Sosyo Ekonomi

Ocak-Haziran 2006-1

Doğrusal Olasılık ve Logit Modelleri ile Parametre Tahmini

M. Emin İnal inal@nigde.edu.tr

Derviş Topuz dervis_topuz@nigde.edu.tr

Okyay Uçan okyayu@nigde.edu.tr

Linear Probability and Parameter Estimation with Logit Models

Abstract

The role of categorical (dummy) variables in regression modelling is investigated in this study. In this respect, it is shown, based on linear probability model and logit model, to what extent the values of dependent dummy variables change with respect to the values of independent variables.

The study aims at determining the probability that the students complete their university programs within the period of four years with respect to their scores from the quantitative section in the Turkish University Placement Test (OYS).

Key Words : Linear Probability, Logit Model, Regression.

JEL Classification Codes : C1, C5, C6.

Özet

Bu çalışmada regresyon modelinde dummy değişkenlerin kullanımı ele alınmaktadır. Bu doğrultuda, bağımsız değişkenlerin değerlerine karşılık bağımlı dummy değişkenlerin alacağı değerlerin nasıl değişmekte olduğu doğrusal olasılık ve logit modelleri bağlamında incelenmektedir.

Bu çalışmada, Türkiye'de Öğrenci Yerleştirme Sınavı'nda (ÖYS) öğrencinin yaptığı sayısal net sayısı, cinsiyet ve bitirdiği lise türü değişkenlerine göre, lisans eğitimini dört yılda tamamlama olasılıkları belirlenmeye çalışılmaktadır.

Anahtar Sözcükler : Doğrusal Olasılık, Logit Modelleri, Regresyon.

1. Giriş

Regresyon analizinde yaygın kullanıma sahip En Küçük Kareler tekniğinin parametre tahmininde kullanılacak varsayımlar, bu yöntemle elde edilen tahmin edicilerin özelliklerini belirlemede yararlı olmaktadır. Fakat günümüz koşullarında elde edilen veri kümeleri için istatistiksel modelin bu varsayımları sağlanmayabilir. Rassal hatalar anakütlenin normallik varsayımı geçersiz olduğunda En Küçük Kareler tekniği ile elde edilecek tahminler sapmalı sonuçlar verebilir. Özellikle uç değerlerin bulunduğu veri kümelerinde hatalara ilişkin dağılım, normal dağılıma uymamaktadır.

Bu çalışmada, **Niğde Üniversitesi İİBF İşletme ve İktisat** bölümlerinden 2000-2001 eğitim- öğretim döneminde mezun olmuş 130 öğrenciden 100'ü kolayda örnekleme yöntemiyle alınmış ve bu öğrencilerin **fakülteden dört yılda mezun olma olasılığı** bağımlı değişken (Y); bağımsız değişkenler, cinsiyetleri (X_1), mezun oldukları lise türü (X_2) ve ÖYS'de yaptıkları sayısal net sayıları (X_3) kullanılarak tahmin edilmeye çalışılmıştır.

Bağımsız değişken olarak, mesela **ÖSS veya ÖYS'de alınan puan türleri**nden birisinin alınmamasının temel nedeni, öğrenciler arasında bu puan türlerine göre çok belirgin farklılıkların olmaması, puanlara ait değişim aralığı değerinin çok düşük olmasıdır.

Bağımsız değişken olarak alınabilecek faktörlerden birisi de, öğrencilerin **fakülteyi kaçıncı sırada tercih ettikleri**dir. Bu değişkenle ilgili değerlerin, sağlıklı bir biçimde öğrenilememesi, bu değişkenin uygulamaya alınmamasını gerektirmiştir.

Öğrencinin geldiği şehir de, bağımsız değişken olarak alınabilecek bir diğer faktör olarak görülebilir. Ancak, şehirlerin sosyal ve kültürel dokularının heterojen olması, kozmopolit bir yapıda olmaları nedeniyle bu değişkenin de bağımsız değişken olarak belirlenmesi çok da anlam ifade etmeyecektir.

Hangi amaçla olursa olsun yapılacak olan çalışmanın özü, verilerden yararlanarak değerlendirmeler yapmaktır. Bu nedenle eldeki veri grubuna uygun analiz teknikleri ele alınarak çözümlemeler yapılmaktadır. Dolayısıyla araştırmacının veri gruplarını iyi bir şekilde tanıyıp birbirinden ayırt etmesi gerekir ki yapacağı çalışma için doğru bir yol izlemiş olsun.

Genel anlamda, değişkenler kendi içinde **nitel** ve **nicel** olmak üzere ikiye ayrılırlar. Nicel değişkenler de kendi içinde **sürekli** ve **kesikli** olmak üzere ikiye ayrılırlar. En basit tanımıyla ölçülebilen değişkene sürekli, sayılabilen değişkene ise kesikli değişken denir. Kesikli değişkenler tam sayı karakterindedir. Sürekli ve kesikli verilerle analizler yapılmasına rağmen bazı durumlarda bu tür verilere uygulanan istatistiksel teknikler yanıltıcı sonuçlar verebilir. Böyle durumlarda dummy verilerden yararlanmak gerekir.

Gözlem değerleri, kategoriler içine ayrıştırılmışsa veri kümesi dummy olarak adlandırılır. Bu bağlamda, ikili veri değerleri sonlu sayıya sahip değişkenlerden oluşur. Bu nedenle bilimsel araştırmaların bir çoğunda araştırmacının ilgilendiği değişkenler diğer bazı değişkenler tarafından etkilenip araştırma süresince farklı değerler almasına neden olmaktadırlar.

Regresyon analizi ekonomi, fizik, biyoloji ve ziraat gibi bilim dallarında kullanılan istatistik yöntemlerden birisi olup, aralarında sebep sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi saptamada kullanılır.

Uygulamada genellikle çeşitli değişkenler arasında bir ilginin bulunduğu görülür. Bu ilginin fonksiyonel olarak ifade edilmesi regresyon teorisinin konusudur. Regresyonda bir veya daha çok bağımsız değişkenle, bir bağımlı değişken açıklanmaya çalışılır (Başar ve Oktay, 2000:71).

Regresyon, anakütle yerine örnekten yola çıkıldığında, açıklayıcı değişkenlerden yararlanarak bağımlı değişkenin tahmin edilmesi ve bu süreçte yapılan hataların ölçülmesidir. Korelasyon çözümlemesi ise bu değişkenler arasındaki ilişkinin yön ve derecesinin saptanmasıdır.

Regresyon modellerinde ele alınan değişkenlerin başlıca özelliklerinden biri, ölçülebilir (kantitatif) olmaları, bu nedenle süreklilik özelliğini içermeleridir. Ancak çok sık rastlanmasa da regresyon modellerinde nitel (kalitatif) değişkenleri de inceleme zorunluluğu ortaya çıkmaktadır. Yabancı kaynaklarda "**Dummy Variable**" veya "**Binary Data**" adları ile rastlanılan bu kalitatif değişken, bu çalışmada **dummy değişken** olarak geçecektir .

Dummy değişkenler regresyon modellerinde tıpkı nicel değişkenler gibi kolaylıkla kullanılabilirler. Gerçekte, bir regresyon modelinin bütün açıklayıcı değişkenleri özünde dummy ya da nitel değişken olabilir. Böyle modellere **varyans çözümlemesi modelleri** denir.

Dummy değişkenlerin doğal bir ölçüm skalası yoktur. Bu nedenle, bu değişkenlerin modele katkısının görülmesi için "bir sınıfa" bu tür değişkenlerin atanması gerekir. Bu sınıfı oluşturan değişkenlere **dummy değişkenler** denir. Dummy değişkenler, çoğunlukla 0 ve 1 değerlerini alır ve bu yeni değişkenlerin sayısal olarak hiç bir önemi yoktur. Çünkü, sadece gözlemlerin hangi kategoriye ait olduklarını gösterirler. Bu değişkenlerin çoğu iki sonucu içerir. Örneğin cinsiyet değişkeni (kadın/ erkek), hasta veya değil gibi iki kategoriye sahip değişkenlerdir. Bazılarının üç kategorisi vardır, yaşanan bölge olarak köy, kasaba, kent gibi. Bu tür değişkenler daha fazla sınıflı da olabilir. Eğer, ilgili değişkenin k tane sınıfı var ise, modelde bu değişken için, k–1 tane dummy değişken tanımlanır (Alpar, 1997:275). Dummy değişkenler açıklayıcı değişkenler olabildiği gibi, bağımlı değişken de olabilir. Bu durumda bağımlı değişken sadece iki değer alır. Bu durumda, olayın varlığı için “1”, yokluğu için de “0” kullanılır.

Dummy değişkenin 0 ve 1 değerini aldığı durumların değişmesi, bu değişkenin katsayısını değiştirmeyip, sadece işaretini değiştirmekte ve sonuç değişmemektedir.

Dummy değişkenlerin kullanıldığı çalışmalarda, genellikle 0 değerini alan kategori; **temel** ya da **karşılaştırma kategorisi** olarak adlandırılır (Alpar, 1997:275).

Bu çalışmanın amacı, regresyon çözümlemesinde dummy bağımlı değişkenlerinin rolünü ele almaktır. Çoğunlukla gölge değişkenler diye adlandırılan nitel değişkenlerin işin içine katılmasıyla doğrusal regresyon çözümlemesinin, gözleme dayalı çalışmalarda karşılaşılan ilginç pek çok sorunun çözümüne yarayabilecek son derece esnek bir araç haline geldiği gösterilmeye çalışılacaktır.

Bu çalışmada, bağımlı değişken (0-1), dummy değişken olarak alınarak, yapı gereği iki değerli (dikotom) özelliğe sahip olduğu durumların olabileceği regresyon modeli uygulanacaktır.

2. Materyal ve Yöntem

2.1. Materyal

Bu çalışmada materyal olarak Niğde Üniversitesi İİBF İşletme ve İktisat Bölümleri Birinci ve İkinci Öğretiminden 2000-2001 eğitim- öğretim döneminde mezun olmuş 130 öğrenciden rasgele seçilen 100’ünün cinsiyetleri (X_1), mezun oldukları lise türü (X_2) ve ÖYS’de yaptıkları sayısal net sayıları (X_3) veri olarak kullanılmıştır. Örneğe alınan öğrencilerle ilgili veriler, fakülte öğrenci işlerinden sağlanmıştır.

Yapılan çalışmada dummy verileri analiz etmek için logit modeli kullanılmıştır. Bu bağlamda, Lojistik Regresyon Analizi sayesinde ilgili gruplar içinde yer alan değişkenlerin ilişkileri ortaya çıkarılmaya çalışılmıştır. İlgili modeller için çözüm de **S-PLUS 2000** ve **SPSS 10.0** paket programlarından yararlanılarak bulunmuştur.

2.2. Yöntem

Nicel veya nitel özellikte bir değişkenin belli bir ölçümsel düzeye erişmesine başka değişkenler etki edebilir. Bu bağlamda "**bağıntı**" en az iki değişik en arasında gözlenen, birlikte, uyumlu, ölçümsel büyüklük değiştirme özelliği olarak tanımlanır. Değişkenlerden biri, diğerindeki ölçümsel farklılaşmalara bağlı olarak farklılaşır önkabulü ile "**bağımlı değişken**" adını alır. Bağımlı değişkenin düzeyini etkilediği varsayılan bir veya daha fazla değişkene ise **bağımsız değişken** denir. Böylece değişkenler arasındaki bağıntı, matematiksel olarak; **y=f (x, z, w,..)** şeklinde belirtilir ve y'nin bağımlı değişken; x, z, w değişkenlerinin bağımsız değişkenlerinin bir fonksiyonu olduğu biçiminde ifade edilir.

Bilimsel araştırmalarda bağıntı soruşturması genellikle iki amaçla yapılır:

2.2.1. Nedensel Öğelerin Araştırılması

Belli bir sonuç değişkenin -bağımlı değişken-, bir dizi, nedensel olduğu düşünülen değişkenden kaynaklandığı, bunların düzeyine göre büyüdüğü veya küçüldüğü düşünülebilir. Araştırıcı, söz konusu nedensel öğelerden hangilerinin gerçekten sonuç değişkenin oluşmasında, doğrudan, tek başına ve istatistiksel olarak da anlamlı düzeyde etkin olduğunu saptamayı amaçlayabilir.

2.2.2. Tahmin Yapmak

Bağıntı değerlendirmesinde bir bağımlı öğe, bir veya daha fazla bağımsız değişkene, parametreleri hesaplanabilen bir matematiksel fonksiyonla ilişkilendirilebilir.

Uygulanan yöntem, bağımlı değişkenin iki durumlu olduğu durumlarda, regresyon analizinde kullanılmasıdır. Burada, sadece iki durumlu bağımlı değişkenli modeller ele alınacaktır. Bunlar,

- Doğrusal Olasılık Modeli
- Logit (Lojistik) Modeli'dir (Tarı, 1999:233).

Bu uygulamada, üçüncü bir grubu oluşturan Probit modeli ele alınmayacaktır. Probit modelinin ele alınmamasının temel nedeni, logit modelin, probit modele göre daha iyi sonuçlar ürettiği düşüncesidir (Gujarati, 1995: 563).

2.2.2.1. Doğrusal Olasılık Modeli

Yapısı gereği bağımlı değişkenlerin kalitatif özelliğe sahip olduğu durumlar için, dummy bağımlı değişkenleri içeren bu yöntemin nasıl olduğunu inceleyebilmek için

$$Y = \alpha + \beta_i X_i + e_i \tag{1}$$

eşitliği ele alınır. Burada;

X_i : Bağımsız (açıklayıcı) değişken

Y_i: Bağımlı dummy değişken

Y=1, Fakülteden dört yılda mezun olma durumunu,

Y=0, Fakülteden beş veya daha fazla yılda mezun olma durumunu göstermektedir.

ei: Hata terimi olup N(0, σ^2) Ortalaması 0 ve varyansı σ^2 olan, bağımsız dağılımlı şans değişkenidir.

Bağımlı değişkenlerin Y_i kalitatif özelliğe sahip olduğu durumlar için açıklayıcı değişken X_i' nin doğrusal bir fonksiyonu olarak tanımlayan (1) nolu modellere **doğrusal olasılık modelleri** denir.

Çünkü, X_i verildiğinde Y_i'nin koşullu beklenen değeri $E(Y_i|X_i)$, X_i veriyken olayın gerçekleşmesinin koşullu olasılığı olarak yorumlanabilir (Tarı, 1999:234). Sapmasız tahmin edicilere ulaşmak için, e_i, ortalaması sıfır olan rassal bir değişken olarak varsayıldığında, $E(e_i) = 0$ olduğu için,

$$E(Y_i|X_i)=\alpha+\beta X_i \tag{2}$$

modeli elde edilir.

Y_i =1 Dört yılda mezun olma durumu (olayın gerçekleşme) olasılığı (P_i),

Y_i =0 Dört yılda mezun olamama (olayın gerçekleşmeme) olasılığı (1-P_i) olarak tanımlanırsa, matematiksel beklenen değer tanımından:

$$E(Y_i)=\sum Y_i P(Y_i)=0\times(1-P_i)+1\times(P_i)=P_i \tag{3}$$

değeri bulunur(Maskie, 2001:48). (2) nolu model ile (3)'nolu modeller karşılaştırılıp model (4) elde edilir.

$$E(Y_i|X_i)=\alpha+\beta X_i \tag{4}$$

olarak da ifade edilebilir.

Böylece, (1) no'lu modelin koşullu beklenen değeri, aslında Y_i'nin koşullu olasılığıdır.

Bağımlı değişkenin bir olasılığı olarak doğrusal olasılık modeli aşağıdaki şekilde yazılabilir (İşyar, 1999:259):

$$P_i = \begin{cases} \alpha + \beta X_i & \text{eger } 0 < \alpha + \beta X_i < 1 \text{ ise,} \\ 1 & \text{eger } \alpha + \beta X_i \geq 1 \text{ ise,} \\ 0 & \text{eger } \alpha + \beta X_i \leq 1 \text{ ise.} \end{cases}$$

P_i olasılığı, 0 ile 1 arasında bulunacağından, $0 \leq E(Y_i \mid X_i) \leq 1$ veya 0 $< \alpha + \beta Xi <$ 1 şeklinde bir sınırlama zorunlu olduğundan koşullu olasılık 0 ile 1 arasında kalmalıdır.

Regresyon modellerinde parametrelerin en küçük kareler yöntemi ile bulunduğu bilinmesine karşın, bağımlı değişkenin dummy değişken olduğu durumda özel sorunlarla karşılaşılabilir. Doğrusal olasılık modelinin sunduğu bu bilgilere rağmen, bu modelin tahmini ve yorumuna ilişkin bazı eleştiriler vardır. Bunlar (Tarı, 1999:237):

2.2.2.1.1. Hata Teriminin (e_i) Normal Dağılımlı Olmaması

En Küçük Kareler yöntemi, hatanın normal olarak dağılmasını gerektirmez. Ancak normal dağılma özelliği önem testleri yapabilmek için gereklidir. Y_i gibi, e_i'ler de yalnızca iki değer (0 ve 1) aldığından, e_i için normalite varsayımı lineer olasılık modellerinde gerekli değildir.

$$\begin{aligned} e_i &= Y_i - \alpha - \beta X_i \\ Y_i = 1 \quad \text{için} \quad e_i &= 1 - \alpha - \beta X_i \end{aligned} \tag{5}$$

$$Y_i = 0 \quad \text{için} \quad e_i = -\alpha - \beta X_i \tag{6}$$

yazılabilir. e_i'nin normal olarak dağıldığını varsaymanın zorluğuna karşın, örnek büyüklüğü sonsuz olarak arttıkça, en küçük kareler yöntemi ile yapılan tahminlerin genellikle normal olarak dağılmaya eğilimli olduğu (asimptotik özellik) gösterilebilir (Gujarati, 1995:543).

2.2.2.1.2. Hata Terimi (e_i) Değişen Varyanslıdır

$E(e_i) = 0$ olmasına karşın e_i'lerin homojen varyanslı oldukları söylenemez (İşyar, 1999:261). Değişen varyanslı (Heteroscedastic) durumunda, en küçük kareler yöntemini kullanmak uygun değildir. Bunun yerine, bu sorunu çözen ve genelleştirilmiş en küçük kareler yönteminin bir ifadesi olan tartılı (ağırlıklı) en küçük kareler yöntemi kullanılabilir (İşyar, 1999:261). e_i'lerin olasılık dağılımı aşağıda görüldüğü gibidir:

Tablo 1
e_i Değerlerine İlişkin Olasılık Dağılımları

Y_i	e_i	Olasılık
1	$1-\alpha-\beta Xi$	P_i
0	$-\alpha - \beta$ Xi	$1-P_i$
Toplam		1

Burada P_i, model (5)'in olasılığını ölçer. X_i'nin sabit olması varsayımına göre e_i'nin olasılık dağılımı Y_i'nin olasılık dağılımına eşittir. Hata payı (e_i)'nin sıfır ortalamaya sabit olduğu varsayımına göre P_i olasılığı ile X_i arasındaki ilişki belirlenebilir.

$$E(e_i) = (1-\alpha-\beta X_i) * P_i + (-\alpha-\beta X_i)(1-P_i) = 0$$

P_i için çözerek,
$$\begin{aligned} P_i &= \alpha + \beta X_i \\ 1 - P_i &= 1 - \alpha - \beta X_i \end{aligned} \tag{7}$$

elde edilir.

Hata payının varyansı ise $E(e_i) = 0$ varsayımı ile

$$Var(e_i)=E(e^2)=(-\alpha-\beta_i X_i)^2(1-P_i)+(1-\alpha-\beta X_i)^2(P_i) \tag{8}$$

$$Var(\mu)=E[\mu-E(\mu)]^2$$

$$=E(\mu)]^2$$

$$Var(\mu)=E(\mu^2)=(-\alpha-\beta i X_i)^2(1-\alpha-\beta_i X_i)+(1-\alpha-\beta_i X_i)^2(\alpha+\beta_i X_i)$$

$$=(\alpha+\beta_i X_i)\,(1-\alpha-\beta_i X_i)$$

$$Var(\mu)=E(Y_i\,|X_i)[1-E(Y_i\,|X_i)]$$

$$=P_i(1-P_i) \tag{9}$$

Burada, $E(Yi|Xi) = \alpha + \beta Xi = Pi$ olmasından yararlanılmıştır (Gujarati,1995:543; Maskie, 2001:50). (9) no'lu model, e_i'nin varyansının değiştiğini gösterir (denklemi, hata payının heterojen varyanslı olduğunu göstermektedir). Çünkü Y_i'nin koşullu beklenen değerine, o da X'in aldığı değere bağlıdır. En sonunda e_i nin varyansı X'e bağlıdır, dolayısıyla sabit değildir. Böylece e_i'nin varyansı X_i'ye bağlı olduğundan homojen varyanslı değildir. Bu nedenle, hetorojen varyans problemini çözmenin çeşitli yolları vardır. Bunlardan en basit olanı,

a. (1) nolu modelin her iki yanını $\sqrt{Wi}$ gibi bir katsayıya (tartı değeri) bölerek,

$$\frac{Yi}{\sqrt{Wi}} = \frac{\alpha}{\sqrt{Wi}} + \beta\frac{Xi}{\sqrt{Wi}} + \frac{ei}{\sqrt{Wi}} \tag{10}$$

modeli elde edilir. Burada:

$$\sqrt{Wi} = \sqrt{E(Yi|Xi)\left[1 - E(Yi|Xi)\right]}$$

$$= \sqrt{Pi(1 - Pi)} \text{ dir.}$$

b. Böylece sabit varyanslılık sağlanır.

c. Gerçek $E(Y_i|X_i)$ bilinmediği için, w_i'ler de bilinmemektedir. Wi'leri tahmin için şu dört adımlı süreç izlenebilir:

Değişen varyans sorununa karşın, (1) nolu modeli klasik en küçük kareler yöntemi ile bulup $\hat{Y} =$ gerçek $E(Y_i \mid X_i)$ nin tahmini elde edilir.

- Sonra w_i' nin tahmini $\hat{w}_i = \hat{Y}_i(1 - \hat{Y}_i)$ değeri bulunur.
- Tahmin edilen $\hat{w}_i$ kullanılarak, verileri (10) nolu model gibi dönüştürülür.
- Dönüştürülen verilere klasik en küçük kareler regresyon modeli uygulanarak parametre değerleri tahmin edilir.

2.2.2.1.3. $P_i = E(Y = 1 \mid X_i)$ Değerinin X_i İle Doğrusal Olarak Arttığının Varsayılması

Örnek büyüklüğü arttıkça hata terimi normal dağılıma yaklaşsa ve değişen varyans durumunda, ağırlıklı en küçük kareler yöntemi kullanılsa da, yukarıdaki c ve ii-nci maddelerdeki sakıncalar ortadan kalkmamaktadır. Bu iki sakıncayı giderebilmek için logit ve probit modelleri geliştirilmiştir. Bu modeller, hem $0 \leq E(Y = 1 \mid X) \leq 1$ şartını sağlayabilmekte ve hem de P_i ile X_i arasındaki ilişkiyi doğrusallıktan kurtarabilmektedirler. Yani, logit ve probit modelleri, farklı bağımsız X değişkeninin olasılığının 0 ile 1 arasında kalmasını sağladıkları gibi; ayrıca, değişik bağımsız değişkene ait belli bir artış karşısında, bu bağımsız değişkenin kullanılma olasılığının değişik miktarda artmasını sağlamaktadırlar.

2.2.2.1.4. R^2 Değerinin Genellikle Küçük Çıkarak, İlişkinin Uyumunu Gösteren Bir Ölçü Olamaması

Geleneksel yolla hesaplanan R^2 , iki uçlu tepki değişkeni modellerinde sınırlı bir yarar sağlar. Belli bir X'e karşılık gelen Y, ya 0 ya da 1'dir. Öyleyse bütün Y değerleri, ya X ekseni ya da 1'in hizasındaki doğru üzerinde yer alır. Genellikle klasik En Küçük Kareler yöntemi ile hesaplanan R^2 , böyle modellerde 1'den çok küçük çıkma eğilimindedir. Çoğu uygulamada R^2 , 0.2 ile 0.6 arasında yer alır. Tahmin edilen Y_i , ya 0'a ya da 1'e yakın çıkacaktır.

Bu nedenle John Aldrich ile Forrest Nelson 'Nitel bağımlı değişkeni olan modellerde, belirlilik katsayısının bir özetleme istatistiği olarak kullanılmasından kaçınılması gerektiğini ileri sürmektedir (Gujarati, 1995:546).

2.2.2.1.5. $0 \leq E(Y_i \mid X_i) \leq 1$ Koşulunun Sağlanamaması

Lineer olasılık modellerinde $E(Y_i \mid X_i)$, X verildiğinde Y_i'nin koşullu olasılığı olduğunda 0 ile 1 arasında bulunmak zorundadır. E $(Y_i|X_i)$'nin bu koşulun dışına çıktığı görülebilir. Bu, lineer olasılık modellerinde, En Küçük Kareler ile tahmin yapmanın en önemli problemidir. **Bu problemi çözmenin iki yolu vardır**. **Birincisi** sıfırdan küçük P_i'leri $P_i = 0$, birden büyük P_i'leri $P_i =1$ olarak almak. **İkinci yol** ise P_i'nin 0 ile 1 arasında kalmasını sağlayacak olan aşağıdaki geliştirilmiş Logit Modeli tekniğini uygulamaktır.

2.2.2.2. Logit Model

Günümüzde nitel değişkenlerden oluşan dummy verileri analiz etmek için çeşitli teknikler kullanılmaktadır. Yapılan bu çalışmada dummy verileri analiz etmek için log-linear modeller kullanılacaktır. Log-linear modeller iki veya daha fazla dummy değişkenin koşullu ilişkisini analiz etmek için geliştirilmiştir. Bununla birlikte, log-linear modeller sayesinde, değişkenlerin oluşturduğu bileşik dağılımı, iki veya daha fazla değişkenin birbirine bağımlı olup olmadığını ve iki veya daha fazla değişken arasındaki ilişkiyi neden-sonuç ilişkisine dayandırmaksızın test etmek mümkündür (Özdamar, 1999:449-450).

Logit modeller, genelleştirilmiş doğrusal modelin belirli koşullar altında oluşturulmuş özel durumlarıdır. Bu durumda yapılacak olan çalışmada, eğer bağımsız değişkenlerin bazısı sürekli veya uygun (ilgili) sınıflar içine ayrıştırılamazsa, o zaman log-linear analiz yerine logistik regresyon kullanılmalıdır. Aynı zamanda eğer değişkenlerin bazısı bağımlı olarak ele alınırsa, o zaman **logit model** uygundur.

Böyle bir durumda 0'la 1 arasında kalma koşulunu sağlayabilmek için logit modelin uygulanması önerilmektedir (Gujarati,1995: 555). Logit model, bağımlı değişkenin tahmini değerlerini olasılık olarak hesaplayarak olasılık kurallarına uygun sınıflama yapma imkanı veren, tablolaştırılmış ya da ham veri setlerini analiz eden bir istatistiksel yöntemdir (Özdamar, 1999:476).

Logit model, bağımsız değişken değeri sonsuza gittiği zaman, bağımlı değişkenin 1'e asimptot olduğu matematiksel bir fonksiyondur.

$$P_i = E(Y = 1 \mid X_i) = \alpha + \beta X_i \tag{11}$$

$$P_i = E(Y_i = 1 \mid X_i) = \frac{1}{1+e^{-(\alpha+\beta X_i)}} = \frac{1}{1+e^{-Z_i}} \tag{12}$$

Burada:

$Z_i = \alpha + \beta Xi$ 'dir.

P_i: açıklayıcı değişken (X_i) hakkında bilgi verirken i-nci bireyin belirli bir tercihi yapma olasılığını ifade etmektedir.

$Z_i = \alpha + \beta Xi$

e = 2,71828'dir (Özdamar, 1999:477).

(12) nolu model logit model olarak adlandırılır. X hangi değerleri alırsa alsın fonksiyondaki eksponansiyel terim daima pozitif olacağı için P_i' nin alt sınırı da 0 olur. Olasılık için gerekli olan $0 \leq P_i \leq 1$ koşulunu bu fonksiyon sağlamış olur. **Logit dağılım fonksiyonu** diye adlandırılan Z_i değişkeni $-\infty$ ile $+\infty$ arasında değer aldıkça P_i de 0 ile 1 arasında değerler alacak ve P_i ile Z_i arasındaki ilişki doğrusal olmayacaktır. Böylece $0 \leq P_i \leq 1$, ve Z_i ile P_i arasındaki ilişkinin doğrusal olmama şartları yerine gelmiş olacaktır. Fonksiyonun belirlenmesi için α ve β parametreleri en küçük kareler yöntemi ile doğrudan tahmin edilemez. Önce ilişki üzerinde bazı işlemler yaparak doğrusal bir ilişki elde etmeye çalışılacaktır. Bu amaçla (11) nolu model, α ve β' ya göre çözülerek , modelin tahmini için;

P_i: Dört yılda mezun olma olasılığı

$1-P_i$: Dört yılda mezun olamama olasılığıdır.

$Pi = \dfrac{1}{1+e^{-Zi}}$ eşitliğinin her iki yanı ($1+e^{-Zi}$) ile çarpılarak

$$(1+e^{-Zi})Pi = 1 \tag{13}$$

elde edilir. Şimdide P_i ile bölüp 1 çıkartılarak,

$$e^{-Zi} = \frac{1}{Pi} - 1 = \frac{1-Pi}{Pi} \tag{14}$$

elde edilir.

$e^{-Zi} = \dfrac{1}{e^{Zi}}$ kullanılarak,

$$e^{Zi} = \frac{Pi}{1-Pi} \tag{15}$$

gösterilebilir. Bu eşitlik bize dört yılda mezun olma olasılığının, dört yılda mezun olamama olasılığına olan oranını verir. Aynı zamanda bu oran bahis oranıdır (odds ratio). (15) nolu modelin e tabanına göre doğal logaritması alınarak,

$$\ln\left(\frac{Pi}{1-Pi}\right) = \ln e^{Zi} = Zi = \alpha + \beta Xi \tag{16}$$

elde edilir. Yani, bahis oranının logaritması $L_{i,}$ yalnız X' e göre değil, (katsayı tahmini bakımından) anakütle katsayılarına göre de doğrusaldır. $L_{i,}$'ye **logit** denir. (16) nolu modellerin adı olan logit modeli de buradan gelir. Bu, parametrelerin tahmininde doğrusal bir ilişki işlemi görebilecek yarı logaritmik bir fonksiyondur.

Modeldeki parametreleri tahmin etmek için Li fonksiyonu,

$$Li = \ln(\frac{P(Y)}{1-P(Y)}) = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_P X_P \tag{17}$$

şeklinde yazılır. α ve $\beta_1, ...\beta_p$ regresyon, katsayılarıdır. P_i=1 ve P_i=0 değerleri logit Li' deki yerine koyulduğunda $\ln(\frac{1}{0})$ ve $\ln(\frac{0}{1})$ değerleri elde edilir ki bunlar anlamsızdır. En küçük kareler yöntemi ile L_i fonksiyonundaki parametrelerin tahmin değerleri bulunamaz, fakat bu parametreler maksimum olabilirlik modeli ile tahmin edilebilir.

Sınıflama ve atama işleminin olduğu, normal dağılım varsayımı, süreklilik varsayımı ön koşulunun olmadığı gibi durumlarda verilerin Logit modeli ile analiz edilmesi gerekir.

Tablo: 2
Üzerinde Çalışılan Verilere Ait Örnek Veri Kümesi

Sıra	Mezuniyet Durumu (Y)	Cinsiyet (X1)	Lise Türü (X2)	Sayısal Net Sayısı (X3)
1	1	2	1	5,75
2	1	1	1	13,50
3	1	2	1	14,00
...	....	....	....	....
98	1	1	2	5,25
99	1	1	1	5,00
100	0	1	1	9,50

Mezuniyet Durumu(Y): 1= Dört yılda 0= Beş ve üzeri yıl
Lise Türü: 1= Klasik 2= Meslek 3= Özel, Anadolu, Fen
Cinsiyet: 1= Bayan 2= Bay

3. Araştırma ve Bulgular

Örnek veri olarak Niğde Üniversitesi İİBF İşletme ve İktisat bölümlerinden alınan mezun olmuş durumda bulunan 100 öğrenciye ait değerler kullanılmıştır.

Verilerin SPSS 10.0 paket programında doğrusal regresyon işlemine tabi tutulmasıyla aşağıdaki değerler elde edilmiştir:

Tablo: 3
En Küçük Kareler Yöntemiyle (n=100) Tahmin Edilen Parametre Değerleri

Değişkenler	β	St. Hata	t	Anlam. Düz.
Sabit	0,334	0,226	1,478	0,143
Sayısal Net	0,02	0,010	1,978	0,051
Cinsiyet	-0,04	0,104	-0,373	0,710
Lise Türü	0,091	0,071	0,127	0,207
R^2	0,060			

Bu değerler ışığında oluşturulan doğrusal regresyon modeli aşağıdaki biçimdedir:

$$Y = \alpha + \beta Xi + ei$$

$$Y = \begin{cases} 1 & \text{Öğrenci dört yılda mezun olmuşsa} \\ 0 & \text{Öğrenci dört yılda mezun olamamışsa} \end{cases}$$

X1=Öğrencinin cinsiyeti

X2=Öğrencinin mezun olduğu lise türü

X3 = Öğrencinin Üniversiteye giriş (ÖYS) sınavındaki sayısal neti

ei= Hata terimi'dir.

$\hat{Y} = 0.334 - 0.04X_1 + 0.091X_2 + 0.02X_3$ denklemi kurulabilir.

Örneğin, Cinsiyeti erkek (2), Lise türü Klasik lise (1) ve sayısal neti 5,75 puan olduğunda, öğrencinin ilgili fakülteden dört yılda mezun olma olasılığı,

$\hat{Y} = 0.334 - 0.04(2) + 0.091(1) + 0.02(5.75)$ =0,46

%46 olarak bulunacaktır. Buna karşılık Cinsiyeti Kız (1), Lise türü klasik lise(1) ve sayısal neti 5,75 puan olduğunda öğrencinin ilgili fakülteden dört yılda mezun olma olasılığı,

$$\hat{Y} = 0.334 - 0.04(1) + 0.091(1) + 0.02(5.75) = 0,50$$

%50 olarak hesaplanacaktır.

Böylece farklı giriş puanları için farklı olasılıklar bulunabilir. $0<E(Yi|Xi)<1$ şartının yerine gelmemesi durumu da söz konusu olabilecektir. Örneğin cinsiyeti: kız(1), Lise türü: klasik lise(1) ve sayısal neti:31.75 puan olan öğrenci için fakülteyi dört yılda bitirme olasılığı,

$$\hat{Y} = 0.334 - 0.04(1) + 0.091(1) + 0.02(31.75) = 1,02$$

%102 Olacaktır. Bu rakam 1' den büyük olduğu için $0<E(Yi|Xi)<1$ şartına aykırıdır.

Lise türü bağımsız değişkeninin ışığında hesaplanabilecek bağımlı değişkene ilişkin olasılık değerleri de şöyle bulunabilir:

Bağımlı Değişken: Mezuniyet durumu (Y_1)

Öğrenci klasik lise mezunu ise X2 =1,

Meslek lisesi mezunu ise X2=2

ve Özel lise, Anadolu veya Fen lisesi mezunu ise X2=3 olur.

En küçük kareler yöntemi ile elde edilen bazı örnek bulgular şöyledir:

a) Cinsiyeti: kız (1), Lise türü: Klasik lise (1)

Sayısal neti: 4 puan

$$\hat{Y} = 0.334 - 0.04(1) + 0.091(1) + 0.02(4) = \%46,5$$

b) Cinsiyeti: kız (1), Lise türü: Meslek lisesi (2)

Sayısal neti: 4 puan

$$\hat{Y} = 0.334 - 0.04(1) + 0.091(2) + 0.02(4) = \%55,6$$

c) Cinsiyeti: kız (1), Lise türü: Özel okul (3)

Sayısal neti: 4 puan

$\hat{Y} = 0.334 - 0.04(1) + 0.091(3) + 0.02(4)$ =%64,7

Buna karşılık; cinsiyet değişkeninin ve lise türü değişkenlerinin değiştirilmesi halinde;

a) Cinsiyeti: Erkek (2), Lise türü: Klasik lise (1)

Sayısal neti: 31.75 puan

$\hat{Y} = 0.334 - 0.04(2) + 0.091(1) + 0.02(31.75)$ =%98

b) Cinsiyeti: Erkek (2), Lise türü: Meslek (2)

Sayısal neti: 31.75 puan

$\hat{Y} = 0.334 - 0.04(2) + 0.091(2) + 0.02(31.75)$ =%107,1

c) Cinsiyeti: Erkek (2), Lise türü: Özel okul (3)

Sayısal neti: 31.75 puan

$\hat{Y} = 0.334 - 0.04(2) + 0.091(3) + 0.02(31.75)$ =%116,2

olacaktır. Burada kullanılan bağımsız değişken değerleri ve bunlar karşılığında bulunan değerlerle ilgili tablo aşağıdadır:

Tablo: 4
En Küçük Kareler Yöntemiyle (n=100) Elde Edilen Bazı Tahmin Sonuçları

Cins.	Lise Türü	Say. Net	Dört Yılda Bit. Olas. (%)
E(2)	Klasik(1)	5,75	46
K(1)	Klasik(1)	5.75	50
K(1)	Klasik(1)	31,75	102
E(2)	Klasik(1)	31,75	98
E(2)	Meslek(2)	31,75	107,1
E(2)	Ö-AL-FL(3)	31,75	116,2
K(1)	Klasik(1)	4,00	46,5
K(1)	Meslek(2)	4,00	55,6
K(1)	Ö-AL-FL(3)	4,00	64,7

Yukarıdaki bulgulardan da görülebileceği gibi, **doğrusal regresyon modelinden elde edilen bazı olasılık değerleri 1 (%100) rakamının sağına düşmektedir. Oysa olasılık değerinin 1'den büyük olamayacağı ortadadır.**

Burada 0<E(Y=1|X)<1 şartını sağlayabilmek için logit modeli uygulanacaktır. Doğrusal olasılık modeline göre öğrencinin fakülteden dört yılda mezun olma olasılığı ile sayısal net arasındaki ilişki

$Pi = E(Y=1 \mid Xi) = \beta_1 + \beta_2 Xi$ şeklinde gösterilmişti. Öğrencinin fakülteden dört yılda mezun olma olasılığı ile sayısal net arasındaki ilişkiyi aşağıdaki gibi gösterilebilir:

$$\hat{L} = 0{,}213 + 0{,}158\,X_1 - 0{,}953\,X_2 + 0{,}095\,X_3$$

O zaman denklem aşağıdaki biçimde olacaktır:

$$\hat{Pi} = \frac{1}{1 + e^{-(0.213 + 0.158 X_i - 0.953 X_2 + 0.095 X_3)}}$$

SPSS paket programın da Logit Li için şu regresyon bulunur:

Tablo: 5
Logit Modeliyle Elde Edilen Tahmin Değerleri

Değişkenler	β	St Hata	Sd	Anlam. Düz.
Sabit	0,213	1,036	1	0,837
Sayısal Net	0,095	0,052	1	0,067
Cinsiyet	0,158	0,465	1	0,734
Lise Türü1	-0,953	0,839	1	0,256
Lise Türü2	-0,595	0,934	1	0,524
Lise Türü3	0	0	2	0,451

Yukarıdaki değerler ışığında, bulunacak olasılık değerlerinin % 100 değerinin ötesine düşüp düşmediği kontrol edilir. Örneğin;

Cinsiyeti: **Kız(1)**, Lise türü: Klasik lise(1) ve sayısal neti: 31.75 puan olan bir öğrenciyi ele alalım. Klasik lise mezunu olan bu öğrencinin fakülteden dört yılda mezun olma olasılığı,

$\hat{P} = \dfrac{1}{1+e^{-2.434}}$ =0,919 % 91,9 olur. Herhangi bir giriş puanı için mezun olma olasılığı tahmin edilebilir.

Örneğin,Cinsiyeti: **Erkek(2)**, Lise türü: Klasik lise(1) ve sayısal neti: 31.75 puan olan bir öğrenciyi ele alalım. Öğrencinin fakülteden dört yılda mezun olma olasılığı,

$\hat{P} = \dfrac{1}{1+e^{-2.5925}}$ =0,9304 % 93,04 olur.

Örneğin;Cinsiyeti: **Kız(1)**, Lise türü: Klasik lise(1) ve sayısal neti: 9 olan bir öğrenciyi ele alalım. Bu öğrencinin fakülteden dört yılda mezun olma olasılığı,

$\hat{P} = \dfrac{1}{1+e^{-0.273}}$ =0,5678 %56,78 olur.

Örneğin; Cinsiyeti: **Erkek(2)**, Lise türü: Klasik lise(1) ve sayısal neti: 9 olan bir öğrenciyi ele alalım. Bu öğrencinin fakülteden dört yılda mezun olma olasılığı,

$\hat{P} = \dfrac{1}{1+e^{-0.431}}$ =0,606 %60,6 olacaktır.

Örneğin; Cinsiyeti: Kız(1), **Lise türü: Meslek lisesi(2)** ve sayısal neti: 9 olan bir öğrenciyi ele alalım. Bu öğrencinin ise, fakülteden dört yılda mezun olma olasılığı, ilgili formül ışığında

$$\hat{P} = \frac{1}{1 + e^{-0.036}}$$ =0,501 %50,1 olacaktır.

Örneğin; Cinsiyeti: Kız(1), **Lise türü: özel lise(3)** ve sayısal neti: 9 olan bir öğrenciyi ele alalım. Bu öğrencinin fakülteden dört yılda mezun olma olasılığı,

$$\hat{P} = \frac{1}{1 + e^{-1.226}}$$ =0,7731 %77,31 olacaktır.

Örneğin; Cinsiyeti: Erkek(2), **Lise türü: özel lise(3)** ve sayısal neti: 9 olan bir öğrenciyi ele alalım. Öğrenci özel lise mezunu (Xi=3) ise, öğrencinin dört yılda fakülteden mezun olma olasılığı,

$$\hat{P} = \frac{1}{1 + e^{-1.384}}$$ =0,7996 %79,96 olacaktır.

Örneğin; Cinsiyeti: Erkek(2), **Lise türü: Klasik lise(1)** ve Sayısal neti: 9 olan bir öğrenciyi ele alalım. Bu öğrencinin fakülteden dört yılda mezun olma olasılığı,

$$\hat{P} = \frac{1}{1 + e^{-0.431}}$$ =0,6061 %60,61 bulunur.

Örneğin; Cinsiyeti: Erkek(2), **Lise türü: Meslek lisesi(2)** ve sayısal neti: 9 olan bir öğrenciyi ele alalım. Öğrenci meslek lisesi mezunu (Xi=2) ise,öğrencinin dört yılda fakülteden mezun olma olasılığı,

$$\hat{P} = \frac{1}{1 + e^{-0.194}}$$ =0,5483 %54,83'tir.

Örneğin;Cinsiyeti: Erkek(2), Lise türü: Klasik lise(1) ve **sayısal neti: 22** olan bir öğrenciyi ele alalım. Öğrenci düz lisesi mezunu (Xi=1) ise, öğrencinin fakülteden dört yılda mezun olma olasılığı,

$$\hat{P} = \frac{1}{1 + e^{-1.666}}$$ =0,8410 %84,1 olur.

Örneğin; Cinsiyeti: Erkek(2), Lise türü: Meslek lisesi(2) ve sayısal neti: 22 olan bir öğrenciyi ele alalım. Görüldüğü gibi, bu öğrencinin fakülteden dört yılda mezun olma olasılığı,

$$\hat{P} = \frac{1}{1 + e^{-1.429}}$$ =0,8067 %80,67'dir.

Örneğin; Cinsiyeti: Erkek(2), Lise türü: Özel lise(3) ve sayısal neti: 22 olan bir öğrenciyi ele alalım. Bu öğrencinin de fakülteden dört yılda mezun olma olasılığı olarak ,

$$\hat{P} = \frac{1}{1 + e^{-2.619}}$$ =0,9321 %93,21 bulunur.

Kullanılan bağımsız değişken değerlerine ve bunlar karşılığında bulunan tüm bu değerlere ilişkin tablo aşağıdadır:

Tablo: 6
Logit Modeliyle Elde Edilen Tahmin Sonuçları

Cins.	Lise Türü	Say Neti	Dört Yılda Bit. Olas.(%)
K(1)	Klasik(1)	31,75	91,90
E(2)	Klasik(1)	31,75	93,04
K(1)	Klasik(1)	9,00	56,78
E(2)	Klasik(1)	9,00	60,60
K(1)	Meslek(2)	9,00	50,10
K(1)	Ö-AL-FL(3)	9,00	77,31
E(2)	Ö-AL-FL(3)	9,00	79,96
E(2)	Klasik(1)	9,00	60,61
E(2)	Meslek(2)	9,00	54,83
E(2)	Klasik(1)	22,00	84,10
E(2)	Meslek(2)	22,00	80,67
E(2)	Ö-AL-FL(3)	22,00	93,21

Bu tablodan da görülebileceği gibi bulunan tüm değerler 0 ile 1 aralığında olmakta ve böylelikle **doğrusal regresyon modelince sağlanamayan olasılık değerlerinin 0 ile 1 aralığında olma kuralı böylelikle sağlanmış olmaktadır.**

X=22 puan ile X=9 puan arasındaki fark $(\Delta x = 13)$ için

$$\Delta \hat{p} = 0.799 - 0.932 = 0.133$$

X= 31.75 ile X=9 arasındaki fark

$(\Delta x = 22.75$ için $\Delta \hat{p} = 0.919 - 567 = 0.352$; değerleri bulunur.

Farklı sayısal net puanları, puandaki belli bir miktar artış (örneğimizde $\Delta X = 13)$ mezun olma olasılığını değişik miktarlarda artırmaktadır. Bu olasılık düşük puanlarda az, orta ve yüksek puanlarda eşit miktarda artmaktadır.

Bu model için elde edilen bulgulara göre de, öğrencinin mezun olduğu lise türü öğrencinin fakülteden dört yılda mezun olma olasılığını büyük ölçüde artırmakta olduğu söylenebilecektir.

Aynı şekilde, öğrencinin cinsiyetinin de fakülteden dört yılda mezun olma durumunu kısmen etkilediği söylenebilecektir. Kız öğrencilerin erkek öğrencilere göre belirgin bir üstünlüğü görülmektedir.

Okullar arasında da başarı konusunda ilk sırayı üçüncü kategori (Anadolu, Fen, Süper Liseler) almakta, bunu meslek liseleri izlemektedir. Öğrencinin fakülteyi dört yılda bitirebilmesi konusunda klasik lisenin etkisi diğerlerine göre minimum olmaktadır.

ÖYS'de yapılan sayısal net sayısı, büyüklüğü oranında öğrencinin fakülteyi dört yılda bitirme olasılığını etkilemektedir.

4. Sonuç ve Öneriler

İki veya daha fazla değişken arasında ilişki olup olmadığı, varsa bunun derecesi veya birinin bir ölçü birimi değişmesine karşılık, ötekinin değişeceği miktar, ya da değişkenlerden birinin, herhangi bir değerine öteki değişkenlerden hangi değerlerin tekabül edeceği gibi problemler regresyon analizi ile çözülebilir.

Dummy değişkenin katsayısı, "1" ile gösterilen karakteristik ile regresyondan dışlanan "0" ile gösterilen karakteristik arasındaki farkı göstermek yerine çeşitli grupların örnek ortalamasına göre farklarını yansıtır.

Regresyonda bir veya birkaç sürekli değişkenin bulunması halinde katsayıların yorumları güçleşir.

Günümüzde Logit modelin parametrelerini tahmin için MINITAB, SYSTAT, STATISTICA, SPSS, SAS, S-PLUS vb. bazı istatistik paket programlar kullanılmaktadır. Önemli olan modeli ve bu model çerçevesinde elde edilecek bulguları iyi yorumlayıp değerlendirebilmektir.

Esas kullanım alanı ve orijini biyolojik bilimler olan bu model, özellikle son yıllarda bilgisayar teknolojisinin ve paket programların gelişmesi ile de önem kazanarak bir çok bilim dalında kullanılır hale gelmiştir.

Regresyon modelinde, bağımlı değişkenin dummy değişken olması durumunda yönteme özgü sorunlarla karşılaşılmaktadır:

1. Açıklanan (bağımlı) değişkeni "**evet-hayır**" ya da "**var-yok**" yanıtlarından oluşan regresyon modelleri, iki uçlu ya da dummy bağımlı değişkenli modeller olarak bilinir. Çok çeşitli alanlarda uygulanabilirler, anket ya da sayım verileriyle yaygın olarak

kullanılırlar.

2. Bu tür modelleri tahminde kullanılan modellerden ikisi, Doğrusal Olasılık Modeli ve Logit modelleridir.

3. Bu iki modelden kullanımı daha basit olan Doğrusal Olasılık Modelidir. Bu çalışma ile de ortaya konulduğu gibi, şu eksiklikleri vardır:

a. Hata teriminin normal dağılmayışı,

b. Değişen varyans,

c. Tahmin edilen olasılığın $0 \leq E(Y_i \mid X_i) \leq 1$ sınırları dışına düşebilmesi,

Bu sorunlar çözülse bile, doğrusal olasılık modeli koşullu olasılıkların, açıklayıcı değişkenle birlikte doğrusal olarak arttığını varsaydığı için mantıksal olarak çok çekici bir model değildir. Açıklayıcı değişkenlerin değerleri sonsuza doğru arttıkça ya da azaldıkça olasılıkların gittikçe düşmesi daha beklenir bir şeydir. Öyleyse burada gerekli olan, birikimli dağılım fonksiyonunun S biçimli özelliğini taşıyan bir olasılık modelidir.

4. Birikimli dağılım fonksiyonu için seçenekler çoksa da uygulamada Logit modelin daha geçerli bir metot olduğu bu çalışma ile de ortaya konulmaktadır.

5. Tahmin edilen olasılıkların 0-1 aralığına düşmesi konusunda logit model daha güvenilir sonuçlar vermektedir.

Bu uygulama ile temel olarak şu gösterilmeye çalışılmıştır: Doğrusal olasılık modelinden ede edilen sonuçlar, bazen 0-1 aralığının dışına düşebilmektedir. Bu durumu ortadan kaldıran ve tüm olasılık değerlerinin 0-1 aralığında yer almasını sağlayacak yöntemlerden birisi logit modeldir.

Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi İşletme ve İktisat Bölümleri'nden 2000-2001 eğitim- öğretim yılında mezun olmuş bulunan 130 öğrenciden 100'ünden elde edilen verilere göre kurulan bu modelden elde edilen bulgular ışığında, **öğrencinin mezun olduğu lise türü**, öğrencinin fakülteden dört yılda mezun olma olasılığını büyük ölçüde artırmakta olduğu söylenebilecektir.

Ayrıca, öğrencinin **cinsiyet**inin de fakülteden dört yılda mezun olma durumunu kısmen etkilediği söylenebilecektir. Bu verilere göre, kız öğrencilerin erkek öğrencilere göre belirgin bir üstünlüğü görülmektedir.

Okullar arasında da başarı konusunda ilk sırayı üçüncü kategori (Anadolu liseleri, özel okullar, fen liseleri) almakta, bunu meslek liseleri izlemektedir. Öğrencinin fakülteyi dört yılda bitirebilmesi konusunda klasik lisenin etkisi diğerlerine göre minimum olmaktadır.

ÖYS'de yapılan sayısal net sayısı da, büyüklüğü oranında öğrencinin fakülteyi dört yılda bitirme olasılığını etkilemektedir.

Bu çalışmayla, logit modelin diğer tahmin yöntemlerine göre daha doğru sonuçlar ürettiği ortaya konulmuştur. Araştırmacıların bu ve buna benzer çalışmalarda, logit modelini kullanmaları daha sağlıklı sonuçlar elde etmelerini sağlayacaktır.

Kaynakça

Alpar, R. (1997), *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş-1*, Kültür Ofset, Ankara.

Başar, A. ve Oktay E. (2000), *Uygulamalı İstatistik-2*, Aktif Yayınevi, İstanbul.

Genceli, M. (1989), *Ekonometride İstatistik İlkeler*, Filiz Kitabevi, İstanbul.

Gujarati, D. N. (1995), *Basic Econometrics*, McGraw –Hill Inc., Int. Eds. 3. ed., Istanbul.

İnal, C. vd. (1983), *İstatistik Terimleri Sözlüğü*, T.D.K. Yayınları, Ankara.

İşyar, Y. (1999), *Ekonometrik Modeller*, VİPAŞ AŞ, 2. bs., Bursa.

Kılıçbay, A. (1986), *Ekonometrinin Temelleri*, İstanbul Ünv. Film Merkezi ve Matbaası, İstanbul.

Krause, A.ve Olson, M. (2000), *The Basics of S and S-PLUS*, 2. eds., New York.

Maskie, G. (2001), "Analisis Model Logistic*", Lintasan Ekonomi*, 18(1), p. 48-53.

Nakip, M. (2003), *Pazarlama Araştırmaları, Teknikler ve (SPSS Destekli) Uygulamalar*, Seçkin Yay., Ankara.

Özdamar, K. (1999), *Paket Programlarla İstatistiksel Veri Analizi*, c. 1, 2. bs., Kaan Kitabevi, Eskişehir.

Püskülcü, H. ve İkiz, F. (1986*), İstatistiğe Giriş*, Ege Ünv. Basımevi, İzmir.

Tarı, R. (1999), *Ekonometri*, Alfa Yayınevi, İstanbul.