

Kukla Değişkenler (Dummy Variables)

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

⇒ Ancak, uygulamalarda bazen nitel değişkenleri de regresyona dahil etmemiz 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 çoğu kez ikili (binary) yapı gösterir: erkek/kadın, arabası 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 verdiğimiz regresyon sonucundan elde edilen tahmin sonuçlarını değiştirmesine rağmen hangi kategorinin sol taraftaki bağımlı değişkene ne kadar etki edeceğini

değiştirmez.

Kadın ve Erkek

female = 1 → erkek = 0 ⇒ $y = \beta_0 + \beta_1 \text{female} + \epsilon$

baz grup

Model 1

male = 0 → erkek = 1 ⇒ $y = \alpha_0 + \alpha_1 \text{male} + \epsilon$

baz grup

Model 2

• Ancak yorumlama yapabilmemiz için hangi kategoriye 1 ve hangi kategoriye 0 dedimizi bilmemiz gerekir

Cinsiyet kuklası → erkek = 0
kadın = 1

Evlilik Kuklası → 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 female = 1

Erkek // female = 0

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

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

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

$$\underline{\text{wage} = \beta_0} \Rightarrow \text{Erkek}$$

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

$$\underline{\text{wage} = \beta_0 + \delta_0} \Rightarrow \text{Kadın}$$

• Eğer δ_0 sıfırdan istatistiki olarak farklı çıkmazsa (t-testi yapacağız) erkek-kadın arasında ücret farklılığı yok demektir.

• $\delta_0 < 0$ ise (istatistiki 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 beklentisi:

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

- Erkek çalışanlar için ücretlerin koşullu beklentisi:

$$E[\text{wage} \mid \text{educ}, \text{female} = 0] = \beta_0 + \beta_1 \text{educ}$$

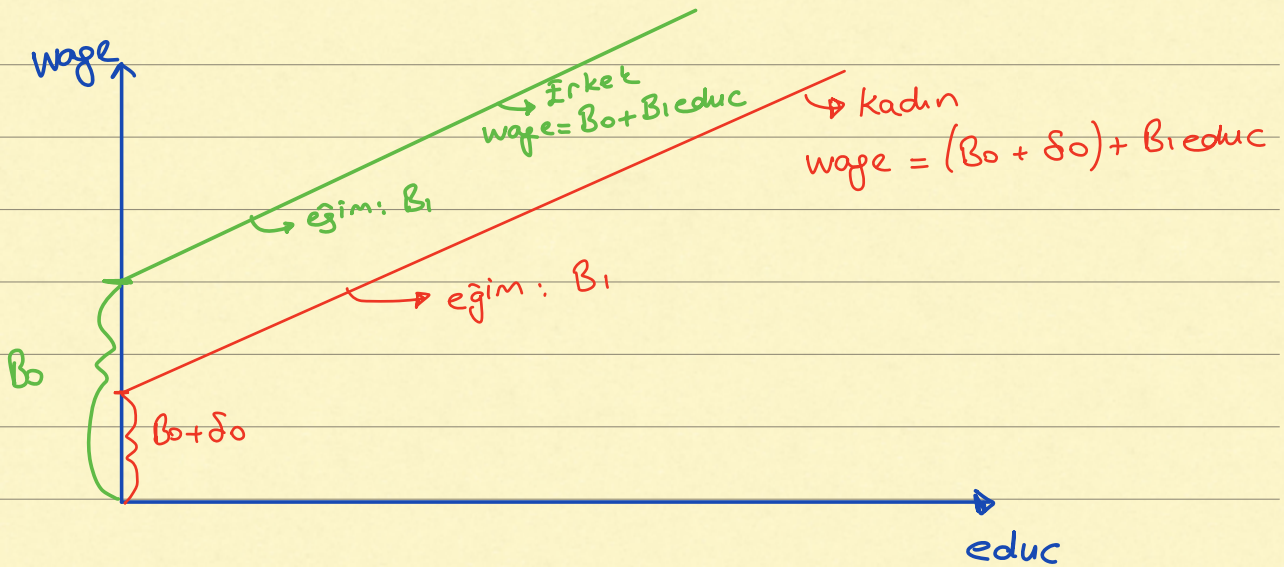
- Kadın çalışanların ücret ortalaması ile erkek çalışanların ücret ortalaması farkını alırsak.

$$= E(\text{wage} \mid \text{educ}, \text{female} = 1) - E(\text{wage} \mid \text{educ}, \text{female} = 0)$$

$$= \beta_0 + \delta_0 + \beta_1 \text{educ} - \beta_0 - \beta_1 \text{educ}$$

$$= \delta_0$$

$\delta_0 < 0$ için Ücret Modelinin ARF-si



- Ücret modelinde β_0 erkekler için olan regresyondaki sabit terimdir.

- Kadınlar için sabit terim $\beta_0 + \delta_0$ dir.

- Bir kukla değişken iki kategoriyi birbirinden

ayırabilmektedir. 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 kullanırsaydık tam oktu doğrusallık problemiyle karşılaşırız.

Bu duruma "kukla değişken tuzağı" (dummy variable trap) derir.

Kişi	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

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

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

Hatalı Model (Kukla Değişken tuzağı olan model)

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

$$\begin{array}{cc} x_1 & x_2 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{array}$$

$$\boxed{x_1 + x_2 = 1} \quad \text{tam lineer}$$

bağıntı
var
tam cıktı
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.

$$\text{wage} = \delta_0 \text{female} + \gamma_0 \text{male} + \beta_1 \text{educ} + u$$

- Modelde sabit terim eklemediğimiz için kukla değişken tutarınca diüsmeyiz.

- Bu modelde her bir grup için bir sabit terim tahmin 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 grup olduğu model \Rightarrow female = 1

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

\rightarrow kadınları sabit terimi $\Rightarrow \beta_0 + \delta_0$
erkeklerin sabit terimi $\Rightarrow \beta_0$

- Kadın çalışanların baz grup olduğu model \Rightarrow male = 1

$$\text{wage} = \alpha_0 + \gamma_0 \text{male} + \beta_1 \text{educ} + u$$

\rightarrow kadınların sabit terimi $\Rightarrow \alpha_0$
erkeklerin sabit terimi $\Rightarrow \alpha_0 + \gamma_0$

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

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

• Kukla deęişkenlerin katsayılarının yorumunu baz göre yapacağız. Bu nedenle hangi grubun baz grup olduğunu bilmemiz çok önemlidir.

• Birden fazla bağımsız deęişken olması kukla deęişkenlerin yorumunu deęiştirmez.

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

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

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

↳ Sonuç: H_0 red edilmiş

Kadın çalışanlara karşı negatif ücret ayrımcılığı yapıyor demektir.

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)
(0.265)
(0.049)
(0.0116)
(0.021)

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

$$\hookrightarrow \begin{aligned} sd_1 &= 4 \\ sd_2 &= 521 \end{aligned}$$

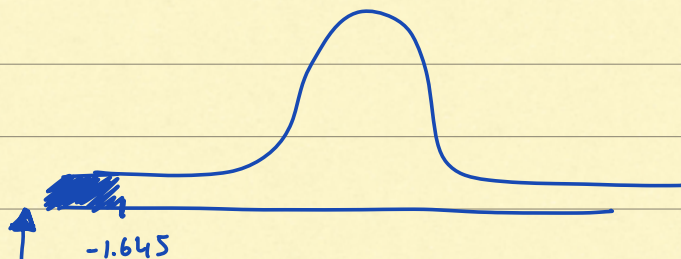
• Kadınlar erkeklere göre ceteris paribus kasulu altında \$1.81 daha az kazanırlar. Baska bir ifade ile aynı eğitim, tecrübe ve kıdemc sahip bir kadın ve erkek arasındaki saatlik ücret farkı \$1.81'dir.

$\alpha = 0.05 \Rightarrow$ 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
 δ_0 istatistiki olarak 0'den küçüktür.

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

- Modelden tüm nicel değişkenleri dışlayıp sadece kukla değişken üzerine regresyon kurarsak

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

$$\hat{\text{wage}} = \frac{7.1}{(0.21)} - \frac{2.51}{(0.303)} \text{female}$$

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

- Erkeklerin sabit terimi $\hat{\beta}_0 = 7.1$ erkek çalışanların ortalama ücret düzeyini verir.
- Kadınların erkeklere maaş farkını ifaden $\hat{\delta}_0 = -2.51$ kadınların ortalama ücretleri ile erkeklerin ortalama ücretleri arasındaki farkı ifade eder.
- Bu sonuçlardan hareketle kadın çalışanların ortalama maaşı $\Rightarrow \hat{\beta}_0 + \hat{\delta}_0 = 7.1 - 2.51 = \boxed{\$4.59}$
- Çalışanları iki ayrı gruba ayırıp aritmetik ortalamalarını hesaplasaydık aynı sonuca ulaşırdık (modelde nicel bağımsız değişkenlerin olmaması durumunda geçerlidir sadece)

İki Nitel Değişken (iki kukla değişken)

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

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

$$\text{wage} = \beta_0 + \delta_0 \text{female} + \delta_1 \text{married} + u$$

$$\hat{\text{wage}} = \frac{6.18}{(0.296)} - \frac{2.29}{(0.302)} \text{female} + \frac{1.34}{(0.310)} \text{married}$$

$$n = 526 \quad \bar{R}^2 = 0.142$$

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

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

- female-in katsayısı ($\hat{\delta}_0$) ise evli olmayan kadınların ortalama ücretleri ile evli olmayan erkeklerin ortalama ücretleri arasındaki farktır \Rightarrow Bu değer $\hat{\delta}_0$ -dur = \$ -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 $\hat{\delta}_1$ -dur \Rightarrow \$ 1.34-dur.

$$\hat{\text{wage}} = \frac{6.18}{(0.296)} - \frac{2.29}{(0.302)} \text{female} + \frac{1.34}{(0.310)} \text{married}$$

$$n = 526 \quad \bar{R}^2 = 0.142$$

- Bu sonuçlardan hareketle evli olmayan kadın çalışanların (female = 1, married = 0) ortalama ücreti şöyle bulunabilir $\Rightarrow \hat{\beta}_0 + \hat{\delta}_0 = 6.18 - 2.29 = \$ 3.89$

- Evli kadın alımlarının ortalama ücreti
 \rightarrow female = 1, married = 1

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

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

- Evli erkekler (female = 0, married = 1) ve evli kadınlar (female = 1, married = 1) arasındaki farkı bakalım.

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

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

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

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

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

$$\ln \text{ wage} = \beta_0 + \delta_0 \text{ female} + \delta_1 \text{ married} + \beta_1 \text{ educ} + \beta_2 \text{ exper} \\ + \beta_3 \text{ exper}^2 + \beta_4 \text{ tenure} + \beta_5 \text{ tenure}^2 + \dots$$

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

(0.099)
(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)

$$n = 526 \quad \bar{R}^2 = 0.4351$$

• Ceteris paribus koşulu altında çalışanın evli olması ortalama ücretlerde fark yaratıyor mu?

$H_0: \delta_1 = 0$ vs. $H_1: \delta_1 \neq 0 \Rightarrow$ t-testi ile çözeceğiz.

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

$\alpha = 0/05$ anlamlılık düzeyi

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

$\hookrightarrow \underline{\underline{1.96}}$

Sonuç: t-ist < t-kritik olduğunda H_0 red edilemez ve çalışanın evli olması maaşlarda bir fark yaratmaz.

• Ceteris paribus koşulu altında çalışanın kadın olması ortalama ücretlerde fark yaratıyor mu?

$H_0: \delta_0 = 0$ vs. $H_1: \delta_0 \neq 0 \Rightarrow$ t-testi ile çözeceğiz.

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

$\alpha = 0/05$ anlamlılık düzeyi

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

↳ 1.96

Sonuç:

H_0 hipotezi red edilir ve çalışanın kadın olması ortalama maaşlarda düşüşe neden olur.



Gök Sayıda Kukla Değişkenler

- Önceki modeli farklı bir şekilde de kurabiliriz.
↳ female ve married kukla değişkenlerini 4 gruba ayırmak olacaktır.

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

(evli kadın) $marrfem = married \times female$

(bekar erkek) $singmale = (1 - married) \times (1 - female)$

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

ör: bekar kadın \Rightarrow $\left. \begin{array}{l} married = 0 \\ female = 1 \end{array} \right\} \begin{array}{l} marrmale = 0 \\ marrfem = 0 \\ singmale = 0 \\ singfem = 1 \end{array}$

↳ singmale (bekar erkek) baz grup olarak seçilmiş olsun.

$$\hat{\ln \text{wage}} = 0.32 + 0.21 \text{ marrmale} - 0.198 \text{ marrfem} - 0.11 \text{ singfem} \\ (0.101) \quad (0.055) \quad (0.058) \quad (0.055) \\ + 0.079 \text{ educ} + 0.027 \text{ exper} - 0.0005 \text{ expersq}$$

$$(0.006)$$

$$(0.005)$$

$$(0.001)$$

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

$(0.006) \qquad (0.0002)$

$$n = 526$$

$$\bar{R}^2 = 0.4525 \quad F = 55.246$$

- $\text{sing male} = 0.32 \Rightarrow$ bekar erkek için ortalama maaş $= 0.32$
 $\text{marr male} = 0.21 \Rightarrow$ evli erkek için ortalama maaş $= \frac{0.32}{+} + 0.21 = 0.53$

$$\text{marr fem} = -0.198 \Rightarrow \text{evli kadın için ortalama maaş} = 0.32 + (-0.198) = \underline{0.122}$$

$$\text{sing fem} = -0.11 \Rightarrow \text{bekar kadın için ortalama maaş} = 0.32 + (-0.11) = 0.21$$

Kukla Değişkenlerde Etkileşim.

- Yukarıdaki kukla değişkenler kullanılarak aynı regresyon-da çok sayıda farklı sabit terimler tahmin edebildik.
- Bu modellerde nicel değişkenlerin eğim katsayıları sabit kabul edilmişti. Şimdi farklı eğimler türetmeyi görelim.
- ÖR: ücret/maaş regresyonunda bir yıllık ilave bir eğitimin erkek ve kadın ücretlerinde yaratacağı etkinin farklı olup olmadığını ölçmek isteyelim.
- Bunun için cinsiyet kukla değişkeni female-i educ değişkeni ile etkileşime gireceğiz: $\text{female} \times \text{educ}$

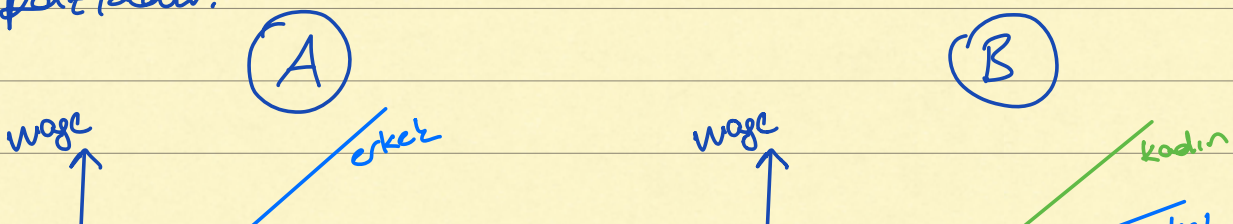
- Böylece eğitim düzeyi arttıkça kadınlarla erkekler arasındaki ücret farkının nasıl değiştiğini öğrenebileceğiz..

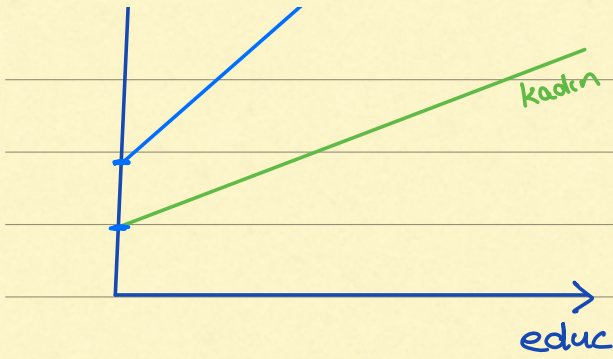
$$\ln \text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \delta_1 \text{female} \cdot \text{educ} + u$$

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

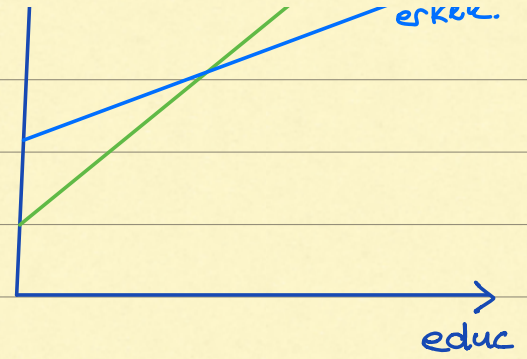
- $\text{female} = 0$ (erkek ise) \Rightarrow değer yerine koyulursa β_0 erkek çalışanların sabit terimini ifade eder.
- β_1 ise erkek çalışanların eğitim katsayısını ifade eder.
- δ_0 kadın çalışanlarla erkek çalışanlar arasındaki sabit terim farkıdır. Öyleyse kadınların sabit terimi $\Rightarrow \beta_0 + \delta_0$
- δ_1 ise kadın çalışanlar ile erkek çalışanlar arasındaki eğitim katsayısı farkıdır. Kadınlar için eğitim değişkeninin eğitim katsayısı $\Rightarrow \beta_1 + \delta_1$

- $\delta_1 > 0$ ise kadın çalışanlar için bir yıllık eğitimin getirisi erkek çalışanlardan daha fazladır.





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



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

• Şekil A'da, kadınlara ait regresyonun erkeklere ait regresyona kıyasla hem sabiti hem de eğimi daha düşüktür.

↳ ① Yani, kadınların saat-başına ücretleri tüm eğitim düzeylerinde erkeklerinkinden daha düşüktür.

② Aradaki fark eğitim düzeyi yükseldikçe artmaktadır.

• Şekil B'de ise, erkeklere kıyasla, kadın kadınlara ait regresyonun sabiti daha düşük, ancak educ değişkeninin eğimi 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)

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

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

- δ_1 istatistiki olarak anlamlı mıdır?

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

- Aynı eğitim düzeyine sahip erkek ve kadın çalışanların ortalama ücretleri aynı mıdır?

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

$\hookrightarrow \underline{\underline{F\text{-testi}}}$

$$\hat{\ln \text{wage}} = 0.389 - 0.227 \text{female} + 0.82 \text{educ}$$

$(0.119) \quad (0.168) \quad (0.008)$

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

(0.0131) (0.005) (0.00011)

$$0.032 \text{ tenure} - 0.00059 \text{ tenure}^2$$

(0.007) (0.00024)

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