LOJİSTİK REGRESYON ANALİZ TEKNİĞİNİN EĞİTİM BİLİMLERİ ARAŞTIRMALARINDA UYGULANABİLİRLİĞİ İLE İLGİLİ BİR ARAŞTIRMA

[1]Derviş TOPUZ **[2]Mehmet ÇAKIR**

[1]ÖğretimGörevlisiNiğdeÜniversitesiFenBilimleri Enstitüsü/ Niğde/Turkey(topuz@nigde.edu.tr)

[2]Doç.Dr.Niğde Üniversitesi Fen Bilimleri Enstitüsü /Niğde/Turkey(mcakir25@hotmail.com)

ÖZET

Son yıllarda kategorik (kesikli) verilerin analizinde lojistik regresyon analiz tekniği gibi teknikler eğitim bilimleri, sosyal bilimler, tıp, biyoloji, halk sağlığı, tarım, veterinerlik, balıkçılık, ormancılık ve gıda, vb. alanlarda çok sık kullanılır olmuş ve geleneksel ki-kare analizi, korelasyon analizi ve birliktelik analizlerinin yetersiz kaldığı durumlarda sıklıkla uygulanır olmuştur.

Lojistik regresyonda, bir veya daha fazla bağımsız değişken ile kesikli bir bağımlı değişken arasındaki ilişki incelenmektedir. Lineer regresyondan farklı olarak, Lojistik regresyonda, bağımlı değişkenin bağımsız değişkenin etkisi ile kaç kat daha fazla ya da yüzde kaç oranda fazla gözlenme olasılığına sahip olduğu tahmin edilmektedir. $X = (x_1, x_2, \dots\dots\dots, x_p)$ şeklinde p tane bağımsız değişken için çoklu Lojistik regresyon modeli,

$$Li = \ln\left(\frac{P(Y)}{1 - P(Y)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots\dots \dots + \beta_p X_p$$

şeklindedir.

Bu çalışmanın amacı, lojistik regresyon analiz tekniğinin eğitim alanındaki çalışmalarda uygulanabilirliliği ile bu tekniğin diğer regresyon tekniklerine kıyasla ne tür avantajlar sağladığını saptamak olmuştur. Mustafa Kemal Üniversitesi (M.K.Ü) 'nin çeşitli Fakülte

ve Yüksek Okullarına 1994-1995 eğitim öğretim yılında giren öğrencilere ait dosyalardan, rasgele örnekleme yöntemi ile seçilen 492 öğrenciye ait dosya verileri kullanılmıştır. Öğrencinin üniversiteye giriş puanı ve mezun olduğu lise türü ile üniversiteden mezun olma olasılığı arasındaki ilişkiler; hem lojistik regresyon hem de alternatif regresyon teknikleri (Doğrusal olasılık modeli, Maksimum olabilirlilik yöntemi) kullanılarak analiz edilmiştir. Bağımsız değişkenler: Öğrencinin üniversiteye giriş puanı (X_1), Lise türü (X_2), bağımlı değişken (Y_i), öğrencinin mezun olma olasılığına etkileri Y_1=1 mezun Y_2= 0 mezun değil şeklinde S-PLUS 2000 for WINDOWS istatistik paket programı kullanılarak değerlendirilmiştir. Analiz bulguları alternatif regresyon modellerine kıyasla lojistik regresyon tekniğinin daha net sonuçlar ürettiğini ortaya koymuştur. Bu çalışma sonucunda eğitim alanındaki çalışmalarda, en azından bu çalışmada örnek olarak seçilen olaya benzer türdeki olayların analizinde Lojistik regresyon tekniğinin daha kullanışlı olduğu kanısına varılmıştır.
Anahtar sözcükler: Lojistik regresyon, Lojistik model, S-Plus, İstatistik, odds.

1. GİRİŞ

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

Regresyon analizi; biyoloji, ziraat, tıp, ekonomi ve fizik gibi bilim dallarında yapılan ve araştırmalarda kullanılan istatistik yöntemlerden birisi olup, aralarında sebep sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi saptamada kullanılır (Püskülcü ve İkiz;1986).

Çeşitli nedenlerle, bir değişkenin değerinin, diğer değişkenlerdeki değişimlere bağlı olarak nasıl etkilendiği istatistiksel analizleri incelenmektedir. Değişkenler arasındaki ilişki bilindiğinde bir değişkenin değerine bakarak diğerinin alacağı değeri tahmin edilebileceği gibi, etki eden faktörler kontrol altına alınabilirse, ilgilenilen değişkenlerin değeri optimum düzeye getirilebilir (Alpar, 1999).
Regresyon analizi, populasyon yerine örneklemden yola çıkıldığında, bağımsız değişkenlerden yararlanarak bağımlı değişkenin tahmin edilmesi ve bu süreçte yapılan hataların ölçülmesidir. Korelasyon analizinde ise bu değişkenler arasındaki ilişkinin yön ve derecesinin saptanmasıdır (Alpar,1997)
Regresyon modellerinde ele alınan değişkenlerin başlıca özelliklerinden biri, süreklilik özelliğini içermeleridir. Ancak çok sık rastlanmasa da regresyon modellerinde nitel (kalitatif) değişkenleri de inceleme zorunluluğu ortaya çıkmaktadır (Kılıçbay,1986). Yabancı kaynaklarda "Dummy Variable" veya "Binary Data" adları ile rastlanılan bu kalitatif değişkenlere Türkçe'de "Kategorik" adı önerilmiştir (İnal, 1983). Kategorik değişkenler regresyon modellerinde tıpkı nicel değişkenler gibi kolaylıkla kullanılabilirler(Genceli,1987). Gerçekte, bir regresyon modelinin bütün açıklayıcı değişkenleri özünde kategorik ya da nitel değişken olabilir. Böyle modeller bağımlı değişkenin değişkenliğini incelemek amacıyla kullanılan modellerdir. Kategorik değişkenlerin doğal bir ölçüm sıkalası yoktur. Bu nedenle, bu değişkenlerin modele katkısının görülmesi için "bir sınıfa" bu tür değişkenlerin atanması gerekir. Bu sınıfı oluşturan değişkenlere kategorik değişkenler denir. Kategorik değişkenler çoğunlukla 0 ve 1 değerlerini alır ve bu yeni değişkenlerin sayısal olarak hiç bir önemi yoktur (Genceli, 1989). Çünkü,

sadece gözlemlerin hangi kategoriye ait olduklarını gösterirler. Bu değişkenlerin çoğu iki sonucu içerir, Örneğin cinsiyet değişkeni erkek-dişi, ölü-canlı, hatalı-hatasız, hasta veya değil vb gibi. Bazılarının üç sınıfı (kategorisi) vardır, yaşanan bölge (köy, kasaba, kent) gibi. Bu tür değişkenler daha fazla sınıflı da olabilir.

Doğrusal olasılık modeli

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

$$Y_i = \beta_0 + \beta_i X_i + e_i \qquad (1)$$

eşitliği ele alınır (Genceli,1987). Burada ;

X_i : Bağımsız (açıklayıcı) değişkenler

Y_i: Bağımlı kategorik değişken (mezun olma durumu)

$Y_i = 1$ ise mezun

$Y_i = 0$ ise mezun değil

β : Regresyon katsayıları

e_i: 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 (linear probability models LPM) denir. Çünkü, X_i verildiğinde Y_i'nin koşullu beklenen değeri $E(Y_i/X_i)$, X_i ye 'i' değeri verirken olayın gerçekleşmesinin koşullu olasılığı olarak yorumlanır (Tarı; 1999).

Sapmasız tahmin edicilere ulaşmak için, e_i, ortalaması sıfır olan şansa bağlı (random) bir değişken olarak varsayıldığında, $E(e_i) = 0$ olduğu için,

$$E(Y_i/X_i) = \beta_0 + \beta X_i \tag{2}$$

modeli elde edilir. $Y_i = 1$ ise olayın gerçekleşme olasılığı (P_i), $Y_i = 0$ ise olayın gerçekleşmeme olasılığı ($1-P_i$) o zaman, $\Sigma p_i = 1$ dir (Tarı,1999).

Bu nedenle, matematiksel beklenen değer tanımından:

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

değeri bulunur.bu değer (2)'nolu modelde yerine konulursa,

$$E(Y_i/X_i) = \beta_0 + \beta X_i = P_i \tag{4}$$

model (4) elde edilir.

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):

$$P_i = \left\{ \begin{array}{lll} \beta_0 + \beta X_i & \text{ise} & 0 < \beta_0 + \beta X_i < 1 \text{ dir} \\ 1 & \text{ise} & \beta_0 + \beta X_i \geq 1 \text{ dir} \\ 0 & \text{ise} & \beta_0 + \beta X_i \leq 1 \text{ dir.} \end{array} \right\}$$

P_i olasılığı, 0 ile 1 arasında bulunacağından, $0 \leq E(Y_i / X_i) \leq 1$ veya 0 $< \beta_0 + \beta X_i < 1$ şeklinde bir sınırlama zorunlu olduğundan koşullu olasılık 0 ile 1 arasında kalmalıdır (Genceli, 1989).

Regresyon modellerinde parametrelerin en küçük kareler yöntemi ile bulunduğunun bilinmesine karşın, bağımlı değişkenin kategorik değişken olduğu durumda özel sorunlarla karşılaşılabilir. Doğrusal olasılık modelinin sunduğu bu bilgilere rağmen, bu modelin tahmini ve yorumuna ilişkin bazı eleştiriler vardır (Tarı; 1999);. Bunlar en küçük kareler yönteminin: Hata terimi(e_i) nin normal dağılımlı olmaması durumu, hata terimi (e_i) nin değişen varyanslı olması durumu, çoklu belirtme katsayısı (R^2) değerinin genellikle küçük çıkarak, ilişkinin uyumunu gösteren bir ölçü olamaması durumu, $P_i = E(Y=1/X_i)$ Değerinin X_i ile doğrusal olarak arttığının varsayılması ve $0 \leq E(Y_i / X_i) \leq 1$ koşulunun sağlanamaması gibi nedenlerden dolayı lojistik regresyon analiz tekniğinin kullanımı önerilmektedir (Gujarati, 1979). Lineer regresyondan farklı olarak, lojistik regresyonda, bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir. Örneğin, biyolojik doz-yanıt(dose-response) çalışmalarında, bir ilacın farklı dozlarda uygulanması halinde ölen zararlıların oranı ölçülmüş olabilir. Burada esas olarak ''gerçekleşme''olasılığı, p, ile ilgilenilmektedir. Burada p olasılığı $[0,1]$ aralığında değer alır ve doğrudan doğruya lineer bir modelle tanımlanamaz. P olasılığının lojistik dönüşümü $\log(\frac{p}{1-p})$ şeklindedir ve $[-\infty, +\infty]$ aralığında değer alır. Bu dönüşüm p nin ''lojit''i olarak adlandırılır. Eğer, ilgili değişkenin k tane sınıfı var ise, modelde bu değişken için, k–1 tane göstermelik değişken tarlımlanır (Alpar, 1997). Kategorik değişkenler açıklayıcı (bağımsız) değişkenler olabildiği gibi, bağımlı değişken de olabilir. Bu durumda bağımlı değişken sadece iki değer alır . Daha açık bir anlatımla, olayın varlığı için "1", yokluğu için de "0" kullanılır (Genceli , 1989).

Logit model

Günümüzde nitel değişkenlerden oluşan kategorik verileri analiz etmek için çeşitli teknikler kullanılmaktadır. Lojistik regresyon analiz tekniği iki veya daha fazla kategorik değişkenin koşullu ilişkisini analiz etmek için geliştirilmiştir (**http://www2.chass**...). Bununla birlikte, lojistik regresyon analiz tekniği 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).

Lojistik regresyon modeli genelleştirilmiş doğrusal modelin belirli koşullar altında oluşturulmuş özel durumudur. Bu durumda yapılacak olan çalışmada, eğer bağımsız değişkenlerin bazısı sürekli veya uygun (ilgili) sınıflar içine ayrıştırılamazsa, lojistik regresyon analiz tekniği kullanılmalıdır. Aynı zamanda değişkenlerin bazısı bağımlı olarak ele alınırsa, o zaman logistik regresyon modellerin kullanılmasının, daha uygun olduğu bildirilmiştir.

Böyle bir durumda 0'la 1 arasında kalma koşulunu sağlayabilmek için lojistik regresyon analiz modelin uygulanması önerilmektedir (Gujarati, 1979). Lojistik regresyon analiz tekniği, 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).

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

$X = (x_1, x_2, \ldots\ldots\ldots, x_p)$ şeklinde p tane bağımsız değişken için çoklu lojistik model,

$$L_i = \ln(\frac{P(Y)}{1-P(Y)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots\dots\dots + \beta_p X_p \quad (5)$$

şeklinde yazılabilir. Burada; β_0 ve $\beta_1, \dots\dots\dots \beta_p$ regresyon, katsayılarıdır.Yalnızca bir tane (p=1) bağımsız değişkenin bulunduğu durum ise basit lojistik regresyon olarak adlandırılır.

$$P_i = E(Y = 1/X_i) = \beta_0 + \beta X_i \quad (6)$$

$$P_i = E(Y_i = 1/X_i) = \frac{1}{1+e^{-(\beta_0+\beta X_i)}} = \frac{1}{1+e^{-Z_i}} \quad (7)$$

Burada:

$Z_i = \beta_0 + \beta X_i$ 'dir.

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

e = 2,71828'dir

(7) nolu model lojistik (birikimli) dağılım fonksiyonu olarak adlandırılır (Öz damar, 1999).

X hangi değerleri alırsa alsın fonksiyondaki eksponansiyel terim daima pozitif olacağı için P_i' nin alt sınırı da 0 olur. Olasılık için gerekli olan $0 \le P_i \le 1$ koşulunu bu fonksiyon sağlamış olur. Lojistik 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 $0 \le P_i \le 1$, ve Z_i ile P_i arasındaki ilişkinin doğrusal olmama şartları yerine gelmiş olacaktır. Fonksiyonun belirlenmesi için β_0 ve β_i regresyon katsayıları en küçük kareler yöntemi ile doğrudan tahmin edilemez. Önce ilişki üzerinde bazı işlemler yaparak doğrusal bir ilişki elde etmeye çalışılmıştır. Bu amaçla (6) nolu modeli β_0 ve β' ye göre çözerek, modelin tahmini için,

P_i: Mezun olma olasılığını ifade ettiğine göre mezun olmama olasılığı ise 1- P_i şöyle gösterilebilir:

$$1 - P_i = P_i = \frac{1}{1 + e^{-Z_i}}$$

eşitliğinin her iki yanını $(1 + e^{-Zi})$ ile çarpılarak

$$(1 + e^{-Z_i})P_i = 1 \qquad (8)$$

elde edilir.

(8) nolu eşitliği P_i ile bölüp 1 çıkartarak,

$$e^{-Z_i} = \frac{1}{P_i} - 1 = \frac{1 - P_i}{P_i} \qquad (9)$$

eşitliği elde edilir.

$$e^{-Zi} = \frac{1}{e^{Z_i}}$$

kullanılarak,

$$e^{Zi} = \frac{P_i}{1 - P_i} \qquad (10)$$

eşitliği ile de gösterilebilir. Bu eşitlik bize herhangi bir olayın olma olasılığının, olmama olasılığına olan oranını verir aynı zamanda bu oran bahis oranıdır. (10) nolu modelin e tabanına göre doğal logaritması alınarak,

$$\ln(\frac{P_i}{1-P_i}) = \ln e^{Z_i} = Zi = \beta_0 + \beta Xi \qquad (11)$$

elde edilir.

Yani, bahis oranının logaritması L_i, yalnız X' e göre değil, (katsayı tahmini bakımından) populasyon katsayılarına göre de doğrusaldır. L_i 'ye logıt denir. (11) nolu modellerin adı olan lojistik regresyon modeli de buradan gelir. Bu parametrelerin tahmininde doğrusal bir ilişki işlemi görebilecek yarı logaritmik bir fonksiyondur.

Lojistik regresyon analizinde katsayı tahminleri maksimum olabilirlik yöntemi ile yapılır. Lojistik regresyonda katsayıların yorumlanması için ''odds'' lar ve ''odds'' oranı''ndan yararlanılır. X'in herhangi bir grubu için y=1 olmasına ilişkin olasılık ''odds'' olarak adlandırılır. ''odds'', lojitin doğal logaritması alınmamış halidir ve örneğin,

X=1 için odds değeri, $\frac{P(x=1)}{1-P(x=1)}$ ve $X = 0$ için odds değeğe $\frac{P(x=0)}{1-P(x=0)}$

şeklinde gösterilir. X=1 için bulunan odd'un X=0 'a ait odds'a oranı,

$\Omega(1,0) = \frac{[p(x=1)/(1-p(x=1)]}{[p(x=0)/(1-p(x=0)]}$ şeklindedir ve odds oranı olarak adlandırılır. Bu oranın logaritması ise log-odds oranı veya log-odds olarak adlandırılır ve lojit fark'a eşittir. Yani $\text{Log}\Omega(1,0) = l_{(1)} - l_{(0)}$ şeklindedir. Basit lojistik regresyonda odds oranı $\Omega = e^{\beta_0}$ ve log-odds yada lojit fark $\log\Omega = \log e^{\beta_0} = \beta_0$ şeklindedir.

P_i=1 ve P_i=0 değerleri logıt Li' deki yerine koyulduğunda $\ln(\frac{1}{0})$ ve $\ln(\frac{0}{1})$ değerleri elde edilir ki bunlar anlamsızdır. En küçük kareler yöntemi ile Li fonksiyonundaki parametrelerin tahmin değerleri bulunamaz, fakat bu parametreler maksimum olabilirlik yöntemi ile tahmin edilebilir.

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

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

Lojistik regresyon analiz tekniği özellikle fen bilimleri, sosyal bilimler, veterinerlik alanlarında ve klinik çalışmalarda yaygın bir şekilde kullanılmaktadır(Şahin,2002).

Bu çalışmanın amacı regresyon yönteminde kategorik değişkenlerden bağımlı değişkenin 1 yada 0 özelliğe sahip olduğu durumların olabileceğini göstere bilmektir. Çoğunlukla kategorik 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 çalışılmış ve lojistik analiz tekniğinin eğitim bilimleri alanındaki araştırmalarda uygulanabilirliği tartışılmıştır.

2. MATERYAL ve YÖNTEM

2.1. Materyal

Bu çalışmada, Mustafa Kemal Üniversitesi (M.K.Ü) 'nin çeşitli Fakülte ve Yüksek Okullarına 1994-1995 eğitim öğretim yılında giren öğrencilere ait dosyalardan, rasgele örnekleme yöntemi ile seçilen dosyalardan Bootstrap örnekleme yöntemi uygulanarak 492 öğrenciye ait dosyalar materyal olarak kullanılmıştır (S-Plus, 2000).

2.2. Yöntem

Bu dosyalardan, öğrencinin üniversite giriş puanı(X_1), mezun olduğu lise türü (X_2) ve mezuniyet(Y_1) durumlarına ait veriler tespit edilerek Tablo 1, Tablo 2 ve Şekil 1- Şekil 4 oluşturulmuştur.

Kategorik verileri analiz etmek için, doğrusal olasılık (en küçük kareler yöntemi) ve lojistik regresyon modeller kullanılmıştır (Genceli 1987 ve 1989, Alpar 1997, Uygu 1990, Özdamar 1999, Gujaraati, 1979 Garson 1999, İşyar 1999). Bu modeller sayesinde ilgili gruplar içinde yer alan değişkenler arasındaki ilişkiler S-PLUS 2000 for WINDOWS istatistik paket programı ile değerlendirilmiştir.

$Y_i = \beta_0 + \beta_i X_i + e_i$ kullanılarak öğrencinin üniversiteye giriş puanının mezuniyet durumuna etkisi en küçük kareler yöntemi ile hesaplanarak elde edilen regresyon katsayı değerleri(β :), Tablo- 3 de verilmiştir. Lise türü ile üniversiteye giriş puanının birlikte mezuniyet durumuna etkileri aynı yöntemle incelenmiş ve regresyon katsayı değerleri Tablo 4 de verilmiştir.

Bu programla, bağımlı (kategorik) değişkenin iki değerli olduğu durumlarda, regresyon analizi yapılmaktadır. Burada, sadece iki durumlu bağımlı değişkenli modeller ele alınmıştır. Lojistik regresyon analiz tekniğine göre yapılan değerlendirmeler Tablo 5- Tablo 8 verilmiştir. Lojistik regresyon modeli ile elde edilen sonuçlar, doğrusal olasılık modeli ile elde edilen sonuçlarla karşılaştırılarak lojistik regresyon analiz tekniğinin eğitim bilimleri araştırmalarında uyulana bilirliği gösterilmeye çalışılmıştır.

3. BULGULAR

Örnek veriler olarak M.K.Ü. nin çeşitli Fakülte ve Yüksek Okullarından alınan 492 öğrenciye ait dosya verileri kullanılmıştır (Tablo 1). Öğrencilerin üniversiteye giriş puanlarının (ÖSYMG.P) aritmetik ortalaması ve standart sapması ($\bar{x} \pm ss$) 406,742 ± 73,141 ve dağılım aralığı 105,037- 506,000 puan olmuştur. Öğrencilerin % 78,5' i (386 öğrenci) dört yıl sonunda mezun olmuş, sadece % 21,5' i mezun olamamıştır. Öğrencilerin lise türü, ÖSYMG. Puanları ve başarı durumlarının dağılımını gösteren histogram ve poligon grafikler Şekil 1- Şekil 3 te gösterilmiştir. Öğrencinin lise türü ve mezuniyet durumuna göre dağılımı tablo-2 de verilmiştir. Bu tablodan da kolayca görülebileceği gibi düz liseden mezun olan öğrencilerin % 92,0 'ı üniversiteden normal zamanda mezun olmuştur. Meslek lisesi mezunlarında bu oran % 85,5 iken özel liselerden mezun olanlarda % 73,3 tür. Diğer liselerden mezun olanların sadece % 50,0 'ı dört yıl sonunda mezun olmuştur (Tablo 2). Belli bir X'e karşılık gelen Y (ya 0 ya da 1). değerlerinin, ya X ekseni ya da 1'in hizasındaki doğru üzerinde yer aldığı hesaplanmıştır (Şekil 4).

3.1. En küçük kareler yöntemi ile hesaplanan parametre değerleri.

Öğrencinin üniversiteye giriş puanının mezuniyet durumuna etkisi en küçük kareler yöntemi(EKKY) ile hesaplanarak elde edilen regresyon katsayı değerleri(β_i), tablo 3 de verilmiştir.

Tablo.3 de görüldüğü gibi, üniversiteye giriş puanının mezun olma olasılığını önemli düzeyde etkilememiştir. Bu bulgulara göre doğrusal olasılık modeli

$Y_i = \beta_0 + \beta_1 X_i + e_i$ $\hat{Y} = -1.828016 + 0.006423$ (ÖSYMG .P) olur.

öğrencinin giriş puanı (X_1) 421.000 puan olduğunda, mezun olma olasılığı

$\hat{Y} = -1.828016 + 0.006423(421.000) = 0.8760$ %87.60 olarak hesaplanmıştır.;

X_1=435.000 puan olduğunda öğrencinin mezun olma olasılığı,

$\hat{Y} = -1.828016 + 0.006423(435.000) = 0.9659$, % 96.59; olmaktadır. Böylece farklı giriş puanları için farklı olasılıklar bulunmuştur.

$0 \leq E(Y_i / X_i) \leq 1$ koşulunun yerine gelmemesi durumunda, örneğin X_i = 506.000 puan olan bir öğrencinin mezun olma olasılığı % 142,20 olacaktır, $\hat{Y} = -1.828016 + 0.006423(506.000) = 1.4220$ den. Bu rakam 1' den büyük olduğu için $0 \leq E(Y_i / X_i) \leq 1$ koşuluna aykırıdır. Öğrencinin mezun olduğu lise türü ile üniversiteye giriş puanının birlikte mezuniyet durumuna etkilerini belirleyen parametre değerleri tablo-4 de görüldüğü gibidir. İlişkiye öğrencinin mezun olduğu lise türünü de eklediğimizde $Y_i = \beta_0 + \beta_i X_i + \beta_2 X_2 + e_i$ modelini elde etmiş oluruz. En küçük kareler yöntemi ile elde edilen bulgular bu denklemde yerine konulursa;

$\hat{Y} = 0.489 + 0.005805\ (\text{ÖSYMG .P}) + 0.04906\ (\text{LISE TÜRÜ})$ olur.

Örneğin giriş puanı X_1= 421.000 puan olan bir öğrenci düz lise mezunu (X_i=1) ise, bu öğrencinin mezun olma olasılığı,

$\hat{Y} = 0.489 + 0.00580(421.000) + 0.04906(1)$

= 2.9798 dir. Öğrenci meslek lisesi mezunu (X_i=2) ise mezun olma olasılığı

$\hat{Y} = 0.489 + 0.00580(421.000) + 0.04906(2)$

=3.0289 olacaktır.

Bu hesaplamalardan da görüldüğü gibi 0<E(Y=1/X)<1 koşulu sağlanamamıştır. Ancak, meslek lisesini bitirmiş bir öğrencinin düz liseyi bitirmiş bir öğrenciye göre mezun olma olasılığını, öğrencinin giriş puanına ek olarak, (3,0289-2,9798=0.0491) % 4.91 oranında artırdığı hesaplanmıştır.

3.2 . Lojistik regresyon yöntemi ile hesaplanan parametre değerleri.

0<E(Y=1/X)<1 koşulunu sağlayabilmek için lojistik regresyon modeli uygulanmış ve hesaplanan parametre değerleri tablo 5 de verilmiştir. Bu tablodan da görüldüğü gibi üniversiteye giriş puanı ile mezun olma arasında önemli bir istatistiksel ilişki bulunmaktadır(t=2.634, p<0.0087).

$L_i = \ln(\frac{P(Y)}{1-P(Y)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots\dots\dots + \beta_p X_p$ modelinde tablo 4 deki bulgular yerine konulduğunda

Lojit Li için $\dot{L}_i = 0.522504 + 0.00653 \ X1$ modeli bulunmuştur. Bu değerler P_i de yerine konulduğunda:

$$\hat{P}_i = \frac{1}{1+e^{-(0.5202504 + 0.0653 \ (ÖSYMG \ .P)}}$$

elde edilir. Bu formülle herhangi bir giriş puanı için mezun olma olasılığı hesaplanarak tablo 6 da verilmiştir. Bu tablodan da görüldüğü gibi giriş puanı arttıkça mezun olma olasılığı artmaktadır.

Örneğin, X=381.000 puan için

$$\hat{Pi} = \frac{1}{1+e^{-0.522504 - 0.00653 \ (381.000)}} = \frac{1}{1+0.0498} = \frac{1}{1.0498} = 0.9525$$

bulunur. Bu öğrenci % 95,25 olasılıkla mezun olacaktır. Giriş puanı X=391.313 olduğunda mezun olma olasılığı % 95,60 yükselmektedir (Tablo 6).

X=391.000 puan ile X=381.000 puan arasındaki fark $(\Delta x = 10)$ için

$$\Delta \hat{p} = 0.9560 - 0.9525 = 0.004$$

X=423.000 ile X=413.000 arasındaki fark

$$\Delta \dot{p} = 0.963 - 0.961 = 0.002 \ ;$$

$X=488.000$ ile $X=478.000$

$\Delta \hat{p} = 0.976 - 0.974 = 0.002$ değerleri bulunmuştur. Farklı giriş (ösym) puanları, puandaki belli bir miktar artış (örneğimizde $\Delta x = 10$) mezun olma olasılığını değişik miktarlarda artırmaktadır. Bu olasılık düşük puanlarda biraz daha çok, orta ve yüksek puanlarda yaklaşık eşit miktarda olmuştur. Giriş puanı ve lise türünün mezuniyet durumuna etkisini belirlemek için Lojistik regresyon modeline X_2 (Lise türü $X_1=1$ Düz lise, $X_2=2$ meslek lisesi mezunu) değişkenini de koyduğumuzda S-Plus bilgisayar paket program bize, maksimum olabilirlik yöntemini uygulayarak, aşağıdaki lojistik Li Regresyonunu vermiştir.

Logit Li için $\hat{L_i} = -0.9554 + 0.0028X_1 + 1.0437X_2$ modeli bulunmuştur. Bu değerler P_i de yerine konulduğunda:

$$\hat{P_i} = \frac{1}{1+e^{-(-0.9554+0.0028(\text{ÖSYMG.P})+1.0437(\text{LİIS TÜRÜ})}}$$

$X_1=381.000$ puan

$X_2=1$ için

$$\hat{P} = \frac{1}{1+e^{0.95454-0.0028(381.000)-1.0437(1)}}$$

$$\hat{P} = \frac{1}{1+e^{-1.1559}} = \frac{1}{1+0.3147} = \frac{1}{1.3147} = 0.7606$$

$X_1=381.000$ puan

$X_2=2$ için

$$\hat{P} = \frac{1}{1+e^{0.95454-0.0028(381.000)-1.0437(2)}}$$

$$\hat{P} = \frac{1}{1+e^{-2.1996}} = \frac{1}{1+0.11084} = \frac{1}{1.1108} = 0.9002$$

Bu model için elde edilen bulgulara göre de, lise türü, öğrencinin mezun olma olasılığını etkilemektedir (Tablo 7). Tabloya göre çok düşük puan alan bir öğrencinin (X_i=358.920) lise türünün özel lise olması durumunda (X_i=3) mezun olma olasılığı pek etkilenmemiştir. Aynı şekilde çok yüksek puan alan bir öğrencinin (Xi=430.000 puan) mezun olma olasılığı zaten oldukça büyük olacağı için giriş puanının beli bir miktarda artması X_i=442.000 puan düzeyine çıkması gibi mezun olma olasılığını çok daha fazla artırmaz. Buna karşılık orta puan düzeyindeki bir (Xi=393.000 puan) puanın belli bir miktar artması(Xi=407.353 puan düzeyine çıkması) mezun olma olasılığını önemli düzeyde arttırmıştır (Altı öğrenciye ait ÖSYMG.P ve lise türü dikkate alınarak hesaplanan olasılıklar örnek olarak tablo 8 de verilmiştir). Bu nedenle ve ayrıca $0<E(Y=1/X)<1$ koşulunu sağladığı için bu tip çalışmalarda, lojistik modelin uygulanmasının yararlı olacağı düşünülmektedir.

SONUÇ VE TARTIŞMA

Bu araştırmada, rasgele örnekleme yöntemi ile seçilen 492 öğrenciye ait dosya verileri kullanılmıştır. Öğrencinin üniversiteye giriş puanı ve mezun olduğu lise türü ile üniversiteden mezun olma olasılığı arasındaki ilişkiler; hem lojistik regresyon tekniği hem de alternatif regresyon teknikleri (Doğrusal olasılık modeli, Maksimum olabilirlilik yöntemi) kullanılarak saptanmıştır. Tahmin edilen olasılığın; en küçük kareler yöntemi ile elde edilen bulgulardan $0 \le E(Y_i / X_i) \le 1$ sınırların dışına çıktığı böylece doğrusal olasılık modelin birçok varsayımları göz ardı edildiği için yapılan çalışmaların yanlış sonuçlar verdiği ve araştırmacıları yanılttığı bu çalışmayla da ortaya konulmuştur. Lojistik regresyon analizi ile elde edilen bulguların $0 \le E(Y_i / X_i) \le 1$ bu sınırların içinde kalması bu modelin varsayımlarının yerine gelmesi ile gerçekleşir. Dolayısıyla, lojistik regresyon analiz yönteminin bu tür çalışmalarda kullanılmasının uygun olacağı kanısına varılmıştır.

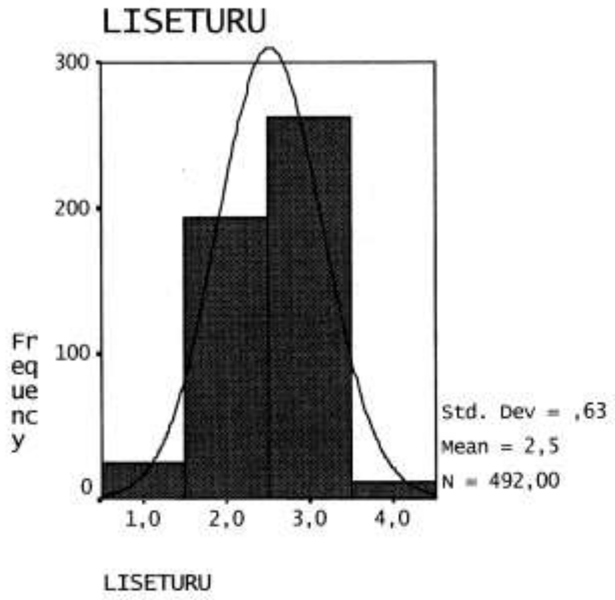
KAYNAKÇA

1. Alpar,R.,(1997). Uygulamalı Çok değişkenli İstatistiksel Yöntemlere Giriş –1. Ankara. Hacettepe Üniversitesi Tıp Fak. Kültür Ofset,
2. Alpar,R.,(1999). Çoklu doğrusal regresyon çözümlemesi. Ankara. Hacettepe Üniversitesi Tıp Dergisi. 30(1):92-94
3. Genceli , M., (1987). Nitel Bağımlı Değişken için Doğrusal olasılık modeli. İstanbul Üniversitesi İktisat Fakültesi Mecmuası, 45: 213-220.
4. Genceli, M., (1989). Ekonometride İstatistik İlkeler.istanbul Filiz kitapevi.
5. Gujariti, D., (1979). Basıc Ekonometrics, McGraw-Hill Book Company
6. http://www2.chass....
7. İnal, C., ve ark. (1983). İstatistik Terimleri Sözlüğü, T.D.K Yayınları, Ankara.Sevinç Basımevi
8. İşyar, Y., (1999). Ekonometrik Modeller. VİPAŞ A.Ş., (Uludağ Üniversitesi Güçlendirme Vakfı), Bursa 695 Sayfa..
9. Kılıçbay, A., (1986). Ekonometrinin Temelleri. İstanbul Üniversitesi Film Merkezi ve Matbaası.Özdamar, K., (1999). Paket Programlar ile İstatistiksel Veri Analizi. Eskişehir Kaan kitapevi.
10. Püskülcü,H.ve İkiz, F., (1986). İstatistiğe Giriş. Bornova,İzmir Ege Üniversitesi Basımevi.
11. S-Plus – 000 for Windows, 1999. http:/www.mathsoft.com/s-plus, Data Aanalysis Products Division, Mathsoft., Inc.Seattle, Washington.USA

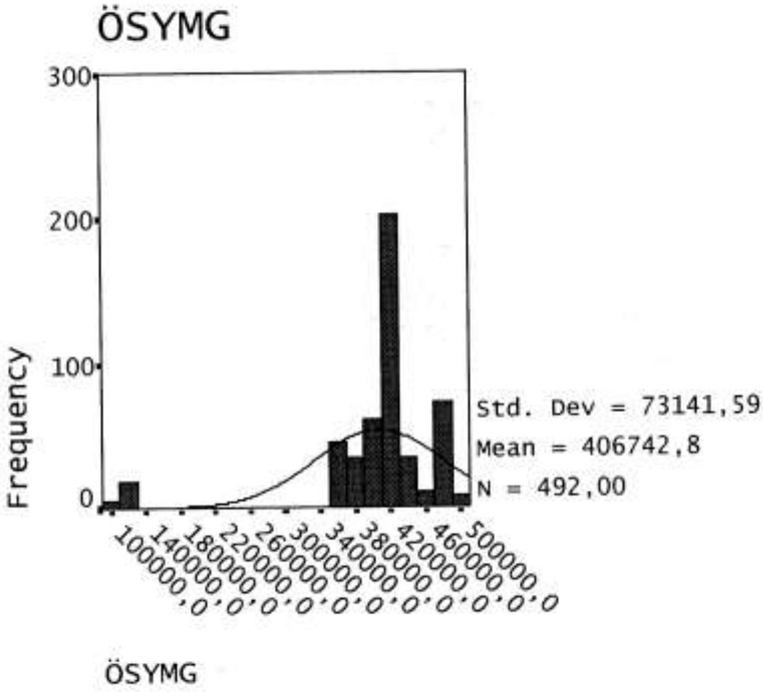
12. Şahin, M.,(2002). Lojistik Regresyon Analizinin Hayvancılık Alanında Uygulanması. III Ulusal Zootekni Bilim Kongresi 14-16 Ekim, Ankara. Ankara Üniversitesi.

14. Tarı,R.,(1999). Ekonometri İstanbul Alfa basım yayım dağıtım.

Şekil.1 Lise türüne göre öğrenci dağılımı

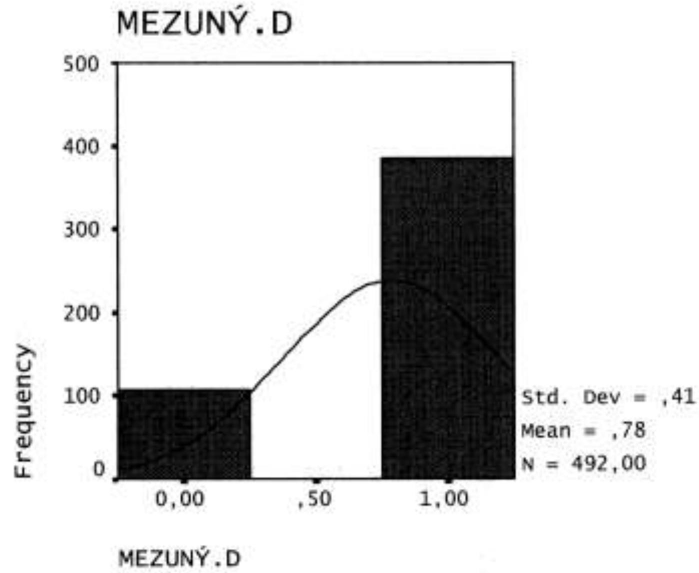


:ekil-2 ÖSYMG.puanına göre öğrencinin dağılımı

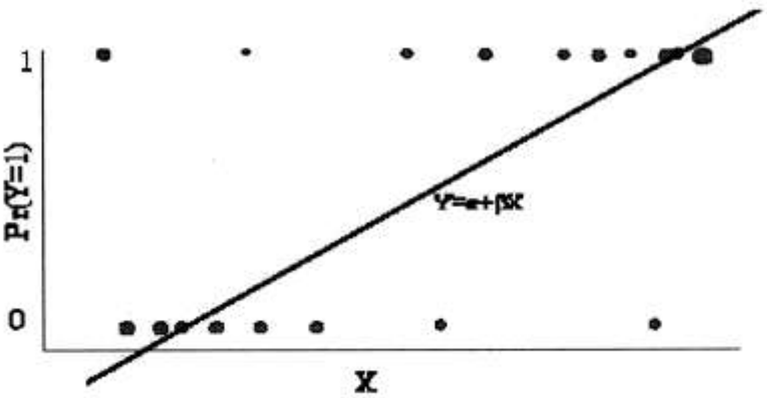


Şekil-3 Mezuniyet durumuna göre öğrencinin

dağılımı



Şekil 4: Bağımsız değişkenin Doğrusal bir fonksiyonu olarak Y'nin beklenen değeri



Tablo 1. Bir Grup Öğrencinin Mezuniyet Durumu, Giriş Puanı ve Lise Türü

Sıra no	Mezuniyet durumu* (Y)	Giriş puanı (X_1)	Lise türü** (X_2)
1	1	391.712	1
2	1	393.569	1
3	0	394.060	1
4	1	393.736	1
5	1	389.174	1
6	1	391.651	1
7	1	398.511	3
.	.	.	
.			
S.	.	.	
.			
490	0	358.710	1
491	1	394.470	1
492	1	381.190	1

(*): Y= 1 ise mezun
Y= 0 ise mezun değil

(**): (X_2 = 1 ise düz lise
=2 ise meslek lisesi
= 3 ise özel lise

Tablo 2. Öğrencinin lise türü ve mezuniyet durumuna göre dağılımı

Mezuniyet durumu		Lise türü				
		Düz	Meslek	Özel	Diğer	Toplam
Mezun değil	Sayı	2	28	70	6	106
	Sütun % si	8,0	14,5	26,7	50,0	21,5
	Toplamın % si	0,4	5,7	14,2	1,2	21,5
Mezun	Sayı	23	165	192	6	386
	Sütun % si	92,0	85,5	73,3	50,0	78,5
	Toplamın % si	4,7	33,5	39,0	1,2	78,5
Toplam	Sütun toplamı(n)	25	193	262	12	492
	Sütun% si toplamı	100	100	100	100	100
	Toplamın % si	5,1	39,2	53,3	2,4	100

Tablo 3. Giriş puanın mezuniyet durumuna etkisi(KKY, n=492)

Değişkenler	$\hat{\beta}$	**Standart Hata**	**t**	Sig
Sabit	-1.828	3.050	-0.59	0.000
ÖSYM.G(X_1)	0.0064	0.007	0.86	0.019

R^2=0. 00149

Tablo 4 . Giriş puanı ve lise türünün mezuniyet durumuna etkisi(EKKY, n=492)

Değişkenler	$\hat{\beta}_{(j)}$	**Standart Hata**	**t**	**Sig**
Sabit	0,489	0,115	4,264	0,000
ÖSYM.G(X_1)	0.00058	0,000	2,295	0,022
Lise Türü(X_2)	0.0490	0,044	1,124	0,262

R^2=0. 117

Tablo 5. Giriş puanın mezuniyet durumuna etkisi(Maksimum olabilirlik Yöntemi,n=492)

Değişkenler	$\hat{\beta}_{(j)}$	**Standart Hata**	**t**	**Sig.**
Sabit	0.523	0.103	5.09	0.00
ÖSYM.G(X_1).	0.006	0.002	2.63	0.008

Tablo 6. Lojistik regresyon modeli ile hesaplanan olasılıklar

Değişkenler	puanlar	P
X_1	381.000	0.9525
X_2	391.313	0.9560
X_3	413.000	0.9616
X_4	423.000	0.9639
X_5	478.000	0.9745
X_6	488.000	0.9761

Tablo 7 . Giriş puanı ve Lise türünün mezuniyet durumuna etkisi (Maksimum Olabilirlik Yöntemi, n=492)

Değişkenler	$\hat{\beta}_{(j)}$	**Standart Hata**	**t**	**Sig**
ÖSYMGP (X_1)	-0.9554	-	0.027	0.005
Lise türü (X_2)	0.0028	2.107031	1.000	0.3179
Sabit	1.0437	2.539944	-0.691	0.49030

ablo 8. Lojistik regresyon modeli ile hesaplanan olasılıklar

Değişkenler	puanlar	Lise türü	P
X_1	381.000	1	0.7606
X_1	381.000	2	0.9002
X_2	391.313	1	0.7605
X_2	391.313	2	0.9002
X_3	478.000	1	0.8115
X_4	478.000	2	0.9214