1

BASİT REGRESYON MODELİ

Hüseyin Taştan[1]

[1]Yıldız Teknik Üniversitesi
İktisat Bölümü

Ders Kitabı:
Introductory Econometrics: A Modern Approach (2nd ed.)
J. Wooldridge

14 Ekim 2012

2

Basit Regresyon Modeli

Basit Regresyon Modeli

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

- y: bağımlı değişken
- x: açıklayıcı değişken
- İki değişkenli regresyon modeli (bivariate regression model), tek açıklayıcı değişkenli model, basit regresyon modeli (simple regression model)
- Amaç: bağımlı değişken y'yi, bağımsız değişken x ile açıklamak.
- Regression teriminin orijini: Galton'un ortalamaya dönüş (regress) yasası

3

Basit Regresyon Modeli

Değişkenlere birbirinin yerine kullanabilen çeşitli isimler verilmiştir:

Terminoloji

y	x
Bağımlı değişken	**Bağımsız değişken**
(Dependent variable)	(Independent variable)
Açıklanan değişken	**Açıklayıcı değişken**
(Explained variable)	(Explanatory variable)
Tepki değişkeni	**Kontrol değişkeni**
(Response variable)	(Control variable)
Tahmin edilen değişken	**Tahmin eden değişken**
(Predicted variable)	(Predictor variable)
Regressand	**Regressor**

4

Basit Regresyon Modeli: $y = \beta_0 + \beta_1 x + u$

Rassal Hata Terimi u (Error term - Disturbance term)

Bağımlı değişken y üzerinde etkili olan x'in dışındaki diğer faktörleri temsil eder. Bu diğer etkenlere gözlenemeyen (unobserved) faktörler denir.

Eğim Katsayısı β_1

- Eğer u'da yer alan diğer faktörler sabit tutulursa, yani $\Delta u = 0$ kabul edilirse x'in y üzerindeki etkisi aşağıdaki gibi bulunabilir:

$$\Delta y = \beta_1 \Delta x$$

- β_1: Diğer faktörler sabitken (ceteris paribus) x'deki bir birim değişmenin y'de yaratacağı değişimi gösterir.

Sabit terim (intercept) β_0

$x = 0$ iken y'nin alacağı değeri gösterir.

Basit Regresyon Modeli Örnekler

Tarımsal üretim ve gübre kullanımı

$$yield = \beta_0 + \beta_1 fertilizer + u$$

yield: soya fasülyesi üretim miktarı
fertilizer: gübre miktarı

Eğim Katsayısı β_1

$$\Delta yield = \beta_1 \Delta fertilizer$$

Ceteris Paribus gübre miktarındaki 1 birimlik değişim, çıktı düzeyini β_1 kadar değiştirir.

Rassal hata terimi u

Soya fasülyesi çıktısını etkileyen, gübre dışındaki, yağmur miktarı, toprağın kalitesi gibi tüm faktörlerin ortak etkisini gösterir.
Ceteris Paribus $\Leftrightarrow$ Holding all other factors fixed $\Leftrightarrow$ $\Delta u = 0$

Basit Regresyon Modeli Örnekler

Basit Ücret Denklemi

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

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

Eğim Katsayısı β_1

$$\Delta wage = \beta_1 \Delta educ$$

Ceteris Paribus eğitim düzeyindeki 1 yıllık değişim, saat başına ücretleri β_1 \$ kadar değiştirir.

Rassal hata terimi u

Ücretleri etkileyen eğitim dışındaki diğer faktörler?
İşgücü piyasasındaki tecrübe, doğuştan gelen yetenek, şu an çalışılan yerdeki kıdem, iş etiği, alınan eğitimin kalitesi, çalışanın cinsiyeti, etnik kökeni, kır ya da kente yaşaması, medeni hali, çocuk sayısı, dış görünüşü vs. gibi çok sayıda faktör ücretleri etkileyebilir.

Doğrusallık (Linearity)

- Basit regresyonun doğrusal (linear) olması şu anlama gelmektedir: x'deki 1 birimlik değişmenin y'de meydana getireceği etki, x'in başlangıç (initial) değeri ne olursa olsun, aynıdır (sabittir).
- Bu sabit etki varsayımı uygulamada çoğu zaman gerçeklere uymaz.
- Örneğin, ölçeğe göre artan ya da azalan getiri doğrusal modelle açıklanamaz.
- Ücret denkleminde, ilave bir yıl eğitimin etkisi önceki eğitim düzey(ler)ine göre daha fazla olacaktır.
- Benzer şekilde tecrübe düzeyi ücretler üzerinde önce artan sonra azalan bir etkiye sahip olabilir.
- Bu ve buna benzer etkilerin kolayca nasıl modelleneceğini daha sonra göreceğiz.

Ceteris Paribus çıkarımlarının yapılabilmesi için gerekli varsayımlar

1. Hata teriminin anakütle ortalamasının sıfır olması

- Basit regresyonda sabit terim (β_0) mevcut olduğu sürece şu varsayımı yapabiliriz:
$$\mathrm{E}(u) = 0$$
- Bu, u'nun içerdiği gözlenemeyen (unobservables) etkilerin dağılımıyla ilgili bir varsayımdır. u'ların bir kısmı $+$, bir kısmı $-$ işaretlidir ve bunlar birbirlerini götürürler diye varsayıyoruz.
- β_0'ı yeniden tanımlayarak bu varsayımın gerçekleşmesini her zaman sağlayabiliriz.

Ceteris Paribus çıkarımlarının yapılabilmesi için gerekli varsayımlar

2. Hata terimi u ile açıklayıcı değişken x'in ilişkisiz olması koşulu

- x'in y üzerindeki etkisini görebilmemiz için u'da içerilen gözlenemeyen faktörlerin aynı -sabit- kalmaları (ceteris paribus koşulu) gerekli idi. Bundan nasıl emin olabiliriz?
- Bunun için x ile u'nun ilişkisiz olması gereklidir. Ancak, u hem x ile, hem de x'in herhangi bir fonksiyonu (örneğin x^2, $\sqrt{x}$ vs.) ile ilişkisiz olmalı.
- Bunun için *Sıfır Koşullu Ortalama (Zero Conditional Mean)* varsayımını yapmamız gerekmektedir:

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

Sıfır Koşullu Ortalama Varsayımı

Hata terimi u ile açıklayıcı değişken x'in ilişkisiz olması koşulu

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

- Burada u ve x tesadüfi değişkenlerdir (random variables). Bu nedenle, x'in verilen herhangi bir değeri için u'nun koşullu dağılımını tanımlayabiliriz.
- x'in verilen belli bir değerine anakütlenin (population) belli bir dilimi (kısmı) karşılık gelir. Buradan hareketle u'nun bu populasyon dilimi içindeki beklenen değerini (ortalamasını) alabiliriz.
- Burada kritik varsayım, u'nun ortalama değerinin x'in alacağı değere bağlı olmamasıdır.
- x'in verilen herhangi bir değeri için gözlenemeyen faktörlerin ortalaması aynıdır ve dolayısıyla da u'nun tüm anakütledeki ortalama değerine (sıfıra) eşittir.

Sıfır Koşullu Ortalama Varsayımı: Örnek

Ücret Denklemi:

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

- Bu regresyonda u çalışanların doğuştan gelen yeteneğini (innate ability) içersin. Bunu abil ile gösterelim.
- $\mathsf{E}(u|x)$ varsayımı tüm eğitim seviyelerinde doğuştan gelen yeteneğin aynı olduğunu söyler:

$$\mathsf{E}(abil|educ = 8) = \mathsf{E}(abil|educ = 12) = \ldots = 0$$

- Eğer eğitimle doğuştan yeteneğin ilişkili olduğunu düşünüyorsak (daha yetenekliler okulda da daha iyiler), bu halde varsayım sağlanamaz.
- Doğuştan yeteneği gözlemleyemediğimiz için de ortalama doğuştan yeteneğin tüm eğitim seviyelerinde aynı olup olmadığını bilemeyiz.

Sıfır Koşullu Ortalama Varsayımı: Örnek

Tarımsal üretim-gübre kullanımı deneyi:

- Soya fasülyesi gübre deneyini hatırlayalım. Arazi eşit parçalara bölünüp rassal olarak herbirine farklı miktarlarda gübre uygulanıyordu.
- Eğer gübre miktarları bu toprak parçalarının özelliklerinden ilişkisiz olarak belirleniyorsa sıfır koşullu ortalama varsayımı sağlanır.
- Ancak, yüksek kaliteli toprak parçalarına yüksek miktarda gübre uygulanırsa hata teriminin beklenen değeri gübre miktarı ile birlikte artar.
- Bu durumda sıfır koşullu ortalama varsayımı sağlanamaz.

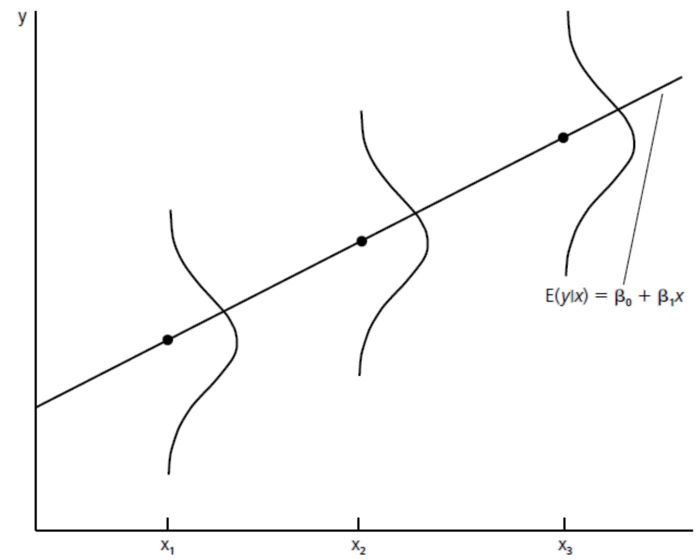
Popülasyon (Anakütle) Regresyon Fonksiyonu (PRF)

- Basit regresyon modelinde y'nin x'e göre koşullu beklenen değerini alalım:

$$
\begin{aligned}
\mathsf{E}(y|x) &= \beta_0 + \beta_1 x + \underbrace{\mathsf{E}(u|x)}_{=0} \\
&= \beta_0 + \beta_1 x
\end{aligned}
$$

- Buna PRF adı verilir. Açıktır ki, bağımlı değişkenin koşullu beklenen değeri x'in doğrusal bir fonksiyonudur.
- PRF'nin doğrusallığı: x'deki 1 birimlik değişime karşılık y'nin koşullu beklenen değeri (koşullu ortalaması) β_1 kadar değişmektedir.
- Verilmiş bir x düzeyinde y'nin dağılımının merkezi $\mathsf{E}(y|x)$'dir.

Popülasyon Regresyon Fonksiyonu (PRF)



Bağımlı değişkenin sistematik ve sistematik olmayan kısımları

- Basit regresyon modelinde

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

$\mathsf{E}(u|x) = 0$ varsayımı geçerliyken bağımlı değişken y'yi iki kısıma ayırmak mümkündür:

- Sistematik kısım: $\beta_0 + \beta_1 x$. Bu y'nin x tarafından açıklanan kısmıdır.
- Sistematik olmayan kısım: u. Bu ise y'nin x tarafından açıklanamayan kısmıdır.

Sıradan En Küçük Kareler Tahmin Edicileri

- Bilinmeyen popülasyon parametreleri (β_0, β_1) verileden hareketle nasıl tahmin edilebilir?
- Popülasyondan n gözlemli (hacimli) bir rassal örneklem (random sample) çektiğimizi düşünelim:

$$\{y_i, x_i : i = 1, 2, \dots, n\}$$

- Regresyon modelini herbir gözlem için aşağıdaki gibi yazabiliriz:

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \dots, n$$

- Bu durumda elimizde iki bilinmeyenli n denklem olacaktır.

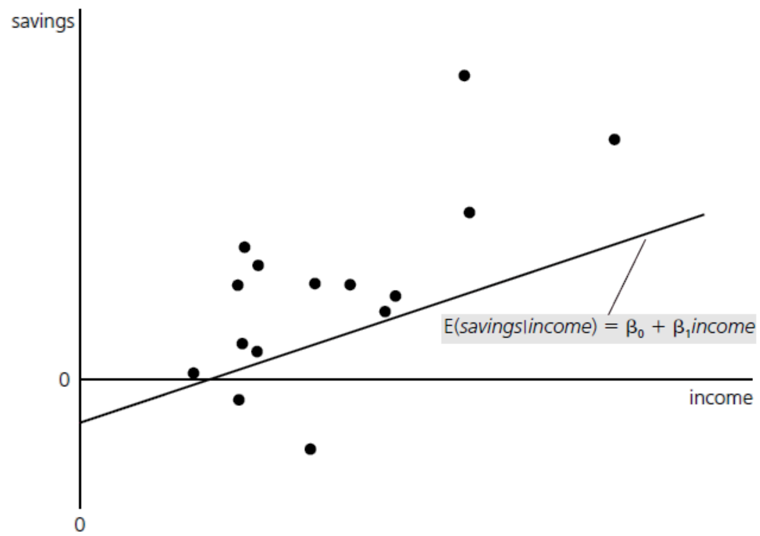
Anakütle Katsayılarının (β_0, β_1) Tahmini

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \ldots, n$$

İki bilinmeyenli n denklem:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + u_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + u_2 \\ y_3 &= \beta_0 + \beta_1 x_3 + u_3 \\ \vdots &= \vdots \\ y_n &= \beta_0 + \beta_1 x_n + u_n \end{aligned}$$

Rassal Örneklem Örnek: Tasarruflar ve Gelir Serpilme Çizimi



Anakütle Katsayılarının (β_0, β_1) Tahmini: Momentler Yöntemi

Ceteris Paribus çıkarımlarının yapılabilmesi için gerekli iki varsayım yapmıştık:

$$\begin{aligned} \mathsf{E}(u) &= 0 \\ \mathsf{Cov}(x, u) &= \mathsf{E}(xu) = 0 \end{aligned}$$

NOT: $\mathsf{E}(u|x) = 0$ ise tanım gereği $\mathsf{Cov}(x, u) = 0$ olur, ancak tersi doğru olmayabilir. $\mathsf{E}(u) = 0$ olduğundan yine tanım gereği $\mathsf{Cov}(x, u) = \mathsf{E}(xu)$. Bu varsayımları ve $u = y - \beta_0 - \beta_1 x$ bilgisini kullanarak **Popülasyon Moment Koşullarını** oluşturabiliriz:

$$\begin{aligned} \mathsf{E}(y - \beta_0 - \beta_1 x) &= 0 \\ \mathsf{E}[x(y - \beta_0 - \beta_1 x)] &= 0 \end{aligned}$$

2 denklem, 2 bilinmeyen

Momentler Yöntemi: Örneklem Moment Koşulları

Popülasyon moment koşulları

$$\begin{aligned} \mathsf{E}(y - \beta_0 - \beta_1 x) &= 0 \\ \mathsf{E}[x(y - \beta_0 - \beta_1 x)] &= 0 \end{aligned}$$

yerine bunların örneklem (sample) analoglarını yazarsak:

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

Örneklem verilerini kullanarak ortaya çıkan bu sistemi $\hat{\beta}_0$ ve $\hat{\beta}_1$ için çözebiliriz. Burada parametrelerin üzerinde şapka olduğuna dikkat edilmelidir. Artık elimizde bilinmeyen sabit sayılar değil, çekilen örnekleme göre değişen rassal değişkenler vardır.

Momentler Yöntemi: Örneklem Moment Koşulları

Toplam operatörünün özelliklerini kullanarak birinci örneklem koşulundan hareketle

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

yazılabilir. Burada $\bar{y}$ ve $\bar{x}$ örneklem ortalamalarıdır.
Buradan,

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

yazılabilir. Bunu ikinci örneklem moment koşulunda yerine koyup çözersek eğim katsayısı tahmincisini bulabiliriz.

Momentler Yöntemi

2. moment koşulunu n ile çarptıktan sonra $\hat{\beta}_0$'yı yerine koyarsak:

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

olur. Bu ifadeyi yeniden düzenleyerek

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

yazabiliriz.

Eğim Katsayısının Tahmincisi

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

Bunu elde ederken aşağıdaki özellikleri kullandık:

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

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

Eğim Katsayısının Tahmincisi

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

1. Eğim katsayısı tahmincisi x ile y'nin örneklem kovaryansının, x'in örneklem varyansına oranıdır.
2. $\hat{\beta}_1$'nın işareti örneklem kovaryansının işaretine bağlıdır. Eğer örneklemde x ile y aynı yönde ilişkiliyse $\hat{\beta}_1$ pozitif işaretli, ters yönde ilişkiliyse negatif işaretli olacaktır.
3. $\hat{\beta}_1$'nın hesaplanabilmesi için x'de yeterli değişkenlik olmalıdır:

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

   Eğer örneklemde tüm x'ler aynı değerleri alıyorsa örneklem varyansı 0 olur. Bu durumda $\hat{\beta}_1$ tanımsız olur. Örneğin, ücret denkleminde tüm çalışanlar 12 yıllık eğitim düzeyine sahipse eğitimin ücretler üzerindeki etkisi hesaplanamaz.

Sıradan En Küçük Kareler (OLS - Ordinary Least Squares)

y'nin modelce tahmin edilen değerleri (fitted values):

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Kalıntı (residual) terimleri gözlenen ile tahmin edilen y değerleri arasındaki farktır:

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \end{aligned}$$

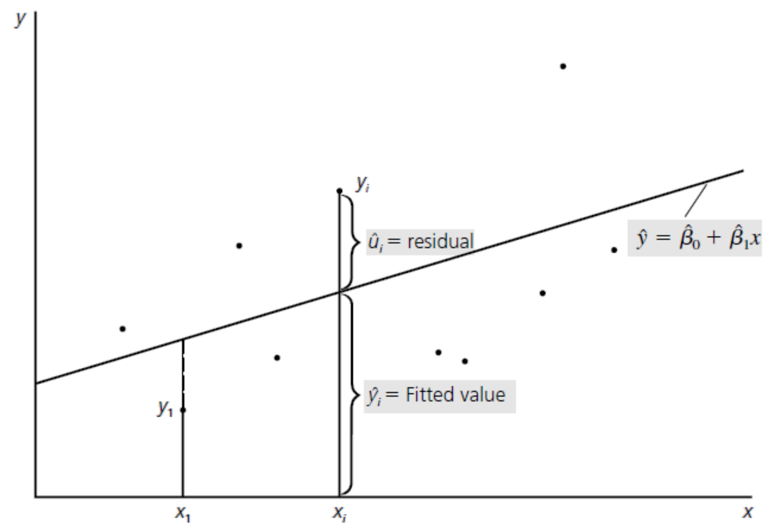
Kalıntı terimi, hata terimi ile karıştırılmamalıdır. u gözlenemeyen bir rassal değişkendir. $\hat{u}$ ise modelce tahmin edilen bir büyüklüktür.

OLS-SEKK amaç fonksiyonu

OLS yöntemi kalıntı kareleri toplamını en küçük yapacak şekilde tahmincileri seçer:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2$$

Kalıntılar



Sıradan En Küçük Kareler (OLS - Ordinary Least Squares)

OLS-SEKK amaç fonksiyonu

OLS yöntemi kalıntı kareleri toplamını (SSR) en küçük yapacak şekilde tahmincileri seçer:

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

OLS birinci sıra koşulları

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

Bu denklem sisteminin çözümü Momentler Yöntemi çözümü ile

Popülasyon ve Örneklem Regresyon Fonksiyonları

Popülasyon Regresyon Fonksiyonu - PRF

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

PRF tektir ve bilinmez.

Örneklem Regresyon Fonksiyonu (Sample Regression Function - SRF)

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

SRF, PRF'nin bir tahmini olarak düşünülebilir. Eğim katsayısının yorumu:

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

ya da

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

Basit Regresyon Örnek: Yönetici Maaşları ve Firma Performansı

- Firma performansı ile yöneticilerin kazançları arasındaki ilişkiyi ölçmek istiyoruz. Model:

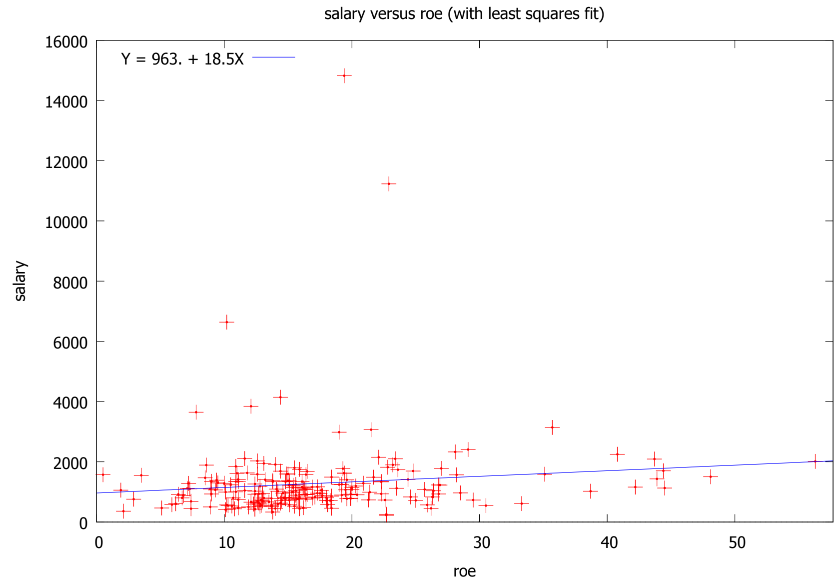
$$salary = \beta_0 + \beta_1 roe + u$$

- salary: yıllık yönetici maaşları (1000 US$), roe: son üç yıla ait sermaye karlılık oranı (return on equity), %
- GRETL ceosal1.gdt dosyasında yer alan $n = 209$ firma verisinden hareketle ÖRF aşağıdaki gibi tahmin ediliyor:

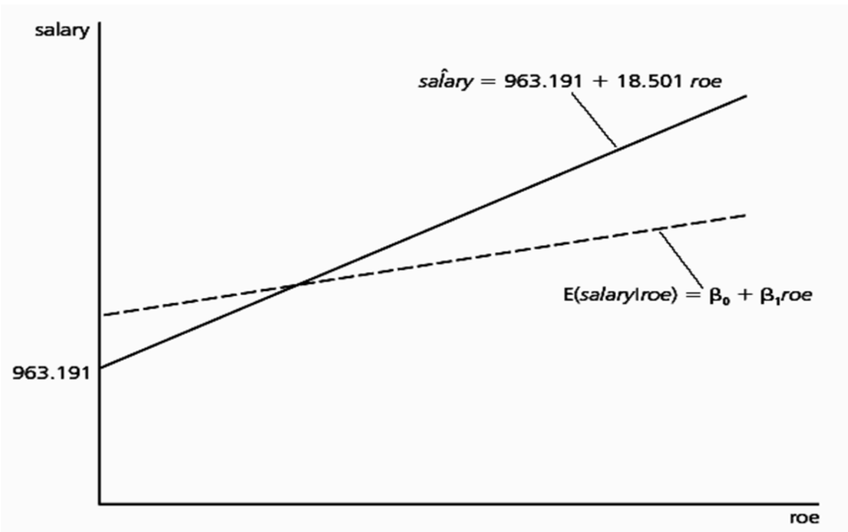
$$\widehat{salary} = 963.191 + 18.501 roe$$

- $\hat{\beta}_1 = 18.501$. Yorum: Sermaye karlılık oranındaki %1 puanlık artış ortalama yönetici maaşlarını yaklaşık 18.501 birim, yani 18,501 US$ arttırır (ceteris paribus).

Yönetici Maaşları - ÖRF



Yönetici Maaşları - ÖRF



Yönetici Maaşları - Tahmin edilen değerler, kalıntılar

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

33

OLS (SEKK) Tahmin Edicilerinin Cebirsel Özellikleri

- OLS kalıntılarının toplamı ve dolayısıyla da örnek ortalaması sıfıra eşittir:
$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \bar{\hat{u}} = 0$$
Bu birinci örneklem moment koşulunun sonucudur.
- Açıklayıcı değişken x ile kalıntı terimleri arasındaki örneklem kovaryansı sıfırdır:
$$\sum_{i=1}^{n} x_i \hat{u}_i = 0$$
Bu da ikinci örneklem moment koşulunun sonucudur.
- $(\bar{x}, \bar{y})$ noktası daima OLS regresyon doğrusu üzerine düşer.
- Tahmin edilen y değerlerinin ortalaması, gözlenen y değerlerinin ortalamasına eşittir: $\bar{y} = \bar{\hat{y}}$

34

Kareler Toplamları (Sum of Squares)

- Her bir i gözlemi için
$$y_i = \hat{y}_i + \hat{u}_i$$
Her iki tarafın örneklem ortalamalarından farklarının karesini alıp toplarsak aşağıdaki büyüklükleri elde ederiz:
- Toplam Kareler Toplamı (SST: Total Sum of Squares)
$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$
- Açıklanan Kareler Toplamı (SSE: Explained Sum of Squares)
$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$
- Kalıntı Kareleri Toplamı (SSR: Residual Sum of Squares)
$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

35

Kareler Toplamları (Sum of Squares)

- SST y'deki toplam değişkenliği verir.
$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$
Var$(y) = SST/(n-1)$ olduğuna dikkat ediniz.
- Benzer şekilde SSE modelce açıklanan kısımdaki değişkenliği verir.
$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$
- SSR ise kalıntılardaki değişkenliğin bir ölçütüdür.
$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$
- y'deki toplam değişkenlik aşağıdaki gibi yazılabilir:
$$SST = SSE + SSR$$

36

Uyum İyiliği (Goodness-of-fit)

- Tanım gereği y'deki toplam değişkenlik aşağıdaki gibi yazılabilir:
$$SST = SSE + SSR$$
- Bu ifadenin her iki tarafını SST'ye bölersek:
$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$
- Açıklanan kısmın değişkenliğinin toplam değişkenlik içindeki payı regresyonun determinasyon (belirlilik) katsayısıdır ve R^2 ile gösterilir:
$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$
- SSE hiç bir zaman SST'den büyük olamayacağı için $0 \leq R^2 \leq 1$
- R^2 y'deki değişkenliğin x tarafından açıklanan kısmının yüzdesini verir. Regresyonun açıklama gücü yükseldikçe R^2 1'e yaklaşır.
- R^2 şu şekilde de hesaplanabilir: $R^2 = Corr(y, \hat{y})^2$

Regresyona Doğrusal Olmayan (Nonlinear) Biçim Ekleme

- Doğrusal (linear) biçim çoğu kez gerçeği yansıtmaz. Bu nedenle doğrusal-olmayan modellerle çalışmak gerekebilir.
- Bağımlı ve bağımsız değişkeni uygun biçimde tanımlayarak basit regresyonu (değişkenlerde) doğrusal-olmayan bir modele dönüştürebiliriz.
- Regresyon modeli hala parametrelerde doğrusaldır. Parametrelerin doğrusal olmayan dönüştürmeleri kullanılmaz.
- Sıkça uygulanan bir yöntem değişkenlerin doğal logaritma dönüştürmesinin kullanılmasıdır. ($\log(y) = \ln(y)$).
- Değişkenlerin başka doğrusal olmayan dönüştürmeleri de kullanılabilir, e.g., karesel kalıp, kübik kalıp, ters kalıp, vs.

Parametrelerde Doğrusallık

- Bir regresyon modelinin doğrusallığını x ve y'nin değil parametrelerin (β)'ların doğrusallığı belirler.
- x ve y'nin $\log x$, $\log y$, x^2, $\sqrt{x}$, $1/x$, $y^{1/4}$ gibi dönüştürmelerini regresyonda kullanabiliriz. Modelimiz hala doğrusaldır.
- Ancak β'ların doğrusal olmayan dönüştürmelerinin yer aldığı modeller doğrusal değildir ve OLS çerçevesinde incelenemez.
- Örneğin aşağıdaki modeller parametrelerde doğrusal değildir:

$$tuketim = \frac{1}{\beta_0 + \beta_1 gelir} + u$$

$$y = \beta_0 + \beta_1^2 x + u$$

$$y = \beta_0 + e^{\beta_1 x} + u$$

Doğal Logaritma Dönüştürmesi İçeren Fonksiyon Kalıpları

Log-Level

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

$$\begin{aligned} \Delta \log y &= \beta_1 \Delta x \\ \%\Delta y &= (100\beta_1)\Delta x \end{aligned}$$

Yorum: x'de meydana gelen bir birim değişime karşılık y, $\%(100\beta_1)$ kadar değişir. Not: $100\Delta \log y = \%\Delta y$

Level-Log

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

$$\begin{aligned} \Delta y &= \beta_1 \Delta \log x \\ &= \left(\frac{\beta_1}{100}\right) \underbrace{100\Delta \log x}_{\%\Delta x} \end{aligned}$$

Yorum: x'de meydana gelen %1 değişime karşılık y, $(\beta_1/100)$ birim kadar değişir (kendi birimi cinsinden).

Doğal Logaritma Dönüştürmesi İçeren Fonksiyon Kalıpları

Log-Log (Sabit Esneklik Modeli)

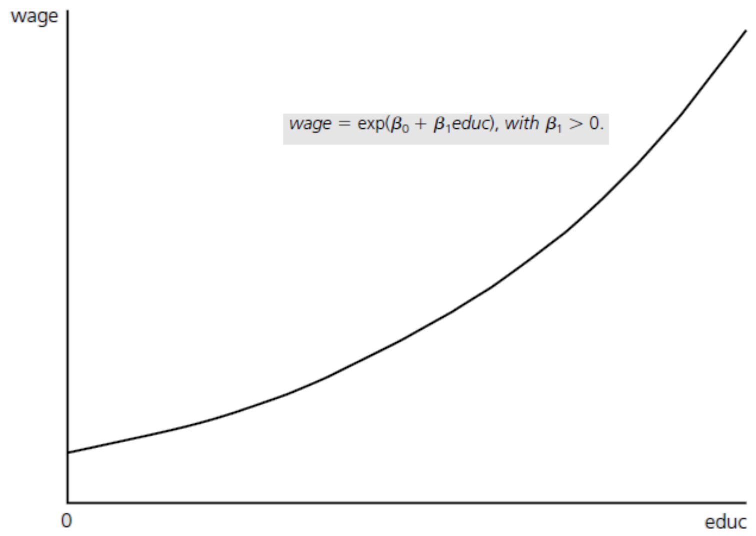
$$\log y = \beta_0 + \beta_1 \log x + u$$

$$\begin{aligned} \Delta \log y &= \beta_1 \Delta \log x \\ \%\Delta y &= \beta_1 \%\Delta x \end{aligned}$$

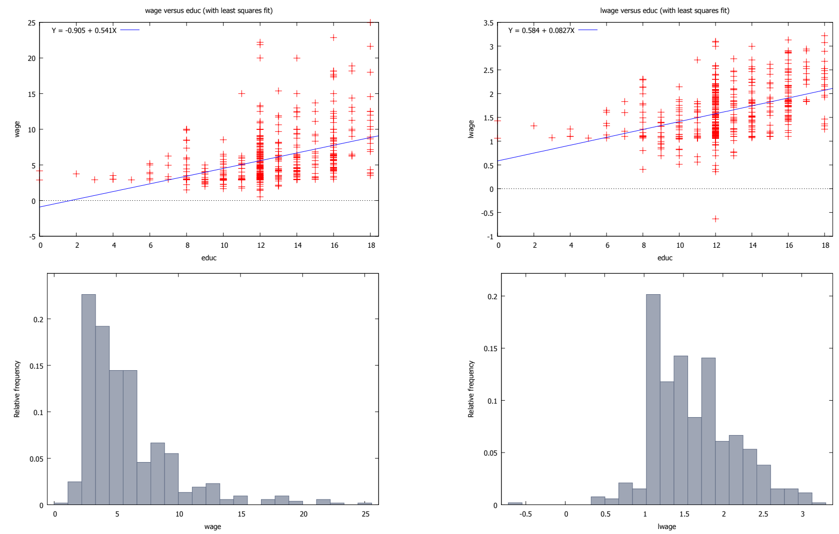
Yorum: Burada β_1 y'nin x'e göre esnekliğini verir. x'de meydana gelen %1 değişime karşılık y'de meydana gelen değişim $\%\beta_1$ olur.

$$\frac{\%\Delta y}{\%\Delta x} = \beta_1$$

Örnek: Ücret-Eğitim ilişkisi, $\log(wage) = \beta_0 + \beta_1 educ + u$



Ücret-Eğitim İlişkisi



Log-Level Basit Ücret Denklemi

$$\widehat{\text{logwage}} = \underset{(0.097)}{0.584} + \underset{(0.008)}{0.083}\,\text{educ}$$

$$n = 526 \quad R^2 = 0.186$$

(standart hatalar parantez içindedir)

- Bu modelde eğim katsayısını 100 ile çarparak % olarak yorumlayabiliriz; $100 \times 0.083 = 8.3$
- Bir yıllık fazladan eğitim ortalama ücretleri yaklaşık %8.3 arttıracaktır. Buna *ilave bir yıllık eğitimin getirisi* (return to another year of education) denir.
- **YANLIŞ**: *ilave bir yıllık eğitim logwage'i %8.3 arttırır*. Burada %8.3 artan logwage değil wage'dir.
- $R^2 = 0.186$ Eğitim logaritmik ücretlerdeki değişkenliğin yaklaşık %18.6'sını açıklayabilmektedir.

Log-Log Örnek: Yönetici Maaşları (ceosal1.gdt)

Model:

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + u$$

Tahmin sonuçları:

$$\widehat{\log(\text{salary})} = \underset{(0.288)}{4.822} + \underset{(0.035)}{0.257}\,\text{log(sales)}$$

$$n = 209 \quad R^2 = 0.211$$

(standart hatalar parantez içindedir)

- Yorum: Satışlarda meydana gelen %1 artış CEO maaşlarında ortalamada %0.257 artışa yol açmaktadır. Bir başka ifadeyle, yönetici maaşlarının satışlara göre esnekliği 0.257 olarak tahmin edilmiştir. Satışlarda yaklaşık %4'lük bir artış maaşları %1 arttırmaktadır.
- $R^2 = 0.211$ Log satışlar, log maaşlardaki değişkenliğin yaklaşık %21.1'ini açıklayabilmektedir.

Doğal Logaritma Dönüştürmesi İçeren Fonksiyon Kalıpları: Özet

Model	Dependent Variable	Independent Variable	Interpretation of β_1
level-level	y	x	$\Delta y = \beta_1 \Delta x$
level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

EKK(OLS) tahmin edicilerinin, $\hat{\beta}_0, \hat{\beta}_1$, istatistiksel özellikleri

- Anakütleden çekilen farklı rasgele örneklerden bulacağımız OLS parametre tahmin değerlerinin dağılımları ne tür özellikler gösterecektir?
- OLS tahmin edicilerinin örnekleme dağılımlarının özellikleri nelerdir?
- İlk olarak sonlu örneklem özelliklerinden sapmasızlık ve etkinliklerini inceleyeceğiz.
- Hatırlarsak sapmasızlık örnekleme dağılımındaki ortalamanın bilinmeyen doğru değere eşit olduğunu söylüyordu.
- Etkinlik ise ilgili tahmin edicinin varyansının o tahmin ediciler kümesi içinde en küçük olduğu anlamına geliyordu.

OLS tahmincilerinin sapmasızlığı

OLS t.e.'nin sapmasızlığı için aşağıdaki varsayımlar gereklidir:

1. SLR.1: Model parametrelerde doğrusaldır: $y = \beta_0 + \beta_1 x + u$
2. SLR.2: Tahminde kullanılan veriler ilgili anakütleden rassal örnekleme yoluyla seçilmiştir.
3. SLR.3: Sıfır Koşullu Ortalama: $\mathsf{E}(u|x) = 0$. Rassal örnekleme için bu varsayım şunu da zorunlu kılar:
$$\mathsf{E}(u_i|x_i) = 0, \quad \forall\, i = 1, 2, \ldots, n$$
4. SLR.4: x'de yeterli değişkenliğin olması. Bu varsayım x'in örneklem varyansının 0 olmamasını gerektirir:
$$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$$

OLS tahmincilerinin sapmasızlığı

TEOREM:$\hat{\beta}$'ların Sapmasızlığı

SLR.1-SLR.4 varsayımları altında OLS tahmin edicileri sapmasızdır:
$$\mathsf{E}(\hat{\beta}_0) = \beta_0, \quad \mathsf{E}(\hat{\beta}_1) = \beta_1$$

İSPAT:

Sapmasızlık OLS tahmin edicilerinin örnekleme dağılımlarının orta noktasının (beklentisinin) bilinmeyen popülasyon parametrelerine eşit olduğunu söyler.

Sapmasızlık Üzerine Notlar

- Sapmasızlık tekrarlanan örneklerden bulunan çok sayıdaki $\hat{\beta}_0$ ve $\hat{\beta}_1$ tahminlerine ait örnek dağılımlarının (sampling distributions) bir özelliğidir.
- Dolayısıyla, tek bir örneklemden (ki çoğu kez böyledir) hareketle hesaplanan betaşapkalarla ilgili olarak hiçbir şey söylemez. Bilinmeyen kitle parametrelerinden çok uzak bir tahmin de elde edebiliriz.
- Yukarıda yaptığımız 4 varsayımdan biri veya bir kaçı sağlanamazsa sapmasızlık özelliği geçerli olmaz.
- Doğrusallığın ve rassal örneklemenin olmaması, u'nun içerdiği faktörlerin x ile ilişkili olmaları (ki, sahte (spurious) korelasyona sebep olur) durumlarında **sapmalı tahmin ediciler** elde edilir.

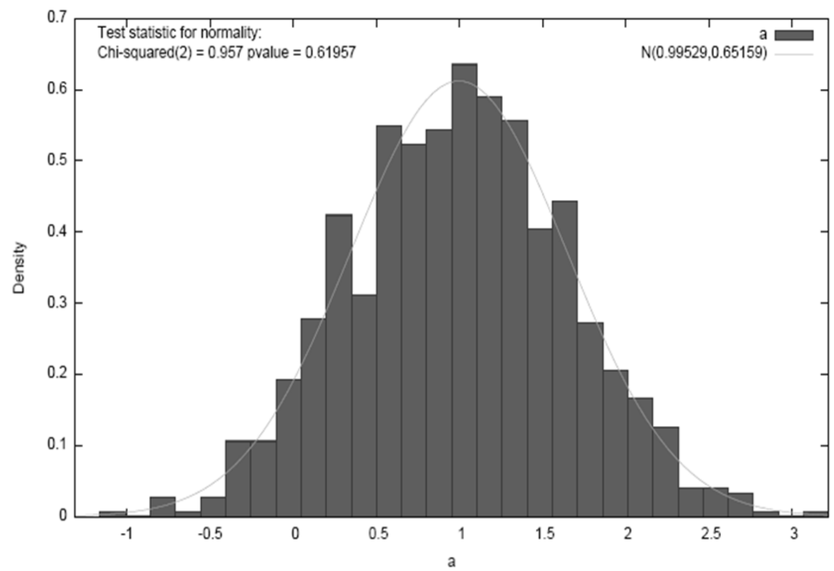
OLS'nin Sapmasızlığı: Basit Bir Monte Carlo Deneyi

- Veri üretim süreci:
$$y = 1 + 0.5x + 2 \times N(0,1)$$
- Gerçek popülasyon parametreleri biliniyor: $\beta_0 = 1$, $\beta_1 = 0.5$, $u = 2 \times N(0,1)$. Burada $N(0,1)$ standart normal dağılımdan çekilen rassal sayıyı ifade etmektedir.
- Benzer şekilde x'ler Uniform dağılımdan çekilsin: $x = 10 \times Unif(0,1)$
- Şimdi bu populasyon modelinde parametre değerlerini bilmediğimizi ve OLS yöntemini kullanarak tahmin edeceğimizi düşünelim.
- GRETL programını kullanarak yukarıda belirtilen dağılımlardan rassal sayılar türetip çok sayıda örnekler elde edebilir ve her örnek için OLS tahminlerini bir dosya içinde kaydedebiliriz.
- Bu basit bir Monte Carlo deneyidir. Tahmin edicilerin, test istatistiklerinin örnekleme dağılımlarının elde edilmesinde sıklıkla kullanılır.

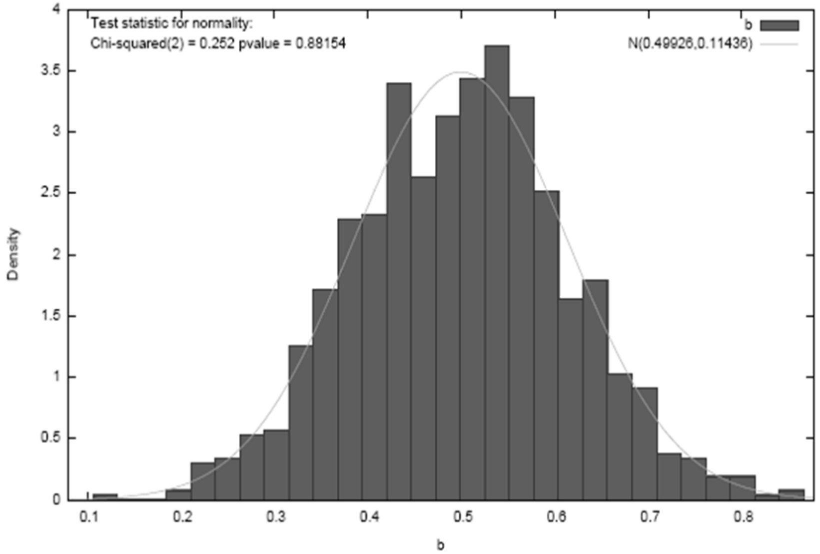
OLS'nin Sapmasızlığı: Basit Bir Monte Carlo Deneyi

```
nulldata 50
seed 123
genr x = 10 * uniform()
loop 1000
    genr u = 2 * normal()
    genr y = 1 + 0.5 * x + u
    ols y const x
    genr a = $coeff(const)
    genr b = $coeff(x)
    genr r2 = $rsq
    store MC1coeffs.gdt a b
Endloop
```

Sabit Terimin OLS t.e.'nin Örnekleme Dağılımı



Eğim Katsayısının OLS t.e.'nin Örnekleme Dağılımı



OLS Tahmincilerinin Varyansı

- OLS t.e., $\hat{\beta}_0$ ve $\hat{\beta}_1$'nın sapmasızlığı için SLR.1-SLR.4 varsayımları gerekli ve yeterliydi.
- OLS tahmincilerinin sonlu örneklem özelliklerinden etkinliğin tesis edilebilmesi va tahmincilerin varyans ve standart hatalarının belirlenebilmesi için bir varsayıma daha ihtiyaç duyulur.
- SLR.5: Sabit Varyans (Homoscedasticity): Bu varsayım gözlenemeyen hata teriminin x'e koşullu varyansının sabit bir sayı olduğunu söyler:

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

- Bu aynı zamanda koşulsuz varyanstır: $\text{Var}(u) = \sigma^2$
- Bu varsayımla birlikte u ve x'in istatistiksel açıdan bağımsız (independent) olduğunu söyleyebiliriz: $\text{E}(u|x) = \text{E}(u) = 0$ ve $\text{Var}(u|x) = \text{Var}(u) = \sigma^2$

OLS Tahmincilerinin Varyansı

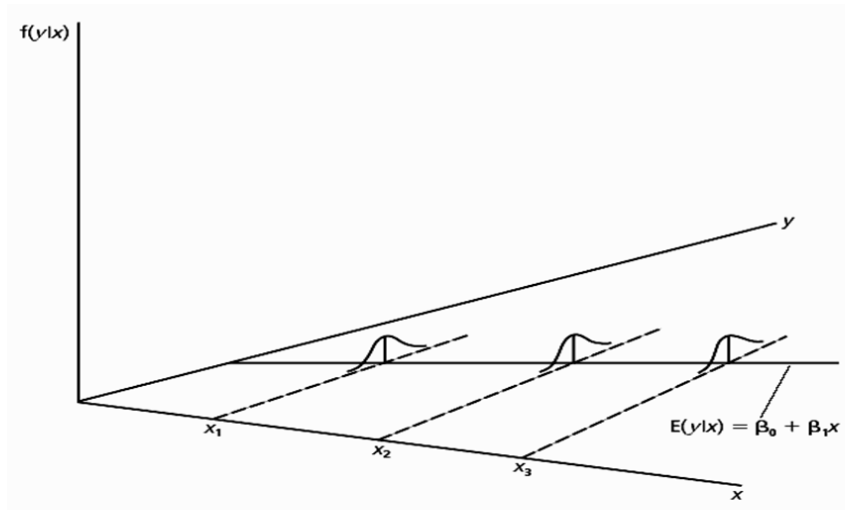
- SLR.3 ve SLR.5 varsayımları y'nin koşullu ortalama ve koşullu varyansı ile de ifade edilebilir:

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

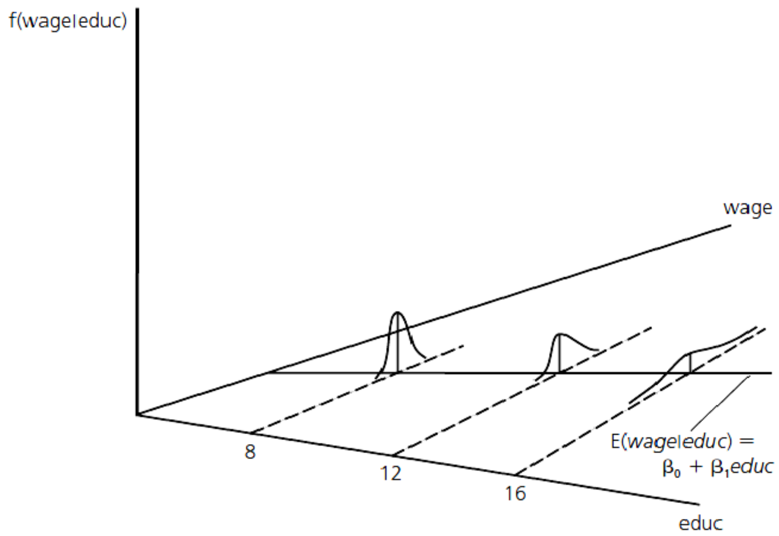
$$\text{Var}(y|x) = \sigma^2$$

- x verilmişken y'nin koşullu beklenen değeri x'in doğrusal bir fonksiyonudur.
- x verilmişken y'nin koşullu varyansı sabittir ve hata teriminin varyansına eşittir.
- σ^2'ye kısaca hata varyansı (error variance) adı verilir.

Sabit Varyans (Homoscedasticity) Varsayımı Altında Basit Regresyon Modeli



Ücretlerin Değişkenliğinin Eğitim Düzeyine Göre Arttığı Durum (Değişen Varyans - Heteroskedasticity)



OLS Tahmincilerinin Örneklem Varyansları

SLR.3 ve SLR.5 varsayımları altında eğim katsayısının OLS tahmincisinin örneklem varyansı:

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

Sabit terimin OLS tahmincisinin örneklem varyansı

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Bu varyans formülleri x değerlerine koşullu olarak türetilmiştir.
- Değişen varyans durumunda (SLR.5 varsayımının sağlanmaması) bu formüller geçersizdir.
- OLS tahmincilerin varyansları hata varyansı ile doğru orantılı, x'deki değişkenlikle ise ters orantılıdır.

Hata Terimi ve Kalıntı Arasındaki Fark

- Hata terimi ile kalıntı terimi birbirine karıştırılmamalıdır.
- Hata terimi tıpkı popülasyon parametreleri gibi gözlenemez ve bilinemez:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- Kalıntı ya da artık terimleri ise (residuals) verilerden hareketle tahmin edilebilir:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

- Kalıntılar hata teriminin bir fonksiyonu olarak yazılabilir:

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = \beta_0 + \beta_1 x_i + u_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i$$

- Sapmasızlık özelliğinden hareketle $\text{E}(\hat{u}) = \text{E}(u) = 0$ olacağı açıktır.

Hata Varyansının Tahmini

- σ^2'nin sapmasız bir tahmincisini bulmak istiyoruz.
- Varsayım gereği $\text{E}(u^2) = \sigma^2$ olduğundan hata varyansının sapmasız bir *tahmin edicisi*:

$$\frac{1}{n}\sum_{i=1}^{n} u_i^2$$

- Ancak u'ları da gözleyemediğimiz için bunu kullanamayız. Bu nedenle varyans tahmini OLS kalıntılarına dayandırılır:

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

- Ancak bu tahminci **sapmalıdır**. Serbestlik derecesi düzeltmesinin yapılması gerekir. Hata varyansının **sapmasız** bir tahmin edicisi şudur:

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

- Serbestlik derecesi (degrees of freedom - dof) = gözlem sayısı

OLS Tahmincilerinin Standart Hataları

- Hata varyansının kareköküne regresyonun standart hatası (standard error of regressin) denir:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2} = \sqrt{\frac{SSR}{n-2}}$$

- $\hat{\sigma}$ *hata kareleri ortalaması karekökü* olarak da adlandırılır (root mean squared error).
- Eğim katsayısının OLS tahmincisinin standart hatası:

$$\mathrm{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \frac{\hat{\sigma}}{s_x}$$

Orijinden Geçen Regresyon

- Bazı durumlarda $x = 0$ iken $y = 0$ olsun isteriz. Örneğin, gelir sıfırken vergi tahsilatı da sıfır olacaktır.
- Modeli sabit terimsiz olarak şu şekilde kurabiliriz: $\tilde{y} = \tilde{\beta}_1 x$. Sabit terimli modelden ayırmak için şapka yerine tilde kullanıldığına dikkat ediniz.
- OLS yöntemi kalıntı kareleri toplamını en küçük yapar

$$\min \sum_{i=1}^{n}(\tilde{y} - \tilde{\beta}_1 x_i)^2$$

- Buradan birinci sıra koşulu:

$$\sum_{i=1}^{n} x_i(\tilde{y} - \tilde{\beta}_1 x_i) = 0$$

- Bu çözülürse orijinden geçen regresyonun eğim tahmincisi bulunur:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$