Basit Doğrusal Regresyon Modeli

Ekonometri I

Dr. Ömer Kara[1]

[1] İktisat Bölümü
Eskişehir Osmangazi Üniversitesi

10 Kasım 2021

Taslak

Motivasyon

Bu bölümde, sırasıyla aşağıdaki konular incelenecektir.

- Basit Doğrusal Regresyon modeli
- Basit Doğrusal Regresyon modelinde Gauss–Markov varsayımları
- Basit Doğrusal Regresyon modelinin tahminine ait yöntemler
- Sıradan En Küçük Kareler parametre tahmincileri
- Determinasyon Katsayısı
- Sıradan En Küçük Kareler parametre tahmincilerinin özellikleri
- Basit Doğrusal Regresyon modelinde Gauss–Markov teoremi

Motivasyon

- Basit Doğrusal Regresyon (BDR) iki farklı değişken arasındaki ilişkiyi incelemek için kullanılır.
- Daha sonra göreceğimiz nedenlerden dolayı, ugulamalı analizde genel bir araç olarak kullanıldığında BDR modelinin kısıtları vardır.
- Buna rağmen, BDR modelinin nasıl yorumlanacağını öğrenmek, sonraki bölümlerde yapacağımız Çoklu Doğrusal Regresyon (ÇDR) modelini temelden anlamak için üzerinde durulması gereken bir konudur.
- BDR modelinin kısıtları ve ÇDR modeli hakkındaki detaylı bilgi "Ekonometri I - Çoklu Doğrusal Regresyon Modeli: Tahmin" konusunda bulunabilir.

Basit Doğrusal Regresyon Modelinin Mantığı

- Uygulamalı ekonometrik analizlerin çoğu şu önermeyle başlar:

Temel Ekonometrik Önerme

y ve x, bir anakütleyi temsil eden iki rassal değişkendir ve biz, "y'yi x cinsinden açıklamak" veya "y'nin x'teki değişikliklerle nasıl değiştiğini incelemekle" ilgileniyoruz.

- "Ekonometri I - Matematik ve İstatistik Gözden Geçirme" konusunda y ve x arasındaki **kesin ilişkiyi** gösteren fonksiyonları incelemiştik.
- Fakat sosyal bilimlerde iki değişken arasındaki ilişki hiçbir zaman kesin değildir.
- Bu nedenle, "y'yi x cinsinden açıklayacak" bir model yazarken üç sorun vardır.
  1. İki değişken arasında hiçbir zaman kesin bir ilişki olmadığına göre, **diğer faktörler**in y'yi etkilemesine nasıl izin verebiliriz?
  2. y ve x arasındaki ilişkiyi belirten **fonksiyonel form** nedir?
  3. y ve x arasında bir **ceteris paribus** ilişkisi yakaladığımızdan nasıl emin olabiliriz?
- Bu sorunları, y'den x'e ilişkin bir denklem yazarak Slayt 6'deki gibi çözebiliriz.

Basit Doğrusal Regresyon Modeli

Model

$$y = \beta_0 + \beta_1 x + u \qquad \text{(İndekssiz)}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \ldots, n \qquad \text{(İndeksli)}$$

- k: bağımsız değişken sayısı $\longrightarrow k = 1$
- $k+1$: bilinmeyen sabit β parametre sayısı $\longrightarrow \beta_0, \beta_1$
- n: gözlem (veri) sayısı $\longrightarrow i = 1, 2, \ldots, n$ ve $s = 1, 2, \ldots, n, \quad i \neq s$
- y: bağımlı değişken
- x: bağımsız değişken
- u: Hata terimi, x dışında modele dahil edilmemiş tüm faktörlerin ortak etkisi
- β_0: Kesim parametresi (1 tane var), sabit terim olarak da adlandırılır
- β_1: x bağımsız değişkeni için eğim parametresi (1 tane var)
- $\mathbf{x}$: Tüm bağımsız değişkenlerin temsili $\longrightarrow \mathbf{x} = \{x\}$
- Yukarıdaki model bazen **anakütle modeli** olarak da bilinir.

Basit Doğrusal Regresyon Modeli

Model

$$y = \beta_0 + \beta_1 x + u \qquad \text{(İndekssiz)}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \ldots, n \qquad \text{(İndeksli)}$$

- u: Bağımlı değişken y üzerinde etkili olan bağımsız değişken x dışındaki diğer gözlenemeyen faktörleri temsil eder.
- β_0: $x = 0$ iken y'nin alacağı değeri gösterir.
- β_1: y'yi etkileyen diğer tüm faktörler, yani u'da içerilen faktörler, sabitken ($\Delta u = 0$), x'deki değişmenin y'de yaratacağı yalın etkiyi/değişmeyi gösterir.

$$\Delta y = \beta_1 \Delta x + \Delta u$$

$$\Delta y = \beta_1 \Delta x \quad \text{eğer} \quad \Delta u = 0$$

  - Parametreleri yorumlarken fonksiyonel forma dikkat edilmelidir.
  - Farklı fonksiyonel formların yorumlamaları hakkındaki detaylı bilgi Slayt **??**'de bulunabilir.
- $\Delta u = 0$ olduğunda, u'nun içinde bulunan tüm gözlenemeyen faktörlerin ayrı ayrı sabit olduğu değil, ortalama olarak değişimin olmadığı kastedilir. Yani, negatif ve pozitif işaretli u'lar birbirini götürdüğünde ortalama olarak değişim olmayacaktır.

Basit Doğrusal Regresyon Modeli

- Regresyon modellerinde değişkenler için kullanılan terminoloji aşağıda verilmiştir.

Tablo 1: Değişkenler Terminolojisi

y	x
Bağımlı Değişken	Bağımsız Değişken
Açıklanan Değişken	Açıklayıcı Değişken
Tepki Değişkeni	Kontrol Değişkeni
Tahmin Edilen Değişken	Tahmin Eden Değişken
Regresand	Regressor

Basit Doğrusal Regresyon Modeli - Örnek 1

Ücret vs. Eğitim Modeli

Bir çalışanın fazladan 1 yıl eğitim aldığında ücretinin ne kadar arttığını araştırmak istediğimizi düşünelim. Yani, eğitimin ücret üzerindeki etkisini ayrıştırmak istiyoruz.

$$wage = \beta_0 + \beta_1 educ + u$$

$wage$: saat başına ücret (dolar); $educ$: eğitim düzeyi (yıl)

- **Eğim Parametresi** β_1: Ceteris paribus eğitim düzeyindeki 1 birimlik değişimin, saat başına ücretinde meydana getireceği değişimi belirtir.

$$\Delta wage = \beta_1 \Delta educ + \Delta u$$

$$\Delta wage = \beta_1 \Delta educ \quad \text{eğer} \quad \Delta u = 0$$

- **Rassal Hata Terimi** u: Saat başına ücreti etkileyen, eğitim düzeyi dışındaki tecrübe, kıdem, doğuştan gelen yetenek, cinsiyet ve yaş gibi tüm faktörlerin ortak etkisini gösterir.
  Ceteris Parıbus $\Leftrightarrow$ Diğer tüm faktörlerin sabit tutulması $\Leftrightarrow$ $\Delta u = 0$

Basit Doğrusal Regresyon Modeli - Örnek 2

Tarımsal Çıktı vs. Gübre Miktarı Modeli

Gübre miktarının üretilen tarımsal çıktı miktarı üzerindeki yalın etkisini araştırmak istediğimizi düşünelim. Yani, kullanılan gübre miktarının üretilen tarımsal çıktı miktarı üzerindeki etkisini ayrıştırmak istiyoruz.

$$output = \beta_0 + \beta_1 fert + u$$

$output$: tarımsal çıktı miktarı; $fert$: gübre miktarı

- **Eğim Parametresi** β_1: Ceteris paribus gübre miktarındaki 1 birimlik değişimin, çıktı miktarında meydana getireceği değişimi belirtir.

$$\Delta output = \beta_1 \Delta fert + \Delta u$$

$$\Delta output = \beta_1 \Delta fert \quad \text{eğer} \quad \Delta u = 0$$

- **Rassal Hata Terimi** u: Çıktı miktarını etkileyen, gübre miktarı dışındaki yağmur miktarı ve toprağın kalitesi gibi tüm faktörlerin ortak etkisini gösterir. Ceteris Paribus $\Leftrightarrow$ Diğer tüm faktörlerin sabit tutulması $\Leftrightarrow$ $\Delta u = 0$

Doğrusal Model

- Regresyon modelinin doğrusal olması şu anlama gelir: x'deki değişmenin y'de meydana getireceği etki, x'in başlangıç değeri ne olursa olsun aynıdır, yani sabittir.
- Uygulamadaki bu sabit etki varsayımı çoğu zaman gerçeklere uymaz. Örneğin:
  - Ölçeğe göre artan ya da azalan getiri doğrusal regresyon modelleriyle açıklanamaz.
  - Slayt 9'de verilen ücret vs. eğitim modelinde, ilave bir yıl eğitimin etkisi önceki eğitim düzey(ler)ine göre aynıdır, fakat gerçekte daha fazla olması beklenir.
  - Tecrübenin ücret üzerindeki etkisini araştıran bir modelde ise gerşekte tecrübe düzeyinin ücretler üzerinde önce artan sonra azalan bir etkiye sahip olması beklenir.
- Doğrusal modellerin bu kısıtına rağmen Ekonometri Teorisi'nde basitliği ve kolay anlaşılması nedeniyle sıklıkla kullanılır.
- Sabit olmayan etkilerin nasıl modelleneceğini daha sonra göreceğiz.

Gauss–Markov Varsayımları

BDR.1: Gözlem Sayısı

Gözlem sayısı n tahmin edilecek anakütle parametre sayısından büyük ya da en azından eşit olmalıdır.

$$n \geq k + 1$$

BDR.2: Parametrelerde Doğrusallık

Model parametrelerde doğrusaldır.

$$y = \beta_0 + \beta_1 x + u \ ✔$$

$$y = \beta_0 + \beta_1 x^2 + u \ ✔$$

$$y = \beta_0 + \beta_1^2 x + u \ ✘$$

$$y = \beta_0 + \sqrt{\beta_1} x + u \ ✘$$

Detay

Gauss–Markov Varsayımları

BDR.3: Rassallık

Tahminde kullanılan n tane gözlem ilgili anakütleden rassal örnekleme yoluyla seçilmiştir. Yani gözlemler stokhastiktir (rassal), yani deterministik (kesin) değildir.

$$\{(x_i, y_i) : i = 1, 2, \ldots, n\}$$

BDR.4: Bağımsız Değişkenin Sabit Olmaması

Örneklemde (ve bu nedenle anakütlede) bağımsız değişken kendi içinde sabit değildir (yeterli değişenlik vardır).

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$$

Detay

Gauss–Markov Varsayımları

BDR.5: Sıfır Koşullu Ortalama

Bağımsız değişkenin herhangi bir değeri verildiğinde, u hata teriminin beklenen değeri sıfıra eşittir.

$$E(u|x) = E(u|\mathbf{x}) = 0$$

Detay

- Yinelenen Beklentiler Kanunu ve koşullu beklenen değerin 5. özelliği kullanılarak Sıfır Koşullu Ortalama varsayımı yeniden tanımlanabilir.

Ek Bilgi

BDR.5: Sıfır Koşullu Ortalama

$$E(u|\mathbf{x}) = E(u) = 0$$

$$Cov(x,u) = 0, \quad Corr(x,u) = 0 \quad \text{ve} \quad E(xu) = 0$$

Sonuç: u ve x ortalama bağımsızdır. Yani u ve x doğrusal olarak ilişkisizdir.

Ek Bilgi

Gauss–Markov Varsayımları

BDR.6: Otokorelasyon Olmaması

Hata terimleri arasında otokorelasyon yoktur.

$$Corr(u_i, u_s|x) = 0, \quad i \neq s$$

$$Corr(u_i, u_s|\mathbf{x}) = 0, \quad i \neq s$$

$$Corr(u_i, u_s) = 0, \quad i \neq s$$

- BDR.6 varsayımı, yatay-kesit verilerindeki rassallık varsayımı (BDR.3) nedeniyle aslında otomatik olarak sağlanır. Fakat çok ekstrem durumlarda gereklidir ve bu nedenle diğer birçok kaynaktan farklı olarak eklenmiştir.
- BDR.6 varsayımı aşağıdaki eşitlikleri de sağlar.

BDR.6: Otokorelasyon Olmaması

$$Cov(u_i, u_s|\mathbf{x}) = 0 \quad \text{ve} \quad Cov(u_i, u_s) = 0, \quad i \neq s$$

$$E(u_i u_s|\mathbf{x}) = 0 \quad \text{ve} \quad E(u_i u_s) = 0, \quad i \neq s$$

Ek Bilgi

Gauss–Markov Varsayımları

BDR.7: Sabit Varyans (Homoscedasticity)

u hata teriminin bağımsız değişken x'e göre koşullu varyansı sabittir.

$$Var(u|x) = \sigma^2$$

$$Var(u|\mathbf{x}) = \sigma^2$$

$$Var(u) = \sigma^2$$

Detay

- BDR.7 varsayımı aşağıdaki eşitlikleri de sağlar.

BDR.7: Sabit Varyans (Homoscedasticity)

$$E(u^2|\mathbf{x}) = \sigma^2 \quad \text{ve} \quad E(u^2) = \sigma^2$$

Ek Bilgi

- σ **regresyonun standart sapması**dır (bilinmiyor, bu nedenle tahmin edilecek).

Gauss–Markov Varsayımları

- Yukarıda verilen **Gauss–Markov Varsayımları** yatay-kesit verisi ile yapılan regresyon için geçerli varsayımlardır.
- Zaman serileri ile yapılan regresyonlarda bu varsayımların değiştirilmesi gerekir.
- Gauss–Markov Varsayımları, **BDR Varsayımları** olarak da anılır.
- Bazı BDR Varsayımlarının detayı ilerleyen slaytlarda konu akışı içinde verilmiştir.
- Gauss–Markov Varsayımları daha sonra **Gauss–Markov Teoremi**'ni oluşturmada kullanılacaktır.
- Gauss–Markov Teoremi ise BDR modelinin **Sıradan En Küçük Kareler Yöntemi** ya da **Momentler Yöntemi** ile tahmini için teorik dayanak sağlamada kullanılacaktır. Bakınız Slayt **??**.

Anakütle Regresyon Fonksiyonu

- **Anakütle Regresyon Fonksiyonu (ARF)**, BDR.5 varsayımı altında, bağımlı değişken y'nin bağımsız değişken x'e göre koşullu ortalamasıdır.

Anakütle Regresyon Fonksiyonu (ARF)

$$y = \beta_0 + \beta_1 x + u \qquad \text{(Model)}$$

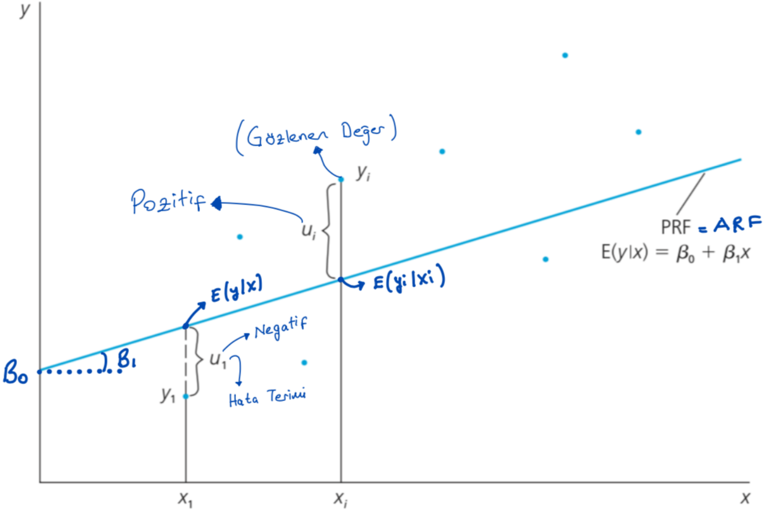
$$E(y|x) = \beta_0 + \beta_1 x \qquad \text{(ARF - İndekssiz)}$$

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x \qquad \text{(ARF - İndekssiz)}$$

$$E(y_i|\mathbf{x}_i) = \beta_0 + \beta_1 x_i \qquad \text{(ARF - İndeksli)}$$

- ARF tektir ve bilinmez.
- ARF, bağımsız değişken x'in doğrusal bir fonksiyonudur.
- $y = \beta_0 + \beta_1 x + u$ modeline ait ARF Şekil 1'de gösterilmiştir.

Anakütle Regresyon Fonksiyonu



Şekil 1: $y = \beta_0 + \beta_1 x + u$ Modeline ARF

Kaynak: Wooldridge (2016)

Anakütle Regresyon Fonksiyonu

- BDR.5 ve BDR.7 varsayımları altında bağımlı değişken y'nin bağımsız değişken x'e göre koşullu dağılımı aşağıdaki gibi ifade edilebilir.

y'nin x'e Göre Koşullu Dağılımı

$$y = \beta_0 + \beta_1 x + u \qquad \text{(Model)}$$

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x \qquad \text{(ARF)}$$

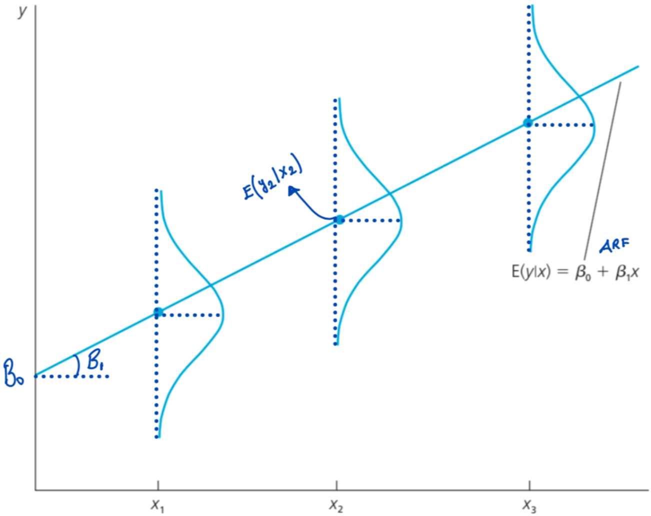
$$Var(y|\mathbf{x}) = \sigma^2$$

$$y|\mathbf{x} \sim (\underbrace{\beta_0 + \beta_1 x}_{\text{Ortalama}}, \underbrace{\sigma^2}_{\text{Varyans}}) \qquad (y|\mathbf{x}\text{'in dağılımı})$$

Ek Bilgi

- Verilmiş bağımsız değişken x düzeyinde bağımlı değişken y'nin dağılımının ortalaması $E(y|\mathbf{x})$ ve varyansı σ^2'dir.
- BDR.5 (sıfır koşullu ortalama) ve BDR.7 (sabit varyans) varsayımları altında, $y = \beta_0 + \beta_1 x + u$ modeline ait ARF ve $y|\mathbf{x}$'in dağılımı Şekil 2'de gösterilmiştir.

Anakütle Regresyon Fonksiyonu ve $y|\mathbf{x}$'in Dağılımı



Şekil 2: $y = \beta_0 + \beta_1 x + u$ Modeline Ait ARF ve $y|\mathbf{x}$'in Dağılımı

Kaynak: Wooldridge (2016)

Örneklem Regresyon Fonksiyonu: Amaç

- BDR tahminindeki asıl amacımız sırasıyla:
  - Öncelikle, iktisat teorisine göre model oluşturmak.
  - Gauss–Markov varsayımları kullanarak ARF'yi oluşturmak.
  - ARF'yi rassal örneklemeyle seçtiğimiz belli sayıdaki veriyi kullanarak tahmin etmektir.
- ARF'nin tahmini ise **Örneklem Regresyon Fonksiyonu**'dur ve bu tahmin örneklemden örnekleme değişir.

Model

$$y = \beta_0 + \beta_1 x + u$$

Anakütle Regresyon Fonksiyonu (ARF)

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x$$

Örneklem Regresyon Fonksiyonu (ÖRF)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- ARF'deki parametreler (β_0, β_1) bilinmeyen sabit sayılarken, ÖRF'deki parametreler $(\hat{\beta}_0, \hat{\beta}_1)$ örneklemden örnekleme değişen rassal değişkenlerdir.

Örneklem Regresyon Fonksiyonu: Amaç

Model

$$y = \beta_0 + \beta_1 x + u \qquad \text{(İndekssiz)}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad \text{(İndeksli)}$$

Anakütle Regresyon Fonksiyonu (ARF)

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x \qquad \text{(İndekssiz)}$$

$$E(y_i|\mathbf{x}_i) = \beta_0 + \beta_1 x_i \qquad \text{(İndeksli)}$$

Örneklem Regresyon Fonksiyonu (ÖRF)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \text{(İndekssiz)}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \text{(İndeksli)}$$

Örneklem Regresyon Fonksiyonu

Örneklem Regresyon Fonksiyonu (ÖRF)

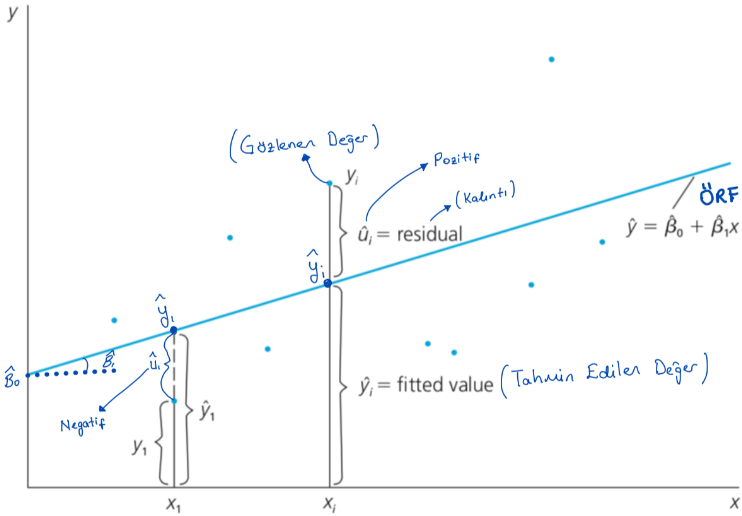
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \text{(İndekssiz)}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \text{(İndeksli)}$$

$$\underbrace{y_i}_{\text{Gözlenen Değer}} = \underbrace{\hat{y}_i}_{\text{Tahmin Edilen Değer}} + \underbrace{\hat{u}_i}_{\substack{\text{Kalıntı (Artık)} \\ \text{Rassal Değil (Deterministik)}}}$$

- $\hat{y}_i$: y_i bağımlı değişkeninin tahmini
- Paramete tahmincileri/tahmin edicileri örneklemden örnekleme değişir, yani rassaldır.
  - $\hat{\beta}_0$: β_0 kesim parametresinin tahmini (1 tane var)
  - $\hat{\beta}_1$: β_1 eğim parametresinin tahmini (1 tane var)
- $\hat{u}_i$: Kalıntı (artık) olarak adlandırılır. Gözlenen değer y_i ile tahmin edilen değer $\hat{y}_i$ arasındakı farkdır. Rassal değildir, tahmin sırasında hesaplanır. Hata terimi u_i'nun örneklem analoğu olarak yorumlanabilir fakat kesinlikle aynı şeyler değildir.
- $y = \beta_0 + \beta_1 x + u$ modeline ait ÖRF Şekil 3'de gösterilmiştir.

Örneklem Regresyon Fonksiyonu



Şekil 3: $y = \beta_0 + \beta_1 x + u$ Modeline Ait ÖRF

Kaynak: Wooldridge (2016)

Örneklem Regresyon Fonksiyonu

- Model, ARF ve ÖRF denklemleri arasında dikkat edilmesi gereken farklar vardır.

Model, ARF ve ÖRF

$$\underbrace{y_i}_{\text{Gözlenen Değer}} = \underbrace{\beta_0 + \beta_1 x_i}_{\substack{E(y_i|\mathbf{x}_i) \\ \text{(Sistemetik Kısım)}}} + \underbrace{u_i}_{\substack{\text{Rassal Hata Terimi} \\ \text{(Sistematik Olmayan Kısım)}}} \quad \text{(Model)}$$

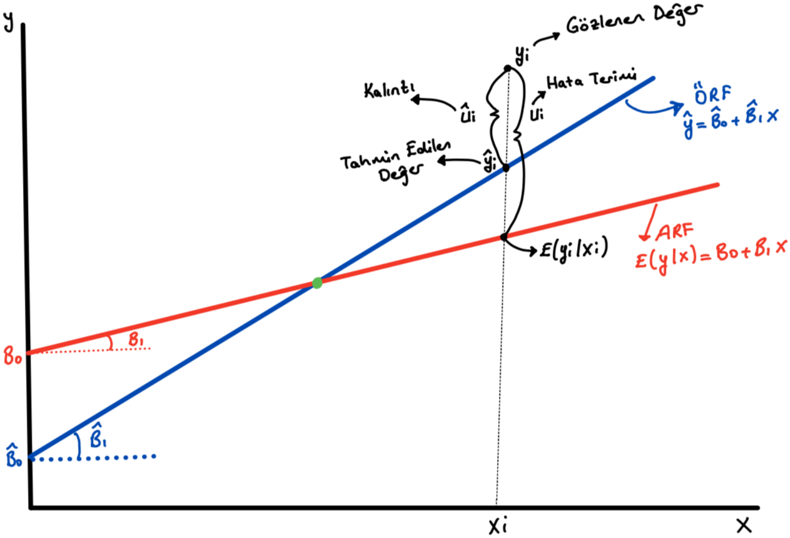
$$E(y_i|\mathbf{x}_i) = \underbrace{\beta_0 + \beta_1 x_i}_{\text{Sistemetik Kısım}} \quad \text{(ARF)}$$

$$\underbrace{\hat{y}_i}_{\text{Tahmin Edilen Değer}} = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\text{Sistemetik Kısmın Tahmini}} \quad \text{(ÖRF)}$$

$$\underbrace{y_i}_{\text{Gözlenen Değer}} = \underbrace{\hat{y}_i}_{\text{Tahmin Edilen Değer}} + \underbrace{\hat{u}_i}_{\substack{\text{Kalıntı (Artık)} \\ \text{Rassal Değil (Deterministik)}}}$$

- $y = \beta_0 + \beta_1 x + u$ modeline ait ARF ve ÖRF Şekil 4'de beraberce gösterilmiştir.

ARF ve ÖRF



Şekil 4: $y = \beta_0 + \beta_1 x + u$ Modeline Ait ARF ve ÖRF

Örneklem Regresyon Fonksiyonu: Tahmin Yöntemleri

Model, ARF ve ÖRF

$$y = \beta_0 + \beta_1 x + u \qquad \text{(Model)}$$

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x \qquad \text{(ARF)}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \text{(ÖRF)}$$

- Örneklem Regresyon Fonksiyonu (ÖRF), iki yöntemle tahmin edilebilir.
  - Sıradan En Küçük Kareler (SEKK) Yöntemi
  - Momentler Yöntemi
- İki yöntem de aynı tahmin sonuçlarını verir.

Sıradan En Küçük Kareler Yöntemi

- **Sıradan En Küçük Kareler Yöntemi**, kalıntı kareleri toplamını (SSR) en küçük yapan parametre tahmincilerini hesaplamaya çalışır.

Örneklem Regresyon Fonksiyonu (ÖRF)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Gözlenen Değer, Tahmin Edilen Değer ve Kalıntı

$$y_i = \hat{y}_i + \hat{u}_i \quad \longrightarrow \quad \hat{u}_i = y_i - \hat{y}_i$$

SEKK Amaç Fonksiyonu

$$\min_{\hat{\beta}_0,\, \hat{\beta}_1} SSR = \min_{\hat{\beta}_0,\, \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2 = \min_{\hat{\beta}_0,\, \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Sıradan En Küçük Kareler Yöntemi

SEKK Birinci Sıra Koşulları

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Sıradan En Küçük Kareler Yöntemi

SEKK Birinci Sıra Koşulları

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \longrightarrow \quad \sum_{i=1}^{n} \hat{u}_i = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \longrightarrow \quad \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

- Birinci sıra koşullarından elde edilen $k + 1$ tane denklemin çözümünden parametre tahmincileri $\hat{\beta}_0$ ve $\hat{\beta}_1$ (toplamda $k + 1 = 2$ tane) bulunur.

Momentler Yöntemi

- Anakütle moment koşulları BDR.5 varsayımı kullanılarak yazılabilir.
- Daha sonra anakütle moment koşullarını kullanarak örneklem moment koşulları elde edilebilir.

BDR.5: Sıfır Koşullu Ortalama

$$E(u|\mathbf{x}) = E(u) = 0$$

$$Cov(x,u) = 0, \quad Corr(x,u) = 0 \quad \text{ve} \quad E(xu) = 0$$

Sonuç: u ve x ortalama bağımsızdır. Yani u ve x doğrusal olarak ilişkisizdir.

Momentler Yöntemi

Anakütle Moment Koşulları ve Örneklem Moment Koşulları

$$\overbrace{E(u) = 0}^{\text{Anakütle}} \longrightarrow \overbrace{\sum_{i=1}^{n} \hat{u}_i = 0}^{\text{Örneklem}}$$

$$E(xu) = 0 \longrightarrow \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

Momentler Yöntemi

- Örneklem moment koşullarından elde edilen $k + 1$ tane denklemin çözümünden parametre tahmincileri $\hat{\beta}_0$ ve $\hat{\beta}_1$ (toplamda $k + 1 = 2$ tane) bulunur.
- SEKK birinci sıra koşulları ve örneklem moment koşulları aslında aynı denklemler kümesini verir.
- Bu nedenle, SEKK Yöntemi ve **Momentler Yöntemi** ile BDR modeli tahmin edildiğinde aynı sonuçlara ulaşılır.
- Genellikle kullanılan yöntem SEKK'dır. Bu nedenle parametre tahmincileri $\hat{\beta}_0$ ve $\hat{\beta}_1$ genellikle **SEKK parametre tahmincileri** ya da **SEKK tahmincileri** olarak adlandırılır.
- Bu yöntemlerin tek çözüm vermesi için BDR.4 (Bağımsız Değişkenin Sabit Olmaması) varsayımının sağlanması gereklidir. Bakınız Slayt 13.

SEKK Parametre Tahmincileri

Ana Model

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad \text{(Model - İndeksli)}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \text{(ÖRF - İndeksli)}$$

- β_0 kesim parametresinin tahmini $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Ek Bilgi

- β_1 eğim parametresinin tahmini, ya da x'in eğim parametresinin tahmincisi, $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Ek Bilgi 1

Ek Bilgi 2

SEKK Parametre Tahmincileri

- β_1 eğim parametresinin tahmini, ya da x'in eğim parametresinin tahmincisi, $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \longrightarrow \hat{\beta}_1 = \frac{\overbrace{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}}^{y \text{ ile } x\text{'in Örneklem Kovaryansı}}}{\underbrace{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}_{x\text{'in Örneklem Varyansı}}}$$

1. $\hat{\beta}_1$, y ile x'in örneklem kovaryansının, x'in örneklem varyansına oranına eşittir.
2. $\hat{\beta}_1$'nın işareti y ile x'in örneklem kovaryansının işaretine bağlıdır. Örneklemde y ve x aynı yönde ilişkiliyse $\hat{\beta}_1$ pozitif işaretli, ters yönde ilişkiliyse negatif işaretli olacaktır.
3. $\hat{\beta}_1$'nın hesaplanabilmesi için x'de yeterli değişiklik olmalıdır. Eğer örneklemde tüm x'ler aynı değerleri alıyorsa x'in örneklem varyansı sıfır olur ve $\hat{\beta}_1$ tanımsız olur.

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$

Sıfır Koşullu Ortalama Varsayımı (BDR.5) Yorumu

Basit Doğrusal Regreson Modeli

$$y = \beta_0 + \beta_1 x + u$$

- BDR modelinde, u'nun x'lerle ilişkisiz olması varsayımı, yani BDR.5:

$$E(u|x) = E(u|\mathbf{x}) = 0$$

- Yani x'in anakütledeki tüm değerleri için u'nun beklenen değeri sıfırdır.
- Ücret vs. eğitim modelinde (Slayt 9) BDR.5 varsayımı aşağıdaki gibi yazılabilir.

$$wage = \beta_0 + \beta_1 educ + u \qquad \text{(Model)}$$

$$E(u|educ) = 0 \qquad \text{(BDR.5)}$$

- Bu ücretleri etkileyen diğer faktörlerin (u) ortalama olarak *educ* ile ilişkisiz olduğu anlamına gelir.
- Örneğin, doğuştan gelen yetenek (ability) u'nun bir parçası ise, ortalama yetenek düzeyi, eğitimin tüm düzeylerinde aynıdır (sabittir).

$$E(ability|educ = 4) = E(ability|educ = 8) = \cdots = 0$$

- Eğer eğitim düzeyi ile doğuştan gelen yeteneğin ilişkili olduğunu düşünüyorsak (örneğin: daha yetenekliler okulda da daha iyiler ve bu nedenle eğitim düzeyleri fazla), bu durumda BDR.5 varsayımı sağlanmaz.

Sıfır Koşullu Ortalama Varsayımı (BDR.5) Yorumu

- Tarımsal çıktı vs. gübre miktarı modelinde (Slayt 10), BDR.5 varsayımı aşağıdaki gibi yazılabilir.

$$output = \beta_0 + \beta_1 fert + u \qquad \text{(Model)}$$

$$E(u|fert) = 0 \qquad \text{(BDR.5)}$$

- Yani, tarımsal çıktı miktarını etkileyen diğer faktörler (yağmur miktarı, toprağın kalitesi vs.), ortalama olarak, $fert$ değişkeniyle ilişkisizdir.
- Eğer yüksek kaliteli toprak parçalarına yüksek miktarda gübre uygulanırsa (yani toprak kalitesi ve gübre miktarı ilişkili), hata terimi u'nun beklenen değeri gübre miktarı ile değişir. Bu durumda BDR.5 varsayımı sağlanmaz.

Regresyonun Yorumu

Model ve ÖRF

$$y = \beta_0 + \beta_1 x + u \qquad \text{(Model)}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta} \qquad \text{(ÖRF)}$$

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x \qquad \text{(Değişim Cinsinden)}$$

- Eğim paramtresi tahmincisi $\hat{\beta}_1$, bağımsız değişken x'in y üzerindeki yalın/kısmi yani ceteris paribus etkisini verir.
- $\hat{\beta}_1$'nın yorumu:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x$$

- x'de meydana gelen 1 birimlik değişmenin y'de meydana getireceği ortalama değişim $\hat{\beta}_1$ kadardır.
  - Parametreleri yorumlarken fonksiyonel forma dikkat edilmelidir.
  - Farklı fonksiyonel formların yorumlamaları hakkındaki detaylı bilgi Slayt **??**'de bulunabilir.

Örnek: CEO Maaşı vs. Karlılık Modeli

CEO Maaşı vs. Karlılık Modeli

$$y = \beta_0 + \beta_1 x + u \quad \longrightarrow \quad salary = \beta_0 + \beta_1 roe + u \qquad \text{(Model)}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \longrightarrow \quad \widehat{salary} = \beta_0 + \beta_1 roe \qquad \text{(ÖRF)}$$

$$\widehat{salary} = 963.191 + 18.501\, roe \qquad \text{(ÖRF)}$$

$$n = 209$$

salary: CEO maaşı (bin dolar); *roe*: şirketin karlılık yüzdesi

- Kesim parametresi $\hat{\beta}_0 = 963.191$ olarak tahmin edilmiştir.
  - $roe = 0$ olduğunda modelce tahmin edilen CEO maaşı $\widehat{salary}$'yi ifade eder. Yani şirketin karlılık yüzdesi *roe* sıfır olduğunda, CEO maaşı \$963191'dir.
- Eğim parametresi $\hat{\beta}_1 = 18.501$ olarak tahmin edilmiştir.
  - *roe*'yı 1 birim arttırdığımızda CEO maaşı *salary* 18.501 birim, yani \$18501 artar.
  - Başka bir ifadeyle, iki CEO'dan birinin *roe* düzeyi diğerinden bir birim fazlaysa, bu iki CEO için tahmin edilen maaş farkı ortalama olarak \$18501'dir.
  - Burada somut iki CEO'dan değil ortalama durumdan bahsedilmektedir.

Örnek: CEO Maaşı vs. Karlılık Modeli

- roe'daki farklı artışların CEO maaşı üzerindeki etkisini tahmin etmek için ÖRF'yi kullanabiliriz. Örneğin, $\Delta roe = 30$ ise

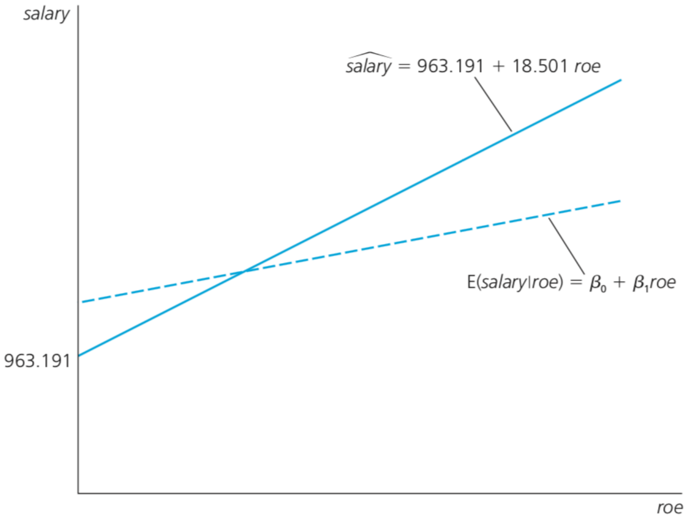
$$\begin{aligned} \Delta\widehat{salary} &= 18.501\Delta roe && \text{(Değişim Cins.)} \\ &= 18.501 \times 30 \\ &= 555.030 \end{aligned}$$

- Farklı roe seviyelerindeki CEO maaşını tahmin etmek için ÖRF'yi kullanabiliriz. Örneğin, $roe = 30$ ise

$$\begin{aligned} \widehat{salary} &= 963.191 + 18.501\,roe && \text{(ÖRF)} \\ &= 963.191 + 18.501 \times 30 \\ &= 1518.221 \end{aligned}$$

  - Ancak bu, şirketi %30 karlılığa sahip olan belirli bir CEO'nun $1518221 kazandığı anlamına gelmez. Bu sadece ÖRF kullanılarak yapılan bir tahmindir.
- Slayt 40'de verilen CEO maaşı vs. karlılık modeline ait
  - ARF ve ÖRF'nin grafiği Şekil 5'de gösterilmiştir.
  - İlk 15 gözlem için bağımsız değişken roe, bağımsız değişken ya da gözlenen değer $salary$, tahmin edilen değer $\widehat{salary}$, ve kalıntı $\hat{u}$ Şekil 6'de gösterilmiştir.

Örnek: CEO Maaşı vs. Karlılık Modeli



Şekil 5: CEO Maaşı vs. Karlılık Modeline Ait ARF ve ÖRF

Kaynak: Wooldridge (2016)

Örnek: CEO Maaşı vs. Karlılık Modeli

TABLE 2.2 Fitted Values and Residuals for the First 15 CEOs

obsno	roe	salary	salaryhat	uhat
1	14.1	1095	1224.058	–129.0581
2	10.9	1001	1164.854	–163.8542
3	23.5	1122	1397.969	–275.9692
4	5.9	578	1072.348	–494.3484
5	13.8	1368	1218.508	149.4923
6	20.0	1145	1333.215	–188.2151
7	16.4	1078	1266.611	–188.6108
8	16.3	1094	1264.761	–170.7606
9	10.5	1237	1157.454	79.54626
10	26.3	833	1449.773	–616.7726
11	25.9	567	1442.372	–875.3721
12	26.8	933	1459.023	–526.0231
13	14.8	1339	1237.009	101.9911
14	22.3	937	1375.768	–438.7678
15	56.3	2011	2004.808	6.191895

Şekil 6: CEO Maaşı vs. Karlılık Modeli Tahminine Ait İlk 15 Gözlem İçin Bazı Değerler

Kaynak: Wooldridge (2016)

Örnek: Ücret vs. Eğitim Modeli

Ücret vs. Eğitim Modeli

$$y = \beta_0 + \beta_1 x + u \quad \longrightarrow \quad wage = \beta_0 + \beta_1 educ + u \qquad \text{(Model)}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \longrightarrow \quad \widehat{wage} = \beta_0 + \beta_1 educ \qquad \text{(ÖRF)}$$

$$\widehat{wage} = -0.9 + 0.54\, educ \qquad \text{(ÖRF)}$$

$$n = 526$$

wage: saat başına ücret (dolar); *educ*: eğitim düzeyi (yıl)

- Kesim parametresi $\hat{\beta}_0 = -0.9$ olarak tahmin edilmiştir.
  - $educ = 0$ olduğunda modelce tahmin edilen ücret $\widehat{wage}$'i ifade eder. Yani çalışanın eğitim düzeyi *educ* sıfır olduğunda, ücreti $-\$0.9$'dur. Ancak ücret negatif olamayacağı için yorumlanması anlamsızdır.
- Eğim parametresi $\hat{\beta}_1 = 0.54$ olarak tahmin edilmiştir.
  - *educ*'u 1 yıl arttırdığımızda ücret *wage* 0.54 birim, yani \$0.54 artar.
  - Başka bir ifadeyle, iki çalışandan birinin *educ* düzeyi diğerinden bir yıl fazlaysa, bu iki çalışan için tahmin edilen ücret farkı ortalama olarak \$0.54'dir.
  - Burada somut iki çalışandan değil ortalama durumdan bahsedilmektedir.

Tahmin Edilen Değer ve Kalıntı

i'inci Gözlem İçin Tahmin Edilen $\hat{y}_i$ Değeri

$$\underbrace{\hat{y}_i}_{\text{Tahmin Edilen Değer}} = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \text{(ÖRF)}$$

- x_i değerini tahmin edilen regresyonda (ÖRF'de) yerine koyarsak tahmin edilen değer $\hat{y}_i$'yi elde ederiz.

Kalıntı (Artık)

$$\underbrace{\hat{u}_i}_{\text{Kalıntı (Artık)}} = \underbrace{y_i}_{\text{Gözlenen Değer}} - \underbrace{\hat{y}_i}_{\text{Tahmin Edilen Değer}}$$

- Gözlenen y_i değeriyle tahmin edilen değer $\hat{y}_i$ arasındaki fark kalıntı $\hat{u}_i$'yi verir.
- $\hat{u}_i > 0$ ise $y_i > \hat{y}_i$, eksik tahmin yapılmıştır.
- $\hat{u}_i < 0$ ise $y_i < \hat{y}_i$, fazla tahmin yapılmıştır.

Tahmin Edilen Değer ve Kalıntıların Cebirsel Özellikleri

1. SEKK kalıntılarının toplamı ve dolayısıyla da örneklem ortalaması sıfıra eşittir.

$$\sum_{i=1}^{n} \hat{u}_i = 0 \quad \text{ve} \quad \bar{\hat{u}} = 0$$

- Bu durum SEKK birinci sıra koşullarından ilkinin (aynı zamanda örneklem moment koşullarından ilkinin) bir sonucudur. Bakınız Slayt 31 ve Slayt 33.
- Anakütledeki hata terimi u'nun örneklemdeki analoğu kalıntı $\hat{u}$ olarak yorumlanabilir fakat kesinlikle aynı şeyler değildir.

$$\underbrace{u}_{\text{Anakütle}} \longrightarrow \underbrace{\hat{u}}_{\text{Örneklem}}$$

$$\underbrace{E(u) = 0}_{\text{Anakütle}} \longrightarrow \underbrace{E(\hat{u}) = 0, \quad \sum_{i=1}^{n} \hat{u}_i = 0 \quad \text{ve} \quad \bar{\hat{u}} = 0}_{\text{Örneklem}}$$

Tahmin Edilen Değer ve Kalıntıların Cebirsel Özellikleri

2. Bağımsız değişken x ile kalıntı $\hat{u}$ arasındaki örneklem kovaryansı ve korelasyon katsayısı sıfırdır.

$$Cov(x, \hat{u}) = 0 \quad \text{ve} \quad Corr(x, \hat{u}) = 0$$

- Bu durum diğer SEKK birinci sıra koşullarının (aynı zamanda diğer örneklem moment koşullarının) bir sonucudur. Bakınız Slayt 31 ve Slayt 33.
- Bağımsız değişken x'le kalıntı $\hat{u}$'nın doğrusal olarak ilişkisizliği çıkarılabilir.

$$Cov(x, u) = 0 \quad \text{ve} \quad Corr(x, u) = 0 \quad \longrightarrow \quad E(xu) = 0 \qquad \text{(Anakütle)}$$

$$Cov(x, \hat{u}) = 0 \quad \text{ve} \quad Corr(x, \hat{u}) = 0 \quad \longrightarrow \quad E(x\hat{u}) = 0 \qquad \text{(Örneklem)}$$

$$\underbrace{E(xu) = 0}_{\text{Anakütle}} \quad \longrightarrow \quad \underbrace{E(x\hat{u}) = 0 \quad \text{ve} \quad \sum_{i=1}^{n} x_i \hat{u}_i = 0}_{\text{Örneklem}}$$

Ek Bilgi

Tahmin Edilen Değer ve Kalıntıların Cebirsel Özellikleri

3. 1. ve 2. cebirsel özelliklerin bir sonucu olarak tahmin edilen değer $\hat{y}$ ile kalıntı $\hat{u}$ arasındaki örneklem kovaryansı ve korelasyon katsayısı sıfırdır.

$$Cov(\hat{y}, \hat{u}) = 0 \quad \text{ve} \quad Corr(\hat{y}, \hat{u}) = 0$$

- Bu özellikten tahmin edilen değer $\hat{y}$ ile kalıntı $\hat{u}$'nun doğrusal olarak ilişkisizliği çıkarılabilir.

$$\underbrace{Cov(\hat{y}, \hat{u}) = 0 \quad \text{ve} \quad Corr(\hat{y}, \hat{u}) = 0}_{\text{Örneklem}} \longrightarrow \underbrace{E(\hat{y}\hat{u}) = 0 \quad \text{ve} \quad \sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0}_{\text{Örneklem}}$$

Ek Bilgi

Tahmin Edilen Değer ve Kalıntıların Cebirsel Özellikleri

4. Tahmin edilen değer $\hat{y}_i$'lerin ortalaması gözlenen değer y_i'lerin ortalamasına eşittir.

$$y_i = \hat{y}_i + \hat{u}_i$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i + \underbrace{\sum_{i=1}^{n} \hat{u}_i}_{=0} \qquad \text{(1. Cebirsel Özellik)}$$

$$n\bar{\hat{y}} = n\bar{y}$$

$$\bar{\hat{y}} = \bar{y}$$

5. $(\bar{x}, \bar{y})$ noktası daima ÖRF'den geçer (ÖRF üzerindedir).

$$(\bar{x}, \bar{y}) \quad \longrightarrow \quad \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Kareler Toplamları

- Her bir i gözlemi için gözlenen değer, tahmin edilen değer ve kalıntı arasındaki ilişki aşağıdaki gibi gösterilebilir.

$$y_i = \hat{y}_i + \hat{u}_i$$

- Her iki tarafın örneklem ortalamalarından sapmalarının karesini alıp toplarsak

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

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left[(\hat{y}_i - \bar{y}) + \hat{u}_i\right]^2 \qquad \text{(1. ve 4. Cebirsel Öz.)}$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}\hat{u}_i^2 + 2\underbrace{\sum_{i=1}^{n}\hat{u}_i\hat{y}_i}_{=0} - 2\bar{y}\underbrace{\sum_{i=1}^{n}\hat{u}_i}_{=0} \qquad \text{(3. Cebirsel Öz.)}$$

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

Kareler Toplamları

- **Toplam Kareler Toplamı**: SST (Total Sum of Squares) y'deki toplam değişkenliği verir.

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

$Var(y) = SST/(n-1)$ olduğuna dikkat edin.

- **Açıklanan Kareler Toplamı**: SSE (Explained Sum of Squares) model tarafından açıklanan kısımdaki, yani tahmin edilen değer $\hat{y}$'lardaki, değişkenliği verir.

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- **Kalıntı Kareleri Toplamı**: SSR (Residual Sum of Squares) model tarafından açıklanamayan kısımdaki, yani kalıntı $\hat{u}$'lardaki, değişkenliği verir.

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

Kareler Toplamları

- y'deki toplam değişkenlik aşağıdaki gibi yazılabilir.

$$SST = SSE + SSR$$

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{SSE}} + \underbrace{\sum_{i=1}^{n}\hat{u}_i^2}_{\text{SSR}}$$

Determinasyon Katsayısı

- y'deki toplam değişkenlik denkleminin her iki tarafını SST'ye bölersek

$$SST = SSE + SSR$$

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

- Açıklanan kısmın değişkenliğinin toplam değişkenlik içindeki payı regresyonun **determinasyon katsayısı** ya da **belirlilik katsayısı**dır (coefficient of determination) ve R^2 ile gösterilir.

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- SSE hiçbir zaman SST'den büyük olamayacağı için $0 \leq R^2 \leq 1$
- R^2, y'deki değişkenliğin x tarafından açıklanan kısmının oranını verir. Regresyonun açıklama gücü yükseldikçe R^2, 1'e yaklaşır.
- R^2'yi yorumlarken, yüzdeye dönüştürmek için genellikle 100 ile çarparız: $100 \times R^2$, y'deki değişkenliğin x tarafından açıklanan kısmının yüzdesini verir.
- R^2 modelin açıklama gücünü (ne kadar iyi fit edildiğini) belirttiği için bazen **uyum iyiliği** (goodness-of-fit) olarak da adlandırılır.
- R^2 şu şekilde de hesaplanabilir: $R^2 = Corr(y, \hat{y})^2$

Determinasyon Katsayısı: Örnek

CEO Maaşı vs. Karlılık Modeli

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \text{(ÖRF)}$$

$$\widehat{salary} = 963.191 + 18.501\, roe \qquad \text{(ÖRF)}$$

$$n = 209, \qquad R^2 = 0.0132$$

salary: CEO maaşı (bin dolar); *roe*: şirketin karlılık yüzdesi

- Determinasyon katsayısı 0.0132 olarak tahmin edilmiştir.
- CEO maaşı *salary*'deki değişkenliğin yaklaşık %1.32'si *roe* değişkeniyle açıklanabilmektedir. Diğer bir deyişle, *salary*'daki değişkenliğin yaklaşık %98.68'i açıklanamamıştır.
- Dışarıda bırakılan birçok faktör (hata terimi u'nun içinde) olduğundan CEO maaşı *salary*'nin küçük bir kısmı açıklanabilmiştir.
- CEO maaşı *salary*'yi etkileyen ve bu modelde yer almayan başka birçok değişken olduğu unutulmamalıdır.

SEKK Parametre Tahmincilerinin Varyansı

BDR.7: Sabit Varyans (Homoscedasticity)

u hata teriminin bağımsız değişken x'e göre koşullu varyansı sabittir.

$$Var(u|x) = \sigma^2$$

$$Var(u|\mathbf{x}) = \sigma^2$$

$$Var(u) = \sigma^2$$

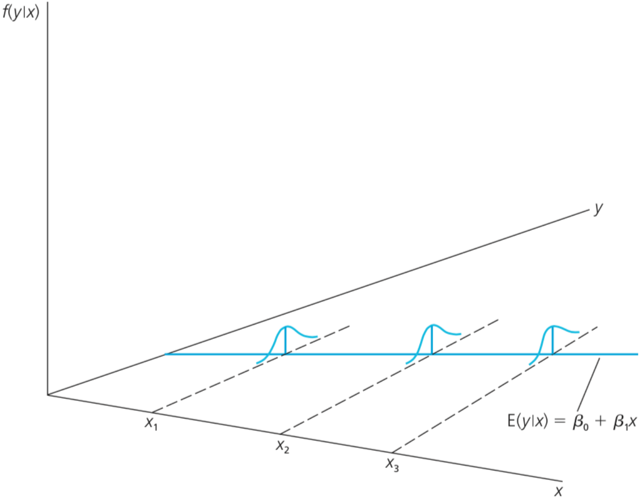
- Bu varsayım SEKK parametre tahmincilerinin varyanslarının ve standart hatalarının türetilmesinde ve etkinlik özelliklerinin belirlenmesinde kullanılır.
  - SEKK parametre tahmincilerinin sapmasızlığı için ÇDR.7 varsayımına ihtiyaç yoktur.
- Örneğin, ücret vs. eğitim modelinde (Slayt 9) bu varsayım, model dışında bırakılan faktörler u'daki değişkenliğin modele dahil edilen $educ$'e bağlı olmadığını söyler.
- BDR.5 ve BDR.7 varsayımları kullanılarak Slayt 20'deki gibi bağımlı değişken y'nin bağımsız değişken x'lere göre koşullu varyansının da sabit olduğu gösterilebilir.

$$Var(y|\mathbf{x}) = \sigma^2$$

Sabit Varyans vs. Değişken Varyans

- **Sabit varyans** (BDR.7) durumunda, $y = \beta_0 + \beta_1 x + u$ modeline ait ARF ve $y|\mathbf{x}$'in dağılımı Şekil 7'de gösterilmiştir (BDR.5 varsayımı sağlanıyorken).
- BDR.7'nin sağlanmadığı duruma **değişen varyans** (heteroscedasticity) denir.
- Değişen varyans durumunda, Slayt 9'de verilen ücret vs. eğitim modeline ait ARF ve $y|\mathbf{x}$'in dağılımı Şekil 8'deki gibi gösterilebilir (BDR.5 varsayımı sağlanıyorken).

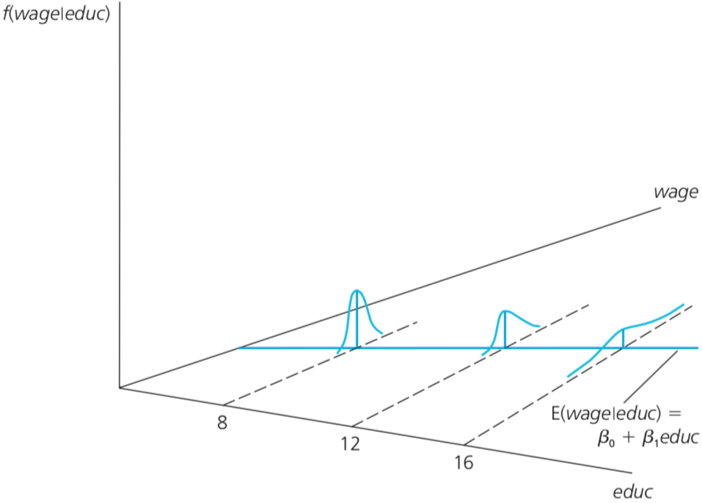
Sabit Varyans



Şekil 7: Sabit Varyans (BDR.7) Durumunda ARF ve $y|\mathbf{x}$'in Dağılımı

Kaynak: Wooldridge (2016)

Değişen Varyans



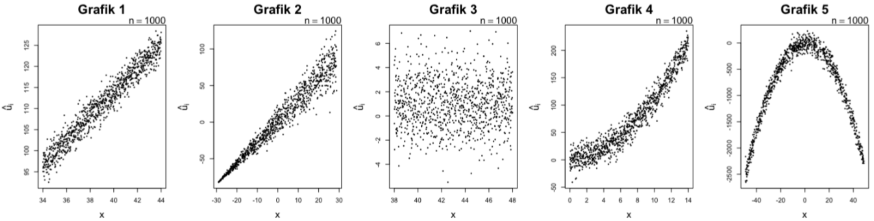
Şekil 8: Değişen Varyans Durumunda ARF ve $y|\mathbf{x}$'in Dağılımı

Kaynak: Wooldridge (2016)

Uygulamada Sabit Varyans vs. Değişken Varyans Belirleme

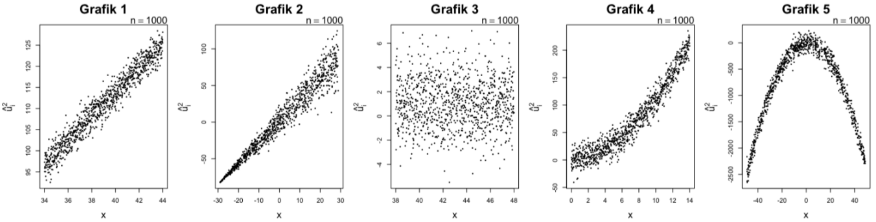
- Uygulamada, hata terimi u gözlenemediği için u'nun x'e göre koşullu varyansı $Var(u|\mathbf{x})$'in sabit olup olmadığını, yani u'daki değişkenliğin $\mathbf{x}$'e göre nasıl değiştiğini, anlamak mümkün olmaz.
- Bunun yerine iki farklı yöntem kullanılabilir.
  - Kalıntı $\hat{u}$ vs. $\mathbf{x}$'in grafiğine bakılır.
  - Kalıntı karesi $\hat{u}^2$ vs. $\mathbf{x}$'in grafiğine bakılır.
- Eğer Şekil 9'deki gibi kalıntı $\hat{u}$ vs. $\mathbf{x}$'in grafiğine bakılırsa, kalıntı $\hat{u}$'daki değişkenliğin bağımsız değişken $\mathbf{x}$'e göre nasıl değiştiği incelenmelidir.
  - Şekil 9'de, sadece Grafik 2'de değişen varyans vardır. Çünkü, $\hat{u}$'daki değişkenlik bağımsız değişken $\mathbf{x}$'e göre değişir.
  - Şekil 9'deki diğer graifklerde sabit varyans vardır.
- Eğer Şekil 10'deki gibi kalıntı karesi $\hat{u}^2$ vs. $\mathbf{x}$'in grafiğine bakılır, kalıntı karesi $\hat{u}^2$'nın bağımsız değişken $\mathbf{x}$'e göre nasıl değiştiği incelenmelidir.
  - Şekil 10'de, sadece Grafik 3'de sabit varyans vardır. Çünkü, $\hat{u}^2$ bağımsız değişken $\mathbf{x}$'e göre değişmez (sabit).
  - Şekil 10'deki diğer graifklerde değişen varyans vardır.

Uygulamada Sabit Varyans vs. Değişken Varyans Belirleme



Şekil 9: Kalıntı $\hat{u}$ vs. $\mathbf{x}$'in Olası Grafikleri

Not: Tüm grafiklerde dikey eksende kalıntılar $\hat{u}_i$ ve yatay eksende ise bağımsız değişken $\mathbf{x}$ vardır.



Şekil 10: Kalıntı Karesi $\hat{u}^2$ vs. $\mathbf{x}$'in Olası Grafikleri

Not: Tüm grafiklerde dikey eksende kalıntılar $\hat{u}_i^2$ ve yatay eksende ise bağımsız değişken $\mathbf{x}$ vardır.

SEKK Parametre Tahmincilerinin Varyansı

Teorem: $\hat{\beta}_1$'nın Varyansı

Gauss–Markov varsayımları (BDR.1 - BDR.7) altında

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{SST_x}, \quad SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Ek Bilgi

- Ekonometrik analizde ana odak $\hat{\beta}_1$ olduğundan, $\hat{\beta}_0$'nın varyansı verilmemiştir.
- σ^2 gözlenemeyen hata terimi u'nun varyansıdır. Bu nedenle σ^2 **hata varyansı**, σ ise **regresyonun standart sapması** olarak adlandırılır.
- SST_x, x'deki örneklem değişkenliğini ifade eder.
- SEKK parametre tahmincilerine ait varyansın olabildiğinde küçük olması istenir, çünkü küçük varyans tahminin hassaslığını arttırır. Bakınız Slayt **??**.
- $Var(\hat{\beta}_j)$, σ^2 ile aynı yönde ilişkilidir. σ^2'yi düşürmenin tek yolu güçlü bağımsız değişkenleri modele eklemektir. Daha büyük bir σ^2, y'yi etkileyen gözlenemeyen hata terimi u'ya ait dağılımın daha fazla yayılmış olduğu anlamına gelir.
- $Var(\hat{\beta}_1)$, SST_x ile ters yönde ilişkilidir. SST_x'yi arttırmanın tek yolu gözlem sayısını arttırmaktır.

SEKK Parametre Tahmincilerinin Varyansı

Teorem: $\hat{\beta}_1$'nın Varyansı

Gauss–Markov varsayımları (BDR.1 - BDR.7) altında

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{SST_x}, \quad SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Hata terimi u gözlenemediği için hata varyansı σ^2 bilinmez.
- Bu nedenle, SEKK parametre tahmincilerinin varyansı $Var(\hat{\beta}_1)$'nın tahmini için öncelikle hata varyansı σ^2'nin tahmin edilmesi gerekir.
- Buradaki önemli nokta, $Var(\hat{\beta}_1)$'nın sapmasız tahmin edilmesi gerekir. Bu nedenle, σ^2'nin de aynı şekilde sapmasız tahmin edilmesi gerekir.

SEKK Parametre Tahmincilerinin Varyansı

Hata Varyansı σ^2

BDR.5 varsayımı altında hata varyansı σ^2 aşağıdaki gibi yazılabilir.

$$Var(u) = \sigma^2 = E(u^2) - \underbrace{E(u)^2}_{=0\ (\text{ÇDR.5})} \qquad \text{(Varyans Formülü)}$$

$$\sigma^2 = E(u^2)$$

- σ^2'nin sapmasız tahmincisi hata terimi u'nun örneklem ortalaması $n^{-1}\sum_{i=1}^{n} u_i^2$'dır.
- Fakat, hata terimi u gözlenemediği için σ^2'nin tahmininde hata terimi u'nun yerine onun örneklem analoğu olan kalıntı $\hat{u}$ kullanılır. $n^{-1}\sum_{i=1}^{n} u_i^2 \longrightarrow n^{-1}\sum_{i=1}^{n} \hat{u}_i^2$
- Fakat $n^{-1}\sum_{i=1}^{n} \hat{u}_i^2$ sapmalı bir tahmincidir. Bu nedenle, σ^2'nin sapmasız tahmincisini hesaplamak bu değerin serbestlik derecesi kullanılarak düzeltilmesi gerekir.

SEKK Parametre Tahmincilerinin Varyansı

Teorem: Hata Varyansı σ^2'nin Sapmasız Tahmini

Gauss–Markov varsayımları (BDR.1 - BDR.7) altında hata varyansı σ^2'nin sapmasız bir tahmincisi:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1} = \frac{SSR}{n-k-1} = \frac{SSR}{n-2} \qquad \text{(BDR'de } k = 2\text{)}$$

- **Serbestik derecesi** (bağımsız bilgi sayısı) $\longrightarrow s.d. = n - (k+1) = n - k - 1$
  - Serbestlik derecesi, SEKK birinci sıra koşullarından ($k + 1$ tane) gelmektedir. Bu koşullar n tane kalıntı $\hat{u}$'nın üzerine $k + 1$ tane kısıt koyar.
  - n tane kalıntıdan $n - (k + 1)$ tanesi biliniyorsa, geriye kalan $k + 1$ kalıntı otomatik olarak bilinecektir. Bu nedenle kalıntıların serbestlik derecesi $n - k - 1$'dir.
- $\hat{\sigma}$ **regresyonun standart sapması** σ'nın bir tahmincisidir ve **regresyonun standart hatası** ya da **ortalama karesel hata** olarak adlandırılır.
- Regresyona yeni bir bağımsız değişken eklendiğinde $\hat{\sigma}$ azalabilir ya da artabilir.
  - Modele yeni bir bağımsız değişken eklendiğinde SSR düşecektir fakat aynı zamanda serbestklik derecesi de 1 düşecektir. SSR payda, serbestlik derecesi ise paydada olduğundan hangi değişimin daha fazla etkiye sahip olduğunu kestiremeyiz.

Kaynaklar

Gujarati, D.N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education.
Stock, J.H. ve M.W. Watson (2015). *Introduction to Econometrics*.
Tastan, H. (2020). *Lecture on Econometrics I. Personal Collection of H. Tastan*. Retrieved from Online.
Wooldridge, J.M. (2016). *Introductory Econometrics: A Modern Approach*. Nelson Education.

Ek Bilgiler

BDR.5: Sıfır Koşullu Ortalama

$$E(u|\mathbf{x}) = 0$$

- Daha önce gördüğümüz Yinelenen Beklentiler Kanunu'nu hatırlayalım.

Yinelenen Beklentiler Kanunu

$$E[E(u|\mathbf{x})] = E(u)$$

Ek Bilgiler

- Yinelenen Beklentiler Kanunu kullanılarak BDR.5 varsayımı yeniden tanımlanabilir.

$$
\begin{aligned}
E[\underbrace{E(u|\mathbf{x})}_{=0}] &= E(u) \\
E[0] &= E(u) \\
0 &= E(u)
\end{aligned}
$$

BDR.5: Sıfır Koşullu Ortalama

$$E(u|\mathbf{x}) = E(u) = 0$$

Yani, hata terimi u'nun bağımsız değişken x'e göre koşullu ve koşulsuz ortalaması sıfırdır.

Ek Bilgiler

- Koşullu beklenen değerin 5. özelliğini kullanarak u ve x arasındaki ilişki hakkında daha fazla yorumda bulunabiliriz.

Koşullu Beklenen Değer: Özellik 5

$$E(u|\mathbf{x}) = E(u) \quad \text{ise} \quad Cov(x,u) = 0 \quad \text{ve} \quad Corr(x,u) = 0$$

Yani, bağımsız değişken x'in her doğrusal fonksiyonu hata terimi u ile ilişkisizdir.

BDR.5: Sıfır Koşullu Ortalama

$$E(u|\mathbf{x}) = E(u) = 0$$

$$Cov(x,u) = 0, \quad Corr(x,u) = 0 \quad \text{ve} \quad E(xu) = 0$$

Sonuç: u ve x ortalama bağımsızdır. Yani u ve x doğrusal olarak ilişkisizdir.

Sunuma Geri Dön

Ek Bilgiler

BDR.5: Sıfır Koşullu Ortalama Varsayımı

$$E(u|\mathbf{x}) = E(u) = 0$$

$$Cov(x,u) = 0, \quad Corr(x,u) = 0 \quad \text{ve} \quad E(xu) = 0$$

$$Cov(x,u) = E(xu) - E(x)\underbrace{E(u)}_{=0} = 0$$

$$= E(xu) = 0$$

Sunuma Geri Dön

Ek Bilgiler

BDR.6: Otokorelasyon Olmaması

$$Cov(u_i, u_s|\mathbf{x}) = 0 \quad \text{ve} \quad Cov(u_i, u_s) = 0, \quad i \neq s$$

$$E(u_i u_s|\mathbf{x}) = 0 \quad \text{ve} \quad E(u_i u_s) = 0, \quad i \neq s$$

$$\begin{aligned} Cov(u_i, u_s|\mathbf{x}) &= E(u_i u_s|\mathbf{x}) - \underbrace{E(u_i|\mathbf{x})}_{=0}\underbrace{E(u_s|\mathbf{x})}_{=0} = 0 \\ &= E(u_i u_s|\mathbf{x}) = 0 \end{aligned}$$

◂ Sunuma Geri Dön

Ek Bilgiler

BDR.7: Sabit Varyans (Homoscedasticity)

$$E(u^2|\mathbf{x}) = \sigma^2 \quad \text{ve} \quad E(u^2) = \sigma^2$$

$$Var(u|\mathbf{x}) = E(u^2|\mathbf{x}) - \underbrace{E(u|\mathbf{x})^2}_{=0} = \sigma^2$$

$$= E(u^2|\mathbf{x}) = \sigma^2$$

Sunuma Geri Dön

Ek Bilgiler

Anakütle Regresyon Fonksiyonu (ARF)

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x \tag{ARF}$$

$$Var(y|\mathbf{x}) = \sigma^2$$

$$y = \beta_0 + \beta_1 x + u$$

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x + \underbrace{E(u|\mathbf{x})}_{=0}$$

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x \tag{ARF}$$

$$y = \beta_0 + \beta_1 x + u$$

$$Var(y|\mathbf{x}) = Var(u|\mathbf{x})$$

$$Var(y|\mathbf{x}) = \sigma^2$$

Sunuma Geri Dön

Ek Bilgiler

Parametre Tahmincileri

β_0 kesim parametresinin tahmini $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$$

- $\hat{\beta}_0$'nın formülü
  - SEKK birinci sıra koşulları ya da örneklem moment koşullarından ilki (Slayt 33)
  - Kalıntı $\hat{u}$'nın denklemi
  - İndeksli haldeki model denklemi

kullanılarak çıkarılabilir.

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i &= \sum_{i=1}^{n} (y_i - \hat{y}_i) = 0 \\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i = 0 \\
&= n\bar{y} - n\hat{\beta}_0 - \hat{\beta}_1 n\bar{x} = 0 \\
&= \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0
\end{aligned}$$

Sonuç: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Sunuma Geri Dön

Ek Bilgiler

Parametre Tahmincileri

β_1 eğim parametresinin tahmini $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- $\hat{\beta}_1$'nın formülü
  - SEKK birinci sıra koşulları ya da örneklem moment koşullarından ikincisi (Slayt 33)
  - Kalıntı $\hat{u}$'nın denklemi
  - β_0 kesim parametresinin tahmini $\hat{\beta}_0$
  - Ortalamadan sapmaların kareleri toplamı

kullanılarak çıkarılabilir.

$$\begin{aligned}\sum_{i=1}^{n} x_i \hat{u}_i &= \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ &= \sum_{i=1}^{n} x_i (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) = 0\end{aligned}$$

Ek Bilgiler

$$\sum_{i=1}^{n} x_i(y_i - (\bar{y} - \hat{\beta}_1\bar{x}) - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} + \sum_{i=1}^{n} \hat{\beta}_1 x_i\bar{x} - \sum_{i=1}^{n} \hat{\beta}_1 x_i x_i = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} x_i(x_i - \bar{x})$$

Sonuç: $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Sunuma Geri Dön

burada

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Ek Bilgiler

Parametre Tahmincileri

β_1 eğim parametresinin tahmini $\hat{\beta}_1$'in alternatif formülü:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Sunuma Geri Dön

burada

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

Ek Bilgiler

Tahmin Edilen Değer ve Kalıntıların Cebirsel Özellikleri - 2

$$Cov(x, \hat{u}) = E(x\hat{u}) - E(x)\underbrace{E(\hat{u})}_{=0} = 0 \qquad \text{(1. Cebirsel Özellik)}$$

$$= E(x_j\hat{u}) = 0$$

ya da

$$Cov(x, \hat{u}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{u}_i - \bar{\hat{u}})}{n-1} = 0$$

$$Cov(x, \hat{u}) = \sum_{i=1}^{n} x_i(\hat{u}_i - \underbrace{\bar{\hat{u}}}_{=0}) = 0 \qquad \text{(1. Cebirsel Özellik)}$$

$$= \sum_{i=1}^{n} x_i\hat{u}_i = 0$$

Sunuma Geri Dön

Ek Bilgiler

Tahmin Edilen Değer ve Kalıntıların Cebirsel Özellikleri - 3

$$
\begin{aligned}
Cov(\hat{y},\hat{u}) &= Cov(\hat{\beta}_0+\hat{\beta}_1 x,\hat{u}) \\
&= \hat{\beta}_1 \underbrace{Cov(x,\hat{u})}_{=0} = 0 && \text{(2. Cebirsel Özellik)} \\
&= E(\hat{y}\hat{u}) = 0 && \text{(Kovaryans formülü ve 1. Cebirsel Özellik)}
\end{aligned}
$$

ve

$$
\begin{aligned}
Cov(\hat{y},\hat{u}) &= \sum_{i=1}^{n}(\hat{y}_i-\bar{y})(\hat{u}_i-\underbrace{\bar{\hat{u}}}_{=0}) = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})\hat{u}_i = 0 && \text{(1. Cebirsel Özellik)} \\
&= \sum_{i=1}^{n}\hat{y}_i\hat{u}_i - \bar{y}\underbrace{\sum_{i=1}^{n}\hat{u}_i}_{=0} = 0 && \text{(1. Cebirsel Özellik)} \\
&= \sum_{i=1}^{n}\hat{y}_i\hat{u}_i = 0
\end{aligned}
$$

Sunuma Geri Dön