# Bölüm 7

# Nitel Tepki Bağlanım Modelleri

## 7.1 Nitel Tepki ve Doğrusal Olasılık Modeli

### 7.1.1 Nitel Bağımlı Değişkenler

- Daha önceki bölümlerde açıklayıcı değişken olarak nicel ya da nitel değişkenler kullanılabileceğini görmüştük.
- Bağımlı değişken ise bir nicel değişken olmak zorunda idi.
- Bu bölümde, bağımlı değişkeni sınırlı değerler alan, örnek olarak bir kukla değişken olabilen modelleri ele alacağız.
- Bu tür modellere özgü bazı tahmin sorunlarını incelerken, almaşık nitel tepki modellerini de tanıtacağız.

Bağımlı değişkenin 0 ve 1 gibi yalnızca iki değer alabileceği modellere ilişkin olarak şu örnek konu başlıkları gösterilebilir:

- Ev sahibi olup olmamayı belirleyen etmenler
- Sendika üyesi olup olmamanın nedenleri
- Bir kredi başvurusunun reddedilip reddedilmeyeceği
- Bir seçimde evet ya da hayır oyunu nelerin belirlediği
- İşgücüne katılıp katılmamanın nelerden etkilendiği
- Kişilerin sigorta yaptırıp yaptırmayacakları
- Şirketlerin hisse senedi çıkartıp çıkartmayacakları

- Şirketlerin ele geçirilmeye hedef olup olmayacakları
- Bir ülkede idam cezasının olup olmayacağı

### 7.1.2 Doğrusal Olasılık Modeli

- Nitel bağımlı değişkene örnek olarak şu modeli ele alalım.

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Burada $X$ hanehalkının gelirini göstermektedir.
- $Y = 1$, aile ev sahibi ise; $Y = 0$, eğer değilse.
- Bir kukla değişken olan $Y$'yi $X$ açıklayıcı değişken(ler)inin doğrusal işlevi olarak belirten yukarıdaki gibi modellere *"doğrusal olasılık modeli"* (linear probability model), ya da kısaca *"DOM"* (LPM) adı verilir.
- Bu modellerde $X_i$ veriliyken $Y_i$'nin koşullu beklenen değeri, olayın gerçekleşme koşullu olasılığı olarak yorumlanabilir.
- Eldeki modele doğrusal olasılık denilme nedenini görmek için $E(u_i) = 0$ varsayımını anımsayalım ve şunu yazalım:

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_i$$

- $Y_i = 1$ olduğunda olayın gerçekleştiğini ve bunun olasılık değerinin de $P_i$ olduğunu söyleyelim. Bu durumda, $Y$'nin olasılık dağılımı şöyledir:

| $Y$ | Olasılık |
|---|---|
| 0 | $1 - P_i$ |
| 1 | $P_i$ |
| Toplam | 1 |

- Beklenen değer tanımından yararlanarak şunu görebiliriz:

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i$$

- $P_i$ bir olasılık olduğu için burada $0 \leq E(Y_i|X_i) \leq 1$ şeklinde bir sınırlama olduğu unutulmamalıdır.

### 7.1.3 DOM Tahminindeki Güçlükler

Yukarıda gördüklerimiz, SEK yönteminin kolaylıkla nitel bağımlı değişkenler için de kullanılabileceği kanısını uyandırıyorsa da durum gerçekte böyle değildir. DOM tahmini, aşağıda gösterilen dört sorunu da beraberinde getirmektedir.

1. Bozukluk terimi $u$'nun normal-dışılığı
2. Bozukluklarda farklıserpilimsellik görülmesi
3. $R^2$'nin yakışma ölçütü olarak kuşkulu değeri
4. $0 \leq E(Y_i|X_i) \leq 1$ koşulunun sağlanamaması

**Bozukluk Terimi $u$'nun Normal-dışılığı**

- DOM tahminininde $u_i$'lerin normal dağılması olanaksızdır.
- Aşağıda da görüldüğü gibi, $Y_i$'ler yalnızca 2 değer aldıkları için, 2 farklı $u_i$ kümesi ortaya çıkar.

| $u_i = Y_i - \beta_1 - \beta_2 X_i$ | | |
|---|---|---|
| $Y_i = 1$ | ise | $u_i = 1 - \beta_1 - \beta_2 X_i$ |
| $Y_i = 0$ | ise | $u_i = -\beta_1 - \beta_2 X_i$ |

- Bu durumda $u_i$'ler normal dağılımı değil, kesikli Bernouilli dağılımını izlerler.
- Diğer taraftan, nokta tahminleri yansız olmayı sürdürürler.
- Ayrıca merkezi limit kanıtsavına göre örneklem büyüklüğü artarken kalıntıların normale yaklaşacağı unutulmamalıdır.
- Dolayısıyla, büyük örneklemlerde $u$'nun normal-dışılığı tahmin ve çıkarsama açısından bir sorun yaratmayabilir.

**Bozukluklarda Farklıserpilimsellik**

- Bozuklukların Bernouilli dağılımına uyduğu bilindiğine göre, DOM tahmininde $u_i$'lerin aynıserpilimsel olduğu varsayımını korumak da olanaksızdır.
- Anımsayacak olursak, ikiterimli dağılımın genel biçimi olan Bernoulli dağılımının ortalaması $p$, varyansı $p(1-p)$'dir.

- Buna göre, doğrusal olasılık modelinin varyansı da şu olur.

$$\text{var}(u_i) = P_i(1 - P_i)$$

- $P_i = E(Y_i|X_i) = \beta_1 + \beta_2 X_i$ olduğuna göre, $u_i$ sonuçta $X_i$ değerlerine bağlıdır ve bu nedenle aynıserpilimsel olamaz.
- Farklıserpilimsellik altında SEK tahminlerinin yansız olmayı sürdürürken enaz varyanslı olamadıklarını anımsayalım.
- Büyük örneklemlerde bu da DOM için sorun olmayabilir. Farklıserpilimsellik ağırlıklı en küçük kareler ya da White ölçünlü hataları kullanılarak aşılmaya çalışılabilir.

### $R^2$'nin Yakışma Ölçütü Olarak Kuşkulu Değeri

- Nitel bağımlı değişkenler, tanım gereği, gözlenen $Y_i$'lerin yalnızca kesikli 0 ve 1 değerlerini alabilmeleri demektir.
- DOM tahmininden elde edilen $\hat{Y}_i$'ler ise yakıştırılan doğru üzerinde farklı ve sürekli değerler alabilirler.
- DOM'ların böyle bir serpilime iyi yakışması beklenemez.
- Bu nedenle, DOM tahmininde $R^2$ genellikle düşük çıkar. Uygulamada genellikle 0.2 ile 0.6 arası değerler beklenir.
- Genel olarak, her türden nitel bağımlı değişken modelinde belirleme katsayısı $R^2$'yi bir özet istatistik olarak kullanmak sakıncalı kabul edilmektedir.

### $0 \leq E(Y_i|X_i) \leq 1$ Koşulunun Sağlanamaması

- DOM tahminine ilişkin asıl ciddi bir sorun, $0 \leq E(Y_i|X_i) \leq 1$ koşulunun sağlanamamasıdır.
- Bu modeller $X$ veriliyken $Y$ olayının gerçekleşme koşullu olasılığını ölçtüğü için, $E(Y_i|X_i)$ değerinin 0 ile 1 arasında yer alması önemlidir.
- SEK yöntemi böyle bir matematiksel sınırlama içermediği için, DOM tahmini sonrasında yakıştırılan değerlerin eksi değerli ya da 1'den büyük çıkmasına sıkça rastlanır.

- Böyle durumlarda eksi değerli $\hat{Y}_i$'leri sıfır, 1'den büyük $\hat{Y}_i$'leri ise 1 varsaymak yoluna gidilebilir.
- Ancak böyle sakıncalı ek varsayımlara gerek yoktur çünkü tahmin edilen olasılıkların 0 ve 1 arasında olmasını güven altına alan almaşık yöntemler bulunmaktadır.

**DOM Açıklayıcı Örnek**

- Açıklayıcı örnek olarak, çoğu ilde tek bir "ticaret ve sanayi odası" varken, bazı büyük illerde ise sanayi odasının ayrı bir kuruluş olarak hizmet verdiğini göz önüne alalım.
- Bir ilde sanayi odası olup olmayacağını sanayi sektöründe faaliyet gösteren firma sayısı ile açıklamak istiyor olalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- $X_i$ burada toplam firma sayısını (100 birim) göstermektedir.
- $Y_i = 1$, ilde sanayi odası var ise; $Y_i = 0$, eğer yok ise.
- Önsel beklentimiz, $\hat{\beta}_2$'nın artı değerli ve $(0{,}1)$ aralığında tahmin edileceği yönündedir.

- TOBB Sanayi Veri Tabanı verilerine dayalı DOM bağlanım sonuçları aşağıdaki gibidir.

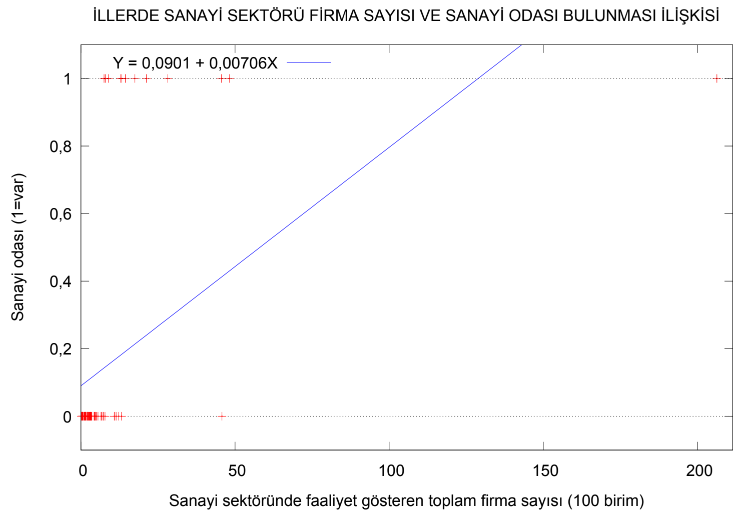
$$\begin{array}{llll} \hat{Y}_i = & 0{,}0901 & +\,0{,}0070\,X_i & \\ \text{öh} & (0{,}0376) & (0{,}0015) & \\ t & (2{,}3933) & (4{,}8092) & r^2 = 0{,}2287 \end{array}$$

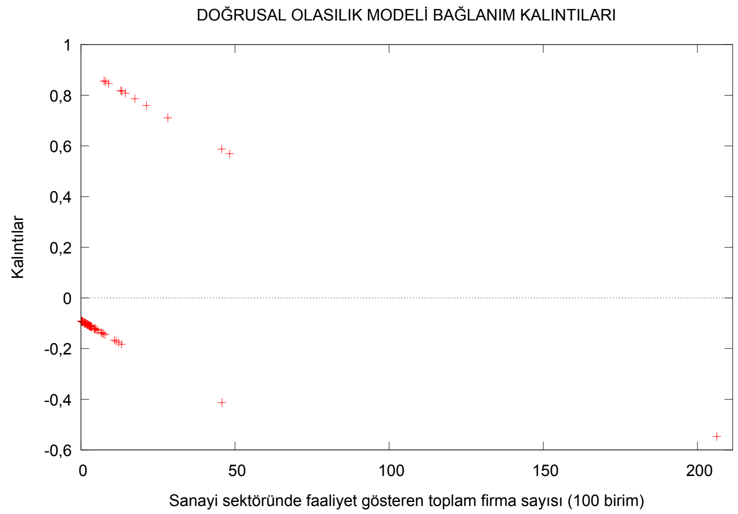
- Sonuçlar, ilde açılacak her 100 yeni firmanın sanayi odası kurulma olasılığını yüzde 0,7 artıracağını göstermektedir.
- İlişki doğrusal tahmin edildiği için, ilde firma sayısı 1000 de olsa 5000 de olsa olasılığın aynı kaldığı varsayılmaktadır.
- Kukla bağımlı değişkene bir doğru yakıştırmak güç olduğu için, $r^2$ değeri beklenildiği gibi düşük bulunmuştur.

- Ayrıca, toplam 20601 sanayi firması bulunan İstanbul için tahmin edilen olasılık 1'den büyük çıkmaktadır:

$$0{,}0901 + (0{,}0070 \times 206{,}01) = 1{,}547$$

- $Y_i$'ye bağlı iki ayrı kalıntı kümesi olduğundan, hataların normal dağılması da söz konusu değildir.

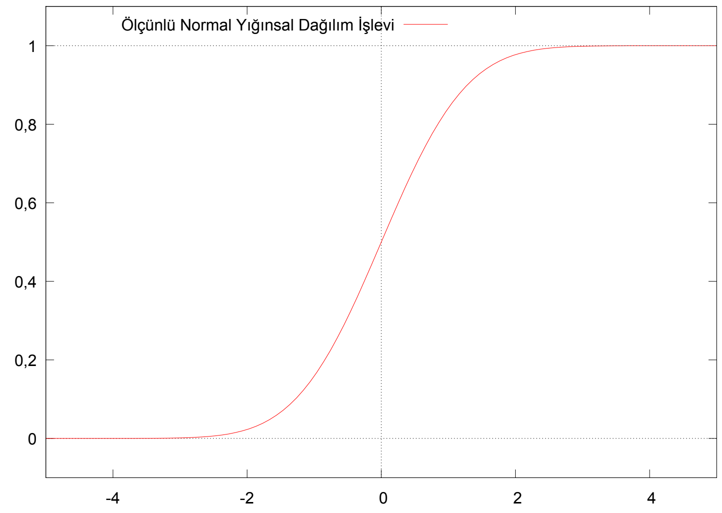


# 7.2 Doğrusal-Dışı Yaklaşım ve Olabirim Modeli

## 7.2.1 Doğrusal Olasılık Modelinin Almaşıkları

- Yukarıda tartışılan teknik sakıncalar bir yana, doğrusal olasılık modelinin en önemli sorunu mantıksaldır.
- DOM tahminine göre $P_i = E(Y = 1|X)$ olasılığı doğrusal olarak artmaktadır. Bu çekici bir varsayım değildir.
- Örnek olarak, belirli bir firma sayısının altında sanayi odası kurulma olasılığı hızla düşer. Aynı şekilde yüksek bir firma sayısı, ilde sanayi odasının bulunma olasılığını artırır ama aşırı yüksek firma sayısının marjinal etkisi azdır.
- Dolayısıyla, bize gereken modelde olasılık asla $[0,1]$ aralığı dışına çıkmazken, ayrıca, başta artarak artmalı ve belirli bir noktadan sonra ise azalarak artan bir yapı göstermelidir.
- Geometrik olarak bu, bir olasılık dağılımına ilişkin *"yığınsal dağılım işlevi"* (cummulative distribution function), ya da kısaca *"YDİ"* (CDF) ile gösterilebilir.
- Bu amaç için yeğlenen bir YDİ, ölçünlü normal dağılımdır.




## 7.2.2 Olabirim Modeli

- Doğrusal olasılık modelinin şöyle olduğunu anımsayalım:

$$P(Y = 1|X_i) = \beta_1 + \beta_2 X_i$$

- Arzuladığımız özellikleri gösteren bir yaklaşım ise şöyledir:

$$P(Y = 1|X_i) = \Phi(\beta_1 + \beta_2 X_i)$$

- Burada $\Phi$, ölçünlü normal dağılıma ait YDİ'dir:

$$\Phi(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2} dz$$

- Yukarıda görülen modele *"olasılık birimi"* (probability unit) teriminin kısaltılmışı olan *"olabirim"* (probit) modeli denir.
- Olabirim modeli, açıklanmaya çalışılan olasılıksal ilişkinin özündeki doğrusal-dışılığı dikkate alırken, aynı zamanda tahmin edilen $\hat{Y}_i$'ların 0 ile 1 arasında kalmasını da sağlar.

**Olabirim Modelinin Tahmin Edilmesi**

- Şimdiye dek gördüğümüz modellerin büyük bir çoğunluğu değişkenlerde doğrusaldı.
- Açıklayıcı değişkenlerin $\ln X_i$, $\sqrt{X_i}$ ya da $1/X_i$ gibi değerler aldığı, "değişkenlerde doğrusal-dışı" modelleri de ele aldık.
- Olabirim modelinde ise, yukarıdakilerden farklı olarak, $\beta_1$ ve $\beta_2$ ölçünlü normal YDİ $\Phi$'ın içinde yer almaktadır.
- Değiştirgelerde doğrusal-dışı olduğu için, böyle bir model SEK yöntemi ile tahmin edilemez.
- Olabirim modelinin tahmin edilmesi, tipik olarak, ençok olabilirlik yöntemi ile olur.
- Ençok olabilirlik yönteminin, $\beta$ değerlerini bulurken eldeki verileri elde etme olasılığını ençokladığını anımsayalım.
- Diğer bir deyişle EO, gözlenen verilere yol açma olasılığı ençok olan değiştirge değerlerinin hesaplanmasıdır.

- Öte yandan, olabirim modelinin EO tahminine yönelik ençoklanmak istenen log olabilirlik işlevine ait basit bir matematiksel gösterim bulunmamaktadır.
- Bu nedenle, tahmin sürecinde bilgisayardan yararlanılarak temelde deneme yanılmaya dayalı yinelemesel bir sayısal hesaplama yöntemi kullanılır.
- Bu işlemler, aralarında Gretl'ın da bulunduğu çoğu modern ekonometri programı içinde hazır gelmektedir. Bu nedenle olabirim modelinin tahmin edilmesi uygulamada basittir.

**Olabirim Modelinde Katsayıların Yorumlanması**

- Olabirim modelinde katsayıların yorumlanmasının önceki modellerden farklı olduğuna dikkat edelim:

$$\Phi(\beta_1 + \beta_2 X_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-z^2/2} dz$$

- Yukarıda verilen gösterim bir olasılıktır ve eksi sonsuzdan $\beta_1 + \beta_2 X_i$ değerine kadar ölçünlü normal YDİ altında kalan sol kuyruk alanını anlatmaktadır.
- Bu nedenle, belli bir $P(Y = 1|X_i)$'yi hesaplamak için önce $I = \beta_1 + \beta_2 X_i$ değerini alırız ve ölçünlü normal YDİ'den $P(Z \leq I)$ olasılığına bakarız.
- Bu noktada, $\Phi$ "s" şeklinde bir eğri olduğu için, $X$ doğrusal olarak artarken olasılığın önce artarak artacağı ve daha sonra da azalarak artacağı unutulmamalıdır.
- Olabirim modelinde katsayıların nasıl yorumlanacağına bir örnek olarak $\beta_1 = -3$, $\beta_2 = 2$, ve $X = 1$ olsun.
- Buna göre $I = -3+2\times1 = -1$ olur. Ölçünlü normal eğri altında $P(Z \leq -1)$ sol kuyruk alanı ise 0,1587'dir. Demek ki burada $Y$ olayının gerçekleşme olasılığı yüzde 15,87'dir.
- Benzer şekilde, $X = 2$ olduğunda $P(Z \leq 1) = 0{,}8413$ ve $X = 3$ olduğunda ise $P(Z \leq 3) = 0{,}9987$ olur.
- Görüldüğü gibi $X = 1$ iken $X$'teki bir birimlik artış olasılığı 0,6826 artırırken, $X = 2$ olduğunda ise bir birimlik benzer artış olasılığı yanlızca 0,1574 artırmaktadır.
- Öyleyse, olabirim modellerinde katsayılar yorumlanırken, seçilen başlangıç $X$ düzeyi önemlidir.

- Bu doğrultuda uygulamada sıkça kullanılan bir değer $X$'in örneklem ortalaması olan $\bar{X}$'dir.
- Çoklu bağlanımdaki yorum için önce şu DOM'a bakalım:

$$P(Y|X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

- Doğrusal modelde $X_2$'deki bir değişimin $Y$ üstündeki etkisi, diğer bir deyişle $X_2$'nin $Y$'ye göre eğimi $\beta_2$'dir.
- Üç değişkenli olabirim modeli ise şöyledir:

$$\Phi(\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}} e^{-z^2/2} dz$$

- Model doğrusal olmadığı için eğimler de sabit değildir ve hem $X_2$'nin hem de $X_3$'ün aldığı değerlere bağlıdır.
- Dolayısıyla, $X_2$'deki bir birimlik değişimin etkisini bulmak için önce $X_2$ ve $X_3$ için birer başlangıç değeri seçilir ve olasılık hesaplanır. Bunun için $\bar{X}_2$ ve $\bar{X}_3$ kullanılabilir.
- Sonra, $X_3$ sabitken $X_2$ bir birim artırılır ve olasılık yeniden hesaplanır. Aradaki fark, seçili $X_2$ ve $X_3$ düzeyi için $X_2$'nin yaklaşık eğimini verir.

**Olabirim Modelinde Çıkarsama**

- Olabirim modelinde kullanılan EO tahmincileri, etkin (enaz varyanslı) olmanın yanı sıra büyük örneklemlerde tutarlı ve normal dağılımlıdır.
- Ancak EO genel olarak bir büyük örneklem yöntemi olduğu için, tahmin edilen ölçünlü hatalar da kavuşmazsaldır.
- Bu nedenle, katsayıların anlamlılığını sınamak için $t$ değeri yerine ölçünlü normal $z$ değeri kullanılır.
- Bu noktada eğer örneklem yeterince büyükse $t$ dağılımının normal dağılıma yakınsadığı da unutulmamalıdır.
- Doğrusal modellerde bağlanımın bütününün anlamlılığını sınayan $F$ sınamasına almaşık olarak, EO tahmininde de *"olabilirlik oranı"* (likelihood ratio) ya da kısaca *"OO"* (LR) sınaması kullanılmaktadır.
- LR sınamasının sıfır önsavı $H_0 : \beta_2 = \beta_3 = \ldots = 0$'dır. Test istatistiği ise $k$ serbestlik derecesi ile $\chi^2$ dağılımına uyar.

### Olabirim Modelinde Yakışmanın İyiliği

- Kukla değişkenler söz konusu olduğunda, yakışmanın iyiliğini ölçmede $R^2$'nin yetersiz olduğunu anımsayalım.
- Olabirim gibi nitel bağımlı değişken modellerinde bu amaç için sıkça kullanılan iki ölçüt vardır.
- Bunlardan ilki *"doğru kestirilen durum sayısı"* (number of cases correctly predicted) değeridir.
- Bu ölçüte göre, $Y_i = 1$ iken tahmin edilen olasılık %50'den yüksekse ya da $Y_i = 0$ iken tahmin edilen olasılık %50'den düşükse, model doğru kestirim yapmıştır.
- İkinci ölçüt ise McFadden $R^2$ ya da *"sahte $R^2$"* (pseudo $R^2$) olarak adlandırılır ve log-olabilirlik istatistiğine dayanır.
- Log-olabilirlik, bağlanım kalıntılarının büyüklüğünü anlatan bir değerdir. McFadden $R^2$, eldeki modele ait log-olabilirliği yalnızca sabit terim içeren almaşık modelinkine oranlar.
- Ölçüt $[0,1]$ aralığındadır ama bildik $R^2$ ile karşılaştırılamaz.

### Olabirim Açıklayıcı Örnek

- Olabirim tahminine ilişkin olarak, Türkiye'de firma sayısına göre illerde sanayi odası bulunup bulunmama olasılığını inceleyen örneğimize dönelim.
- Sonuçlar şöyledir:

$$\begin{array}{llll} \hat{Y}_i = & -\ 1{,}8108 & +\ 0{,}0825\ X_i & \\ \text{öh} & (0{,}2802) & (0{,}0203) & \\ z & (-6{,}4620) & (4{,}0678) & \text{McFadden } R^2 = 0{,}3834 \end{array}$$

'Doğru kestirilen' durum sayısı = 71 (88,8%)

| | Kestirilen 0 | Kestirilen 1 |
|---|---|---|
| Gözlenen 0 | 67 | 1 |
| 1 | 8 | 4 |

Olabilirlik oranı: Ki-kare(1) = 25,9288 ($p$-değeri = 0,0000)

- 1,96'dan büyük $z$ değerleri, $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın her ikisinin de anlamlı olduğunu göstermektedir.
- Katsayıları doğrudan yorumlamak güçtür. $\hat{\beta}_2$'nın artı işaretli olmasından, firma sayısındaki artışın sanayi odası olma olasılığını artırdığı yönünde kaba bir yorum yapabiliriz.
- Verilere göre illerdeki ortalama firma sayısı 849'dur.
- Buna göre ortalama ilde sanayi odası olma olasılığı şudur:

$$\Phi(-1{,}8108 + 0{,}0825 \times 8{,}49) = \Phi(-1{,}1104) = 0{,}1335$$

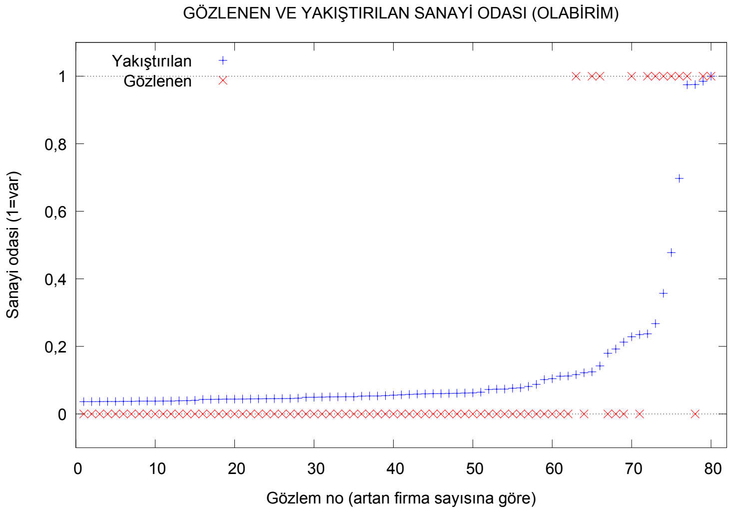
- Yukarıdaki $\Phi(I) = P(Z \leq I)$ değeri ölçünlü normal dağılım çizelgesinden hesaplanabilir.
- Veri biriminin 100 firma olduğunu anımsayalım. $X = 9{,}49$ olduğunda olasılık da $0{,}1520$ olur. Demek ki ortalama ilde kurulacak 100 yeni şirket olasılığı $0{,}0185$ kadar artıracaktır.
- Bu değer 849 ile 949 arasındaki ortalama eğimdir. Gretl, $\bar{X} = 849$ noktasındaki eğimi 0,0178 olarak kullanıcıya verir.

*(... devam)*

- Farklı $X_i$ değerleri için $P(Y = 1|X_i)$ olasılığı benzer şekilde bulunabilir.
- Örnek olarak, $X = 45{,}00$ ise tahmin edilen olasılık şudur:

$$\Phi(-1{,}8108 + 0{,}0825 \times 45{,}00) = \Phi(1{,}9039) = 0{,}9431$$

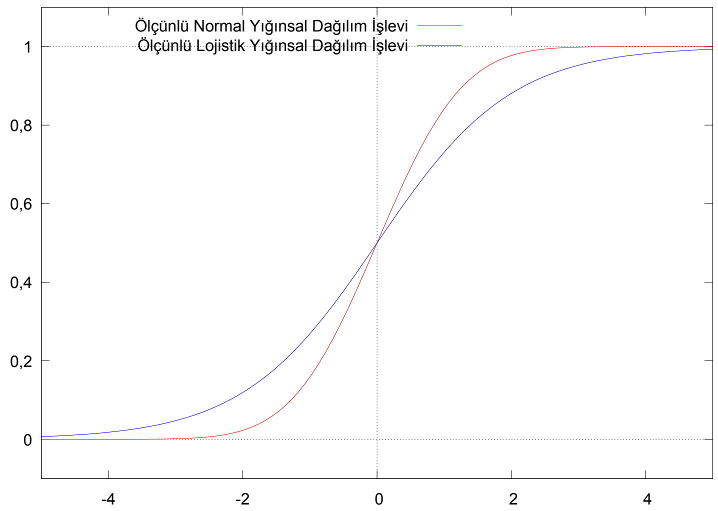
- Demek ki yaklaşık 4500 sanayi firması bulunan İzmir ya da Bursa gibi bir ilde sanayi odası bulunma olasılığı %94'tür.
- Yakışmanın iyiliğine ilişkin olarak, sahte $R^2$ bize modelin açıklama gücünün yaklaşık %38 olduğunu anlatmaktadır.
- Diğer yandan, modelin 81 ilden 78'indeki durumu doğru olarak kestirebildiği de görülmektedir.


GÖZLENEN VE YAKIŞTIRILAN SANAYİ ODASI (OLABİRİM)
Yakıştırılan
Gözlenen
Sanayi odasi (1=var)
1
0,8
0,6
0,4
0,2
0
0
10
20
30
40
50
60
70
80
Gözlem no (artan firma sayısına göre)

## 7.3 Diğer Nitel Tepki Modelleri

### 7.3.1 Logbirim Modeli

- Olabirim modeline almaşık olarak sıklıkla kullanılan bir model *"logbirim"* (logit) modelidir.
- Logbirim sözcüğü, *"logaritmik birim"* (logarithmic unit) teriminin kısaltılmasından gelmektedir.
- Logbirim bağlanımının olabirimden en büyük farkı ölçünlü normal YDİ yerine ölçünlü lojistik YDİ'yi kullanmasıdır.
- Şekilsel olarak ölçünlü normal YDİ ile ölçünlü lojistik YDİ birbirlerine benzer. Aralarındaki en büyük fark ise lojistik eğrinin normal eğriye göre daha kalın kuyruklu olmasıdır.
- Diğer bir deyişle, logbirim modelinde $P_i$ koşullu olasılığı 0 ya da 1 değerine biraz daha yavaş bir hızla yaklaşır.




- İki değişkenli logbirim modeli şu şekilde gösterilir:

$$P(Y = 1|X_i) = F(\beta_1 + \beta_2 X_i)$$

- Burada $F$, ölçünlü lojistik dağılıma ait YDİ'dir:

$$F(X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}$$

- Görüldüğü gibi, lojistik YDİ matematiksel olarak normal YDİ'ye göre daha yalındır.
- Uygulamada olabirim modeli gibi logbirim modeli de ençok olabilirlik yaklaşımı ile ve sayısal yöntemler kullanılarak tahmin edilir.
- Logbirim modelinde katsayıların yorumlanması, olabirimle aynıdır. Farklı $X$ değerleri için olasılıklar hesaplanır ve $X$'ler değişince olasılıkların ne kadar değiştiğine bakılır.
- İstatistiksel çıkarsama ve yakışmanın iyiliği konuları da yine olabirim modeli ile benzer şekildedir.
- Gerçek şu ki iki model çoğu zaman birbirine oldukça yakın sonuçlar üretir.
- Eskiden, logbirim bağlanımını yeğlemenin bir nedeni daha hızlı hesaplanabilmesiydi. Hızlı bilgisayarlar bu üstünlüğü ortadan kaldırmıştır.
- Bugün iki model arasındaki seçim çoğunlukla hangisinin yazılım tarafından daha iyi desteklendiği ile ilgilidir.
- Gretl her iki modeli de eşit düzeyde destekler.

**Logbirim Açıklayıcı Örnek**

- Firma sayılarına göre sanayi odası bulunup bulunmama olasılığını bir de logbirim modeli ile tahmin edelim.
- Sonuçlar aşağıdaki gibidir:

$$\begin{array}{llll} \hat{Y}_i = - & 3{,}3083 & +\,0{,}1857\,X_i & \\ \text{öh} & (0{,}6350) & (0{,}0622) & \\ z & (-5{,}2099) & (2{,}9867) & \text{McFadden } R^2 = 0{,}3861 \end{array}$$

'Doğru kestirilen' durum sayısı = 72 (90,0%)

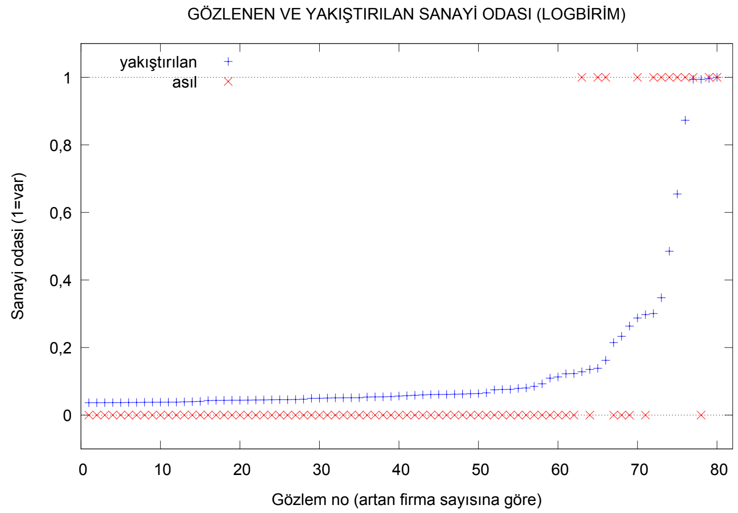
| | Kestirilen 0 | Kestirilen 1 |
|---|---|---|
| Gözlenen 0 | 67 | 1 |
| 1 | 7 | 5 |

Olabilirlik oranı: Ki-kare(1) = 26,1155 ($p$-değeri = 0,0000)

- $\hat{\beta}_2$'nın anlamlı ve artı işaretli olmasına bakarak, kabaca, firma sayısının ilde sanayi odası bulunma olasılığı üzerinde olumlu etkisi olduğu yorumunu yapabiliriz.
- Tahmin edilen katsayılar farklı olsa da logbirim modeli olabirim modeline benzer sonuçlar göstermektedir.
- Örnek olarak, 849 firma bulunan ortalama ildeki sanayi odası bulunma olasılığı şudur:

$$\frac{1}{1+e^{-(-3{,}3083+0{,}1857\times 8{,}49)}} = 0{,}1504$$

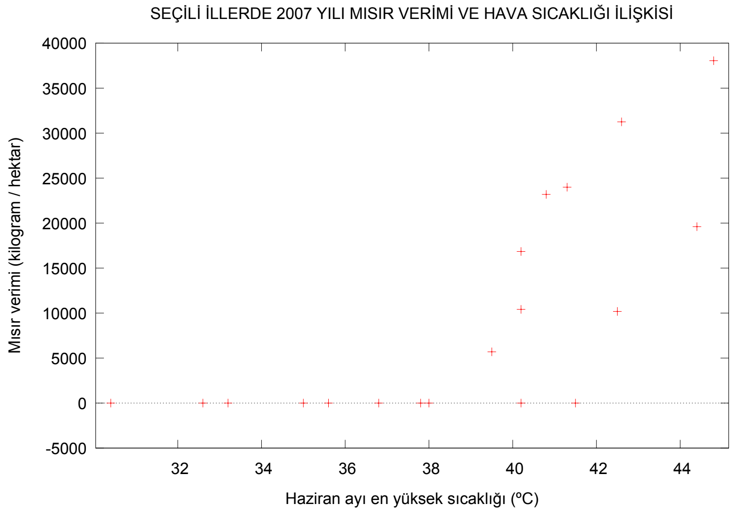
- Demek ki olabirim modeli tarafından %13,4 tahmin edilen olasılığı log birim modeli de %15,0 olarak bulmuştur.
- Sahte $R^2$, doğru kestirilen durum sayısı, ve olabilirlik oranı istatiği de önceki olabirim tahminleri ile neredeyse aynıdır.




### 7.3.2 Tobirim Modeli

- Olabirim ve logbirim modellerin bir uzantısı da 1958 yılında Nobel ödüllü ekonomist James Tobin tarafından geliştirilen ve bu nedenle *"tobirim"* (tobit) olarak adlandırılan modeldir.
- Bu modeli açıklayabilmek için, illerdeki ortalama sıcaklık ve mısır bitkisi üretiminin verimliliği ilişkisini ele alalım.
- Bir sıcak iklim bitkisi olan mısır tarımı Türkiye'nin her ilinde yapılmadığı için, burada iki gözlem kümesi söz konusudur.

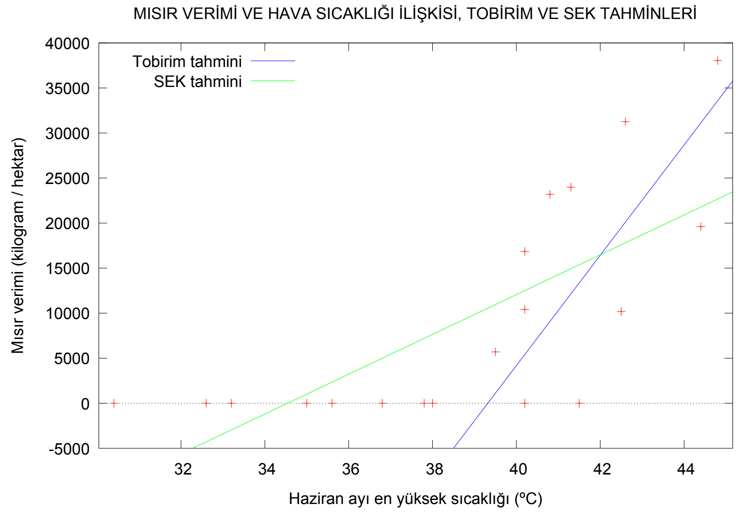
- Birinci kümede hem sıcaklık hem de verimlilik bilgisinin bulunduğu iller yer almaktadır.
- İkinci kümeyi ise iklim ya da diğer nedenler yüzünden mısır tarımı yapılmayan ve bu nedenle yalnızca sıcaklık bilgisinin olduğu iller oluşturmaktadır.
- Bağımlı değişkene ait bilginin yalnızca bazı gözlemler için bulunabildiği böyle bir veri setine *"sansürlü"* (censored) örneklem adı verilir.




- Ekonometrik olarak, tobirim modeli şöyle gösterilebilir:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i && Y_i > 0 \text{ ise,} \\ Y_i &= 0 && \text{eğer değilse.} \end{aligned}$$

- İkinci satırda $Y_i$ aslında artı ve eksi değerler alabilecekken, mısır tarımı yapılmadığı için sıfır olarak gözlenmektedir.
- Öyleyse gerçekte tam göremediğimiz bir *"örtük değişken"* (latent variable) $Y_i^*$ vardır. Bizim elimizdeki $Y_i$ ise bir *"sınırlı bağımlı değişken"* (limited dependent variable) olmaktadır.
- Bu nedenle, yalnızca birinci satır kullanılır ve tek bir SEK bağlanımı ile tahmin edilirse, sonuçlar yanlı ve tutarsız olur.
- $Y_i = 0$ gözlemlerini dışlayan bir bağlanım ile tüm verileri kullanan bir bağlanımın sonuçlarının farklı çıkacağı açıktır.
- Böyle bir durumda yansız ve tutarlı tahminciler üreten tobirim modeli, uygulamada ençok olabilirlik yöntemi ile kolaylıkla tahmin edilebilmektedir.




### 7.3.3 İleri Model ve Konular

**Sıralı Logbirim ve Olabirim**

- Olabirim ve logbirim modellerine ilişkin örneklerimizdeki $Y$ değişkenleri yalnızca iki değer alabilen kuklalar idi.
- Ancak kimi zaman bağımlı değişken için ikiden fazla değer söz konusu olabilir. Üstelik bunlar genellikle sayısal değil, *"sırasal"* (ordinal) niteliktedir.
- Örnek olarak, anketlerde "tümüyle katılıyorum," "kısmen katılıyorum," "katılmıyorum" gibi yanıtlar yaygındır.
- Benzer şekilde kişilerin öğrenim düzeyi "ilköğretim," "lise," "üniversite," "yüksek lisans" değerlerini alabilir.
- Uygulamada bunlar veri setine $0, 1, 2, \ldots$ biçiminde işlenir. Ancak gerçekte bir seçeneğin diğerinin bir fazlası ya da iki katı olduğu söylenemez.
- $Y$ bağımlı değişkeninin ikiden fazla seçenek alabildiği ve bunların belli bir sırayı izlediği durumları incelemek için *"sıralı"* (ordered) olabirim ve logbirim yöntemleri kullanılır.
- Başta gördüğümüz iki değersel modellerin genellemesi olan bu modellerde her bir sonucun gerçekleşme olasılığı yine normal ve lojistik YDİ'ler kullanılarak bulunur.
- Bu yöntemlerin hesaplanması ve yorumlanması biraz daha karmaşık olduğu için burada ele alınmayacaktır.

**Çokterimli Logbirim ve Olabirim**

- İkiden fazla kesikli değer alan ancak bu değerlerin doğal bir sıra izlemediği değişkenlere *"kesikli seçim"* (discrete choice) değişkeni adı verilir.
- Örnek olarak ulaşım aracı için "otomobil," "otobüs," "tren" gibi seçenekler belirlenebilir.
- Aynı şekilde kişilerin işi "serbest," "ücretli," "memur" gibi farklı ulamlara ayrılabilir.
- $Y$ bağımlı değişkeninin aldığı değerler eğer mantıksal bir sıra izlemiyor ise, bu durumda *"çokterimli"* (multinomial) olabirim ve logbirim yöntemleri kullanılır.
- Tekil seçenek olasılıklarının olabirim ya da logbirim olarak belirtildiği bu modeller de yine ileri konular arasındadır.

**Süre Modelleri**

*"Süre modeli"* (duration model) adı verilen bir model sınıfı vardır. Sistem aksaklıklarını inceleyen ve *"Sağkalım çözümlemesi"* (survival analysis) de denilen çalışmalarda sıkça yararlanılan bu modellerin konusu şunlar olabilir:

- İşsiz kalma süresini belirleyen etmenler
- Bir grevin ne kadar uzayabileceği
- Hastaların iyileşme süreleri
- Bir makinenin ömrünün ne olacağı

Tüm bu örneklerde temel konu zamandır ve bir rastsal değişken olarak modellenir. Tahmin süreci ise yine uygun bir olasılık dağılım işlevi ya da bunun YDİ'sinin kullanılmasına dayalıdır. Gretl gibi ekonometri yazılımları tarafından sunulan tahmin seçenekleri arasında çeşitli süre modelleri de bulunmaktadır.

# Önümüzdeki Dersin Konusu ve Ödev

**Ödev**
Kitaptan *Bölüm 15* “Qualitative Response Regression Models” okunacak.

**Önümüzdeki Ders**
Eşanlı Denklem Modelleri

# UADMK Açık Lisans Bilgisi



A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 BY-NC-SA