

Kukla Değişkenler (Dummy Variables)

* Önceki bölümlerde bağımlı ve bağımsız değişkenlerimiz her zaman nüch bilgiler içeriyordu. ÖR: ücretler, iş deneyimi, ev fiyatları, oda sayısı

⇒ Ancak, uygulamalarda bazı nitel değişkenleri de regresyona dahil etmeniz gerekebilir.

ÖR: bireylerin cinsiyeti, ırk, din, doğduğu bölge eğitim (lise, üniversite, yüksek eğitim)

⇒ Nitel değişkenler ikili/kukla (binary / dummy) değişkenlerle ifade edilebilir.

Nitel Bilgi:

- Nitel enformasyon genellikle ikili (binary) yapı gösterir: erkek/kadın, arabaşı olan /olmayan yerli/yabancı, ölüm cezası olan /olmayan ülkeler.
- Bu tür nitel bilgileri ikili değişkenlerle ifade ederiz.
↳ "0" / "1" kullanarak ifade edeceğiz.
- Hangi kategoriye 1 ya da 0 verdığınız regresyon sonucundan elde edilen tahmin sonuçlarını değiştirmeye rağmen hangi kategorinin sol tarafındaki bağımlı değişkene ne kadar etki edeceğini

değiştirmez.

Kadın ve Erkek

$$\text{female} = 1 \rightarrow \text{erkek} = 0 \Rightarrow \text{model 1}$$

bat grup

$$y = \beta_0 + \beta_1 \text{female} + \epsilon$$

→ bat grup

$$\text{male} = 0 \quad \text{erkek} = 1 \Rightarrow \text{model 2}$$
$$y = \alpha_0 + \alpha_1 \text{male} + \epsilon$$

• Ancak yorumlama yapabilmemiz için hangi kategoriye 1 ve hangi kategoriye 0 dediğimizi bilmemiz gereklidir.

bat grup

Cinsiyet Kuklesi \rightarrow erkek = 0
kadın = 1

Evlilik Kuklesi \rightarrow evli = 1

evli değil = 0

~~gender~~

Kisi	wage	educ	exper	female	married
1	3.10	11	2	1	0
2	3.24	12	22	0	1

Tek Kukla Değişkenli Regresyon

• ÖRNEK:

$$\text{wage} = \beta_0 + \beta_1 \text{female} + \beta_2 \text{educ} + u$$

Kadın için $\text{female} = 1$

Erkek // $\text{female} = 0$

- δ_0 : Aynı eğitim düzeyine sahip bir erkekle bir kadının maaşları arasındaki farkı ifade eder.

$$\text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + u$$

$$\bullet \text{female} = 0 \rightarrow \text{Erkek} \Rightarrow \frac{\text{wage} = \beta_0 + \delta_0 \cdot 0}{\text{wage} = \beta_0} \Rightarrow \text{Erkek}$$

$$\bullet \text{female} = 1 \rightarrow \text{Kadın} \Rightarrow \frac{\text{wage} = \beta_0 + \delta_0 \cdot 1}{\text{wage} = \beta_0 + \delta_0} \Rightarrow \text{Kadın}$$

- Eğer δ_0 sıfırdan istatistikî olarak farklı çıkmazsa (t -testi yapacağınız) erkek-kadın arasında ücret farklılığı yok demektir.

- $\delta_0 < 0$ ise (istatistikî olarak) aynı eğitim düzeyinde kadın çalışanlar erkeklerden ortalamada daha kazanmaktadır.

model

$$\text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + u$$

- Kadın çalışanlar için ücretlerin koşullu beklenisi:

$$\mathbb{E}(\text{wage} | \text{educ}, \text{female} = 1) = \beta_0 + \delta_0 + \beta_1 \text{educ}$$

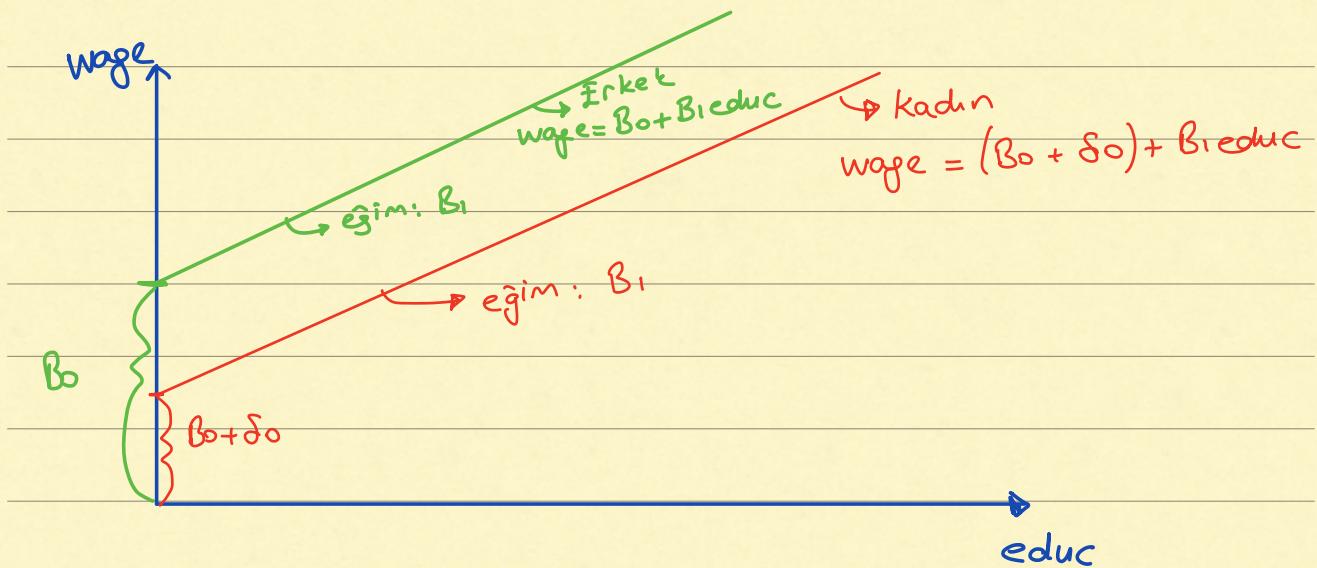
- Erkek Çalışanlar için ücretlerin koşullu beklenisi:

$$\mathbb{E}(\text{wage} | \text{educ}, \text{female} = 0) = \beta_0 + \beta_1 \text{educ}$$

- Kadın çalışanların ücret ortalaması ile erkek çalışanların ücret ortalaması farkını alırsak.

$$\begin{aligned}
 &= \mathbb{E}(\text{wage} | \text{educ}, \text{female} = 1) - \mathbb{E}(\text{wage} | \text{educ}, \text{female} = 0) \\
 &= \beta_0 + \delta_0 + \beta_1 \text{educ} - \beta_0 - \beta_1 \text{educ} \\
 &= \delta_0
 \end{aligned}$$

$\delta_0 < 0$ iain Ücret Modelinin ARF'si



- Ücret modelinde β_0 erkekler için olan regresyonel sabit terimidir.
- Kadınlar için sabit terim $\beta_0 + \delta_0$ dir.
- Bir kukla değişken iki katemorisi birbirinden

ayırılmaktadır. Bu nedenle regresyona kadın ve erkekler için ayrı ayrı kukla değişken koymaya gerek yoktur.

- Gerek kukla değişken sayısı = kategori sayısı - 1
- Cinsiyet için iki kukla değişken kullanılsaydı tam noktlu doğrusallık problemiyle karşılaşırız. Bu duruma "kukla değişken tuzagi" (dummy variable trap) denir.

Kisi	wage	educ	female	male	female + male
1	3.21	10	1	0	1
2	3.52	11	0	1	1
3	3.65	12	1	0	1

↳ Olası yapılabilecek modeller

$$\textcircled{1} \quad \text{wage} = \beta_0 + \beta_1 \text{educ} + \delta_0 \text{female} + u$$

$$\textcircled{2} \quad \text{wage} = \beta_0 + \beta_1 \text{educ} + \delta_0 \text{male} + u$$

Hatalı Model (Kukla Değişken tuzagi olan model)

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \delta_0 \text{female} + \delta_1 \text{male} + u$$

$$\begin{matrix} x_1 & x_2 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{matrix}$$

$$x_1 + x_2 = 1$$

$$x_1 + x_2 = 1$$

$$x_1 + x_2 = 1$$

tam linear

bağıntı
var
tam açıklı
doğrusal
bağıntı

- Kukla değişken değeri 0 ise baz grubu ifade eder.

- Diğer bir alternatif ise, Regresyonda hiç sabit terim kullanmayıp her bir kategori için bir kukla değişken kullanmaktır.

$$Wage = \delta_0 \text{female} + \gamma_0 \text{male} + B_1 \text{educ} + u$$

- Modelde sabit terim eklemediğimiz için kukla değişken tuzağına düşmeyez.

- Bu modelde her bir grup için bir sabit terim tahrin edilecektir \rightarrow kadın $\Rightarrow \delta_0$
erkek $\Rightarrow \gamma_0$

- Sabit terimsiz regresyonda R^2 hesaplanamaz. Hipotez testlerinin yapılması daha zor olur.

- Bu nedenle biz ilk versiyonu kullanmaya çalışacağız.

- Erkek çalışanların baz grubu olduğu model $\Rightarrow \text{female}_1 = 1$
 $\text{Wage} = \delta_0 + \delta_1 \text{female} + B_1 \text{educ} + u$
 \hookrightarrow kadınların sabit terimi $\Rightarrow \delta_0 + \delta_0$
erkeklerin sabit terimi $\Rightarrow \delta_0$

- Kadın çalışanların baz grubu olduğu model $\Rightarrow \text{male} = 1$

$$\text{Wage} = \delta_0 + \gamma_0 \text{male} + B_1 \text{educ} + u$$
$$\hookrightarrow$$
 kadınların sabit terimi $\Rightarrow \delta_0$

$$\hookrightarrow$$
 erkeklerin sabit terimi $\Rightarrow \delta_0 + \gamma_0$

$$\textcircled{1} \quad \alpha_0 = \beta_0 + \delta_0$$

$$\textcircled{2} \quad \beta_0 = \alpha_0 + \gamma_0$$

- Kukla değişkenlerin katsayılarının yorumunu baz göre yapacağız. Bu nedenle hangi grubun baz grub olduğunu bilmeniz çok önemlidir.

- Birden fazla bağımsız değişken olması kukla değişkenlerin yorumunu değiştirmez.

$$\text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + u$$

δ_0 : Aynı eğitim, tecrübe ve kıdem düzeyine sahip bir kadın ile erkek arasındaki sabit terim formunu ifade eder.

- $H_0: \delta_0 = 0$ vs $H_1: \delta_0 < 0$

↳ Sonuç: H_0 red edilmiş

Kadın çalışanlara karşı negatif ücret

ayrimcılığı yapılmıyor denir.

t-testi kullan.

Örnek: $\text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure}$

$$\hat{Wage} = -1.57 - 1.81 \text{female} + 0.572 \text{educ} + 0.025 \text{exper} + 0.141 \text{tenure}$$

$(0.725) \quad (0.265) \quad (0.049) \quad (0.0116) \quad (0.021)$

$$n = 526 \quad R^2 = 0.364 \quad F\text{-ist} = 74.398$$

↳ $sd_1 = 4$
 $sd_2 = 521$

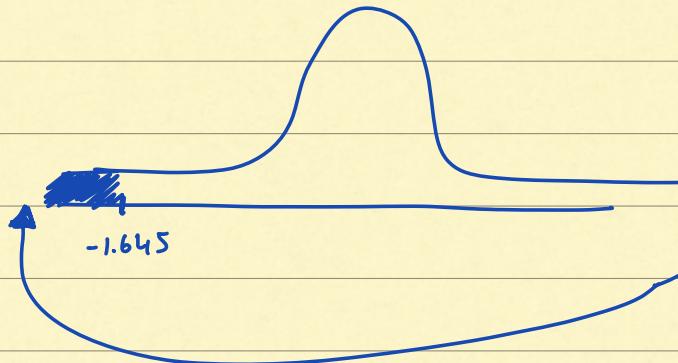
- Kadınlar erkeklerle göre ceteris paribus koşulu altında \$1.81 daha az kazanırlar. Baska bir ifade ile aynı eğitim, tecrübe ve kıdemle sahip bir kadın ve erkek arasındaki saatlik ücret farkı \$1.81'dir.

$\alpha = 0.05 \Rightarrow$ anlamlılık düzeyi

$$H_0: \delta_0 = 0 \quad \text{vs} \quad H_1: \delta_0 < 0$$

$$t\text{-ist} = \frac{\hat{\delta}_0 - \delta_0}{se(\hat{\delta}_0)} = \frac{-1.81 - 0}{0.265} = -6.830$$

$$t_{\alpha, n-k-1} = t_{0.05, 521} = 1.645$$



H_0 red edilir
 $\hat{\delta}_0$ istatistikî olarak 0'den küçüktür.

- Sonuç olarak $\hat{\beta}_0 = -1.57 \Rightarrow$ erkeklerin sabit terimidir

- Modelden tüm nicel değişkenleri displayip sadece kukla değişken üzerine regresyon kurorsak

$$wage = \hat{B}_0 + \delta_0 \text{female} + u$$

$$\hat{wage} = 7.1 - 2.51 \text{female}$$

(0.21) (0.303)

$$n = 526 \quad \bar{R}^2 = 0.1140$$

- Erkeklerin sabit terimi $\hat{B}_0 = 7.1$ erkek çalışanların ortalaması ücret düzeyini verir.
- Kadınların erkeklerce maaş farkını ifade eden $\hat{\delta}_0 = -2.51$ kadınların ortalaması ücretleri ile erkeklerin ortalaması ücretleri arasındaki farkı ifade eder.
- Bu sonuçlardan hemenlikle kadın çalışanların ortalaması maaşı $\Rightarrow \hat{B}_0 + \hat{\delta}_0 = 7.1 - 2.51 = \4.59
- Çalışanlar, iki ayrı grubu ayırip aritmetik ortalamalarını hesaplayarak aynı sonuca ulaşır (modelde nicel bağımsız değişkenlerin olmaması durumunda geçerlidir sadece)

İki Nitel Değişken (iki kukla değişken)

female = 1 \Rightarrow kadın ise (erkek baz grubu \Rightarrow female = 0)

married = 1 \Rightarrow evli ise (evli olmayan baz grubu \Rightarrow married = 0)

$$wage = \hat{B}_0 + \hat{\delta}_0 female + \hat{\delta}_1 married + u$$

$$\hat{wage} = 6.18 - 2.29 \text{ female} + 1.34 \text{ married}$$

(0.296) (0.302) (0.310)

$$n = 526 \quad R^2 = 0.142$$

- Bu regresyonda baz grub "evli olmayan erkekler"dir.

$female = 0$ ve $married = 0 \Rightarrow$ Sabit terim = \$ 6.18
evli olmayan erkekler $\Rightarrow \hat{B}_0$

- female'in katsayısı ($\hat{\delta}_0$) ise evli olmayan kadınların ortalama ücretleri ile evli olmayan erkeklerin ortalama ücretleri arasındaki farktır \Rightarrow Bu değer $\hat{\delta}_0$ dir = \$ -2.29

- Married'in katsayısı ($\hat{\delta}_1$) ise evli olan erkeklerin ortalama ücretleri ile evli olmayan erkeklerin ortalama ücretleri arasındaki farktır \Rightarrow Bu değer $\hat{\delta}_1$ dir = \$ 1.34 dir.

$$\hat{wage} = 6.18 - 2.29 \text{ female} + 1.34 \text{ married}$$

(0.296) (0.302) (0.310)

$$n = 526 \quad R^2 = 0.142$$

- Bu sonuçlardan herkelle evli olmayan kadınların ($female = 1$, $married = 0$) ortalama ücreti söyle bulunabilir $\Rightarrow \hat{B}_0 + \hat{\delta}_0 = 6.18 - 2.29 = \$ 3.89$

- U

- Evli kadın aileşanlarının ortalaması ücreti
↳ $\text{female} = 1$, $\text{married} = 1$

$$\hat{\text{wage}} = 6.18 - 2.29 \text{ female} + 1.34 \text{ married}$$

$$\begin{aligned}\hat{\text{wage}} &= 6.18 - 2.29 + 1.34 \\ &= \$ 5.23\end{aligned}$$

- Evli erkekler ($\text{female} = 0$, $\text{married} = 1$) ve evli kadınlar ($\text{female} = 1$, $\text{married} = 1$) arasındaki farkı baktık.

$$\hat{\text{wage}} = 6.18 - 2.29 \text{ female} + 1.34 \text{ married}$$

$$\begin{aligned}\hat{\text{wage}} &= 6.18 - 2.29 + 1.34 \\ &= \$ 5.23 \Rightarrow \text{evli kadınlar}.\end{aligned}$$

$$\hat{\text{wage}} = 6.18 - 2.29 \text{ female} + 1.34 \text{ married}$$

$$\begin{aligned}\hat{\text{wage}} &= 6.18 + 1.34 \\ &= \$ 7.52 \Rightarrow \text{evli erkekler}.\end{aligned}$$

$$\text{evli erkek} - \text{evli kadın} = 7.52 - 5.23 = 2.29 \$$$

$$\ln \text{wage} = \beta_0 + \beta_1 \text{female} + \beta_2 \text{married} + \beta_3 \text{educ} + \beta_4 \text{exper}$$

$$\beta_5 \text{expersq} + \beta_6 \text{tenure} + \beta_7 \text{tenuresq} + \epsilon$$

$$\ln wge = 0.42 - 0.29 \text{female} + 0.05 \text{married} + 0.08 \text{educ} + 0.03 \text{exper}$$

(0.098) (0.036) (0.040) (0.007) (0.005)

$$- 0.0005 \text{expersq} + 0.03 \text{tenure} - 0.0006 \text{tenursq}$$

(0.0001) (0.007) (0.0002)

$n = 526$

$$R^2 = 0.4351$$

- Ceteris paribus koşulu altında çalışanın erli olması ortalaması ücretlerde fark yaratıyor mu?

$H_0: \delta_1 = 0$ vs. $H_1: \delta_1 \neq 0 \Rightarrow t\text{-testi ile çözüceğiz.}$

$$t\text{-ist}: \frac{\hat{\delta}_1 - \delta_1}{se(\hat{\delta}_1)} = \frac{0.05}{0.04} = \underline{\underline{1.25}}$$

$\alpha = 0/05$ anlamlılık düzeyi

$t\text{-kritik} \Rightarrow t_{\alpha/2}, n-k-1 = t_{0.025}, 526-7-1 = t_{0.025}, 518$

$\underline{\underline{1.96}}$

Sonuç: $t\text{-ist} < t\text{-kritik}$ olduğundan H_0 red edilemez ve çalışanın erli olması maaşlarda bir fark yaratmaz.

- Ceteris paribus koşulu altında çalışanın kadın olması ortalaması ücretlerde fark yaratıyor mu?

$H_0: \delta_0 = 0$ vs. $H_1: \delta_0 \neq 0 \Rightarrow t\text{-testi ile çözüceğiz.}$

$$t\text{-ist}: \frac{\hat{\delta}_0 - \delta_0}{se(\hat{\delta}_0)} = \frac{-0.29}{0.036} = \underline{\underline{-8.05}}$$

$\underline{\underline{\alpha = 0/05}}$ anlamlılık düzeyi

$t\text{-kritik} \Rightarrow t_{\alpha/2}, n-k-1 = t_{0.025}, 526-7-1 = t_{0.025}, 518$

$\rightarrow \frac{1.96}{\sqrt{2}}$

Sonuç:

H_0 hipotezi red edilir ve
calisanın kadın olması ortalamaya
maaslarında düşüse neden
olsur.



Cok Sayıda Kukla Değişkenler

- Önceki modeli farklı bir şekilde de kura biliirdik.
 \hookrightarrow female ve married kukla değişkenlerini 4 gruba ayırmak olacaktır.

(evli erkek) marrmale = married \times (1 - female)

(evli kadın) marrfem = married \times female

(bekar erkek) sing male = (1 - married) \times (1 - female)

(bekar kadın) singfem = (1 - married) \times female

ÖR: bekar kadın \Rightarrow married = 0 }
 female = 1 }
 marr male = 0
 marrfem = 0
 sing male = 0
 singfem = 1

\hookrightarrow sing male (bekar erkek) bazı grup oluştur
seçilmiş olsun.

$$\ln \hat{wage} = 0.32 + 0.21 \text{marrmale} - 0.198 \text{marrfem} - 0.11 \text{singfem}$$

(0.101) (0.055) (0.058) (0.055)

$$+ 0.079 \text{ educ} + 0.027 \text{ exper} - 0.0005 \text{ expersq}$$

(0.006)

(0.005)

(0.001)

$$+ 0.029 \text{ tenure} - 0.0005 \text{ tenure}^2$$

(0.006) (0.0002)

$$n = 526 \quad R^2 = 0.4525 \quad F = 55.246$$

- sing male = 0.32 \Rightarrow bekor erkek iain _{ortalama maas} = 0.32
- marr male = 0.21 \Rightarrow evli erkek iain _{ortalama maas} = $\frac{0.32}{0.21} = 0.53$

$$\text{marr fem} = -0.198 \rightarrow \text{evli kadın iain} \\ \text{ortalama maas} = 0.32 + (-0.198) \\ = 0.122$$

$$\text{sing fem} = -0.11 \Rightarrow \text{bekor kadın iain} \\ \text{ortalama maas} = 0.32 + (-0.11) \\ = 0.21$$

Kukla Değişkenlerde Etkileşim.

- Yukarıdaki kukla değişkenler kullanılarak aynı regresyonda çok sayıda farklı sabit terimler tahmin edebildik.
- Bu modellerde nicel değişkenlerin eğim katsayıları sabit kabul edilmişti. Şimdi farklı eğimler türetmeye görelim.
- ÖR: ücret/maas regresyonunda bir yıllık ilave bir eğitimin erkek ve kadın ücretlerinde yaratacağı etkinin farklı olup olmadığını ölçmek isteyelim.
- Bunun iain cinsiyet kukla değişkeni female-i educ değişkeni ile etkileşime girmeyeceğiz: female \times educ

- Böylece eğitim düzeyi artıktan kadınlarla erkekler arasındaki ücret farkının nasıl değiştigini söyleyebiliriz..

$$\ln \text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \delta_1 \text{female.educ} + u$$

$$\ln \text{wage} = (\beta_0 + \delta_0 \text{female}) + (\beta_1 + \delta_1 \text{female}) \text{educ} + u$$

- $\text{female} = 0$ (erkek ise) \Rightarrow değer yerine koymursa β_0 erkek çalışanların sabit terimini ifade eder.
- β_1 ise erkek çalışanların eğim kat sayısını ifade eder.
- δ_0 kadın çalışanların erkek çalışanlar arasındaki sabit terim farkıdır. Öyleyse kadınların sabit terimi $\Rightarrow \beta_0 + \delta_0$
- δ_1 ise kadın çalışanlar ile erkek çalışanlar arasındaki eğim kat sayısını farkıdır. Kadınlar için eğitim değişkeninin eğim kat sayısı $\Rightarrow \beta_1 + \delta_1$
- $\delta_1 > 0$ ise kadın çalışanlar için bir yıllık eğitimin getirişi erkek çalışanlardan daha fazladır.

(A)

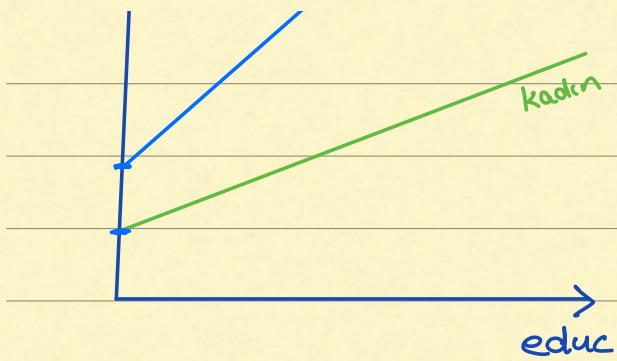
wage
↑

erkek
/

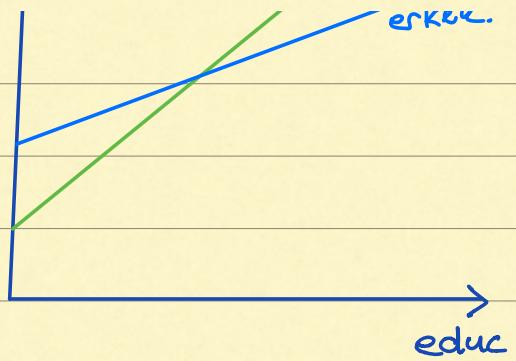
(B)

wage
↑

kadın
/.../



$$\delta_0 < 0 ; \delta_1 < 0$$



$$\delta_0 < 0 ; \delta_1 > 0$$

- Sekil A-de, kadınlara ait regresyonun erkeklerle ait regresyona kıyasla hem sabiti hem de eğini daha düşüktür.
- . ↳ ① Yani, kadınların saat - başına ücretleri tüm eğitim düzeylerinde erkeklerinkinden daha düşükür.
- ② Aradaki fork eğitim düzeyi yükseldikçe ortmaktadır.

• Sekil B-de ise, erkeklerle kıyasla, kadın kadınlara ait regresyonun sabiti daha düşük ancak educ değişkeninin eğini daha yüksektir.
 ↳ düşük eğitim düzeylerinde kadınlar, yüksek eğitim düzeylerinde ise erkekler daha az kazanmaktadır.

$$\ln \text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \delta_1 \text{female} \cdot \text{educ} + u$$

erkek (female=0)

$$\hookrightarrow \ln \text{wage} = \beta_0 + \beta_1 \text{educ} + u$$

kadın (female=1)

$$\hookrightarrow \ln \text{wage} = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) \text{educ} + u$$

$$\ln \text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \underline{\delta_1 \text{female} \cdot \text{educ}} + u$$

• δ_1 istatistikî olarak anlamlı mudīr?

$$H_0: \delta_1 = 0 \quad \text{vs.} \quad H_1: \delta_1 \neq 0 \Rightarrow \underline{\text{t-test}}$$

- Aynı eğitim düzeyine sahip erkek ve kadın çalışanların ortalaması ücretleri aynı mudīr?

$$H_0: \delta_0 = \delta_1 = 0 \quad \text{vs.} \quad H_1: H_0 \text{ doğru değil}$$

F-testi

$$\begin{aligned} \hat{\ln \text{wage}} &= 0.389 - 0.227 \text{female} + 0.82 \text{educ} \\ &\quad (0.119) \quad (0.168) \quad (0.008) \end{aligned}$$

$$-0.056 \text{female.educ} + 0.029 \text{exper} - 0.00058 \text{exper}^2$$

$$(0.0131) \quad (0.005) \quad (0.00011)$$
$$0.032 \text{tenure} - 0.00059 \text{tenure}^2$$
$$(0.007) \quad (0.00024)$$

$$n = 526 \quad R^2 = 0.441$$