

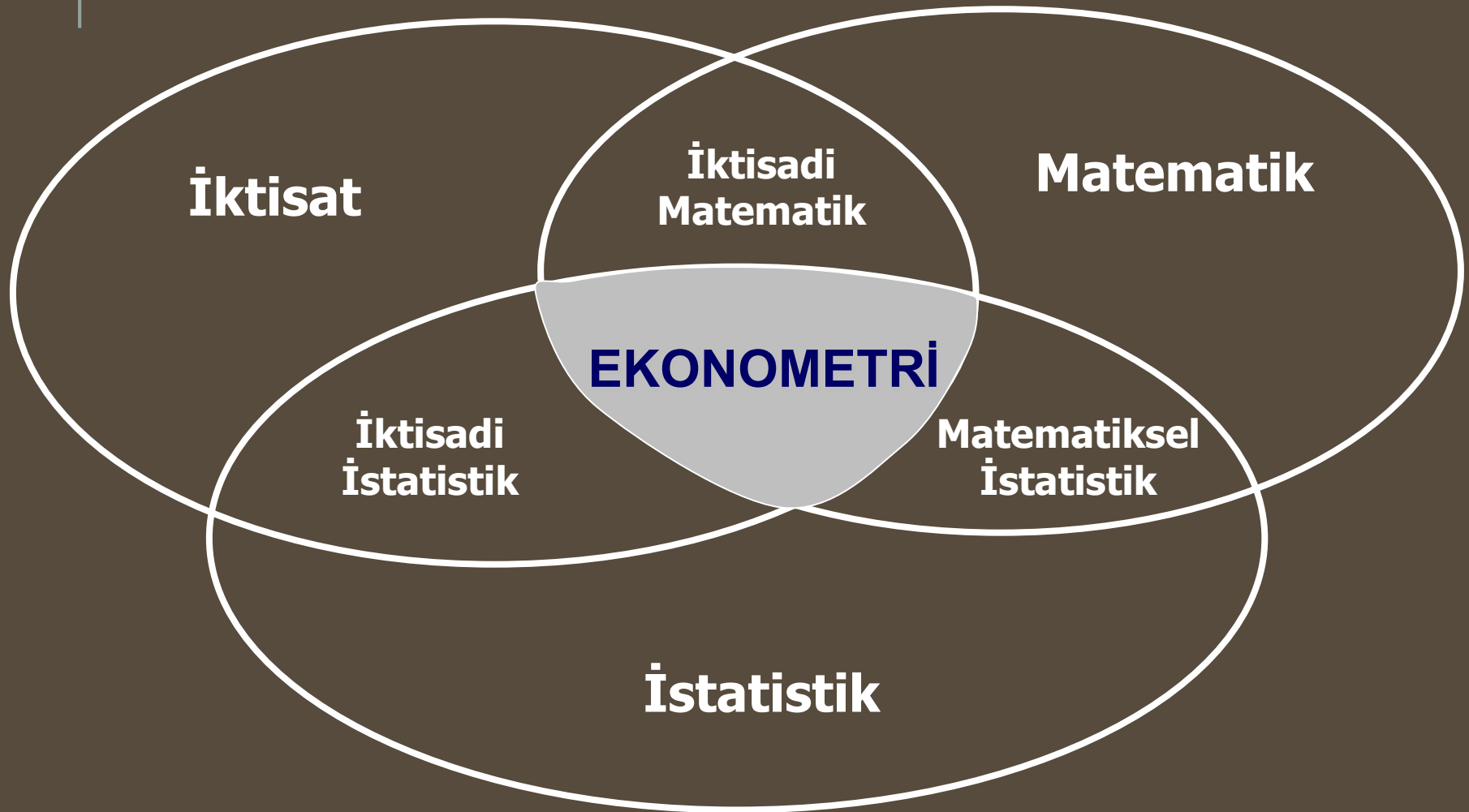
EKONOMETRİNİN TEMEL KAVRAMLARI

Doç. Dr. Orhan GÜNDÜZ

EKONOMETRİ NEDİR?

- İktisat teorisinin, matematik ve istatistik yöntemlerle *kanıtlanması (test edilmesi)*,
- İktisadi problemlere, matematik ve istatistik teorisi yardımıyla *çözüm bulmak*,
- İktisadi olayları, matematik ve istatistiği kullanarak *analiz etme, yorumlama, çıkarsamada bulunma ve öngörü yapma*

EKONOMETRİ NEDİR?



NEDEN EKONOMETRİ?

İktisat bilimi tek başına genellikle niteliksel (kalitatif) açıklamalar ve hipotezler oluşturduğu, herhangi bir sayısal ölçüm ya da değer sunmadığı,

İktisadi matematik, iktisat teorisinin sadece matematiksel bir formda (denklemlerle) ifade edilmesine ve iktisadi teorinin ölçülebilirliği ya da test edilebilirliği ile ilgilenmediği,

İktisadî istatistik temelde sadece iktisadî verinin toplanması, işlenmesi ve grafiklerle ya da tablolarla sunulması ile ilgilendiğinden dolayı

EKONOMETRİ NEDİR?

Diyagramdaki temel unsurun iktisat bilimi ve iktisadi teori olduğu unutulmamalıdır.

İktisat bilimi ekonometriye hipotez verir

İktisadi test (fiyat teorisi, rekabet piyasaları gibi)



İstatistik test (t , F , R^2)



Ekonometrik test (Multicollinearity, otokorelasyon gibi)



EKONOMETRİ BİLİMİNİN AMACI

İktisadi ilişkilerin düzeylerini, sonuçlarını ve katsayılarını gerçeğe en yakın bir şekilde tahmin etmektir.

İktisadi ilişki "MODEL"

Gerçeğe en yakın "İstatistikî Tümevarım"

İKTİSADİ MODEL

İktisadi bilgilerden hareketle, değişkenler arasında kurulan matematiksel ilişki "iktisadi model"

İSTATİSTİKİ TÜMEVARIM

Örnekten hareketle istatistiki analiz metotlarıyla ana kütlenin özelliklerinin gerçeğe yakın tanımlanması istatistiki tümevarımdır.

Temel Kavramlar

Tüketici No	Cinsiyet (K: Kadın, E: Erkek)	Aylık Gelir (TL)	Yıllık Giyim Harcaması (TL)	Evdeki Birey Sayısı
1	E	750	340	3
2	E	500	120	4
3	K	350	250	2
4	E	400	100	3
5	K	250	120	3
6	K	375	300	4
7	E	1500	1080	3
8	E	600	150	4
9	K	2800	1300	5
10	K	425	275	4

Veri: Analizi yapılacak sayı veya gerçeklerdir. Örneğin 1. tüketiciye ait aylık gelir olan 750, veridir.

Veri seti: Bir araştırma için toplanan verilerdir. Yanda, 10 tüketiciye ait veri setidir

Denek (Öge): Hakkında veri toplanan birey veya nesnedir. Her tüketici, bir denektir. Buna göre 10 denek bulunmaktadır

Değişken: Söz konusu deneklerin bir özelliğidir. Çizelgede 4 değişken bulunmaktadır: Cinsiyet (Kadın, Erkek), Aylık Gelir (Milyon TL), Yıllık Giyim Harcaması (Milyon TL), Evdeki Birey sayısı

Gözlem: Tek bir deneğe ait tüm değişkenlere ait veriler, bir gözlemdir. Örneğin 2 no'lu tüketiciye ait; cinsiyeti, aylık geliri, yıllık giyim harcaması ve evdeki birey sayısı verileri, bu tüketiciye ait gözlemdir

NOMİNAL ÖLÇEK

Aralarında büyüklük, sıralama v.b ilişki olmayan genellikle nitel verileri birbirinden ayırmaya yarayan bir numaralama veya sayısal etiketleme ölçeğidir. (okul numarası, futbolcuların sırt numarası, mesleki durum, medeni durum, cinsiyet gibi)

SIRALI ÖLÇEK

Verilerin belli bir ölçüte göre büyükten küçüğe veya küçükten büyüğe sıralanmasıdır (gelir düzeyi, eğitim düzeyi, Likert ölçeği gibi)

ARALIK ÖLÇEK

Ölçüm, tartım v.b yöntemlerle elde edilen ve iki değer arasında değerleri birbirine eşit olan, başlangıcı ve bitiş noktası olan verilerdir.

Örneğin hava sıcaklığı 0°C iken, sıcaklık yok denemez.

Saat 24:01 de başladığınız sıcaklık ölçümü 24 saat sonrasına kadar çok farklı dereceler alabilir.

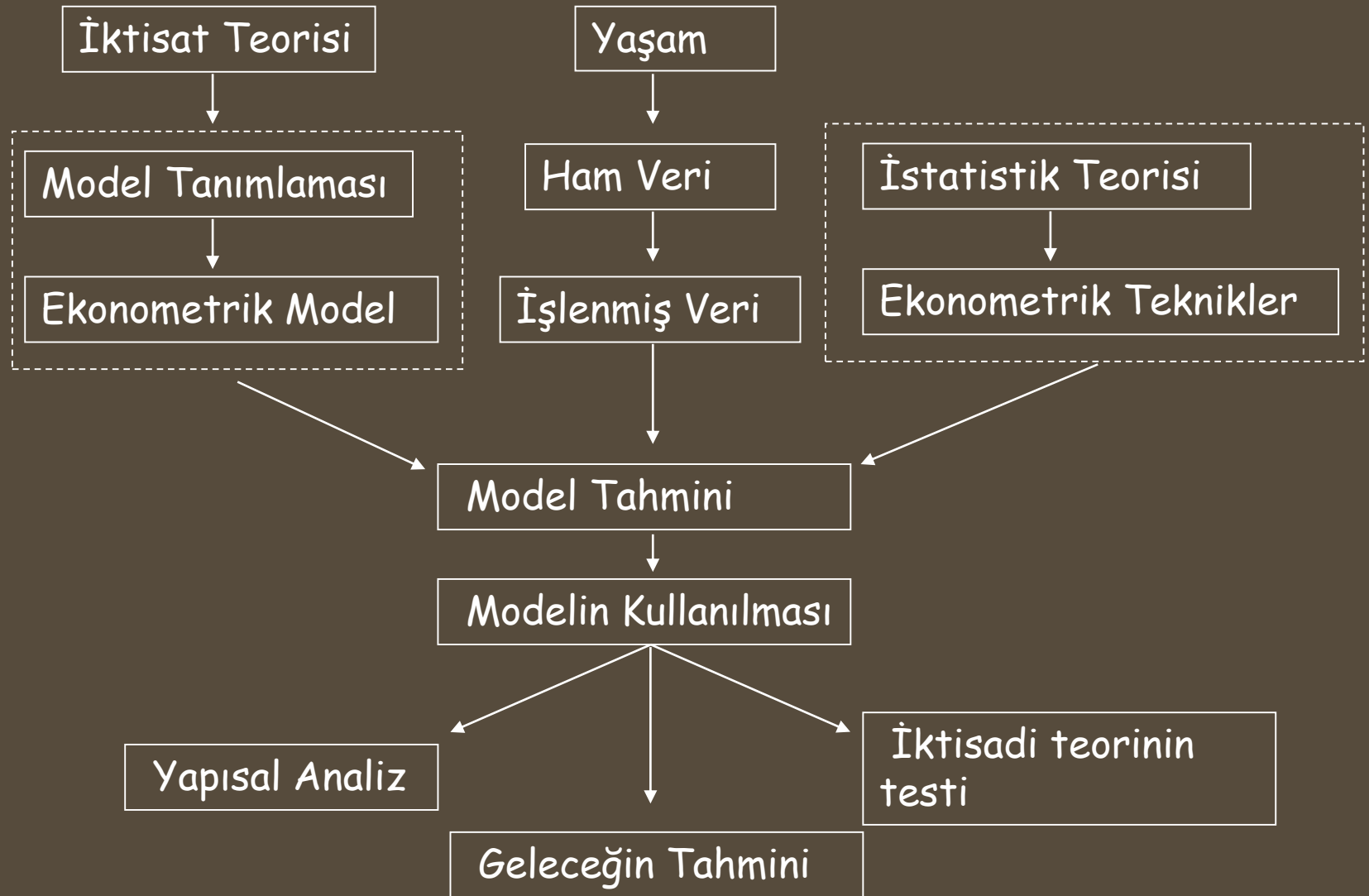
Yılın günleri de bu verilere örnek verilebilir.

ORANSAL ÖLÇEK

Gözlemlerin aldığı değerlerin, oransal olarak karşılaştırılabildiği veri tipidir. Ölçüm veya tartımla elde edilebilir. Başlangıç ve bitiş noktaları kesindir. Veriler içerisinde 0 (sıfır) var ise yokluk ifade eder. Yine, 2 sayısı 1 sayısının 2 katıdır; Fiyat, miktarı, boy, ağırlık verilebilecek örneklerdir.

EKONMETRİK ARAŞTIRMANIN AŞAMALARI

Ekonometrik Yaklaşım



Ekonometrik arařtırmanın ařamaları

İktisadi teori ya da hipotezin ortaya konulması



Bireylerin geliri arttıkça harcamaları da artar, ancak bu artış gelirdeki kadar olmaz (marjinal tüketim eğilimi)

Buğday fiyatları arttıkça ekim alanları artar

Döviz kurundaki artış pamuk ihracatını azaltmaktadır

Akciğer kanserine yakalanma oranı sigara içen bireylerde daha yüksektir

Tamamlayıcı malın fiyatının artması ürün talep miktarında azalmaya neden olur

Ekonometrik araştırmanın aşamaları

İktisadi teori ya da hipotezin ortaya konulması

Model tanımlama

Kesin (Deterministik) Model

Stokastik (olasılıklı) model

- Model değişkenlerinin belirlenmesi (Araştırılan konuya uygun değişkenler ve veriler kullanmak)

- Bağımlı, Bağımsız Değişken Ayırımının Yapılması

- Model katsayılarının İşaret ve Büyüklüklerinin tartışılması

- Modelin Matematiksel Şeklinin Belirlenmesi (Doğrusal mı? doğrusal olmayan fonksiyonel ilişkimi?

Doğrusal model
Tam logaritmik
Yarı logaritmik
Üssel
Ters

Matematiksel modelin seçimi için serpilme diyagramına bakmak gerekir.

Verilerin dağılım şekli matematiksel fonksiyon tiplerinden hangisine uyuyorsa uygun fonksiyon tipine karar verilir

Bağımsız değişken sayısı birden fazla olduğunda serpilme diyagramı net bilgi vermeyeceğinden her fonksiyon tipi denenerek yüksek R^2 , F en yüksek veya AIC, SIC, hata teriminin standart hatası

($s = \sqrt{\frac{\sum e^2}{n-k}}$) en düşük olan seçilir.

İki değişken arasındaki ilişki



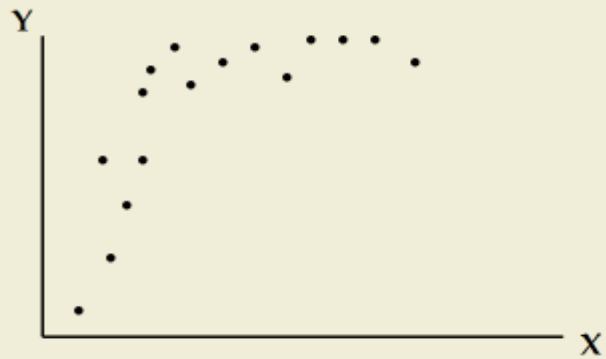
(a) Doğrusal

İki değişken arasındaki ilişki



(b) Doğrusal

İki değişken arasındaki ilişki



(c) Eğrisel

İki değişken arasındaki ilişki

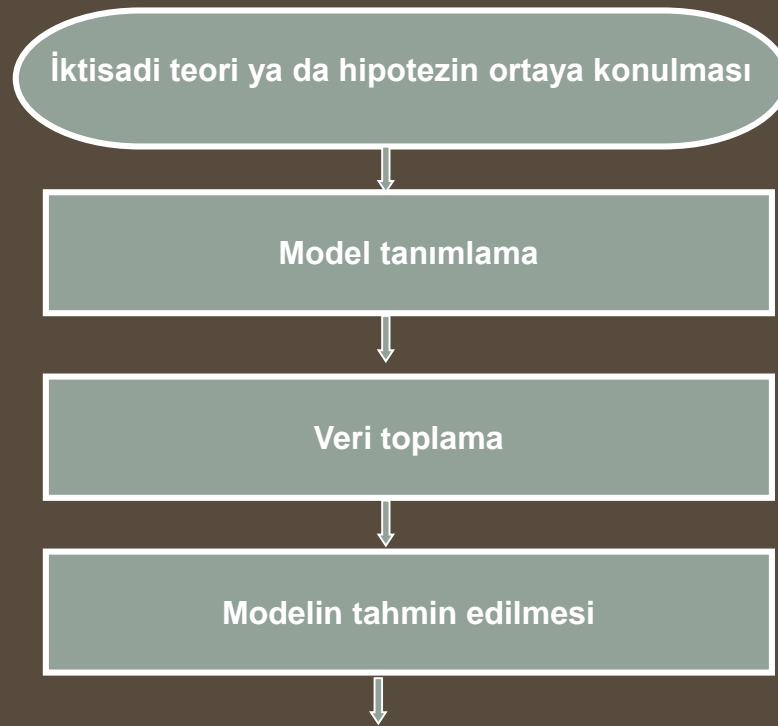


(d) İlişki yok

Ekonometrik araştırmanın aşamaları

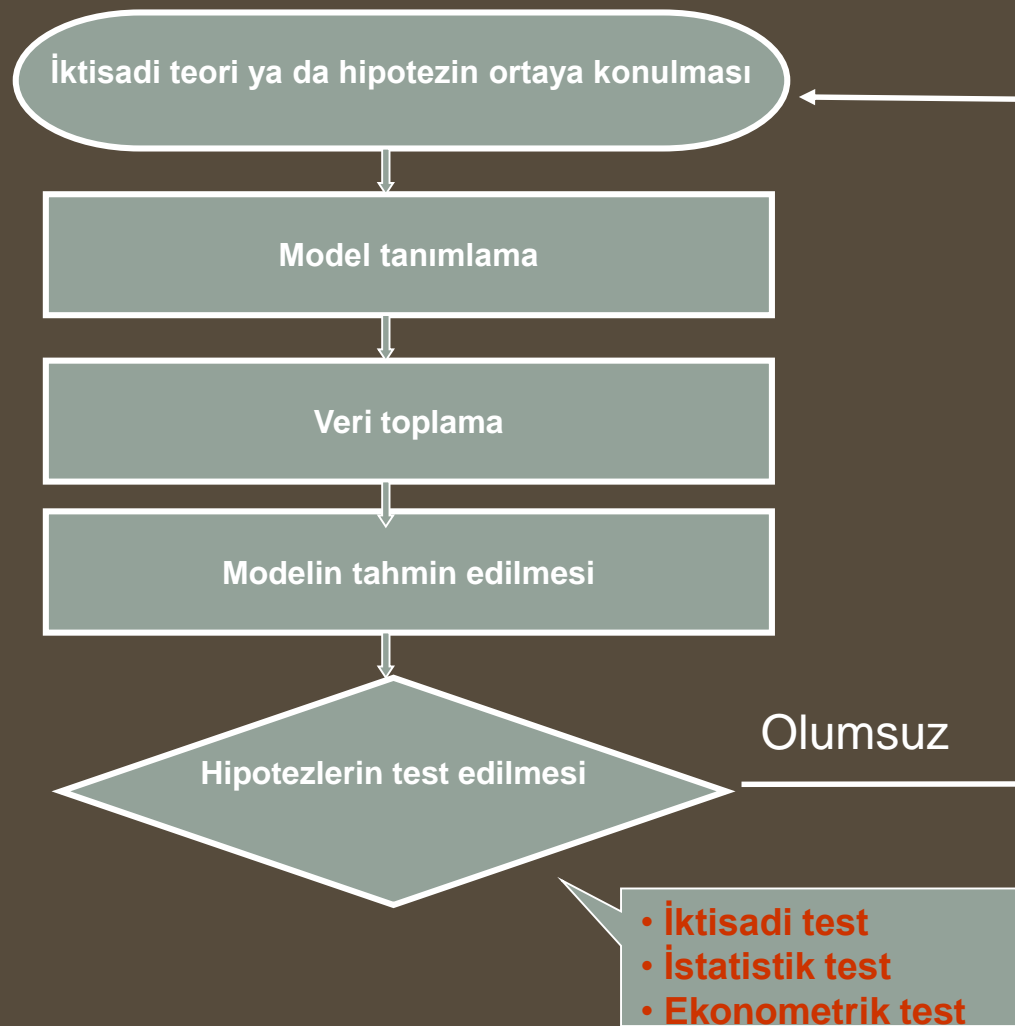


Ekonometrik araştırmanın aşamaları

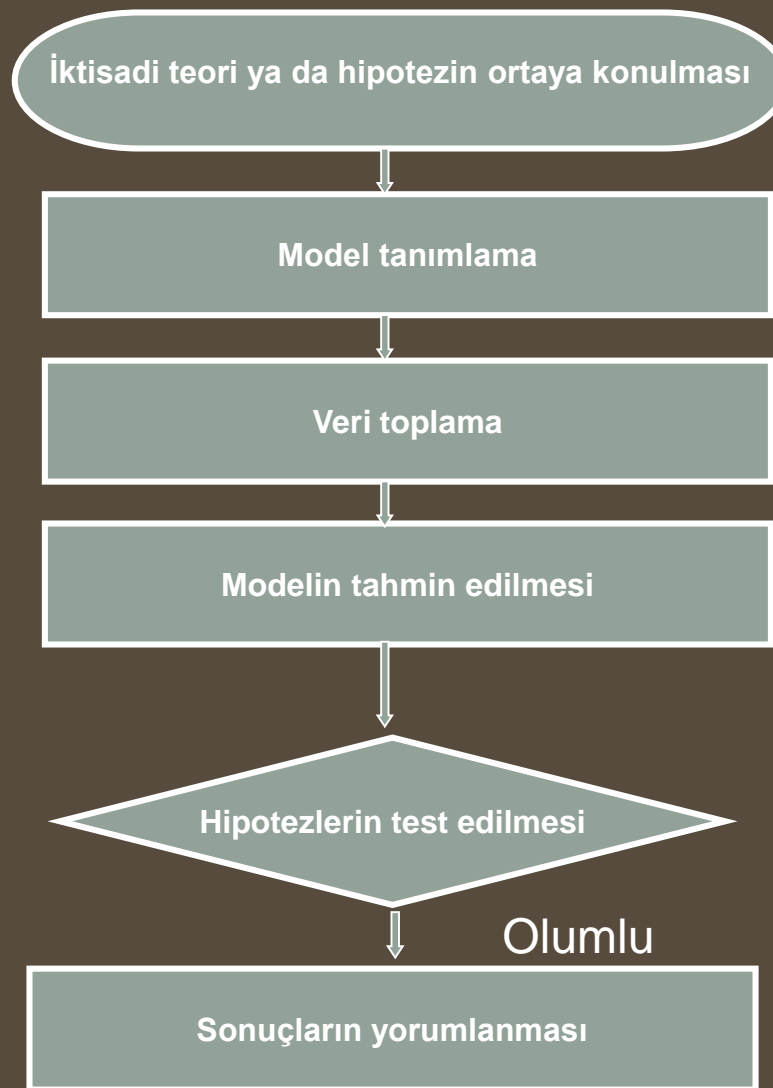


- En küçük kareler yöntemi (EKK)
- Dolaylı en küçük kareler yöntemi (DEKK)
- 2 Aşamalı EKK
- 3 aşamalı EKK
- En yüksek olabilirlik yöntemi

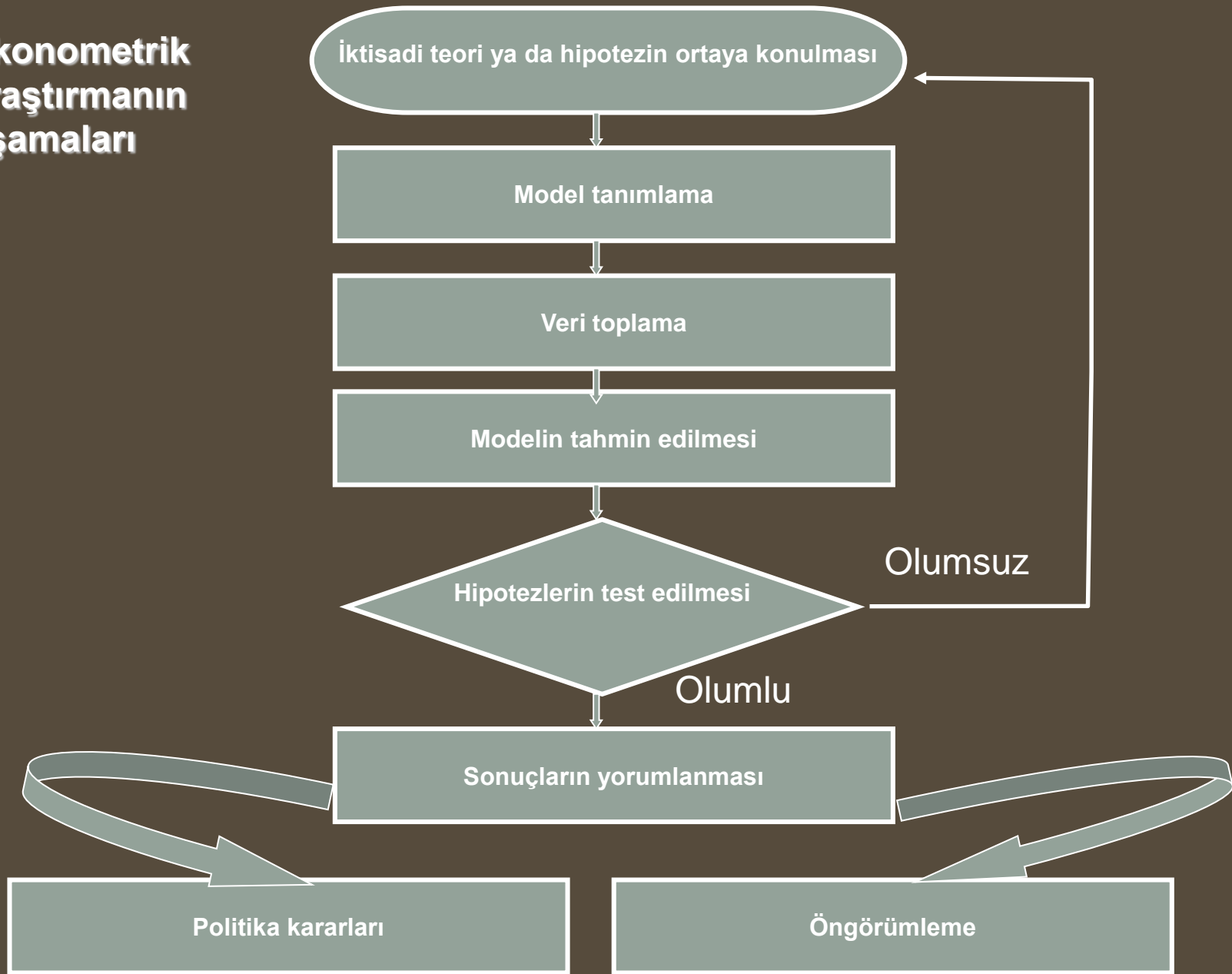
Ekonometrik araştırmanın aşamaları



Ekonometrik araştırmanın aşamaları



Ekonometrik araştırmanın aşamaları





REGRESYON ANALİZİ

REGRESYON ANALİZİ

Regresyon, iki yada daha çok değişken arasındaki ortalama ilişkinin matematiksel bir fonksiyonla incelenmesidir.

Değişkenler

- 1) Bağımlı değişken (açıklanan)
- 2) Bağımsız değişken (açıklayan)

Regresyon analizi ekonometri biliminin amaçlarına ulaşmada kullanılan en yaygın yöntemdir. Çünkü bu analiz tahmin etme, test etme ve öngörü yapabilmeye olanak sağlamaktadır.

Değişkenler arasındaki ilişkiler aşağıdaki gibi sınıflandırılabilir:

- i) Belirleyici (deterministik) ilişkiler
- ii) DeneySEL (ampirik, stokastik) ilişkiler

Kesin (Deterministik) Model

Değişkenler arasında kesin bir ilişki olduğunu varsayan modeller, kesin (deterministik) modeller

Örneğin arz miktarı y 'nin, fiyat düzeyi x 'in tam bir buçuk katı

$$y = 1.5x \quad \text{ise}$$

fiyat ile miktar arasında kesin bir ilişkiyi temsil eder. Bu ilişkide hata payı yoktur.

Stokastik (Olasılıklı) Model

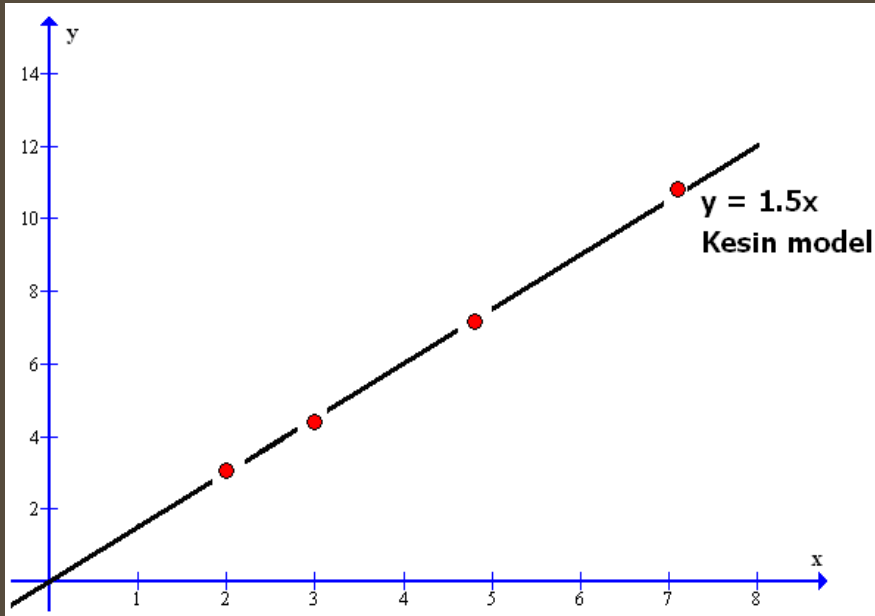
Eğer arz miktarında, önemli fakat ele alınmayan değişkenlerin veya tesadüfi olguların yol açtığı **açıklanmayan değişimlerin olacağına inanıyorsak**, kesin model yerine tesadüfi hataya yer veren modelden yararlanmamız gerekir. Olasılıklı model hem kesin ögeyi hem de "tesadüfi hata" ögesini içerir. Örneğin eğer arz miktarı y 'nin, fiyat düzeyi x ile:

$$y = 1.5x + \text{Tesadüfi Hata}$$

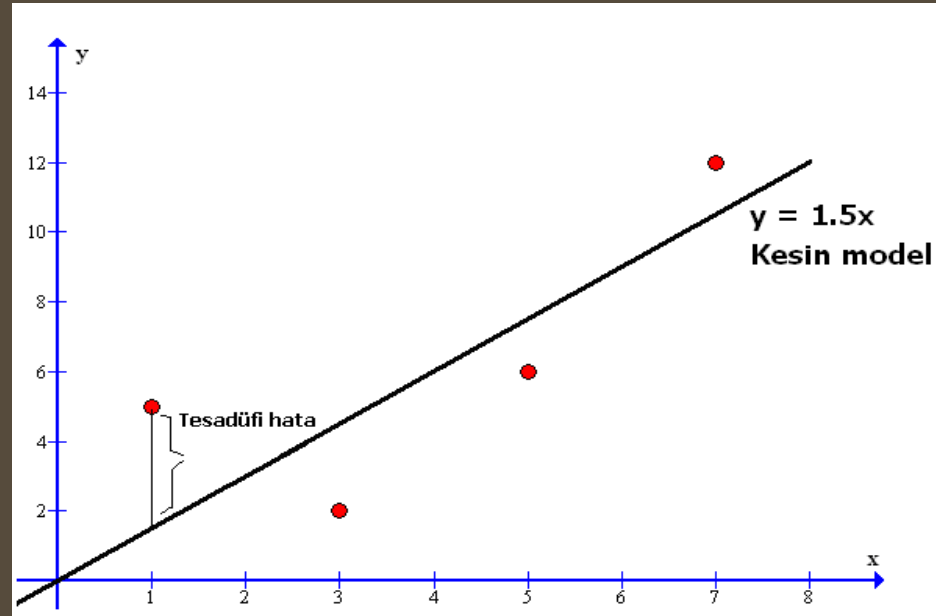
şeklinde bir ilişkisi olduğunu düşünüyorsak, x ile y arasında *olasılıklı* bir ilişki olduğunu anlarız. Görüldüğü gibi, olasılıklı modelin kesin ögesi

$1.5x$ 'tir.

Kesin (Deterministik) ve Stokastik (Olasılıklı) Model



Kesin Model: $y=1.5x$



Olasılıklı Model: $y=1.5x + \text{Tesadüfi hata}$

y	x
1	8
5	12
10	16
15	20
20	24
25	28
30	32

Kesin model

y	x
$1 \pm u$	8
$5 \pm u$	12
$10 \pm u$	16
$15 \pm u$	20
$20 \pm u$	24
$25 \pm u$	28
$30 \pm u$	32

Stokastik model

$u = \text{gerçek değerler} - \text{tahmin edilen değerler}$

u (hata terimi-tesadüfi hata) ortaya çıkış nedenleri

1. Modele alınmayan bağımsız değişkenler
2. İnsanların sürekli davranış değişkenlikleri (ekonomi ilmine ait tüm varsayımlar sabit iken insan kararlarının değişme olasılığı)
3. Matematiksel modelin yanlış seçilmesi (doğrusal yerine doğrusal olmayan model seçilmesi gibi)
4. Toplulaştırma yanlışlıkları (farklı davranışlar gösteren ekonomik birimlerin verilerinin tek bir veri olarak modele dahil edilmesi gibi)
5. Ölçme yanlışlıkları (yanlış örnekleme, verilerin toplanması aşamasında meydana gelen hatalar, hatalı veriler)

Stokastik ilişkilerin incelenmesi regresyon analizinin kapsamına girmektedir. Esasında regresyon analizi tesadüfi hata (hata terimi) nın analizine dayanır.

Regresyon analizinde değişkenler iki grup altında incelenir:

- Bağımsız değişkenler (açıklayıcı değişkenler)
- Bağımlı değişkenler

Bağımlı değişken: Modelin ifade ettiği olay tarafından belirlenirken,

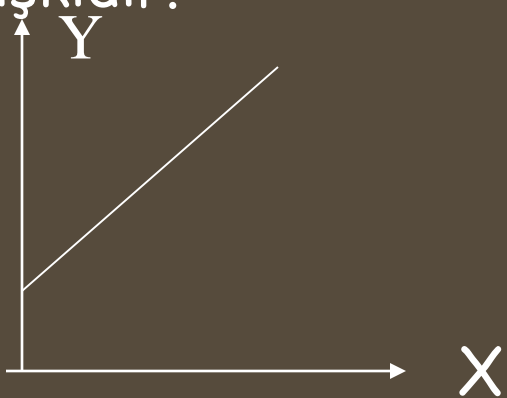
Bağımsız değişken: Modelin ifade edilen olaydan bağımsız olan verileridir.

Örneğin kişilerin gelirlerinin değişmesi, harcama miktarlarının da değişmesine neden olur. Bu durumda gelir bağımsız değişken, harcama miktarı ise bağımlı değişkendir.

Regresyon analizinde genellikle bağımsız değişkenler (X) , bağımlı değişkenler (Y) ile gösterilirler.

Basit doğrusal regresyondaki basit kelimesi iki değişken arasındaki ilişkiyi açıklamak için kullanılır. Doğrusal kelimesi, kurulan modelin parametreleri açısından doğrusal bir model olmasındandır.

İki değişken arasındaki en basit ilişki, bir doğru ile açıklanabilen ilişkidir.



Genel olarak bir doğrunun matematik gösterimi:

$Y = \beta_0 + \beta_1 X$ şeklindedir. Burada β_1 ,

eğimdir ve X 'teki 1 birimlik değişimin Y 'de yaptığı değişikliği gösterir.

β_0 ise X 'in değeri 0 olduğunda Y 'nin almış olduğu değerdir ve Y ekseninin kesme noktası (sabit) olarak isimlendirilir.

EN KÜÇÜK KARELER (EKK) YÖNTEMİ İLE TAHMİN

Gözlemleri en iyi açıklayan doğrunun belirlenmesi için çeşitli yöntemler ileri sürülebilir. Fakat günümüzde en çok kullanılan yöntem "En Küçük Kareler" adı verilen yöntemdir. Bu yöntem gözlemlerin belirlenen doğrudan uzaklıklarının (hata terimlerinin) karelerinin toplamının en küçük yapılmasına dayanır.

EKK YÖNTEMİNİN VARSAYIMLARI-İLKELERİ

1. Hata terimi stokastiktir (u_i ların değeri önceden bilinmez, tesadüfi olarak ortaya çıkar)
2. $E(u_i)=0$ (hata teriminin ortalaması sıfırdır)

0 (sıfır) olmaması
halinde ne?????

EKK YÖNTEMİNİN VARSAYIMLARI-İLKELERİ

1. Hata terimi stokastiktir (u_i ların değeri önceden bilinmez, tesadüfi olarak ortaya çıkar)
2. $E(u_i)=0$ (hata teriminin ortalaması sıfırdır)
3. u_i ler normal dağılım gösterir. (Ancak bu şekilde t, F testleri ve güven aralıkları belirlenebilir)
4. u_i ler sabit varyanslıdır

Eeee bu ne
demek şimdi?????

Sabit Varyans (Homoscedasticity)

$$E(u_i^2) = \sigma_u^2$$

Hata terimi varyansı, açıklayıcı değişkene (X_i) göre değişmez, sabit kalır. Bu varsayım sağlanmadığında normal dağılım gerçekleştiremeyeceğinden istatistik hipotez testleri geçerliliğini kaybeder.

Değişen varyansın ortaya çıkmasının çeşitli nedenleri vardır. Bu nedenler arasında (1) Önemli açıklayıcı değişkenlerin model dışında tutulması, (2) mevsimsellik gösteren zaman serisinin modelde bağımlı değişken olarak kullanılması, (3) bağımlı değişkenin ölçümünün veya tanımının yanlış yapılması ve bu hatanın bağımlı değişken(ler)e göre değişmesi, (4) benzer olmayan ana kütleler üzerinde çalışılması.

Değişen varyans problemini saptamada kullanılan birkaç istatistik test bulunmaktadır.

- ❖ Grafik yöntem,
- ❖ Park testi,
- ❖ Goldfeld-Quandt testi,
- ❖ Glejser testi,
- ❖ Spearman sıra korelasyon testi,
- ❖ Breusch-Pagan-Godfrey testi ve
- ❖ White nR^2 testi.

EKK YÖNTEMİNİN VARSAYIMLARI-İLKELERİ

1. Hata terimi stokastiktir (u_i ların değeri önceden bilinmez, tesadüfi olarak ortaya çıkar)
2. $E(u_i)=0$ (hata teriminin ortalaması sıfırdır)
3. u_i ler normal dağılım gösterir. (Ancak bu şekilde t, F testleri ve güven aralıkları belirlenebilir)
4. u_i ler sabit varyanslıdır
5. Hata terimleri arasında içsel bağıntı (otokorelasyon) yoktur.
6. Bağımsız değişken ile hata terimleri arasında ilişki yoktur
7. Bağımsız değişkenler arasında bir ilişki yoktur veya çok düşüktür (Multicollinearity-Çoklu doğrusal bağıntı)
8. Bağımsız değişken değerleri kesin ve hepsi aynı olmamalıdır
9. Model doğru kurulmalıdır
10. Gözlem sayısı, değişken sayısından büyük olmalıdır ($n>k$)

EN KÜÇÜK KARELER (EKK) YÖNTEMİ İLE TAHMİN

$$Y = \beta_0 + \beta_1 X + e$$

modelinde hata terimi:

Niçin karesi
alınır???

$$e = Y - \beta_0 - \beta_1 X$$

olarak yazılabilir. Bu ifadenin **karesi** alınıp tüm gözlemler için toplanırsa:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y - \beta_0 - \beta_1 X)^2$$

ifadesi elde edilir. EKK yöntemine göre bu ifadeyi minimize eden b_0 ve b_1 değerleri β_0 ve β_1 'in tahmincileri olur.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y - \beta_0 - \beta_1 X)^2$$

İfadesini minimize eden parametre tahmincilerinin değerlerini bulabilmek için eşitliğin β_0 ve β_1 'e göre türevleri alınıp 0'a eşitlenir.

β_0 'a göre türev alınırsa;

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^n e_i^2 = \frac{\partial}{\partial \beta_0} \sum_{i=1}^n (Y - \beta_0 - \beta_1 X)^2$$

$$= -2 \sum_{i=1}^n (Y - \beta_0 - \beta_1 X)$$

β_1 'e göre türev alınırsa;

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^n e_i^2 = \frac{\partial}{\partial \beta_1} \sum_{i=1}^n (Y - \beta_0 - \beta_1 X)^2$$

$$= -2X \sum_{i=1}^n (Y - \beta_0 - \beta_1 X)$$

Her iki denklemi de 0'a eşitlersek;

$$-2 \sum_{i=1}^n (Y - b_0 - b_1 X) = 0$$

$$\sum_{i=1}^n (Y - b_0 - b_1 X) = 0$$

$$-2 \cdot \sum_{i=1}^n X \cdot (Y - b_0 - b_1 X) = 0$$

$$\sum_{i=1}^n X \cdot (Y - b_0 - b_1 X) = 0$$

$$-2 \sum_{i=1}^n (Y - b_0 - b_1 X) = 0$$

$$\sum_{i=1}^n (Y - b_0 - b_1 X) = 0$$

$$-2 \cdot \sum_{i=1}^n X \cdot (Y - b_0 - b_1 X) = 0$$

$$\sum_{i=1}^n X \cdot (Y - b_0 - b_1 X) = 0$$

Parantezleri açarsak;

$$\sum Y - n \cdot b_0 - b_1 \sum X = 0$$

$$\sum XY - b_0 \sum X - b_1 \sum X^2 = 0$$

Bu denklemlere doğrunun NORMAL DENKLEMLERİ denir.

Normal denklemler alt alta yazılıp birlikte çözüldüklerinde b_0 ve b_1 tahmincileri bulunur.

$$\sum Y = n \cdot b_0 + b_1 \sum X$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2$$

$$b_1 = \frac{\sum XY - \frac{(\sum X) \cdot (\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

şeklindeki formüller yardımıyla da tahminciler bulunabilir.

REGRESYON ANALİZİ ÖRNEK UYGULAMA

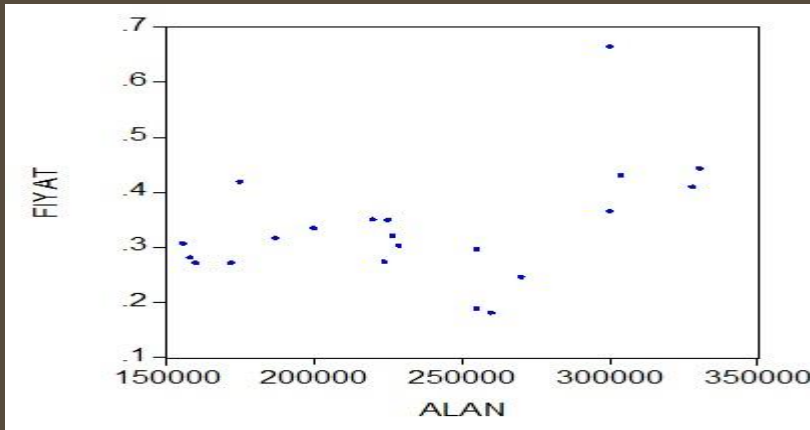
İktisadi olay: Domates fiyatı ile domates ekim alanları arasındaki ilişki

Model: doğrusal

Bağımlı değişken (y) : Domates ekim alanı

Bağımsız değişken (x): domates üretici eline geçen fiyat (t-1)

Tahmin yöntemi: klasik EKK



Ekim alanı(ha)	Fiyat (P_{t-1}) \$
171873	0.271
155616	0.307
160000	0.271
175000	0.418
187000	0.316
158000	0.281
200000	0.334
220000	0.350
225000	0.349
223875	0.273
255000	0.296
260000	0.181
255000	0.188
270000	0.245
228714	0.302
226667	0.320
300000	0.365
330507	0.443
304000	0.430
328000	0.409
300000	0.664

	Ekim alanı(ha) Y	Fiyat X	Y ²	X ²	X*Y
	171873	0,271	29540328129	0,073441	46577,58
	155616	0,307	24216339456	0,094249	47774,11
	160000	0,271	25600000000	0,073441	43360
	175000	0,418	30625000000	0,174724	73150
	187000	0,316	34969000000	0,099856	59092
	158000	0,281	24964000000	0,078961	44398
	200000	0,334	40000000000	0,111556	66800
	220000	0,35	48400000000	0,1225	77000
	225000	0,349	50625000000	0,121801	78525
	223875	0,273	50120015625	0,074529	61117,88
	255000	0,296	65025000000	0,087616	75480
	260000	0,181	67600000000	0,032761	47060
	255000	0,188	65025000000	0,035344	47940
	270000	0,245	72900000000	0,060025	66150
	228714	0,302	52310093796	0,091204	69071,63
	226667	0,32	51377928889	0,1024	72533,44
	300000	0,365	90000000000	0,133225	109500
	330507	0,443	1,09235E+11	0,196249	146414,6
	304000	0,43	92416000000	0,1849	130720
	328000	0,409	1,07584E+11	0,167281	134152
	300000	0,664	90000000000	0,440896	199200
TOPLAM	4934252	7,013	1,22253E+12	2,556959	1696016
ORTALAMA	234964,4	0,333952			

$$\sum Y = n.b_0 + b_1 \sum X$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2$$

$$4934252 = 21 \cdot b_0 + 7,013 \cdot b_1$$

$$1696016 = 7,013 \cdot b_0 + 2,556959 \cdot b_1$$

$$4934252 = 21 \cdot b_0 + 7,013 \cdot b_1$$

$$1696016 \cdot (-2,99) = -(2,99 \cdot 7,013 \cdot b_0) + (-2,99 \cdot 2,556959 \cdot b_1)$$

$$4934252 = 21 \cdot b_0 + 7,013 \cdot b_1$$

$$-5078616,28 = -21b_0 + (-7,66b_1)$$

$$-144364,28 = 0,647 \cdot b_1$$

$$b_1 = 224537.2$$

$$4934252 = 21 \cdot b_0 + 7,013 \cdot 224537.2$$

$$b_0 = 159948.6$$

Dependent Variable: ALAN
 Method: Least Squares
 Sample: 1992 2012
 Included observations: 21

Tablodaki katsayıları ve istatistik
 testleri yorumlayalım

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	159948.6	34164.53	4.681716	0.0002
FIYAT	224537.2	95063.97	2.361959	0.0290
R-squared	0.171509	Mean dependent var		234964.4
Adj R-squared	0.127904	S.D. dependent var		56195.69
S.E. of regres	52478.96	Akaike info criterion		24.66460
Sum squar resid	5.23E+10	Schwarz criterion		24.76408
Log likelihood	-256.9784	F-statistic		3.933251
Durbin-Watson	0.413692	Prob(F-statistic)		0.061984

$$ALAN = 159948.6 + 224537.2 * FIYAT (t-1)$$

$$ALAN = 159948.6245 + 224537.2492 * FIYAT (t-1)$$

1999 yılı için tahmin edilen Y değeri

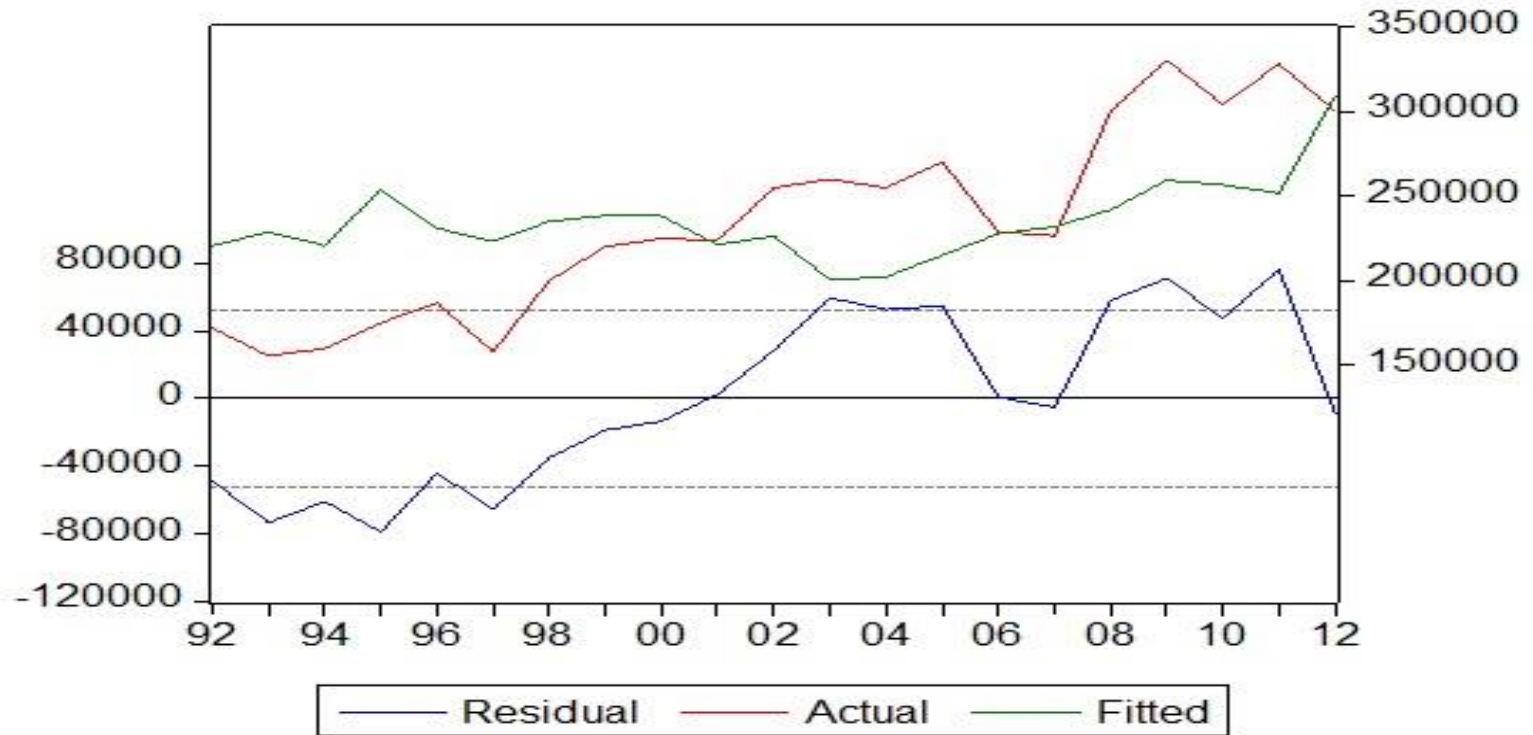
$$Alan (1999) = 159948.6245 + 224537.2492 * FIYAT (1998)$$

$$= 159948.6245 + 224537.2492 * 0.3502$$

$$Tahmin = 238581,57$$

$$Gerçekleşen = 220000$$

$$u (hata) = 220000 - 238581.57 = -18581.57 \text{ hektar}$$



REGRESYON DENKLEMİNİN İNCELENMESİ

Regresyon denklemini incelerken genellikle bizi en çok ilgilendiren soru incelediğimiz iki değişken arasında gerçekten bir ilişki olup olmadığı sorusudur. Bu soru aslında basit doğrusal regresyonda β_1 'in değerinin 0 olup olmadığının araştırılmasıdır. Bu araştırmayı yaparken istatistiksel testler kullanmak gerektiğinden hata terimi ve parametre tahmincilerinin dağılışları hakkında bazı varsayımlarda bulunmak gerekir.

Hata terimi e 'ler, ortalaması 0 ve varyansı s^2 olan birbirinden bağımsız normal dağılışlar gösterirler.

$$E(e)=0 \quad \text{Var}(e)=s^2$$

- *Tahminin Standart Hatası ve Varyansı*

Tahminin standart hatası s , noktaların regresyon doğrusu etrafındaki dağılımlarının ortalama bir ölçüsünü verir.

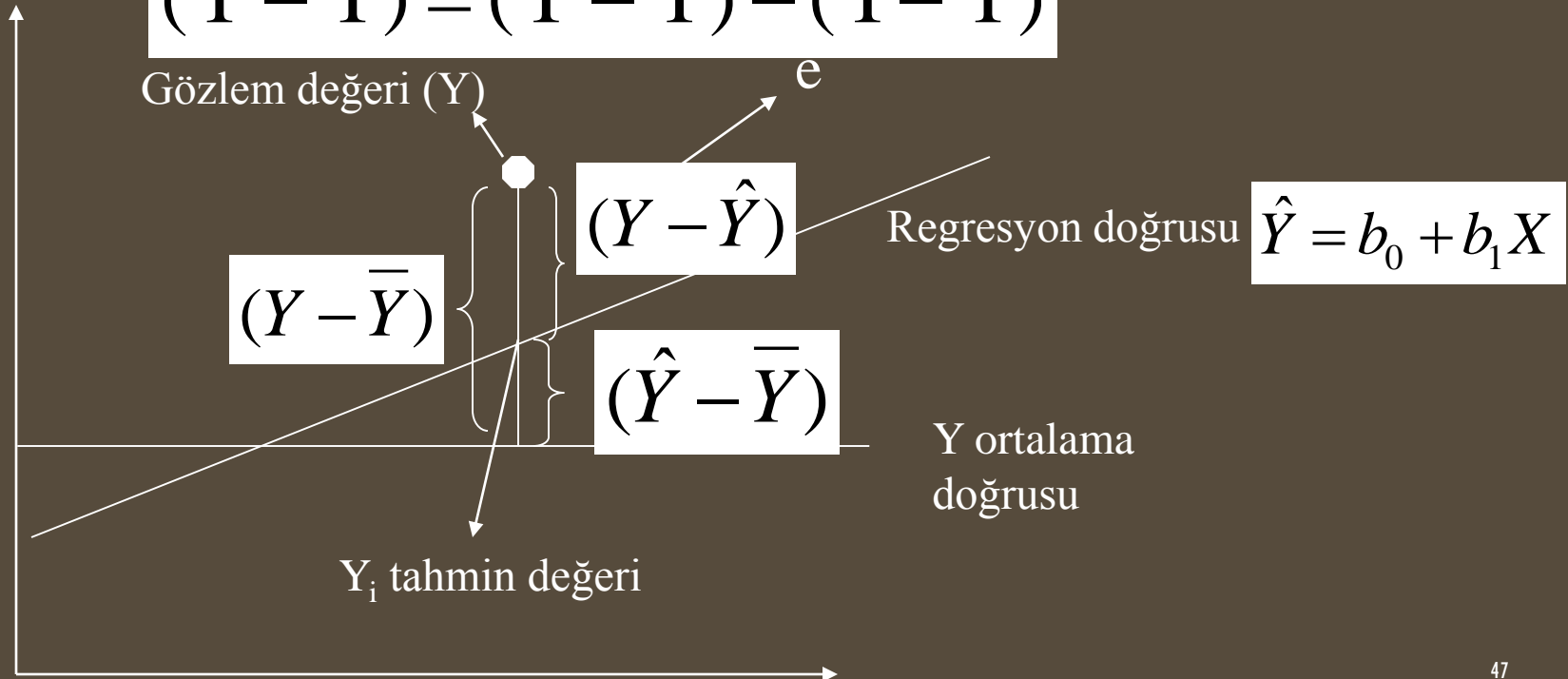
$$s = \sqrt{\frac{\sum e^2}{n-k}}$$

$$s^2 = \frac{\sum e^2}{n-k}$$

Tahminlenen Regresyonun Duyarlılığı

Regresyon denklemi tahminlendikten sonra bu denklemin ilişkiyi ne derece açıkladığı ve bu denklem kullanılarak yapılacak tahminlerin ne derece hassas olacağının araştırılması gerekir. Bunun için gözlenen değerler ile tahmini değerleri arasındaki farkı yazıp y'lerin ortalamasını buna ekleyip çıkarırsak aşağıdaki ifadeyi elde ederiz. Bu ifadenin grafiksel karşılığı şekilde görülmektedir.

$$(Y - \hat{Y}) = (Y - \bar{Y}) - (\hat{Y} - \bar{Y})$$



$$(Y - \hat{Y}) = (Y - \bar{Y}) - (\hat{Y} - \bar{Y})$$

Daha sonra her iki tarafın kareleri alınıp tüm gözlemler için toplanırsa;

$$\sum (Y - \hat{Y})^2 = \sum (Y - \bar{Y})^2 - \sum (\hat{Y} - \bar{Y})^2$$

İfade tekrar düzenlenirse:

$$\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y - \hat{Y})^2$$

Ortalama etrafındaki
kareler toplamı
(genel KT)

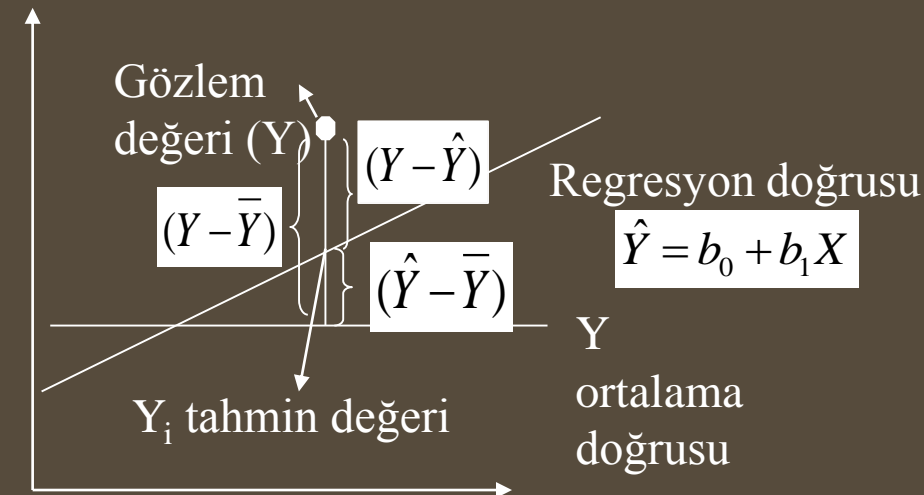
Regresyon kareler
toplamı

Regresyondan
sapmalar (hata)
kareler toplamı

Eğer gözlenen değerlerin hepsi tahmin edilen doğru üzerinde olsaydı, hata kareler toplamı “0” olacak ve uyumun çok iyi olduğu söylenebilecektir. Bu bilgiyi kullanarak, regresyon doğrusunun ne derece iyi tahminlenmiş olduğunu regresyon kareler toplamının ortalama etrafındaki kareler toplamına oranına bakarak söyleyebiliriz. Bu orana BELİRLEME KATSAYISI adı verilir ve R^2 ile gösterilir.

$$R^2 = \frac{\text{regresyon kareler toplamı}}{\text{genel kareler toplamı}} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$$

R^2 'nin 1'e yaklaşan değerleri bize uyumun iyi olduğunu belirtir. ($0 < R^2 < 1$)



Şekilden de görüldüğü gibi, regresyon kareler toplamının büyümesi, gözlem değerinin tahminlenmiş regresyon doğrusuna yaklaşması anlamına gelmektedir ve bu da belirleme katsayısını arttırır.

Hesaplama kolaylığı açısından kareler toplamaları formülleri aşağıdaki şekilde de kullanılabilir:

Genel kareler toplamı (GKT)

$$\sum (Y - \bar{Y})^2 = \sum Y^2 - (\sum Y)^2 / n$$

Regresyon kareler toplamı (RKT)

$$\sum (\hat{Y} - \bar{Y})^2 = b_1^2 \left[\sum X^2 - (\sum X)^2 / n \right]$$

Hata kareler toplamı (HKT)

$$\sum (Y - \hat{Y})^2 = \text{ilk ikisinin farkı}$$

Katsayıların Standart Hataları

$$E(b_0) = \beta_0$$

$$Var(b_0) = \frac{s^2}{\sum X^2 - (\sum X)^2 / n}$$

$$s(\hat{b}_0) = s \cdot \sqrt{\frac{\sum X^2}{n \sum x^2}}$$

$$E(b_1) = \beta_1$$

$$Var(b_1) = \frac{s^2 \sum X^2}{n [\sum X^2 - (\sum X)^2 / n]}$$

$$s(\hat{b}_1) = \frac{s}{\sqrt{\sum x^2}}$$

Katsayıların Güven Aralıkları

$$\hat{b}_0 \pm t_{tab} s(\hat{b}_0)$$

$$\hat{b}_1 \pm t_{tab} s(\hat{b}_1)$$

Parametrelerin teker teker anlamlılığı testi:

Sabit terim β_0 'ın testi için hipotezler:

$H_0 : \beta_0 = 0$ test istatistiği:

$$t = \frac{\hat{b}_0 - b_0}{\sqrt{\text{Var}(b_0)}}$$

$H_1 : \beta_0 \neq 0$

Eğim katsayısı β_1 'in testi için hipotezler:

$H_0 : \beta_1 = 0$ test istatistiği:

$$t = \frac{\hat{b}_1 - b_1}{\sqrt{\text{Var}(b_1)}}$$

$H_1 : \beta_1 \neq 0$

t istatistiği değerleri genelde paket programlar tarafından hesaplanıp verilmektedir. Hesaplanan test istatistikleri (n-2) serbestlik dereceli t dağılışı değeri ile kontrol edilir.

Regresyon doğrusunun tüm parametrelerinin istatistiksel açıdan anlamlı olup olmadığını test etmek için önce *Varyans Analizi Tablosu* aşağıdaki şekilde oluşturulur:

Varyasyon kaynağı	Serbestlik derecesi	Kareler toplamı	Kareler ortalaması
Regresyon	1	RKT	RKO=RKT/1
Hata	n-2	HKT=GKT-RKT	HKO=HKT/(n-2)
Genel (toplam)	n-1	GKT	

Daha sonra sabit terim dışındaki parametrelerin 0'dan farklı olup olmadığı hipotezi test edilir.

$$H_0 : \beta_1=0$$

Test İstatistiği:

$$F = \frac{\text{Regresyon Kareler Ortalaması}}{\text{Hata Kareler Ortalaması}}$$

$$H_1 : \beta_1 \neq 0$$

TAHMİNİN VARYANSI VE GÜVEN ARALIĞININ BULUNMASI

Regresyon denkleminin elde edilmesinin en önemli amaçlarından biri bağımsız değişkenin herhangi bir değeri için Y 'nin alacağı değerin tahminlenmesidir.

$\hat{Y}_k = b_0 + b_1 X_k$ şeklinde hesaplanan bu tahminin, varyansı ve o noktadaki gerçek değer için güven aralıklarının bulunması istenir. Bu tahminin varyansı:

$$s_{\hat{y}_k}^2 = \left[\frac{1}{n} + \frac{(X_k - \bar{X})^2}{\sum X^2 - (\sum X)^2 / n} \right] \cdot s^2$$

Hatanın varyansı

yada

$$s_{\hat{y}_k} = s \sqrt{1 + \frac{1}{n} + \frac{(X_k - \bar{X})^2}{\sum x^2}}$$

X_k noktasında Y 'nin alacağı ortalama değer için güven % $(1 - \alpha)$ 'lık güven aralığı:

$$\hat{Y}_k \pm t_{[\alpha/2, n-2]} \cdot s_{\hat{y}_k}$$

ÇOKLU REGRESYON ANALİZİ

$$Y = b_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k + u$$

EKKY varsayımları çoklu regresyon analizinde de geçerlidir.

KORELASYON KATSAYISI

Korelasyon katsayısı, regresyon modeli ile bulunan tahmini \hat{Y} değerlerinin, gerçek değerlere uygunluğunu ölçmede kullanılır. Korelasyon katsayısı -1 ile 1 arasında değişir.

Katsayının -1 çıkması, iki değişken arasında ters yönlü tam bir ilişkinin olduğunu, 1 çıkması ise doğru yönlü tam bir ilişkinin olduğunu ifade eder.

Katsayının -1'e doğru yaklaşması, değişkenler arasında ters yönlü kuvvetli bir ilişkiyi gösterirken, 1'e yaklaşması değişkenler arasında doğru yönlü kuvvetli bir ilişkiyi ifade eder.

Korelasyon katsayısının işareti, regresyon doğru veya eğrisine ait eğim katsayısının işaretidir.

Korelasyon katsayısının karesi, belirleme katsayısını (determinasyon katsayısını) verir.