ÇOKLU REGRESYON ANALİZİNDE NİTEL DEĞİŞKENLER

Hüseyin Taştan[1]

[1]Yıldız Teknik Üniversitesi
İktisat Bölümü

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

1 Aralık 2012

Regresyon Analizinde Nitel Bilgi

- Önceki bölümlerde bağımlı ve bağımsız değişkenlerimiz her zaman nicel (quantitative) bilgiler içeriyorlardı: ücretler, iş deneyimi, ev fiyatları, oda sayısı vs.
- Ancak, uygulamada çoğu kez nitel (qualitative) değişkenleri de regresyona dahil etmek zorundayız.
- Örnek nitel değişkenler: bireylerin cinsiyeti, ırkı, dini, doğduğu bölge (kuzey/ güney ya da batı/doğu gibi), eğitimi (lise/yüksek gibi)
- Bu değişkenler ikili/kukla (binary/dummy) değişkenlerle ifade edilebilir.

Nitel Değişkenler: Ders Planı

- Nitel Bilginin Betimlenmesi
- Tek Kukla Açıklayıcı Değişken
- Çok Kategorili Kukla Değişkenler
- Etkileşimli Kukla Değişkenler
- Bağımlı Değişkenin Nitel Olduğu Durum (Doğrusal Olasılık Modeli)

Nitel Bilgi

- Nitel enformasyon çoğu kez ikili (binary) yapı gösterir: erkek/kadın, arabası olan/olmayan, yerli/yabancı, ölüm cezası olan/olmayan ülkeler, vb.
- Bu tür enformasyonu ikili (binary) değişkenlerle ifade edeceğiz. Bunlara “sıfır/bir” (0/1) değişkenler ya da “gölge ya da kukla (dummy)” değişken de denir.
- Hangi kategoriye 1 ya da 0 dediğimiz regresyon sonuçlarını değiştirmez, ancak yorum yapabilmemiz için hangi kategoriye 1 hangisine 0 dediğimizi bilmemiz gerekir.
- Örnek: Cinsiyet kuklası: kadın=1, erkek=0,
- Evlilik kuklası: evli=1, evli değil=0.

Örnek Nitel Bilgi Veri Seti: Ücretler

person	wage	educ	exper	female	married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

Tek Kukla Değişkenli Regresyon

- İkili (nitel) enformasyonu regresyona nasıl dahil edeceğiz?
- x'lerden birisi kukla değişken olsun. Örnek:

$$wage = \beta_0 + \delta_0 \; female \; + \beta_1 \; educ + u$$

- Kadın çalışanlar için $female = 1$ erkekler için $female = 0$ değerini almaktadır.
- δ_0: Aynı eğitim düzeyine sahip bir erkekle bir kadının saat ücretleri arasındaki fark.
- Eğer δ_0 sıfırdan farklı çıkmazsa (t testi yapacağız) erkek-kadın arasında ücret farklılığı yok demektir.
- $\delta_0 < 0$ ise aynı eğitim düzeyinde kadın çalışanlar erkeklerden ortalamada daha az kazanmaktadırlar.

Tek Kukla Değişkenli Regresyon

$$wage = \beta_0 + \delta_0 \; female \; + \beta_1 \; educ + u$$

- Kadın çalışanlar için ücretlerin koşullu beklentisi:

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

- Erkek çalışanlar için

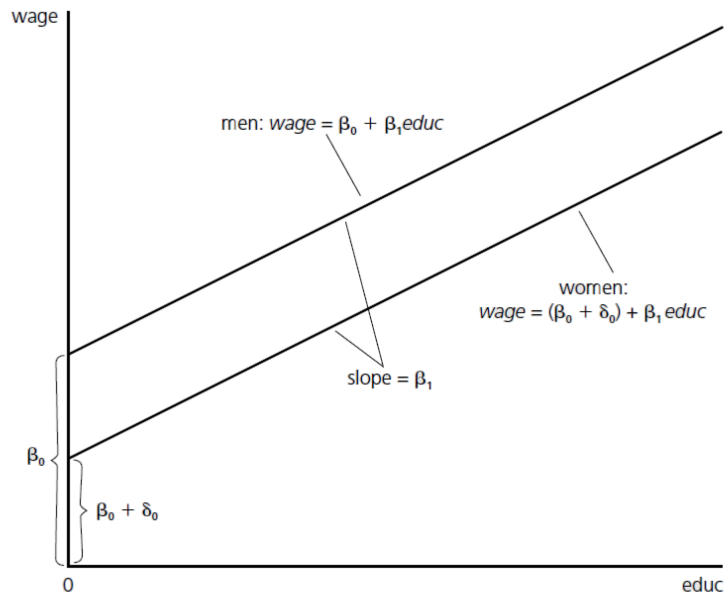
$$\mathsf{E}(wage|female = 0, educ) = \beta_0 + \beta_1 \; educ$$

- Kadın çalışanların ücret ortalaması ile erkeklerin ücret ortalaması arasındaki farkı alırsak:

$$\mathsf{E}(wage|female = 1, educ) - \mathsf{E}(wage|female = 0, educ)$$

$$= \beta_0 + \delta_0 + \beta_1 \; educ - (\beta_0 + \beta_1 \; educ) = \boldsymbol{\delta_0}$$

$\delta_0 < 0$ için Ücret Denklemi



Kukla Değişkenler

- Ücret denkleminde β_0 erkek çalışanların regresyonunun sabit terimidir.
- Kadın çalışanların sabit terimi ise $\beta_0 + \delta_0$ olur.
- Bir kukla değişken iki kategoriyi birbirinden ayırabilmektedir. Bu nedenle regresyona kadın ve erkekler için ayrı ayrı kuklalar koymaya gerek yoktur.
- Gerekli kukla sayısı $=$ kategori sayısı -1
- Cinsiyet için iki kukla kullansaydık tam çoklu-doğrusallık problemi ortaya çıkar. Buna "kukla değişken tuzağı" denir.
- Kukla değişkenin 0 değerini aldığı gruba baz grubu (ya da karşılaştırma grubu) adı verilir. Önceki ücret denkleminde baz grup erkek çalışanlardır.
- Female değişkeninin katsayısı baz grupla farkı (δ_0) verecektir.

Kukla Değişkenler

- Erkek çalışanların baz grup olduğu model: $female = 1$

$$wage = \beta_0 + \delta_0 \; female \; + \beta_1 \; educ + u$$

- Kadın çalışanların baz grup olduğu model: $male = 1$

$$wage = \alpha_0 + \gamma_0 \; male \; + \beta_1 \; educ + u$$

- Bu durumda kadın çalışanların sabit terimi: $\alpha_0 = \beta_0 + \delta_0$
- Erkek çalışanların sabit terimi: $\alpha_0 + \gamma_0 = \beta_0$
- Kukla değişkenlerin katsayılarının yorumunu baz gruba göre yapacağız. Bu nedenle hangi grubun baz grup olduğunu bilmemiz şarttır.

Kukla Değişkenler

- Diğer bir alternatif, regresyonda hiç sabit (intercept) kullanmayıp her bir kategori için bir kukla kullanmaktır.

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

- Modelde sabit terim olmadığı için kukla değişken tuzağına düşmeyiz.
- Bu modelde her grup için bir sabit terim tahmin edilecektir.
- Ancak sabit terimsiz bir regresyonda R^2 hesaplanamaz.
- Ayrıca testlerin yapılması biraz daha zordur.
- Bu nedenlerle yukarıdaki formülasyonu tercih etmiyoruz.

Kukla Değişkenler: Birden Fazla Nicel Değişken

- Birden fazla açıklayıcı değişken olması kukla değişkenlerin yorumunu değiştirmez:

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

- δ_0: Aynı eğitim, tecrübe ve kıdem düzeyine sahip bir kadın çalışan ile erkek çalışan arasındaki sabit terim farkı.
- $H_0 : \delta_0 = 0$ boş hipotezi $H_1 : \delta_0 < 0$ lehine reddedilirse kadın çalışanlara karşı (negatif) ücret ayrımcılığı yapılıyor demektir.
- Bunu standart t testiyle sınayabiliriz.

Örnek: Ücret Denklemi

$$\widehat{\text{wage}} = \underset{(0.725)}{-1.57} - \underset{(0.265)}{1.81}\ \text{female} + \underset{(0.049)}{0.572}\ \text{educ} + \underset{(0.0116)}{0.025}\ \text{exper} + \underset{(0.021)}{0.141}\ \text{tenure}$$

$$n = 526 \quad R^2 = 0.364 \quad F(4, 521) = 74.398 \quad \hat{\sigma} = 2.9576$$

- Kadın çalışanlar erkek çalışanlara göre ceteris paribus 1.81$ daha az kazanmaktadır. Başka bir ifadeyle aynı eğitim, tecrübe ve kıdem düzeyine sahip bir kadın ile erkek çalışan arasındaki saatlik ücret farkı 1 dolar 81 cent'tir.
- -1.57 erkek çalışanların sabit terimidir. educ, exper ve tenure aynı anda 0 olduğunda tahmin edilen ücret düzeyidir. Bu değişkenler hiçbir zaman 0 olmadığından negatif olmasının bir önemi yoktur.

Kukla Değişkenler

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

$$\widehat{\text{wage}} = \underset{(0.21)}{7.1} - \underset{(0.303)}{2.51}\ \text{female}$$

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

- Bu regresyonda sabit terim, 7.1$ erkek çalışanların ortalama ücret düzeyini vermektedir.
- Female'in katsayısı ise kadınların ortalama ücretleri ile erkeklerin ortalama ücretleri arasındaki farktır. Bu fark 2.51$'dır.
- Bu sonuçlardan hareketle kadın çalışanların ortalama ücretlerini bulabiliriz: $7.1 - 2.51 = 4.59\$$
- Çalışanları iki gruba ayırıp aritmetik ortalamalarını hesaplasaydık aynı sonuca ulaşırdık (modelde nicel x değişkenlerinin olmadığı durumda geçerli).

Nitel Değişkenler

- Gruplar arasında fark olup olmadığı standart t testiyle hesaplanabilir.
- Bu testin geçerli olması için grupların varyansının aynı olması gerekir (sabit varyans)
- Bazı durumlarda kukla değişkenler cinsiyet gibi önceden belirlenmiş (predetermined) grupları değil bireylerin seçimlerini temsil edebilirler (örneğin araba sahibi olma, bilgisayar sahibi olma, vs.)
- Bu durumda nedenselliğin yönünü tayin edebilmemiz için bağımlı değişkeni etkileyen tüm faktörleri kontrol etmemiz gerekebilir.
- Örnek: üniversite başarısı ve bilgisayar sahipliği: GPA ortalamasının bilgisayar sahibi olan öğrenciler arasında olmayanlara göre daha yüksek çıktığını düşünelim. Acaba "*bilgisayar sahibi olmak üniversite başarısını arttırır*" diyebilir miyiz? Nedenselliğin yönü nedir?

Kukla Değişkenler

- Modele iki nitel değişken ekleyelim: $female = 1$ çalışan kadınsa; $married = 1$ çalışan evliyse

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

$$n = 526 \quad \bar{R}^2 = 0.1429$$

- Bu regresyonda baz grup "evli olmayan erkek çalışanlar"dır ($female = 0, married = 0$). Sabit terim, 6.18, bu grubu temsil etmektedir.
- Female'in katsayısı ise kadınların ortalama ücretleri ile evli olmayan erkeklerin ortalama ücretleri arasındaki farktır. Bu fark 2.51$'dır.
- Bu sonuçlardan hareketle evli olmayan kadın çalışanların ($female = 1, married = 0$) ortalama ücretlerini bulabiliriz: $6.18 - 2.29 = 3.89\$$

Kukla Değişkenler

$$\widehat{\text{wage}} = \underset{(0.296}{6.18} - \underset{(0.302)}{2.29}\ \text{female} + \underset{(0.310)}{1.34}\ \text{married}$$
$$n = 526 \quad \bar{R}^2 = 0.1429$$

- Benzer şekilde evli kadın çalışanların ($female = 1, married = 0$) ortalama ücretleri: $6.18 - 2.29 + 1.34 = 5.23\$$
- Evli erkeklerle evli kadınlar arasındaki ücret farkı: $(6.18 + 1.34) - (6.18 - 2.29 + 1.34) = 2.29\$$ erkekler ortalamada daha fazla kazanıyor.
- Nicel değişkenler ekleyerek (eğitim, tecrübe, kıdem, vs.) ceteris paribus yorumu yapabileceğimiz modeller geliştirebiliriz.

Ücret Denklemi

$$\widehat{\text{lwage}} = \underset{(0.098)}{0.42} - \underset{(0.036)}{0.29}\ \text{female} + \underset{(0.040)}{0.05}\ \text{married} + \underset{(0.007)}{0.08}\ \text{educ}$$
$$+ \underset{(0.005)}{0.03}\ \text{exper} - \underset{(0.0001)}{0.0005}\,\text{expersq} + \underset{(0.007)}{0.03}\ \text{tenure} - \underset{(0.0002)}{0.0006}\,\text{tenursq}$$
$$n = 526 \quad \bar{R}^2 = 0.4351$$

- Diğer faktörler kontrol edildikten sonra çalışanın evli olması ortalama ücretlerde bir fark yaratıyor mu?
- married katsayısı 0.05 olarak tahmin edilmiş. t oranı: $0.05/0.04 = 1.25$. Boş hipotez reddedilemez.
- Eğitim, tecrübe, kıdem ve cinsiyet kuklası kontrol edildikten sonra evli çalışanlar ile evli olmayanlar arasındaki ortalama ücretler aynıdır.
- Cinsiyet kuklasının katsayısı azaldı. Ancak işareti hala negatif ve istatistik bakımından anlamlı.

Çok Sayıda Kukla Değişkenler

- Önceki modeli farklı bir şekilde de kurabilirdik. $female$ ve $married$ kuklaları çalışanları 4 gruba ayırmaktadır.
- Bu gruplar için ayrı ayrı kuklalar oluşturabilirdik:

$$marrmale = married \times (1 - female)$$
$$marrfem = married \times female$$
$$singfem = (1 - married) \times female$$
$$singmale = (1 - married) \times (1 - female)$$

- $marrmale$ evli erkek çalışanları, $marrfem$ evli kadın çalışanları, $singfem$: bekar kadın çalışanları ve $singmale$ bekar erkek çalışanları temsil etmektedir.
- Bunlardan birini baz grup seçerek regresyonu yeniden tahmin edebiliriz.
- Baz grup: $singmale$ olsun.

Çok Sayıda Kukla Değişkenler

$$\widehat{\text{lwage}} = \underset{(0.101)}{0.32} + \underset{(0.055)}{0.21}\ \text{marrmale} - \underset{(0.058)}{0.198}\,\text{marrfem} - \underset{(0.055)}{0.11}\ \text{singfem}$$
$$+ \underset{(0.006)}{0.079}\,\text{educ} + \underset{(0.005)}{0.027}\,\text{exper} - \underset{(0.0001)}{0.0005}\,\text{expersq} + \underset{(0.006)}{0.029}\,\text{tenure}$$
$$- \underset{(0.0002)}{0.0005}\,\text{tenursq}$$
$$n = 526 \quad \bar{R}^2 = 0.4525 \quad F(8, 517) = 55.246 \quad \hat{\sigma} = 0.39329$$

Kukla Değişkenlerle Etkileşim

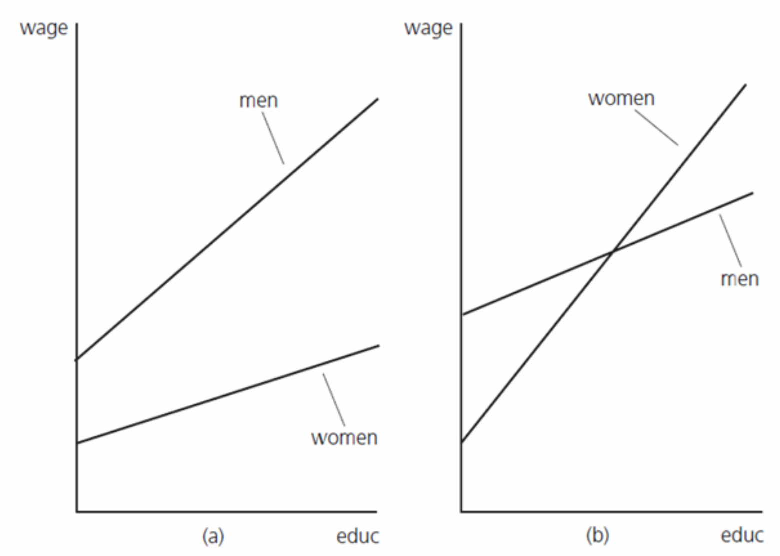
- Yukarıda kukla değişkenler kullanarak aynı regresyonda çok sayıda farklı sabit terimler tahmin edebildik.
- Bu modellerde nicel değişkenlerin eğim katsayıları sabit kabul edilmişti.Şimdi farklı eğimler türetmeyi görelim.
- Örneğin, ücret regresyonunda bir yıllık ilave bir eğitimin erkek ve kadın ücretlerinde yaratacağı etkinin farklı olup olmadığını ölçmek isteyelim.
- Bunun için cinsiyet kukla değişkenini ($female$) eğitim ($educ$) ile etkileşimli olarak modele ekleyeceğiz: $female \times educ$.
- Böylece eğitim düzeyi arttıkça kadınlarla erkekler arasındaki ücret farkının nasıl değiştiğini ölçebileceğiz.

Kukla Değişkenlerle Etkileşim: Ücret Denklemi

$$\log(wage) = (\beta_0 + \delta_0 female) + (\beta_1 + \delta_1 female) \times educ + u$$

- $female = 0$ değeri yerine konursa, β_0'ın erkek çalışanların sabit terimi olduğu kolayca görülür.
- β_1 ise erkek çalışanların eğim katsayısıdır.
- δ_0 kadın çalışanlar ile erkek çalışanlar arasındaki sabit farkıdır. Öyleyse kadınların sabit terimi $\beta_0 + \delta_0$
- δ_1 ise kadın çalışanlar ile erkek çalışanlar arasındaki eğim katsayısı farkıdır. Kadınlar için eğitim değişkeninin eğim katsayısı: $\beta_1 + \delta_1$
- $\delta_1 > 0$ ise kadın çalışanlar için bir yıllık eğitimin getirisi erkek çalışanlardan daha fazladır diyebiliriz.

Cinsiyet ve Eğitim Etkileşimi: sol: $\delta_0 < 0, \delta_1 < 0$; sağ: $\delta_0 < 0, \delta_1 > 0$



Cinsiyet ve Eğitim Etkileşimi

- Şekil(a)'da, kadınlara ait regresyonun erkeklere ait regresyonunkilere kıyasla hem sabiti, hem de eğimi daha düşüktür.
- Yani, kadınların saat-başına ücretleri tüm eğitim düzeylerinde erkeklerinkinden daha düşüktür.
- Aradaki fark eğitim düzeyi yükseldikçe artmaktadır.
- Şekil (b)'de ise, erkeklere kıyasla, kadın regresyonunun sabiti daha düşük, ancak educ değişkeninin eğimi daha yüksektir. Bu şu anlama gelir: düşük eğitim düzeylerinde kadınlar, yüksek eğitim düzeylerinde ise erkekler daha az kazanmaktadır.

Cinsiyet ve Eğitim Etkileşimi

$$\log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female \cdot educ + u$$

- Modeli yukarıdaki gibi formüle edip tahmin edebiliriz. Bunun için modele $female$ kuklasının yanı sıra $female \times educ$ etkileşim değişkeninin eklenmesi yeterli olacaktır.
- Etkileşim değişkeni erkek çalışanlar için 0, kadın çalışanlar için ise $educ$ değerini alacaktır.
- $H_= : \delta_1 = 0,\ H_1 : \delta_1 \neq 0$. Boş hipotez: "Bir yıllık fazladan eğitimin getirisi kadın ve erkek çalışanlar için aynıdır"
- $H_= : \delta_0 = 0, \delta_1 = 0$: "Aynı eğitim düzeyine sahip kadın ve erkek çalışanların ortalama ücretleri aynıdır". F testi ile sınanabilir.

Cinsiyet ve Eğitim Etkileşimi

$$\begin{aligned}
\widehat{\log(wage)} = {} & \underset{(.119)}{.389} - \underset{(.168)}{.227}\ female + \underset{(.008)}{.082}\ educ \\
& - \underset{(.0131)}{.0056}\ female{\cdot}educ + \underset{(.005)}{.029}\ exper - \underset{(.00011)}{.00058}\ exper^2 \\
& + \underset{(.007)}{.032}\ tenure - \underset{(.00024)}{.00059}\ tenure^2 \\
& n = 526,\ R^2 = .441.
\end{aligned}$$

Kukla Değişkenlerle Etkileşim

- Bir yıllık ilave eğitimin getirisi erkek ve kadınlar için aynıdır. $Female \cdot educ$ etkileşim değişkeninin katsayısının t istatistiği $-0.0056/0.0131 = -0.43$'dür ve istatistiksel olarak anlamsızdır.
- Regresyonda etkileşim değişkeni bulunduğundan, $female$ değişkeninin katsayısı, $educ = 0$ iken erkek ve kadınlar arasındaki ücret farkını ölçecektir.
- Örnekte eğitim yılı sıfır olan kişi bulunmadığı için ve ayrıca $female$ ile $female \cdot educ$ değişkenleri arasında çoklu-bağıntı (multicollinearity) olduğundan standart hatası yüksek, dolayısıyla da t değeri düşüktür (-1.35).

Kukla Değişkenlerle Etkileşim

- Bu nedenle, $female$'in katsayısını şöyle tahmin edeceğiz: Etkileşim değişkeninde $educ$ yerine onun ortalamasından farkını kullanacağız. $educ$ değişkeninin ortalaması 12.5 yıldır.
- Yeni etkileşim değişkenimiz $female \times (educ - 12.5)$ olacaktır.
- Bu değişiklikten sonra regresyonu yeniden tahmin edeceğiz. Bu durumda, sabit terim $educ = 12.5$ iken geçerli olan ücret farkını gösterecektir.
- Yani ortalama eğitim düzeyinde ücret farkı olup olmadığını göreceğiz.
- F testi sonucu, bu katsayıların ikisinin birden sıfıra eşit olmadığını gösteriyor. O halde sabit terim erkek ve kadın için farklıdır.

Kukla Değişkenlerle Etkileşim

```
. gen femeduc1=female*(educ-12.5)

. reg  lwage female educ femeduc1 exper expersq tenure tenursq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  7,   518) =   58.37
       Model |  65.4081534     7  9.34402192           Prob > F      =  0.0000
    Residual |   82.921598   518  .160080305           R-squared     =  0.4410
-------------+------------------------------           Adj R-squared =  0.4334
       Total |  148.329751   525   .28253286           Root MSE      =   .4001

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.296345   .0358358    -8.27   0.000    -.3667465   -.2259436
        educ |   .0823692   .0084699     9.72   0.000     .0657296    .0990088
    femeduc1 |  -.0055645   .0130618    -0.43   0.670    -.0312252    .0200962
       exper |   .0293366   .0049842     5.89   0.000      .019545    .0391283
     expersq |  -.0005804   .0001075    -5.40   0.000    -.0007916   -.0003691
      tenure |   .0318967    .006864     4.65   0.000      .018412    .0453814
     tenursq |    -.00059   .0002352    -2.51   0.012    -.001052    -.000128
       _cons |    .388806   .1186871     3.28   0.001     .1556388    .6219732
------------------------------------------------------------------------------

. test  female femeduc1

 ( 1)  female = 0
 ( 2)  femeduc1 = 0

       F(  2,   518) =   34.33
            Prob > F =    0.0000
```


Kukla Bağımlı Değişken: Doğrusal Olasılık Modeli

- Şimdiye kadar bağımlı değişken y niceldi (quantitative), örneğin, ücretler, başarı notu, logaritmik fiyatlar, vs.
- Bağımlı değişkenin nitel (qualitative) olduğu, $y = \{0, 1\}$ değerlerini aldığı, modeller tanımlamak isteyebiliriz.
- Örneğin, ev sahibi olanlar/olmayanlar, evli olanlar/olmayanlar, çalışanlar/çalışmayanlar, vs.
- Bu durumda modeli nasıl yorumlamak gerekir? OLS ile tahmin edilebilir mi?
- $y = 1$ olduğu duruma başarı, $y = 0$ olduğu duruma yenilgi diyelim. Açıktır ki bu durumda y bir Bernoulli değişkenidir.

Kukla Bağımlı Değişken: Doğrusal Olasılık Modeli

- Standart varsayımlar sağlanıyorsa koşullu beklenen değer aşağıdaki gibi yazılabilir:

$$\mathsf{E}(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

- y sadece 0 ya da 1 değerini aldığından bu beklenen değer aşağıdaki gibi yazılabilir (ikili tepki (binary response) modeli):

$$\begin{aligned} \mathsf{E}(y|x) &= P(y = 1|x) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \end{aligned}$$

- Başarı olasılığına $p(x) = P(y = 1|x)$ diyelim. Yukarıdaki ifade $p(x)$'in x'lerin doğrusal bir fonksiyonu olduğunu söylemektedir.
- Yenilgi olasılığı: $P(y = 0|x) = 1 - P(y = 1|x)$

Kukla Bağımlı Değişken: Doğrusal Olasılık Modeli

- y'nin ikili değerler aldığı bu modele doğrusal olasılık modeli (linear probability model) adı verilir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- Bu modelde x'ler nitel ya da nicel olabilir.
- Eğim katsayıları artık başarı olasılığındaki değişme olarak yorumlanır:

$$\Delta P(y = 1|x) = \beta_j \Delta x_j$$

- OLS örneklem regresyon fonksiyonu

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

- Burada $\hat{y}$ başarı olasılığının OLS tahminidir.

Örnek: Kadınların Emek Piyasasına Katılımı, mroz.gdt

- y değişkeni $inlf$ eğer örneğe giren evli kadınlar 1975 yılının herhangi bir ayında işgücüne katılmışsa (ücretli çalışmışsa) 1, hiç katılmamışsa 0 değerini almaktadır.
- Açıklayıcı değişkenler şunlardır:
- $nwifeinc$: kocanın geliri (bin$),
- $kidslt6$: 6 yaşından küçük çocuk sayısı,
- $kidsge6$: 6-18 yaş arasındaki çocuk sayısı,
- $educ, exper, age$
- Model

$$\widehat{inlf} = \hat{\beta}_0 + \hat{\beta}_1 nwifeinc + \hat{\beta}_2 educ + \ldots + \hat{\beta}_7 kidsge6$$

Kadınların İşgücüne Katılımı, mroz.gdt

Model 1: OLS, using observations 1–753
Dependent variable: inlf

	Coefficient	Std. Error	t-ratio	p-value
const	0.585519	0.154178	3.7977	0.0002
nwifeinc	−0.00340517	0.00144849	−2.3508	0.0190
educ	0.0379953	0.00737602	5.1512	0.0000
exper	0.0394924	0.00567267	6.9619	0.0000
expersq	−0.000596312	0.000184791	−3.2270	0.0013
age	−0.0160908	0.00248468	−6.4760	0.0000
kidslt6	−0.261810	0.0335058	−7.8139	0.0000
kidsge6	0.0130122	0.0131960	0.9861	0.3244

Mean dependent var	0.568393	S.D. dependent var	0.495630
Sum squared resid	135.9197	S.E. of regression	0.427133
R^2	0.264216	Adjusted R^2	0.257303
$F(7, 745)$	38.21795	P-value(F)	6.90e–46

Kadınların İşgücü Piyasasına Katılımı

- $kidsge6$ hariç tüm değişkenler istatistiksel olarak anlamlı çıkmıştır. Katsayıların işaretleri iktisat teorisi ve sağduyu ile uyumludur.
- Katsayıların yorumları: Örneğin $educ$ değişkeninin katsayısı, 0.038, diğer faktörler sabitken, ilave bir yıl eğitimin işgücüne katılma $(inlf = 1)$ olasılığını 0.038 kadar arttırdığı anlamına gelmektedir.
- $nwifeinc$ katsayısının yorumu: kocanın yıllık maaşı 10 birim (yani 10000\$) artarsa, kadınların işgücüne katılma olasılığı 0.034 düşer.
- $exper$ karesel biçimde alınmış: İşareti $+$, karesinin işareti $-$ olduğu için, deneyim işgücüne katılma şansını arttırıyor, ancak bu artış giderek azalıyor.

Kadınların İşgücü Piyasasına Katılımı

- Beklendiği üzere kadının işgücüne katılma olasılığını belirleyen en önemli faktör küçük çocuğunun olmasıdır. $kidslt6$ değişkeninin katsayısı -0.262'dir.
- Yani,ceteris paribus, 6 yaşından küçük çocuk sayısında 1 birimlik (1 çocuk) artış kadının işgücüne katılma olasılığını 0.262 kadar azaltmaktadır.
- Örnekte 6 yaşından küçük en az bir çocuğu olan kadınlar toplamın beşte biri kadardır.

Doğrusal Olasılım Modelinin (LPM) Yetersizlikleri

- Bağımsız değişkenlerin belli değerleri için sıfırdan küçük ya da birden büyük olasılık tahminleri ($\hat{y}$) verebilir, ki bu olasılık ilkeleri ile çelişir.
- Önceki örnekte içerilen 753 kadından sadee 16'sı için $\widehat{inlf} < 0$, 17'si için de $\widehat{inlf} > 0$ çıkmıştır.
- Ancak, LPM modelinin asıl yetersizliği bu değildir. Modelde, olasılığın x değerlerine doğrusal bir şekilde bağlı olması büyük mahsur yaratmaktadır.
- Örneğimizde bir kadının ilave bir küçük çocuğa sahip olması onun işgücüne katılma şansını 0.262 azaltmaktadır.İlişki doğrusal olduğu için, bu 0.262 rakamı, sıfırıncı çocuktan 1.ci çocuğa geçişte de, diyelim, 3.cü çocuktan 4.cü çocuğa geçişte de aynıdır.
- Oysa bu gerçekçi değildir. İlişki büyük olasılıkla doğrusal-olmayan (nonlinear) niteliktedir. Yani ilk çocukta annenin evde kalma olasılığı 2.ci çocuk olduğunda evde kalma olasılığından daha yüksektir.

Doğrusal Olasılım Modelinin (LPM) Yetersizlikleri

- Bu yetersizliklerine rağmen doğrusal olasılık modeli (linear probability model, LPM) faydalı bir modeldir ve uygulamada sıkça kullanılmaktadır.
- Özellikle,bağımsız değişkenler (x'ler) kendi ortalamalarına yakın değerler aldıklarında LPM oldukça başarılı tahmin vermektedir.
- Önceki örnekte kadınların %96'sının ya hiç çocuğu yoktur ya da 1 tane altı yaşından küçük çocuğa sahiptir.
- Dolayısıyla $kidslt6$ değişkeninin katsayısı (-0.262) aslında bize ilk çocuğun annenin işgücüne katılma şansını ne kadar azalttığını vermektedir.
- Bu katsayı tahminini, diyelim, 4.cü çocuktan 5.ci çocuğa geçiş vb. değerler için kullanmamak gerekir.

Doğrusal Olasılım Modelinin (LPM) Yetersizlikleri

- Doğrusal olasılık modeli MLR.5: Sabit Varyans varsayımını sağlamaz.
- Hata terimi sabit varyanslı değildir. y'nin ikili değerler alan bir Bernoulli değişkeni olduğunu hatırlarsak varyans aşağıdaki gibi türetilebilir:

$$\text{Var}(u|x) = \text{Var}(y|x) = p(x) \cdot [1 - p(x)]$$

- $p(x)$, x'lerin lineer kombinasyonu olduğundan $\text{Var}(u|x)$ değişkendir (heteroscedastic).
- Bu durumad OLS sapmasız, tutarlı, ancak etkin değildir. Standart hatalar geçerli olmaz.
- OLS'den daha etkin tahmin yöntemleri bulunabilir. Örneğin Ağırlıklandırılmış En Küçük Kareler.