat{Y}_t = & -1{,}7203 & + \ 0{,}8327 & \ln X_{2t} \\ \text{öh} & (1{,}4369) & (0{,}1845) & \\ t & (-1{,}1972) & (4{,}5142) & r^2 = 0{,}3070 \end{array}$$

- Kuramsal beklentinin aksine α_2 burada artı değerlidir ve 0,83 gibi yüksek, gerçek dışı bir büyüklüktedir.
- Demek ki belli bir model "doğru" olarak kabul ediliyorsa bir ya da birkaç değişkeni çıkartarak modeli değiştirmek yanlı tahminlere yol açmaktadır.
- Yanlış belirtilen bir model anakütle katsayı tahminlerinin yanlı olması gibi ciddi bir soruna neden olabilmektedir.

8.3 Çokterimli Bağlanım Modelleri

- Çoklu bağlanımın bir şekli de *"çokterimli"* (polynomial) bağlanım modelleridir.
- Şimdiye kadar ele aldığımız tüm örneklerde bağlanım işlevinin değişkenlerde doğrusal olduğunu varsamıştık.
- Gerçek hayatta bu varsayımın geçerli olmadığı pek çok durum düşünülebilir.
- Örnek olarak, gelir düzeyi yükseldikçe doğurganlığın da düştüğü bilinen bir olgudur.
- Düşük gelir düzeylerinde çocuk bir tür sosyal güvence olarak düşünülebildiği için doğurganlık hızı yüksektir.
- Gelir arttıkça ortalama çocuk sayısı da azalır ancak ilişki doğrusal değildir. Belli bir gelirden sonra çocuk sayısının sıfır ya da eksi değerlere ulaşacağını beklemeyiz.
- İki değişken arasındaki doğrusal olmayan bir ilişkiyi incelemenin bir yolu çokterimli SEK modelidir.
- Genel olarak, r'inci dereceden çokterimli bağlanım modeli şöyle gösterilir:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_r X^r$$

- Buradaki tek açıklayıcı değişken olan X, farklı kuvvetlerle gösterildiği için bu model bir çoklu bağlanım modelidir.
- Çokterimli modeller β katsayılarında doğrusal oldukları için SEK yöntemi ile tahmin edilebilirler.
- Bu modelde X ve X'in kuvvetleri arasındaki ilişki güçlü olmakla birlikte doğrusal olmadığı için, KDBM'nin "çoklueşdoğrusallık yoktur" varsayımı çiğnenmemiş olur.
- Doğrusal modellerde β terimlerinin Y'nin farklı X'lere göre sabit eğimini verdiğini anımsayalım.
- Değişkenlerde doğrusal-dışı olan çokterimli modellerde ise katsayıların yorumlanması biraz daha karmaşıktır.

- Bu modellerde ele alınan ilişki eğrisel olduğu için, eğim de X'in düzeyinine göre değişir.
- Bu nedenle, X'deki bir birimlik artışın Y üzerindeki etkisini bulmak için, önce bir başlangıç X düzeyi seçilir ve buna karşılık gelen $\hat{Y}$ değeri hesaplanır.
- Daha sonra X bir birim artırılır ve $\hat{Y}$ yeniden hesaplanır.
- Aradaki fark, seçili X düzeyindeki ortalama eğimi verir.

Çokterimli Bağlanım Açıklayıcı Örnek

- Çokterimli bağlanım modeline bir örnek olarak, Türkiye'de illerdeki gelir ve doğurganlık ilişkisini *"kareli"* (quadratic) bir işlev çerçevesinde ele alalım.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

- Burada Y ortalama çocuk sayısını, X ise kişi başına düşen gayri safi yurtiçi hasılayı göstermektedir.
- Görüldüğü gibi bu modelde Y ve X değişkenleri arasındaki ilişkiyi tanımlayan iki ayrı β_1 ve β_2 bulunmaktadır.
- Kabaca, β_1 ilişkinin yönünü gösterirken β_2'nin ise eğriselliği anlattığını söyleyebiliriz.
- Önsel beklentimiz, X artarken Y'nin de azalacağı ancak bu azalmanın giderek yavaşlayacağı yönündedir. Buna göre β_1 eksi, β_2 ise artı değer almalıdır.

Modeli 2000 yılı Türkiye verilerine yakıştırdığımızda aşağıdaki bulguları elde ediyoruz:

$$\begin{array}{lcccl} \hat{Y}_i = & 5{,}9486 & -\ 0{,}0030\ X_i & +\ 4{,}978\text{e-}07\ X_i^2 & \\ \text{öh} & (0{,}3835) & (0{,}0004) & (9{,}727\text{e-}08) & R^2 = 0{,}5196 \\ t & (15{,}5094) & (-7{,}2485) & (5{,}1179) & \bar{R}^2 = 0{,}5073 \end{array}$$

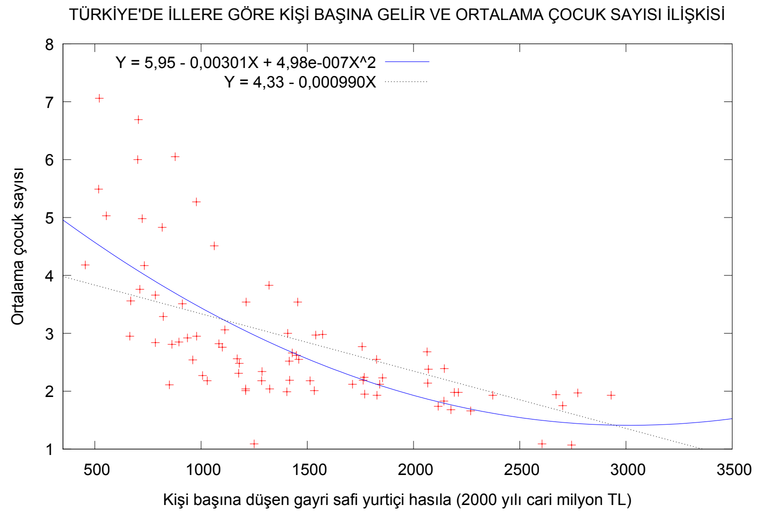
- Katsayıların işaretleri beklentilerimiz ile örtüşmektedir.
- İlişki doğrusal olsaydı, $\hat{\beta}_2$ anlamlı çıkmayacaktı. $\hat{\beta}_2$'nın anlamlı olması doğrusal-dışılığı onaylayıcı niteliktedir.
- Gelir 1000 TL olduğunda ortalama çocuk sayısı şudur:

$$\hat{Y} = 5{,}9486 - (0{,}0030 \times 1000) + (4{,}978\text{e-}07 \times 1000^2) = 3{,}44$$

- Gelir 1100 TL olduğunda çocuk sayısı ise şöyledir:

$$\hat{Y} = 5{,}9486 - (0{,}0030 \times 1100) + (4{,}978\text{e-}07 \times 1100^2) = 3{,}24$$

- Demek ki $X = 1000$ olduğunda, gelir düzeyindeki 100 TL kadar bir artış ortalama çocuk sayısını 0,2 düşürmektedir.



Uygulamaya İlişkin İki Nokta

Son olarak, çokterimli modeller kullanılırken özellikle iki noktaya dikkat etmek önemlidir:

1. Öncelikle doğrusal-dışı ilişki tanımlanmalıdır. Araştımacı, X ve Y arasındaki ilişkinin neden doğrusal olmayabileceğini sorgulamalıdır. Daha sonra, uygun bir işlev biçimi seçmek için iktisat kuramı temel alınmalıdır.

2. İkinci olarak, uygun bir çokterimli model belirtilip tahmin edildikten sonra bunun ilişkiyi iyi anlattığı ve doğrusal modelden üstün olduğu doğrulanmalıdır. Bunun için tahmin edilen bağlanım işlevinin çizdirmek ve bağlanımın verilere iyi yakışıp yakışmadığına bakılabilir. Ayrıca, anakütle bağlanım işlevinin doğrusal olduğu sıfır önsavı istatistiksel yöntemler kullanılarak sınanmalıdır.

Bu çıkarsama yöntemleri ise bir sonraki konuda ele alınacaktır.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 7* "Multiple Regression Analysis: The Problem of Estimation" okunacak.

Önümüzdeki Ders
Çoklu Bağlanım Çözümlemesi: Çıkarsama Sorunu

Bölüm 9

Çoklu Bağlanım Çözümlemesi - Çıkarsama Sorunu

9.1 T Sınamaları

9.1.1 Çoklu Bağlanımda Önsav Sınaması

- Bu bölümde daha önce iki değişkenli bağlanım modelleri için ele almış olduğumuz aralık tahmini ve önsav sınaması kavramlarını çok değişkenli modellere genişleteceğiz.
- Bilindiği gibi amacımız yalnızca bağlanım katsayılarını tahmin etmek değil, aynı zamanda bu katsayılara ilişkin çeşitli çıkarsamalar ve önsav sınamaları da yapmaktır.
- Bu doğrultuda u_i hatalarının sıfır ortalama ve σ^2 sabit varyanslı normal dağılıma uydukları varsayımını çoklu bağlanım modelleri için de sürdüreceğiz.

İkili bağlanım modelinin basit dünyasından dışarı çıkıldığında önsav sınaması aşağıdaki gibi farklı şekiller almaktadır:

1. Tek bir kısmi bağlanım katsayısına ilişkin önsav sınaması,
2. Tahmin edilen bağlanım modelinin bütününün sınanması,
3. İki ya da daha çok katsayının eşitliğinin sınanması,
4. Katsayıların belli sınırlamalara uygunluğunun sınanması,
5. Modelin farklı veri setlerindeki kararlılığının sınanması,
6. Bağlanım modellerinin işlev biçimlerinin sınanması.

İzleyen bölümde bu sınama çeşitleri ayrı ayrı ele alınacaktır.

- Farklı önsav sınama biçimlerini göstermek için, Türkiye'de 81 ile ait ve 2000 yılı verileri kullanılarak tahmin edilmiş şu modeli ele alalım:

$$\begin{array}{llllll} \hat{Y}_i = & 7{,}3778 & +\,1{,}4718\,X_{2i} & -\,0{,}2014\,X_{3i} & \\ \text{öh} & (1{,}0689) & (0{,}3850) & (0{,}0717) & R^2 = 0{,}3139 \\ t & (6{,}9021) & (3{,}8223) & (-2{,}8078) & \bar{R}^2 = 0{,}2963 \end{array}$$

- Burada
  - Y ilin aldığı göçün toplam il nüfusuna oranını (%),
  - X_2 cari fiyatlarla kişi başına düşen GSYH'yi (1000 TL),
  - X_3 erkek nüfustaki işsizlik oranını (%)

  göstermektedir.

- Sonuçlara göre, milli gelirdeki 1000 liralık artış ilin göç alma yüzdesini yaklaşık 1,5 puan yükseltirken işsizlikteki benzer bir artış ise % 0,2'lik eksi yönlü bir etkiye neden olmaktadır.

- Katsayılar anlamlıdır ve önsel beklentilerle de uyumludur.

9.1.2 Tek Bir Katsayının Sınanması

- $u_i \sim N(0, \sigma^2)$ varsayımı altında, herhangi bir tekil bağlanım katsayısına ilişkin önsavlar için t sınamasını kullanabiliriz.

- Örnek olarak, erkek işsizlik oranının göç alma üzerinde bir doğrusal etkisi olmadığı varsayımını şöyle sınarız:

$$H_0 : \beta_3 = 0, \qquad H_1 : \beta_3 \neq 0$$

$$t = \frac{\hat{\beta}_3 - \beta_3^*}{\text{öh}(\hat{\beta}_3)} = \frac{-0{,}2014 - 0}{0{,}0717} = -2{,}8089$$

- $\alpha = 0{,}05$ seçilirse, 78 (81-3) sd ile kritik $t_{\alpha/2} = 1{,}9908$ olur.

- Hesaplanan t değeri kritik t değerini aştığı için, istatistiksel olarak β_3'ün anlamlı olduğunu ya da diğer bir deyişle sıfırdan anlamlı ölçüde uzak olduğunu söyleyebiliriz.

- Bilindiği gibi önsav sınamasına diğer bir yaklaşım da güven aralığı yöntemidir.

- Örnek olarak β_2'nin yüzde 95 güven aralığı şöyledir:

$$\begin{aligned}
\hat{\beta}_2 - t_{\alpha/2}\text{öh}(\hat{\beta}_2) &\leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\text{öh}(\hat{\beta}_2) \\
1{,}4718 - 1{,}9908(0{,}3850) &\leq \beta_2 \leq 1{,}4718 + 1{,}9908(0{,}3850) \\
0{,}7053 &\leq \beta_2 \leq 2{,}2383
\end{aligned}$$

- 81 gözlemli 100 farklı örneklem seçilir ve $\hat{\beta}_2 \pm t_{\alpha/2}\text{öh}(\hat{\beta}_2)$ gibi böyle 100 güven aralığı bulunursa, bunlardan 95'inin anakütledeki gerçek β_2'yi içermesi beklenir.

9.1.3 İki Katsayının Eşitliğinin Sınanması

- Şimdi de β_2 ve β_3 eğim katsayılarının birbirine eşit olup olmadığını sınamak istediğimizi varsayalım.
- Bunun için sıfır ve almaşık önsavları iki şekilde yazabiliriz:

$$H_0 : \beta_2 = \beta_3$$
$$H_1 : \beta_2 \neq \beta_3$$

$$H_0 : (\beta_2 - \beta_3) = 0$$
$$H_1 : (\beta_2 - \beta_3) \neq 0$$

- Milli gelir ve işsizlik oranı katsayılarının eşit olmasını doğal olarak beklemiyoruz. Dolayısıyla örneğimizde bu sınama iktisat bağlamında gereksizdir.
- Diğer yandan, uygulamada bu tür önsav sınamasına sıkça başvurulur.
- Örnek olarak Y bir mala olan talebi, X_2 ve X_3 de sırasıyla tüketicinin gelir ve servetini gösteriyor olsun.
- Log-doğrusal model için yukarıdaki sıfır önsavları talebin gelir ve servet esnekliklerinin aynı olduğu anlamına gelir.
- İki katsayı tahmininin eşitliğini sınamak için *t* sınaması yöntemi kullanılabilir:

$$t = \frac{(\hat{\beta}_2 - \hat{\beta}_3) - (\beta_2 - \beta_3)^*}{\text{öh}(\hat{\beta}_2 - \hat{\beta}_3)}$$

- Klasik varsayımlar altında $n-k$ sd (örneğimizde $k=3$) ile *t* dağılımına uyan yukarıdaki istatistik şöyle de yazılabilir:

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) - 2\text{cov}(\hat{\beta}_2, \hat{\beta}_3)}}$$

- Yukarıda, $\text{öh}(\hat{\beta}_2 - \hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) - 2\text{cov}(\hat{\beta}_2, \hat{\beta}_3)}$ ve H_0'a göre $\beta_2 - \beta_3 = 0$ özdeşliklerinden yararlanılmıştır.
- Üçlü bağlanım örneğimiz için $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = 0{,}0104$'tür.
- $\text{var}(\hat{\beta}_2)$ ve $\text{var}(\hat{\beta}_3)$ değerleri ise ölçünlü hataların karesi alınarak 0,1482 ve 0,0051 olarak bulunur.
- Buna göre $\beta_2 - \beta_3 = 0$ sınamasını şöyle yaparız:

$$t = \frac{1{,}4718 + 0{,}2014}{\sqrt{0{,}1482 + 0{,}0051 - 0{,}0208}} = 4{,}5966$$

- Eldeki değer 78 sd ve çift kuyruklu sınama için hesaplanan $t = 1{,}9908$ kritik t değerini aştığı için, X^2 ve X^3'e ait katsayı değerlerinin aynı olduğu sıfır önsavı reddedilir.

9.2 F Sınamaları

9.2.1 Bağlanımın Bütününün Anlamlılık Sınaması

- Türkiye örneğimize dönelim ve β_2 ve β_3'ün aynı anda sıfır olduğunu öneren $H_0 : \beta_2 = \beta_3 = 0$ önsavını ele alalım.
- Bu sıfır önsavının sınanmasına, bağlanıma ilişkin *"bütünün anlamlılığı"* (overall significance) sınaması adı verilir.
- Bu sınama tekil anlamlılık sınamalarından farklıdır.
- Bunun nedeni şudur: $\hat{\beta}_2$ ve $\hat{\beta}_3$ gibi farklı katsayılar için tekil anlamlılık sınaması yaparken, her bir sınamanın farklı ve bağımsız bir örnekleme dayandığı varsayılır.
- Diğer yandan, verili bir örneklemde cov$(\hat{\beta}_2, \hat{\beta}_3) = 0$ geçerli olmayabilir. Diğer bir deyişle, $\hat{\beta}_2$ ile $\hat{\beta}_3$ ilişkili olabilirler.
- Bu durumda, $\hat{\beta}_2$ ile $\hat{\beta}_3$'nın aynı anda $[\hat{\beta}_2 \pm\ t_{\alpha/2}\text{öh}(\hat{\beta}_2)]$ ve $[\hat{\beta}_3 \pm\ t_{\alpha/2}\text{öh}(\hat{\beta}_3)]$ aralıklarında bulunma olasılığı $(1-\alpha)^2$ değildir.
- Anakütle kısmi bağlanım katsayılarının aynı anda sıfır olduğu yönündeki ortak önsavı sınamak için varyans çözümlemesi yöntemi kullanılabilir:

$$\begin{array}{ccc} \sum y_i^2 & = & \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} \quad + \sum \hat{u}_i^{\,2} \\ \text{TKT} & = & \text{BKT} \quad + \text{KKT} \end{array}$$

- Buna göre aşağıdaki VARÇÖZ çizelgesini düzenleyebiliriz:

Değişimin Kaynağı	KT	sd	OKT (KT/sd)
Bağlanımdan (BKT)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	$k-1$	$\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{k-1}$
Kalıntılardan (KKT)	$\sum \hat{u}_i^{\,2}$	$n-k$	$\frac{\sum \hat{u}_i^{\,2}}{n-k} = \hat{\sigma}^2$
Toplamlarından (TKT)	$\sum y_i^2$	$n-1$	

- Burada k, sabit terim ile birlikte tahmin edilen toplam anakütle katsayılarının sayısıdır.
- Üçlü model için, hata teriminin normal dağıldığı varsayımı ve $\beta_2 = \beta_3 = 0$ sıfır önsavı altında şu istatistik hesaplanır:

$$F = \frac{(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})/(k-1)}{\sum \hat{u}_i^{\,2}/(n-k)} = \frac{\text{BKT}/sd}{\text{KKT}/sd}$$

- Yukarıda verilen değişkenin $(k - 1)$ ve $(n - k)$ sd ile F dağılımına uyduğu gösterilebilir.
- Buna göre, hesaplanan F istatistiğinin p değeri yeterince küçükse H_0 reddedilir.

Bütünün Anlamlılık Sınaması Açıklayıcı Örnek

Bağlanımın bütününün anlamlılığının sınanmasına örnek olarak Türkiye için gelir, işsizlik, ve göç alma modelimize dönelim ve aşağıdaki VARÇÖZ çizelgesini oluşturalım:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	205,022	2	102,511
Kalıntılar	448,080	78	5,74461
Toplam	653,102	80	

- F değeri çizelgeden aşağıdaki gibi hesaplanır:

$$F = \frac{102{,}511}{5{,}74461} = 17{,}8448$$

- Yüzde 5 anlamlılık düzeyinde ve 2 ile 78 sd için kritik değer $F_{0{,}05}(2, 78) = 3{,}1138$'dir.
- Hesaplanan F değeri anlamlı olduğu için H_0 reddedilir.

Çoklueşdoğrusallığın Etkisi

- Göstermiş olduğumuz yöntemle hesaplanan F istatistiği çoğu zaman yüksek çıkar.
- Tüm bağlanım katsayıları tek tek istatistiksel olarak anlamlı değilken, F değerinin anlamlı çıkması olasıdır.
- Bu durum açıklayıcı değişkenler kendi aralarında yüksek derecede ilinti gösteriyorsa karşımıza çıkabilir.
- Bu sorunu ileride çoklueşdoğrusallık başlığı altında ayrıntılı biçimde ele alacağız.
- Şimdilik F ve t sınama sonuçlarını yorumlarken dikkatli olmak gerektiğini vurgulamakla yetiniyoruz.

R^2 ve F Arasındaki İlişki

- Belirleme katsayısı R^2 ile varyans çözümlemesindeki F değeri arasında yakın bir ilişki vardır.
- k değişkenli durumda ve $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ sıfır önsavı altında şu gösterilebilir:

$$
\begin{aligned}
F &= \frac{n-k}{k-1}\frac{\text{BKT}}{\text{KKT}} \\
&= \frac{n-k}{k-1}\frac{\text{BKT}}{\text{TKT}-\text{BKT}} \\
&= \frac{n-k}{k-1}\frac{\text{BKT}/\text{TKT}}{1-(\text{BKT}/\text{TKT})} \\
&= \frac{n-k}{k-1}\frac{R^2}{1-R^2}
\end{aligned}
$$

- Burada $R^2 = \text{BKT}/\text{TKT}$ tanımı kullanılmıştır.
- Eşitliğe göre R^2 ile F aynı yönde değişirler.
- Tahmin edilen bağlanımın bütün olarak anlamlılığının ölçüsü olan F demek ki aynı zamanda $H_0 : R^2 = 0$ sınamasına eşdeğerdir.
- F sınamasının R^2 cinsinden gösterilmesinin üstün yanı hesaplama kolaylığıdır. Tek gereken R^2 değeridir.
  - VARÇÖZ çizelgesini R^2 ile aşağıdaki gibi düzenleyebiliriz:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	$R^2(\sum y_i^2)$	$k-1$	$R^2(\sum y_i^2)/(k-1)$
Kalıntılar	$(1-R^2)(\sum y_i^2)$	$n-k$	$(1-R^2)(\sum y_i^2)/(n-k)$
Toplam	$\sum y_i^2$	$n-1$	

9.2.2 Bir Açıklayıcı Değişkenin Marjinal Katkısı

- Bir açıklayıcı değişkenin marjinal katkısına bakmak için, X_2 ve X_3 gibi iki değişkeni modele sırayla ekleyelim.
- Burada görmek istediğimiz, eklenen değişkenin KKT'yi eskiye oranla ne ölçüde azalttığıdır.
- Çoğu görgül çalışmada, çeşitli olası X değişkenleri içinden KKT'yi çok azaltmayanları modele eklememek yeğlenebilir.

- Aynı şekilde KKT'yi önemli ölçüde azaltan, diğer bir deyişle R^2'yi "anlamlı" biçimde yükselten değişkenler de modelden çıkartılmak istenmez.
- Bu yüzden bir X değişkeninin marjinal katkısı uygulamada önemli bir konudur.
- Ek bir açıklayıcı değişkenin KKT'yi anlamlı biçimde azaltıp azaltmadığını bulmak için yine varyans çözümlemesinden yararlanılabilir.
- Türkiye'de illerin aldığı göç örneğimize dönelim ve şu ikili bağlanımı tahmin edelim:

$$\begin{array}{llll} \hat{Y}_i = & 4{,}8832 & +\,1{,}8800\, X_{2i} & \\ \text{öh} & (0{,}6197) & (0{,}3717) & \\ t & (7{,}8797) & (5{,}0574) & r^2 = 0{,}2446 \end{array}$$

- Bulgular, kişi başına düşen GSYH'yi gösteren X_2'nin Y'yi anlamlı biçimde etkilediğini göstermektedir.
- İkili bağlanıma ait VARÇÖZ çizelgesi aşağıdaki gibidir:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	159,732	1	159,732
Kalıntılar	493,370	79	6,24519
Toplam	653,102	80	

- Şimdi, ildeki erkek işsizlik oranını gösteren X_3 değişkenini modele eklemek istediğimizi varsayalım.
- Üçlü bağlanıma ait VARÇÖZ çizelgesi ise şöyle idi:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	205,022	2	102,511
Kalıntılar	448,080	78	5,74461
Toplam	653,102	80	

- KKT'deki $(493{,}370 - 448{,}080 = 45{,}290)$ birimlik azalmanın istatistiksel olarak anlamlı olup olmadığını bulmak istiyoruz.
- X_2'nin katkısı biliniyorken, X_3'ün marjinal katkısı şu sınama istatistiği ile ölçülebilir:

$$F = \frac{Q_3/sd}{Q_4/sd} = \frac{(\text{KKT}_{eski} - \text{KKT}_{yeni})/m}{\text{KKT}_{yeni}/(n-k)}$$

- Burada m, yeni modele eklenen değişken sayısını gösterir.
- Elimizdeki örnek için F istatistiği şu şekilde hesaplanır:

$$F = \frac{(493{,}370 - 448{,}080)/1}{448{,}080/78} = 7{,}884$$

- Bulunan istatistik anlamlıdır. İldeki erkek işsizlik oranını eklemek KKT'yi anlamlı biçimde azaltmaktadır.
- Eldeki F oranı R^2 değerlerini kullanarak da bulunabilir:

$$F = \frac{(R^2_{yeni} - R^2_{eski})/m}{(1 - R^2_{yeni})/(n-k)}$$

- Örneğimiz için:

$$F = \frac{(0{,}3139 - 0{,}2446)/1}{(1 - 0{,}3139)/78} = 7{,}878$$

- Bu da yuvarlama hataları dışında önceki değer ile aynıdır.

Yeni Bir Değişken Ne Zaman Eklenmeli?

- Araştırmacılar çoğu zaman aynı bağımlı değişkeni içeren ama açıklayıcı değişkenleri farklı olan modeller arasında seçim yapmak durumunda kalırlar.
- Böyle durumlardaki genel eğilim en yüksek $\bar{R}^2$'yi seçmek yönündedir.
- Diğer yandan, yeni eklenen bir değişkenin katsayısının t değeri mutlak olarak 1'den büyük olduğu sürece $\bar{R}^2$ artar.
- Diğer bir deyişle yeni eklenen bir değişkene ilişkin $F(= t^2)$ değeri 1'den büyükse, bağlanım $\bar{R}^2$ değeri de yükselir.
- Demek ki $\bar{R}^2$ değerini yükselttiği halde KKT'yi istatistiksel olarak anlamlı ölçüde azaltmayan bir ek değişkenin modele eklenmesi konusunda dikkatli olunmalıdır.

9.2.3 Sınırlamalı Enküçük Kareler Yöntemi

- İktisat kuramı zaman zaman belli bir bağlanım modelindeki katsayılar için bir takım doğrusal sınırlamalar öngörebilir.
- Örnek olarak Cobb-Douglas üretim işlevini ele alalım:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i}$$

- Burada Y üretim, X_2 emek girdisi, X_3 de sermaye girdisidir.
- Modelin log-doğrusal biçimdeki gösterimi şöyledir:

$$\ln Y_i = \beta_1' + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

- Burada β_1', $\ln \beta_1$'dir.
- Eğer ölçeğe göre sabit getiri söz konusu ise, iktisat kuramı aşağıdaki doğrusal sınırlamayı öngörür:

$$\beta_2 + \beta_3 = 1$$

- $\beta_2 + \beta_3 = 1$ gibi bir doğrusal sınırlamanın geçerli olup olmadığı, t sınaması yöntemi kullanılarak görülebilir.
- Bunun için, önce model tahmin edilir ve $H_0 : \beta_2 + \beta_3 = 1$ önsavı bildik yolla sınanır:

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{\sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2\text{cov}(\hat{\beta}_2 \hat{\beta}_3)}}$$

- Bulunan t değeri eğer seçili anlamlılık düzeyindeki kritik t değerinden büyükse, H_0 reddedilir.
- t sınaması yaklaşımı, *"sınırlamasız"* (unrestricted) bağlanım bulunduktan sonra sınama yapmaya dayandığı için yeğlenmeyen bir yöntemdir.

- Daha doğru bir yaklaşım *"sınırlamalı enküçük kareler"* (restricted least squares) yöntemidir.
- Bu yönteme göre, $(\beta_2 = 1 - \beta_3)$ denkleme en başta koyulur ve *"sınırlamalı"* (restricted) model aşağıdaki gibi türetilir:

$$\begin{aligned}
\ln Y_i &= \beta_1' + \ (1-\beta_3)\ln X_{2i} + \beta_3 \ln X_{3i} \ + u_i \\
&= \beta_1' + \ln X_{2i} + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i \\
\ln Y_i - \ln X_{2i} &= \beta_1' + \qquad \beta_3(\ln X_{3i} - \ln X_{2i}) \qquad + u_i \\
\ln(Y_i/X_{2i}) &= \beta_1' + \qquad \beta_3 \ln(X_{3i}/X_{2i}) \qquad + u_i
\end{aligned}$$

- Burada (X_{3i}/X_{2i}) sermaye/emek oranını, (Y_i/X_{2i}) ise çıktı/emek oranını gösteren önemli iktisadi büyüklüklerdir.
- Tanımlanan sınırlamanın geçerli olup olmadığı iki bağlanımın karşılaştırılması ile bulunur:

$$F = \frac{(\text{KKT}_s - \text{KKT}_{sz})/m}{\text{KKT}_{sz}/(n-k)} = \frac{(R^2_{sz} - R^2_s)/m}{(1-R^2_{sz})/(n-k)}$$

- m burada doğrusal sınırlama sayısını, sz ve s ise sınırlamasız ve sınırlamalı bağlanımları göstermektedir.
- *Dikkat:* Sınırlamasız ve sınırlamalı modellerde bağımlı değişken farklı ise, R^2_{sz} ve R^2_s'nin birlikte kullanılabilmesi için gerekli dönüşümün yapılmış olması önemlidir.

Sınırlamalı Enküçük Kareler Açıklayıcı Örnek

- Örnek olarak, Tayvan tarım kesimi için Cobb-Douglas üretim modelini ölçeğe göre sabit getiri sınırlaması ile tahmin edelim:

$$\begin{array}{llll}
\ln\widehat{(Y_i/X_{2i})} = & 1{,}7086 & +\,0{,}61298\ln(X_{3i}/X_{2i}) & \\
\text{öh} & (0{,}4159) & (0{,}0933) & r^2 = 0{,}7685
\end{array}$$

- Sınırlamasız model için R^2 değeri, gerekli dönüştürmeden sonra $0{,}8489$ olarak bulunur ve şu F istatistiği hesaplanır:

$$F = \frac{(R^2_{sz} - R^2_s)/m}{(1-R^2_{sz})/(n-k)} = \frac{(0{,}8489 - 0{,}7685)/1}{(1-0{,}8489)/12} = 6{,}385$$

- F çizelgesinden, gözlenen değerin %5 düzeyinde anlamlı olduğu görülür ve $H_0 : \beta_2 + \beta_3 = 1$ sıfır önsavı reddedilir.
- Eğer sınırlamanın geçerli olduğuna karar verilmiş olsaydı, sınırlamalı model için tahmin edilen $0{,}61298$ değeri β_3'ü gösterdiği için β_2 de $0{,}38702$ olarak kolayca bulunurdu.
- Şimdi de sınırlamasız bağlanım bulgularına bir göz atalım:

$$\begin{array}{lllll} \widehat{\ln Y}_i = -3{,}3384 & + 1{,}4988 \ln X_{2i} & + 0{,}4899 \ln X_{3i} & R^2 = 0{,}8890 \\ \text{öh} \quad (2{,}4495) & (0{,}5398) & (0{,}1020) & \bar{R}^2 = 0{,}8705 \end{array}$$

- Yukarıda emek girdisi esnekliğinin istatistiksel olarak anlamlı olmadığı görülüyor.
- Bu örnek, yalnızca tahmin edilen katsayılar ile yetinmeyip biçimsel sınama da yapmanın daha iyi bir çözümleme için gerekliliğini göstermesi bakımından önemlidir.

Genel F Sınaması

- Bir açıklayıcı değişkenin marjinal katkısı bölümünde söz edilen "yeni" model aslında sınırlamasız modeldir. Buna göre "eski" model de $\beta_3 = 0$ varsayımı ile sınırlamalı olur.
- Aslında, bağlanım bütününün anlamlılığını sınamaya ilişkin formüldeki payın BKT olmasının nedeni de buradaki "süper sınırlamalı" modelin KKT'sinin $\text{TKT}_{sz} = \text{TKT}$ olmasıdır.
- Ele almış olduğumuz örneklerden de anlaşılacağı gibi, F sınaması yöntemi k değişkenli bağlanım modelindeki m anakütle katsayısının sınanması için genel bir yöntemdir.
- Örnek olarak

$$\begin{array}{ll} H_0: \ \beta_2 = \beta_3 & (m = 1), \\ H_0: \ \beta_3 + \beta_4 + \beta_5 = 3 & (m = 1), \\ H_0: \ \beta_3 = \beta_4 = 0 & (m = 2), \\ H_0: \ \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 & (m = 4) \end{array}$$

gibi pek çok farklı önsav F sınaması ile sınanabilir.

Genel F sınamasının adımları aşağıdaki gibi özetlenebilir:

1. Birincisi geniş sınırlamasız model ve diğeri de daha dar sınırlamalı model olmak üzere iki model vardır.

2. Bunlardan ikincisi, birinciden bazı değişkenler çıkarılarak ya da çeşitli doğrusal sınırlamalar getirilerek elde edilir.

3. Daha sonra; sınırlamasız ve sınırlamalı modeller verilere yakıştırılır ve KKT$_{sz}$ ve KKT$_s$ toplamları ya da R^2_{sz} ve R^2_s belirleme katsayıları bulunur.

4. Eğer bağımlı değişkenler farklıysa, R^2_{sz} ve R^2_s kullanmak için bunları önce birbirleriyle uyumlandırmak gereklidir.

5. KKT$_{sz}$ için serbestlik derecesi $(n - k)$'dir. KKT$_s$ için ise serbestlik derecesi toplam kısıtlama sayısı m'dir.

6. Son olarak, formülü verilen F istatistiği hesaplanır ve bu değer $F_\alpha(m, n - k)$'den büyükse sıfır önsavı reddedilir.

9.3 Diğer Sınama ve Konular

9.3.1 Chow Sınaması

- Eğer model katsayıları zaman içerisinde sabit kalmayıp değişime uğruyorlar ise bu duruma *"yapısal değişim"* (structural change) denir.

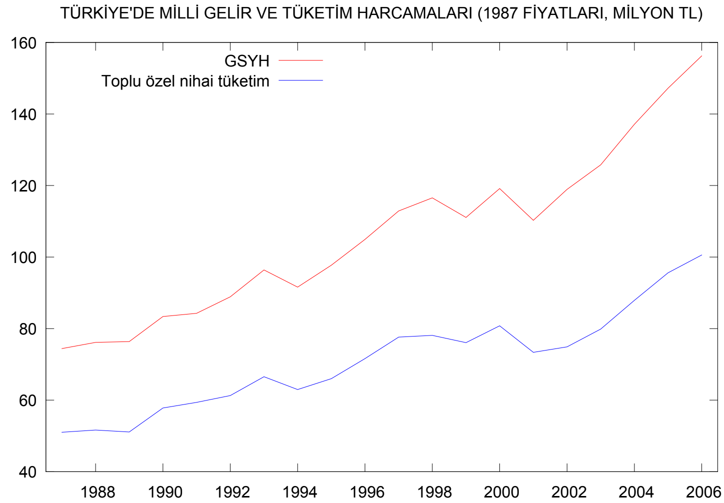
- Yapısal değişime örnek neden olarak
  2001 yılında dalgalı kur rejimine geçiş,
  1999 yılı vergi yasası reformu,
  1990-1991 Körfez Savaşı,
  1973-1977 OPEC petrol ambargosu
  gibi ulusal ya da küresel etmenler gösterilebilir.

- Yapısal değişim konusu özellikle zaman serileri içeren bağlanım modellerinde önemlidir.

- Bir yapısal değişimin varlığını görebilmeye örnek olarak, 1987 ve 2006 yılları arasında Türkiye'de toplam tüketim harcamaları ve GSYH (1987 fiyatları, milyon TL) örneğimizi anımsayalım:

Çizelge: Türkiye'de Tüketim ve GSYH (1987–2006)

Yıl	C	Y	Yıl	C	Y
1987	51.019	74.416	1997	77.620	112.892
1988	51.638	76.143	1998	78.113	116.541
1989	51.105	76.364	1999	76.077	111.083
1990	57.803	83.371	2000	80.774	119.147
1991	59.366	84.271	2001	73.356	110.267
1992	61.282	88.893	2002	74.894	118.923
1993	66.545	96.391	2003	79.862	125.778
1994	62.962	91.600	2004	87.897	137.110
1995	66.011	97.729	2005	95.594	147.200
1996	71.614	104.940	2006	100.584	156.249



- Türkiye'de tüketim harcamaları ve milli gelir arasındaki ilişkiyi incelemek istiyoruz.
- Elimizde 1987 ve 2006 yılları arasını kapsayan bir SEK bağlanımını tahmin etmek için yeterli veriler bulunmaktadır.
- Diğer yandan, tasarruf ve gelir arasındaki ilişkinin 20 yıl boyunca aynı kaldığını varsaymak fazla inandırıcı olmaz.
- Örnek olarak, Şubat 2001 ve öncesinde yaşanan olaylar sonrasında Cumhuriyet tarihindeki en büyük ekonomik krizlerden birinin ortaya çıkmış olduğunu biliyoruz.
- Buna dayanarak, 2001 ve sonrası dönemin yapısal olarak farklı olup olmadığını görmek istediğimizi varsayalım.
- Yapısal kararlılığı sınamak için örneklemi 2001 öncesi ve 2001 ve sonrası olarak iki döneme ayırabiliriz.
- Böylece elimizde tahmin edilebilecek üç ayrı bağlanım olur:

$$\begin{array}{ll} \text{1987-2000 dönemi:} & Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad (n_1 = 14) \\ \text{2001-2006 dönemi:} & Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad (n_2 = 6) \\ \text{1987-2006 dönemi:} & Y_t = \alpha_1 + \alpha_2 X_t + u_{3t} \quad (n_3 = 14 + 6 = 20) \end{array}$$

- Yukarıdaki üçüncü bağlanım, tüm gözlemleri kapsamakta ve 1987-2006 aralığı içinde yapısal bir değişim olmadığını varsaymaktadır.

- Öyleyse üçüncü model, $\lambda_1 = \gamma_1$ ve $\lambda_2 = \gamma_2$ koşullarından dolayı bir sınırlamalı model olarak düşünülebilir.
- Üç bağlanıma ait bulgular aşağıdaki gibidir:

$$\begin{array}{llll}
\hat{Y}_t = & 1{,}5027 & +\,0{,}6679X_t & R^2 = 0{,}9937 \\
t & (1{,}0156) & (43{,}5400) & \text{KKT}_1 = 8{,}9210 \\
\\
\hat{Y}_t = & 1{,}0706 & +\,0{,}6358X_t & R^2 = 0{,}9835 \\
t & (0{,}1947) & (15{,}4361) & \text{KKT}_2 = 10{,}3487 \\
\\
\hat{Y}_t = & 8{,}0344 & +\,0{,}5934X_t & R^2 = 0{,}9854 \\
t & (4{,}3310) & (34{,}8352) & \text{KKT}_3 = 55{,}0062
\end{array}$$

- Sonuçlar, tasarruf ve gelir arasındaki ilişkinin iki alt döneme ait tahminlerinde farklılıklar olduğunu göstermektedir.
- Buna göre üçüncü bağlanımın uygun ve güvenilir olmadığı düşünülebilir.
- Yapısal değişimin varlığını sınamak için kullanılabilecek yöntemlerden biri Chow sınamasıdır.
- Bu sınama, bildiğimiz F sınamasından farklı olmamakla birlikte geliştiricisi Gregory Chow'un adıyla anılır.
- Chow sınamasının gerisinde iki önemli varsayım vardır:
- *Varsayım 1:* Birinci ve ikinci modellere ait hata terimleri aynı sabit varyans ile normal dağılmaktadırlar:

$$u_{1t} \sim N(0, \sigma^2) \text{ ve } u_{2t} \sim N(0, \sigma^2)$$

- *Varsayım 2:* u_{1t} ve u_{2t} aynı zamanda bağımsız dağılırlar.

Verilen varsayımlar altında Chow sınaması şöyle yapılır:

1. Birinci modelden sd'si $(n_1 - k)$ olan KKT_1 bulunur.
2. İkinci modelden sd'si $(n_2 - k)$ olan KKT_2 bulunur.
3. İki bağlanıma ait hata terimleri bağımsız kabul edildiği için, $\text{KKT}_{sz} = \text{KKT}_1 + \text{KKT}_2$ olarak hesaplanır.
4. Tüm gözlemlerin kullanıldığı 3. model tahmin edilir ve KKT_3 ya da KKT_s bulunur.

5. Yapısal değişim yoksa KKT_s ve KKT_{sz} istatistiksel olarak farklı olmamalıdır. Sınamak için şu istatistik hesaplanır:

$$F = \frac{(\text{KKT}_s - \text{KKT}_{sz})/k}{(\text{KKT}_{sz})/(n_1 + n_2 - 2k)} \sim F_{[k,(n_1+n_2-2k)]}$$

- Örneğimize dönecek olursak F istatistiğini şöyle buluruz:

$$F = \frac{(55{,}0062 - 19{,}2697)/2}{(19{,}2697)/(16)} = 14{,}8363$$

- Gözlenen değer 2 ve 22 sd için yüzde 1 kritik F değeri olan 6,23'ten büyük olduğu için; $H_0 : \lambda_1 = \gamma_1,\ \lambda_2 = \gamma_2$ reddedilir.
- Demek ki Chow sınaması 2001 yılında Türkiye'nin bir yapısal değişim geçirdiği savını desteklemektedir.

Chow sınaması ile ilgili şu noktalara dikkat edilmelidir:

- Chow sınaması, birden fazla yapısal değişimin varlığını sınamak için genellenebilir.
- Örnek olarak, örneklemi üç ayrı döneme bölüp dört farklı bağlanım tahmini yapmak ve daha sonra KKT_s'yi de $\text{KKT}_1 + \text{KKT}_2 + \text{KKT}_3$ olarak hesaplamak olanaklıdır.
- Chow sınamasında *"yapısal kırılma"* (structural break) noktasının hangi dönemde yer aldığının bilindiği varsayılır.
- Chow sınaması iki bağlanımın farklı olup olmadığını söyler ancak farkın sabit terimden mi, X_t'nin katsayısından mı, ya da aynı anda her ikisinden mi kaynaklandığını bildirmez.
- Yapısal değişimin kaynağının ne olduğunu anlamak için kukla değişkenlere dayanan farklı bir yaklaşım gereklidir.
- Ayrı dönemlere ait hata varyanslarının sabit olduğu varsayımının ayrıca sınanması gerekli olabilir.

9.3.2 MWD Sınaması

- Doğrusal ve log-doğrusal model arasında bir seçim yapma zorunluluğu, görgül çalışmalarda sık sık ortaya çıkar.
- Böyle bir model seçimi için MacKinnon, White ve Davidson (1983) tarafından önerilen MWD sınaması kullanılabilir.
- MWD sınaması şu sıfır ve almaşık önsavları içerir:

$$\begin{aligned} H_0 &: \text{ Doğrusal model} \\ H_1 &: \text{ Log-doğrusal model} \end{aligned}$$

MWD sınamasının adımları aşağıdaki gibidir:

1. Doğrusal model tahmin edilir ve $\hat{Y}$ bulunur.
2. Log-doğrusal model tahmin edilir ve $\widehat{\ln Y}$ bulunur.
3. $Z_1 = \ln \hat{Y} - \widehat{\ln Y}$ değişkeni türetilir.
4. Y'nin X'lere ve Z_1'e göre bağlanımı hesaplanır. Eğer Z_1'in katsayısı bilindik t sınaması ile istatistiksel olarak anlamlı çıkarsa, H_0 reddedilir.
5. $Z_2 = \exp(\widehat{\ln Y}) - \hat{Y}$ değişkeni türetilir.
6. $\ln Y$'nin $\ln X$'lere ve Z_2'ye göre bağlanımı hesaplanır. Eğer Z_2'nin katsayısı t sınaması ile anlamlı bulunursa, H_1 savı reddedilir.

Karmaşık gibi görünse de MWD sınamasının mantığı basittir:

- Eğer doğrusal model gerçekten doğru modelse, dördüncü adımda hesaplanan Z_1 değeri anlamlı olmamalıdır.
- Çünkü böyle bir durumda doğrusal modelin $\hat{Y}$ kestirimleri (karşılaştırma yapabilmek için logları alındıktan sonra) ile log-doğrusal modelin kestirimleri farklı çıkmamalıdır.
- Aynı yorum H_1 almaşık önsavı için de geçerlidir.

MWD Sınaması Açıklayıcı Örnek

- Örnek olarak 1971-1975 arası dönem için ABD'nin Detroit şehrindeki gül talebini ele alalım:

$$\text{Doğrusal model: } Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_t$$

$$\text{Log-log model: } \ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + v_t$$

- Burada
  Y satılan gül miktarını (düzine),
  X_2 ortalama toptan gül fiyatını (dolar),
  X_3 ise ortalama toptan karanfil fiyatını (dolar)
  göstermektedir.

- Beklentiler α_2 ile β_2'nin eksi, α_3 ve β_3'ün ise artı değerli olması yönündedir.

- Bağlanım bulguları aşağıdaki gibidir:

$$\begin{array}{lllll} \widehat{Y}_t & = 9734{,}2176 & - \, 3782{,}1956 X_{2t} & + \, 2815{,}2515 X_{3t} & R^2 = 0{,}7710 \\ t & (3{,}3705) & (-6{,}6069) & (2{,}9712) & F = 21{,}84 \end{array}$$

$$\begin{array}{lllll} \widehat{\ln Y}_t & = 9{,}2278 & - \, 1{,}7607 \ln X_{2t} & + \, 1{,}3398 \ln X_{3t} & R^2 = 0{,}7292 \\ t & (16{,}2349) & (-5{,}9044) & (2{,}5407) & F = 17{,}50 \end{array}$$

- Görüldüğü gibi hem doğrusal hem de log-doğrusal model verilere iyi yakışmıştır.

- Katsayılar beklenen işaretleri taşımaktadır ve t değerleri de istatistiksel olarak anlamlıdır.

- Önce modelin doğrusal olup olmadığını sınayalım:

$$\begin{array}{llllll} \widehat{Y}_t & = 9727{,}5685 & - \, 3783{,}0623 X_{2t} & + \, 2817{,}7157 X_{3t} & + \, 85{,}2319 Z_{1t} \\ t & (3{,}2178) & (-6{,}337) & (2{,}8366) & (0{,}0207) \\ R^2 = & 0{,}7707 & & & F = 13{,}44 \end{array}$$

- Z_1'in katsayısı anlamlı olmadığına göre, modelin gerçekte doğrusal olduğunu öne süren önsavı reddetmiyoruz.

- Şimdi de gerçek modelin log-doğrusallığını sınayalım:

$$\begin{array}{lcllll}
\widehat{\ln Y}_t = & 9{,}1486 & -\ 1{,}9699 \ln X_{2t} & +\ 1{,}5891 \ln X_{3t} & -\ 0{,}0013 Z_{2t} \\
t & (17{,}0825) & (-6{,}4189) & (3{,}0728) & (-1{,}6612) \\
R^2 \ = & 0{,}7798 & & & F = 14{,}17
\end{array}$$

- Z_2'ye ait t değeri $-1{,}6612$'dir. Dolayısıyla log-doğrusallık varsayımı da $\%5$ anlamlılık düzeyinde reddedilemez.
- Örneğin de gösterdiği gibi bazı durumlarda modellerin ikisi de reddedilmeyebilmektedir.

9.3.3 Diğer Bazı Sınama ve Konular

Almaşık Sınamalar

- Görüldüğü gibi doğrusal bağlanım modelleri çerçevesinde çeşitli önsavları sınamak için t ve F sınamalarından yararlanılabilmektedir.
- Doğrusal modellerin basit dünyasından çıkıldığında ise doğrusal ve doğrusaldışı her modelde kullanılabilecek önsav sınamalarına gereksinim duyulur.
- Bu amaç için sıkça kullanılan üç yöntem şunlardır:

"Wald sınaması" (Wald test)
"Olabilirlik oranı" (likelihood ratio), kısaca *"OO"* (LR)
"Lagrange çarpanı" (Lagrange multiplier), kısaca *"LÇ"* (LM)

- Bu üç sınama kavuşmazsal olarak eşdeğerdir ve üçünün de sınama istatistiği χ^2 dağılımına uyar.
- Diğer yandan, doğrusal modellerdeki her türlü sınama için F yeterlidir ve Wald, OO ve LÇ'ye bakmaya gerek yoktur.
- Dolayısıyla bu sınama üçlüsünü şimdilik ele almayacağız.

Çoklu Bağlanım ve Kestirim

- Tahmin edilen bir bağlanım işlevi, belli bir X_0 değerine karşılık gelen Y'yi kestirmek için kullanılabilir.
- İki farklı kestirim türü vardır: *"Ortalama kestirimi"* (mean prediction) ve *"bireysel kestirim"* (individual prediction).
- Ortalama kestirimi, belli X_0 değerlerine karşılık gelen $E(Y|X_0)$ koşullu olasılık değerinin kestirilmesini içerir.

- Bireysel kestirim ise X_0'ın karşılığı olan tekil $Y|X_0$ değerinin kestirilmesi demektir.
- Ortalama kestirimi, anakütle bağlanım işlevindeki noktanın kestirimidir ve varyansı bireysel kestirimden daha küçüktür.
- Çoklu bağlanımda kestirim değerlerinin varyans ve ölçünlü hata formülleri karışık olduğu için bunları daha sonra dizey gösterimi ile ele alacağız.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 8* "Multiple Regression Analysis: The Problem of Inference" okunacak.

Önümüzdeki Ders
Kukla Değişkenlerle Bağlanım

Bölüm 10

Kukla Değişkenlerle Bağlanım

10.1 Nitel Değişkenlerle Bağlanım

- Bağlanım çözümlemelerinde bağımlı değişken, sayısal büyüklükler yanında nitel değişkenlerden de etkilenebilir.

  Nicel Değişkenler

  Gelir, üretim, fiyat, maliyet, enflasyon, işsizlik oranı, yaş, boy, çocuk sayısı, ...

  Nitel Değişkenler

  Cinsiyet, ırk, din, savaş, coğrafi bölge, hükümet politikalarında değişme, grev, ...

- Görgül çalışmalarda karşılaşılan pek çok sorunu çözmede nitelik bildiren *"kukla"* (dummy) değişkenlerden yararlanılır.
- Bu değişkenler aynı zamanda *"nitel"* (qualitative) değişken, *"ulamsal"* (categorical) değişken veya *"gösterge"* (indicator) değişkeni olarak da adlandırılmaktadırlar.
- Kukla değişkenler bir veri sınıflandırma aracıdır.
- Nitel özellikleri nicel olarak gösterebilmek için, niteliğin varlık ya da yokluğunu gösteren 1 ve 0 değerlerini alırlar.
- Örnek olarak bir kimsenin üniversite mezunu olduğu 1 ile, olmadığı ise 0 ile gösterilebilir.

- Kuklaların mutlaka 0 ya da 1 değerleri almaları gerekmez. Doğrusal ilişkili herhangi bir sayı çifti kullanılabilir.
- Diğer yandan, yorumlamada sağladığı kolaylıktan dolayı uygulamada {0, 1} çifti yeğlenmektedir.
- Bağlanım modellerinde kukla değişkenlerin kullanılması ek bir zorluk getirmemektedir.
- Kukla içeren bağlanımların hesaplanması bilindik şekilde olur. Katsayıların anlamlılığı da t istatistiği ile sınanabilir.

10.1.1 VARÇÖZ Modelleri

- Bir bağlanım modelindeki tüm açıklayıcı değişkenlerin birer kukla değişken olması mümkündür.
- Böyle modellere *"VARÇÖZ"* (ANOVA) modelleri de denir.
- VARÇÖZ modelleri özellikle toplumbilim, psikoloji, eğitim, pazar araştırması gibi alanlarda yaygındır.

Örnek olarak, Şubat-Mart 2011 döneminde Ankara Çankaya'da satılığa çıkartılan ikinci el daire fiyatlarının üç ayrı semtte nasıl farklılık gösterdiğini inceleyen aşağıdaki modeli ele alalım.

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i$$

- Y burada satış ilanı verilen apartman dairesinin 1000 TL olarak fiyatıdır.
- Kukla değişkenler ise D ile gösterilmektedir.
- $D_{2i} = 1$, daire Dikmen'de ise; $D_{2i} = 0$, eğer değilse. $D_{3i} = 1$, daire Kavaklıdere'de ise; $D_{3i} = 0$, eğer değilse.
- *Not:* Eğer $D_{2i} = 0$ ve $D_{3i} = 0$ olursa daire Kavaklıdere ya da Dikmen'de değil, üçüncü seçenek olan Cebeci semtinde bulunuyor demektir.
- Bu modelin bize gösterdiği şey şudur:

$$\begin{aligned}
\text{Dikmen'de ortalama fiyat:} &\quad E(Y_i|D_{2i}=1, D_{3i}=0) = \beta_1 + \beta_2 \\
\text{Kavaklıdere'de ortalama fiyat:} &\quad E(Y_i|D_{2i}=0, D_{3i}=1) = \beta_1 + \beta_3 \\
\text{Cebeci'de ortalama fiyat:} &\quad E(Y_i|D_{2i}=0, D_{3i}=0) = \beta_1
\end{aligned}$$

- Bağlanıma ilişkin tahmin sonuçları aşağıdaki gibidir:

$$
\begin{array}{llllll}
\hat{Y}_i = & 91{,}4615 & +\,47{,}2746\,D_{2i} & +\,94{,}1218\,D_{3i} & \\
\text{öh} & (11{,}6983) & (13{,}6480) & (16{,}8850) & \\
t & (7{,}8184) & (3{,}4638) & (5{,}5743) & R^2 = 0{,}3491
\end{array}
$$

- D_2 ve D_3 anlamlı olduğu için, Ankara'nın bu üç semtinde daire fiyatlarının farklı olduğunu söyleyebiliriz.
- Buna göre Cebeci'deki ortalama bir dairenin fiyatı yaklaşık 91 bin TL iken, Dikmen ve Kavaklıdere'deki ortalama fiyat ise sırası ile $91 + 47 \approx 140$ ve $91 + 94 \approx 185$ bin liradır.

Birkaç Önemli Nokta

Kukla değişken kullanımı ile ilgili dikkat edilmesi gereken bazı noktalar şunlardır:

1. Bir nitel değişkende m sayıda sınıf ya da *"ulam"* (category) varsa, $(m - 1)$ kukla değişken kullanılmalıdır. Aksi halde *"kukla değişken tuzağı"* (dummy variable trap) denilen *"tam eşdoğrusallık"* (exact collinearity) oluşur. Ancak, sıfır noktasından geçen bağlanımlarda ulam sayısı kadar kukla değişken koymak mümkündür.
2. "Dikmen" ve "Dikmen değil" gibi seçeneklere hangi değerin atanacağı isteğe bağlıdır. Eğer Dikmen değil $= 1$ olursa β_2 katsayısı da eksi çıkar. Demek ki kukla değişken içeren modelleri yorumlarken, 1 ve 0 değerlerinin nasıl verildiğini bilmek önemlidir.
3. 0 değeri verilen ulam, *"yazın"* (literature) içerisinde farklı adlarla karşımıza çıkabilmektedir:

Türkçe	İngilizce
"Taban ulam"	(Base category)
"Kıyas ulamı"	(Benchmark category)
"Karşılaştırma ulamı"	(Comparison category)
"Denetim ulamı"	(Control category)
"Gönderi ulamı"	(Reference category)
"Atlanan ulam"	(Omitted category)

Örneğimizde taban ulam Cebeci semtidir. Taban ulam, diğerlerini karşılaştırmada kullanılan sınıftır. Taban ulamı gösteren ya da ölçen terim de β_1 sabit terimidir. D_2 ve D_3 kukla değişkenlerine gelen β_2 ve β_3 katsayılarına ise *"sabit terim farkı"* (constant term difference) adı verilir.

10.1.2 KOVÇÖZ Modelleri

- Birçok iktisadi araştırmada, yalnızca kukla değişkenlerin kullanıldığı VARÇÖZ modellerine çok sık rastlanmaz.
- Bunun yerine nitel ve nicel değişkenlerin birlikte olduğu *"kovaryans çözümlemesi"* (analysis of covariance) ya da *"KOVÇÖZ"* (ANCOVA) modelleri yeğlenir.
- Bu modellerde nicel değişkenlere *"denetim değişkeni"* (control variable) de denir.

KOVÇÖZ modellerine bir örnek olarak Ankara'daki daire fiyatları örneğimizi geliştirelim.

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i$$

- Y_i burada satılığa çıkartılan apartman dairesinin fiyatıdır.
- X_i ise dairenin metrekare cinsinden büyüklüğüdür.
- $D_{2i} = 1$, daire Dikmen'de ise; $D_{2i} = 0$, eğer değilse. $D_{3i} = 1$, daire Kavaklıdere'de ise; $D_{3i} = 0$, eğer değilse.
- *Not:* Cebeci'yi taban olarak ulamlandırmayı sürdürüyoruz.
- Bağlanım tahminleri aşağıdaki gibidir:

$$\begin{array}{lllll}
\hat{Y}_i = & -27{,}7639 & +\ 15{,}1179\ D_{2i} & +\ 72{,}4160\ D_{3i} & +\ 1{,}24994\ X_i \\
\text{öh} & (15{,}6787) & (9{,}7259) & (11{,}4113) & (0{,}1431) \\
t & (-1{,}7708) & (1{,}5544) & (6{,}3460) & (8{,}7354) \\
R^2 = & 0{,}7217 & & &
\end{array}$$

- Yeni modeldeki sabit terim fark katsayısının Kavaklıdere için anlamlıyken Dikmen için anlamlı olmadığını görüyoruz.
- Diğer bir deyişle, metrekare sabitken, Dikmen'deki daire fiyatlarının Cebeci ile aynı olduğu reddedilmemekte, Kavaklıdere için ise 72 bin liralık bir fark gözlenmektedir.
- Bulguların Dikmen için farklı çıkmasının nedeni, ilk baştaki modelde X'i hesaba katmamış olmamızdır.
- Sonuçlar Ankara'nın bu üç semtindeki dairelerin metrekare fiyatının yaklaşık 1250 lira olduğunu göstermektedir.
- Burada sabit terimleri farklı olan ancak aynı β_4 eğimini paylaşan 3 farklı bağlanımı ele aldığımıza dikkat ediniz.

10.2 Kukla Değişken Kullanım Şekilleri

10.2.1 Chow Sınamasının Kukla Almaşığı

- Önceki örnekte, nitel değişkenlerin sabit terimi etkilediği ama eğim katsayısını etkilemediği varsayılmıştı.
- Diğer yandan, eğer farklı ulamların eğim katsayısı da farklı ise sabit terim farklarını sınamanın pek anlamı yoktur.
- Birden fazla bağlanımın aynı olup olmadığını sınamak için çok adımlı Chow sınamasının kullanılabildiğini biliyoruz.
- Farklı bağlanımları sabit terimler, eğimler ya da her ikisi yönünden ayırt edebilen daha genel bir sınama yöntemi kukla değişkenler ile olanaklıdır.
- Türkiye için tüketim harcamaları ve milli gelir verilerimizi anımsayalım:

Çizelge: Türkiye'de Tüketim ve GSYH (1987–2006)

Yıl	C	Y	Yıl	C	Y
1987	51.019	74.416	1997	77.620	112.892
1988	51.638	76.143	1998	78.113	116.541
1989	51.105	76.364	1999	76.077	111.083
1990	57.803	83.371	2000	80.774	119.147
1991	59.366	84.271	2001	73.356	110.267
1992	61.282	88.893	2002	74.894	118.923
1993	66.545	96.391	2003	79.862	125.778
1994	62.962	91.600	2004	87.897	137.110
1995	66.011	97.729	2005	95.594	147.200
1996	71.614	104.940	2006	100.584	156.249

Türkiye'deki 1994 krizini anımsayalım. Verileri 1994 öncesi ve sonrası olarak ikiye ayıralım ve şu iki modeli inceleyelim:

$$\begin{aligned} &\text{1987-1993 dönemi: } Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}, \ n_1 = 7 \\ &\text{1994-2006 dönemi: } Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}, \ n_2 = 13 \end{aligned}$$

Yukarıdaki iki model dört farklı olasılık sunmaktadır:

1. Eğer $\lambda_1 = \gamma_1$ ve $\lambda_2 = \gamma_2$ ise, iki bağlanım sabit terim ve eğim olarak aynıdır: *"Çakışan"* (coincident) bağlanımlar.
2. Eğer $\lambda_1 \neq \gamma_1$ ve $\lambda_2 = \gamma_2$ ise, iki bağlanım yalnızca sabit terimler yönünden farklıdır: *"Koşut"* (parallel) bağlanımlar.
3. Eğer $\lambda_1 = \gamma_1$ ve $\lambda_2 \neq \gamma_2$ ise, iki bağlanım aynı sabit terimli ama farklı eğimlidir: *"Uyumlu"* (concurrent) bağlanımlar.

4. Eğer $\lambda_1 \neq \gamma_1$ ve $\lambda_2 \neq \gamma_2$ ise, iki bağlanım bütünüyle farklıdır: *"Benzemez"* (dissimilar) bağlanımlar.

- Elimizdeki iki modeli karşılaştırabilmek için tüm n_1 ve n_2 gözlemlerini toplayıp aşağıdaki bağlanımı tahmin edelim:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

- $E(u_t) = 0$ varsayımı ile şu iki bağlanımı buluruz:

$$\begin{aligned} E(Y_t|D_t = 0, X_t) &= \quad \alpha_1 + \beta_1 X_t \\ E(Y_t|D_t = 1, X_t) &= \quad (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t \end{aligned}$$

- Y_t ve X_t farklı yıllar için tüketim ve geliri göstermektedir.
- $D_t = 0$ 1994 öncesi, $D_t = 1$ ise 1994 ve sonrası dönemdir.
- α_2 sabit terim farkıdır.
- β_2 ise eğim katsayısı farkı olup, ikinci dönem işlevinin eğim katsayısının ilk ya da temel döneme ait eğim katsayısından ne kadar farklı olduğunu gösterir.
- Model tahmini şu sonuçları vermektedir:

$$\begin{array}{llllll} \hat{Y}_t & = & -4{,}7884 & +\,16{,}2163\ D_t & +\,0{,}7455\ X_t & -\,0{,}1796\ D_t X_t \\ \text{öh} & & (6{,}9547) & (7{,}5961) & (0{,}0836) & (0{,}0874) \\ t & & (-0{,}6885) & (2{,}1348) & (8{,}9146) & (-2{,}0556) \\ R^2 & = & 0{,}9887 & & & \end{array}$$

- Buna göre 1987-94 dönemi tasarruf-gelir bağlanımı şudur:

$$\hat{Y}_t = -4{,}7884 + 0{,}7455\ X_t$$

- 1994-2006 dönemi tasarruf-gelir bağlanımı ise şöyledir:

$$\begin{aligned} \hat{Y}_t &= (-4{,}7884 + 16{,}2163) + (0{,}7455 - 0{,}1796)\ X_t \\ &= \qquad 11{,}4279 \qquad + \qquad 0{,}5659 \qquad X_t \end{aligned}$$

- Sabit terim farkı ve eğim farkının her ikisinin de istatistiksel olarak anlamlı bulunması, bu iki bağlanımın "benzemez" olduğunu göstermektedir.

Kukla değişken yönteminin Chow sınamasına üstünlükleri şunlardır:

1. Kukla değişken yaklaşımı, tek bir bağlanım tahmini içerdiği için uygulama yönünden basittir.
2. Kukla değişkenler, iki bağlanımın farklı olup olmadığının yanı sıra farkın sabit terimden mi yoksa eğimden mi kaynaklandığını da göstermektedir.
3. Tek bağlanım olması önsav sınamalarında kolaylık sağlar.
4. Verilerin bir arada kullanılması serbestlik derecesini arttırır. *Dikkat:* Modele eklenen her kukla değişkenin serbestlik derecesini bir azalttığı unutulmamalıdır.

10.2.2 Karşılıklı Etkileşim

Kukla değişkenlerin bir diğer kullanım alanı da açıklayıcı değişkenler arası karşılıklı etkileşimi incelemektir.
Ankara örneğimize dönelim ve şimdi de şu modeli ele alalım:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$$

- Y_i burada evin fiyatını, X_i ise m^2 alanını göstermektedir.
- $D_{2i} = 1$, kot daire ise; $D_{2i} = 0$, eğer değilse. $D_{3i} = 1$, su deposu bulunuyorsa; $D_{3i} = 0$, eğer yoksa.
- Model tahmin sonuçları aşağıdaki gibidir.

$$\begin{array}{lccccc}
\hat{Y}_i & = & 1{,}2103 & -\,46{,}2989\,D_{2i} & +\,20{,}4479\,D_{3i} & +\,1{,}2023\,X_i \\
\text{öh} & & (19{,}0367) & (12{,}8606) & (9{,}7909) & (0{,}1633) \\
t & & (0{,}0636) & (-3{,}6001) & (2{,}0885) & (7{,}3639) \\
R^2 & = & 0{,}5942 & & &
\end{array}$$

- Bulgular kot dairelerin yaklaşık 46 bin lira ucuz olduğunu, apartmanda su deposu bulunmasının ise ortalama daire fiyatını yaklaşık 20 bin TL yükselttiğini göstermektedir.
- Tahmin etmiş olduğumuz modeldeki üstü kapalı varsayım, D_2 ve D_3'ün fark etkilerinin birbirinden bağımsız olduğudur.
- Diğer bir deyişle, su deposu olsa da olmasa da kot dairenin fark etkisinin aynı olduğu kabul edilmektedir.
- Belli bir uygulamada bu varsayım savunulamayabilir.

- D_2 ve D_3 gibi iki ayrı nitel değişken arasında var olabilecek karşılıklı etkileşim şu şekilde ele alınır:

$$\hat{Y}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i$$

- Burada
  α_2 kot dairenin fark etkisini,
  α_3 su deposu bulunmasının fark etkisini,
  α_4 kot daire ve su deposu olmasının birlikte fark etkisini
  göstermektedir.

- Karşılıklı etkileşimi öneren model tahminleri şöyledir:

$$\begin{array}{llllll}
\hat{Y}_i = & 1{,}1340 & -\ 44{,}7608\ D_{2i} & +\ 21{,}1225\ D_{3i} & -\ 4{,}3179\ (D_{2i}D_{3i}) & +\ 1{,}2014\ X_i \\
\text{öh} & (19{,}2074) & (16{,}1315) & (10{,}7339) & (26{,}9206) & (0{,}1648) \\
t & (0{,}0590) & (-2{,}7747) & (1{,}9678) & (-0{,}1604) & (7{,}2912) \\
R^2 = & 0{,}5944 & & & &
\end{array}$$

- *"Etkileşim kuklası"* (interaction dummy) α_4'ün istatistiksel olarak anlamlı olup olmadığı yine t sınamasıyla bulunabilir.

- Sonuçlar, bir apartmanda su deposu bulunmasının kot daire fiyatlarını da diğer daireler ile aynı şekilde artırdığını göstermektedir.

10.2.3 Parça-Yollu Doğrusal Bağlanım

- Kukla değişkenlerin bir diğer kullanım alanı da *"parça-yollu bağlanım"* (piecewise regression) modelleridir.

- Bu modellere yönelik olarak, Ankara'daki satılık daireler örneğimizdeki fiyat-metrekare ilişkisini göz önüne alalım.

- Daire fiyatlarının *"eşik"* (threshold) düzeyi denilen bir X^* değeri öncesinde ve sonrasında farklı şekilde değiştiğini varsayalım.

- Örnek olarak, daire fiyatları metrekareye göre doğrusal olarak artsın ancak X^* eşik düzeyinden sonra daha dik bir eğimle artıyor olsun.

- Buna göre, elimizdeki model iki farklı parçadan oluşan bir doğrusal bağlanım modelidir.

- Bu tür modeller daha genel bir tür olan *"kama işlevleri"* (spline functions) yaklaşımına bir örnektir.

- Parça-yollu bağlanımı açıklamak için şu modele bakalım:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i$$

- Y_i burada dairenin fiyatını, X_i de metrekare genişliğini göstermektedir.
- X^* değeri genişliğin eşik düzeyidir ve önceden bellidir.
- $D_i = 1$, eğer $X_i \geq X^*$ ise;$D_i = 0$, eğer $X_i < X^*$ ise.
- $E(u_i) = 0$ varsayımı altında şunu görebiliriz:

  Eşiğe kadar: $E(Y_i|D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i$
  Eşik sonrası: $E(Y_i|D_i = 1, X_i, X^*) = \alpha_1 - \beta_2 X^* + (\beta_1 + \beta_2) X_i$

- Buna göre β_1 parça-yollu bağlanımın birinci parçasının, $(\beta_1 + \beta_2)$ ise ikinci parçasının eğimini vermektedir.
- Kırılma yoktur diyen önsav için $\hat{\beta}_2$'nın p değerine bakılır.
- Verilerden, $X^* = 120m^2$ sonrasında fiyatların değişiyor olabileceğini çıkardığımızı varsayalım.
- Fiyat (Y) ve genişlik (X) verilerini bir parça-yollu doğrusal bağlanım modeline yakıştırırsak şu bulguları elde ederiz:

$$\begin{array}{llllll}
\hat{Y}_i = & -0{,}4698 & + 1{,}1967\, X_i & + 0{,}2490\, (X_i - X^*) D_i & \\
\text{öh} & (33{,}2861) & (0{,}3203) & (0{,}5660) & \\
t & (-0{,}0141) & (3{,}7365) & (0{,}4400) & R^2 = 0{,}4822
\end{array}$$

- Dairelerin metrekare fiyatı yaklaşık 1200 TL kadardır.
- 120 metrekare üstünde fiyat $(1200 + 250)$ olmakla birlikte, aradaki fark istatistiksel olarak anlamlı değildir.
- Öyleyse $(X - X^*)D$ değişkeni modelden çıkartılabilir.

10.3 Kukla Değişkenlere İlişkin Konular

10.3.1 Mevsimsel Çözümlemeler

- Günlük, aylık ya da üç aylık verilere dayanan birçok zaman serisi; *"mevsimsel örüntü"* (seasonal pattern) ya da *"düzenli salınımsal hareket"* (regular oscillatory movement) gösterir.
- Buna örnek olarak yeni yıl öncesi mağaza satışlarını ya da bayram öncesi hanelerin artan para talebini gösterebiliriz.
- Bir zaman serisinin çeşitli bileşenleri üzerinde ayrı ayrı yoğunlaşmak için, mevsimsel bileşenin çıkarılması istenir.
- TÜFE ve ÜFE gibi önemli iktisadi zaman serileri genellikle *"mevsimsel ayarlamalı"* (seasonally adjusted) yayınlanır.
- *"Mevsimsellikten arındırma"* (deseasonalization) işleminin çeşitli yolları vardır ve bunlardan birisi de kukla değişkenler yöntemidir.
- Konuya ilişkin olarak, Türkiye'de inşaat kesimi için üç aylık üretim ve toplam maliyet endeksleri verilerini ele alalım.

Çizelge: İnşaat Kesiminde Üretim ve Maliyet (2005=100)

Dönem	Üretim	Maliyet	Dönem	Üretim	Maliyet
2005Ç1	78,07	98,44	2008Ç1	100,01	138,79
2005Ç2	105,47	98,65	2008Ç2	128,49	153,78
2005Ç3	114,56	100,97	2008Ç3	128,62	142,16
2005Ç4	101,90	101,93	2008Ç4	105,08	136,62
2006Ç1	90,34	105,63	2009Ç1	81,24	135,44
2006Ç2	126,56	118,86	2009Ç2	101,53	136,62
2006Ç3	136,43	119,73	2009Ç3	106,09	137,37
2006Ç4	120,15	119,72	2009Ç4	97,61	137,47
2007Ç1	101,83	124,53	2010Ç1	88,76	142,26
2007Ç2	135,27	125,75	2010Ç2	121,29	142,77
2007Ç3	142,52	125,89	2010Ç3	128,60	145,52
2007Ç4	120,08	126,56	2010Ç4	115,21	147,81

- İnşaat üretim faaliyetlerinde mevsimsel bir etki olup olmadığını görmek için aşağıdaki modeli inceleyelim:

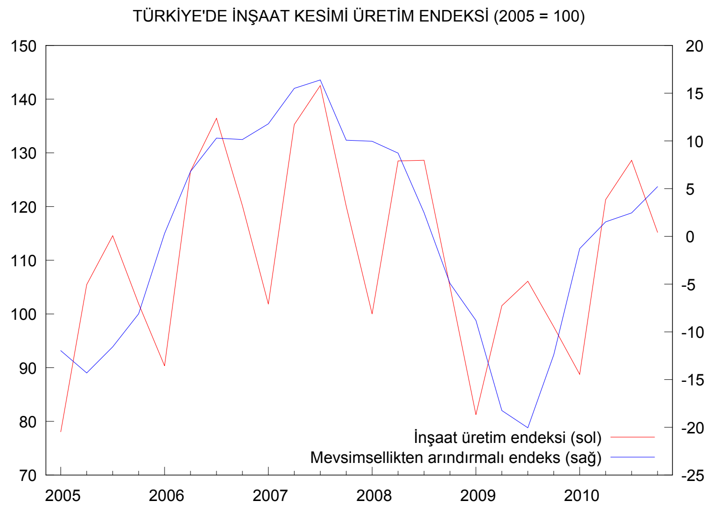
$$\hat{Y}_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_i$$

- $D_{2t} = 1$, ikinci üç ay ise; $D_{2t} = 0$, eğer değilse. $D_{3t} = 1$, üçüncü üç ay ise; $D_{3t} = 0$, eğer değilse. $D_{4t} = 1$, dördüncü üç ay ise; $D_{4t} = 0$, eğer değilse.

- Burada mevsim değişkeni dört ulamdan oluştuğu için üç farklı kukla değişken kullanılmıştır.
- Nicel bir değişken kullanmayıp, Y_t'nin yalnızca sabit terime göre bağlanımını hesapladığımıza dikkat ediniz.
- Tahmin sonuçları aşağıdaki gibidir:

$$\begin{array}{lllll} \hat{Y}_t = & 90{,}0422 & +\,29{,}7242\,D_{2t} & +\,36{,}0968\,D_{3t} & +\,19{,}9633\,D_{4t} \\ \text{öh} & (4{,}7960) & (6{,}7826) & (6{,}7826) & (6{,}7826) \\ t & (18{,}7744) & (4{,}3824) & (5{,}3220) & (2{,}9433) \\ R^2 = & 0{,}6183 & & & \end{array}$$

- İnşaat üretim endeksinin taban dönem olan kış döneminde ortalama 90 düzeyinde olduğunu görüyoruz.
- Endeks bahar döneminde 30 puan yükselmekte, yazın bu yükselişini sürdürmekte, ve güz döneminde gerilemektedir.
- Yukarıdaki bağlanım tahminine ait kalıntılar, inşaat üretim endeksinin mevsimsellikten arındırmalı bir serisini verir.
- Bu kalıntılardan daha sonra serinin *"eğilim bileşeni"* (trend component), *"çevrimsel bileşen"* (cyclical component) ve *"rastsal bileşen"* (random component) unsurları bulunabilir.
- *Dikkat:* Sözü edilen bu mevsimsellikten arındırma işlemi her zaman serisi için uygun değildir.



- Şimdi, önceki yılın aynı dönemine ilişkin toplam maliyeti (X_{t-4}) bir nicel değişken olarak modele ekleyelim.Tahmin sonuçları şöyledir:

$$\begin{array}{llllll}
\hat{Y}_t = & 145{,}4342 & + 32{,}9016\, D_{2t} & + 38{,}0662\, D_{3t} & + 20{,}9031\, D_{4t} & - 0{,}4396\, X_{t-4} \\
\text{öh} & (15{,}7171) & (5{,}6225) & (5{,}5994) & (5{,}5901) & (0{,}1262) \\
t & (9{,}2533) & (5{,}8517) & (6{,}7983) & (3{,}7393) & (-3{,}4831) \\
R^2 = & 0{,}8020 & & & &
\end{array}$$

- İkinci, üçüncü, ve dördüncü çeyreklerdeki üretimin ilk çeyrekten yüksek olduğunu bu modelde de görüyoruz.
- Ayrıca, mevsimsel etkiler gözönüne alındığında, maliyet endeksindeki 1 puanlık artışın üretim endeksinde yaklaşık 0,44 puanlık bir azalmaya yol açtığı anlaşılmaktadır.
- Bu noktada ilginç bir soru X_t'nin de Y_t gibi bir mevsimsel örünüm sergileyip sergilemediği sorusudur.
- Az önce ele almış olduğumuz modelin bir özelliği, bağımlı değişken Y_t'yi mevsimsellikten arındırırken aynı zamanda X_t'yi de mevsimsellikten arındırmasıdır.
- Bunu görmek için ilk bağlanımı tahmin edelim ve kalıntıları saklayalım. Bu, mevsimsellikten arındırmalı $Y_{2,t}$ olsun.
- Şimdi de aynı modeli bu sefer de maliyet bağımlı değişken olacak şekilde tahmin edip kalıntıları saklayalım. Bu da mevsimsellikten arındırmalı $X_{2,t}$ olsun.
- $Y_{2,t}$ ve $X_{2,t-4}$ bağlanıma birlikte sokulursa, $X_{2,t-4}$'ün eğim katsayısının önceki beş değişkenli bağlanımdaki X_{t-4} ile aynı olduğu görülür. Yani bir taşla iki kuş vurmuş oluyoruz.

10.3.2 Yarı-Logaritmasal İşlevler

- Eğitim deneyimi (yıl) ve cinsiyete (1 = erkek) göre öğretim görevlisi işe başlama ücretlerini (yıllık, bin dolar) gösteren şu varsayımsal verileri ele alalım:

Ücret	Deneyim	Cinsiyet
23,0	1	1
19,5	1	0
24,0	2	1
21,0	2	0
25,0	3	1
22,0	3	0
26,5	4	1
23,1	4	0
25,0	5	0
28,0	5	1
29,5	6	1
26,0	6	0
27,5	7	0
31,5	7	1
29,0	8	0

- Verileri şu log-doğ modeline yakıştırmak istiyor olalım:

$$\ln Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + u_i$$

- Y_i başlama ücreti, X_i ise eğitim deneyimidir.
- $D_i = 1$, eğer erkekse; $D_i = 0$, kadınsa.
- β_2 katsayısı burada X_i'deki bir birimlik değişmeye karşılık Y_i'deki göreli değişmeyi göstermektedir.
- Göreli değişme 100 ile çarpılır ise yüzde değişme olur.
- Ancak yukarıdaki açıklama, değişkenin yalnızca sürekli bir değişken olması durumunda geçerlidir.
- Kukla değişkenin ortalama Y_i'deki göreli etkisini bulmak için, tahmin edilen $\hat{\beta}_3$ katsayısının e tabanına göre ters logaritmasının alınması ve bundan 1 çıkartılması gerekir.
- Örnekteki modeli tahmin edersek şunu buluruz:

$$\begin{array}{lllll} \widehat{\ln Y_i} = & 2{,}9298 & +0{,}0546\, X_i & +0{,}1341\, D_i \\ t & (481{,}524) & (48{,}3356) & (27{,}2250) \\ & & & R^2 = 0{,}9958 \end{array}$$

- Buna göre cinsiyet farkı dikkate alındığında ortalama işe başlama ücreti, deneyim yılı başına % 5,46 artmaktadır.
- Ancak, D_i'nin katsayısına bakarak ücretlerin erkekler için yüzde 13,41 daha fazla olduğunu söylemek doğru olmaz.

- 0,1341'in ters logaritması alınır ve bundan 1 çıkartılırsa 0,1435 bulunur. Demek ki erkek öğretim görevlisi ücretleri kadınlara göre yüzde 14,35 daha yüksektir.

10.3.3 İleri Çalışma Konuları

Rastsal Değiştirge Modelleri

- Ele almış olduğumuz modellerde β anakütle katsayılarının bilinmeyen ama sabit büyüklükler olduğunu anımsayalım.
- Kukla değişkenlere ilişkin ileri konulardan biri de *"rastsal değiştirge"* (random parameter) modelleridir.
- Yazında çeşitli biçimlerde karşımıza çıkan bu modeller, β değiştirgelerinin de rastsal olduğunu varsayar.

Değiştirilen Bağlanım Modelleri

- İki bağlanımın hem sabit terim farkı hem de eğim farkı kullanılarak karşılaştırıldığı kukla değişken modellerinde, kırılma noktasının bilindiği örtük olarak varsayılır.
- Diğer yandan, kırılma noktasının örneğin 1994'te mi ya da başka bir dönemde mi olduğu çoğu zaman bilinemez.
- Dolayısıyla, bir diğer ileri çalışma konusu da *"değiştirilen bağlanım"* (switching regression) modelleridir.
- Bu modeller, kırılma noktasının da rastsal olmasına izin vererek bağlanımın *"yinelemeli"* (iterative) olarak tahmin edilmesini sağlarlar.

Dengesizlik Modelleri

- Pazarın dengeye gelmediği, arzın talebe eşit olmadığı durumlar için özel tahmin yöntemleri gerekir.
- Örnek olarak bir malın talebi, fiyat ve çeşitli değişkenlerin bir işlevi olarak modellenirken, aynı malın arzı da yine fiyat ve diğer değişkenlerin bir işlevi olarak modellenebilir.
- Arzda yer alan değişkenler taleptekilerden farklı olursa, gerçekte alınıp satılan mal miktarı arzın talebe eşitlendiği noktada olmayabilir ve bu da dengesizliğe yol açar.

- İşte böyle durumları kukla değişkenler yardımıyla ele alan modellere de *"dengesizlik"* (disequilibrium) modelleri denir.

Farklıserpilimsellik ve Özilintinin Etkisi

- Türkiye'de 1994 sonrası tüketimde yapısal bir değişiklik olup olmadığını inceleyen örneği anımsayalım:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_i$$

- Kukla değişken kullanılan böyle bir modelde örtük olarak *"aynıserpilimsellik"* (homoscedasticity), diğer bir deyişle var(u_{1i}) = var$(u_{2i}) = \ldots = \sigma^2$ varsayımı söz konusudur.
- Eğer bu varsayım sağlanamıyorsa tutarsız sonuçlar elde edilmesi olasıdır.
- Öyleyse, kukla değişkenli modellerde *"farklıserpilimsellik"* (heteroscedasticity) sorununun olmadığı doğrulanmalıdır. (*Not:* Bunun için Chow yerine Wald sınaması yapılabilir.)
- Bu tür modellerde özilinti olmadığı varsayımı da önemlidir. Bu konu daha sonra ele alınacaktır.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 9* "Dummy Variable Regression Models" okunacak.

Önümüzdeki Ders
Doğrusal Bağlanım Modeline Dizey Yaklaşımı