MULTİNOMİNAL LOJİSTİK REGRESYON ANALİZİNİN İSTİHDAMDAKİ İŞGÜCÜNE UYGULANMASI

Yrd. Doç. Dr. Nuran BAYRAM*

Özet

Bu çalışmada amaç, ikili ve multinominal lojistik regresyon analizinin önce teorik yapılarını ortaya koymak, daha sonra ücretsiz çalışan işgücü ile gelir karşılığı çalışan işgücünü çeşitli bağımsız değişkenlerle karşılaştırılarak mevcut durumu açıklanmaya çalışmaktır. Bu amaçla, Devlet İstatistik Enstitüsünden elde edilen Ekim 1999 Hanehalkı İşgücü Anketi veri grubu kullanılarak uygulama yapılmıştır. Yapılan uygulamada, istihdamda olanlar ve yeni çalışabilir nüfus tanımına göre (DİE) 15-64 yaş arası toplam 20.123 veri kullanılmıştır.

Anahtar Kelimeler: *İkili Lojistik Regresyon Analizi, Multinominal Lojistik Regresyon Analizi, Bahis Oranı, En Çok Olabilirlik, İşgücü*

1. GİRİŞ

Yapılacak analizlerde verilerin dayandığı ölçme düzeyinin bilinmesi yararlanılacak analiz tekniğinin belirlenmesi açısından büyük bir öneme sahiptir. Nitekim kategorik değişken adı verilen sınıflayıcı veya sıralayıcı ölçme düzeyinde elde edilmiş değişkenler söz konusu olduğunda başvurulabilecek analiz tekniklerden biri de lojistik regresyon analizidir.

Lojistik regresyon analizinin temel amacı, diğerlerinde olduğu gibi, bir veya birden fazla bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi modellemektir (Hosmer&Lemeshow, 2000). Hemen ekleyelim ki, klasik doğrusal regresyon analizlerinde bağımlı değişken sürekli iken, lojistik regresyon analizinde bağımlı değişken kategoriktir. Bunun yanı sıra, bağımsız

* Yrd. Doç. Dr., Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü

değişkenlerin tümü kategorik değişken, sürekli değişken veya kategorik ve sürekli değişkenlerin bir karması olabilir.

Analizde tek bir bağımsız değişkene yer verildiğinde lojistik regresyon, birden fazla bağımsız değişkene yer verildiğinde ise çoklu lojistik regresyon söz konusudur. Öte yandan, bağımlı değişken sadece iki kategoriye sahip olduğunda ikili lojistik regresyon, buna karşılık sınıflayıcı ölçme düzeyinde ölçülmüş ikiden fazla kategoriye sahip olduğunda multinominal lojistik regresyondan söz edilir.

Lojistik regresyon analizi çeşitli varsayım (normallik, ortak kovaryansa sahip olma gibi) ihlalleri durumunda diskriminant analizi ve çapraz tablolara bir alternatif oluşturmaktadır (Tatlıdil, 1996). Bağımsız değişkenler kategorik ve sürekli değişkenlerin karması olduğunda, çok değişkenli normallik varsayımı geçerli olmaz. Bu gibi durumlarda bağımsız değişkenlerin dağılımı hakkında herhangi bir varsayımda bulunmayan lojistik regresyon analizi kullanılabilir (Sharma, 1996). Analizin kullanım rahatlığının yanı sıra çözümlemeden elde edilen modelin matematiksel olarak çok esnek ve kolay yorumlanabilir olması analize olan ilgiyi arttırmaktadır.

2. İKİLİ LOJİSTİK REGRESYON ANALİZİ

Herhangi bir regresyonda bağımsız değişkenin değeri verildiğinde, bağımlı değişkenin ortalama değeri $E(Y|x)$ olarak gösterilir ve koşullu ortalama olarak adlandırılır. Burada Y, bağımlı değişken değeri, x de bağımsız değişken değeri olmak üzere $E(Y|x)$, x verildiğinde Y'nin koşullu ortalamasını (beklenen değerini) ifade eder. Doğrusal regresyonda bu ortalamanın x'e göre doğrusal bir eşitlik ile ifade edilebileceği varsayılır (Hosmer&Lemeshov, 2000).

$$E(Y|x) = \beta_0 + \beta_1 x \tag{1}$$

İki kategoriye sahip veri için koşullu ortalama en az sıfır ve en fazla bir olmaktadır: $0 \leq E(Y|x) \leq 1$. Bağımlı değişken iki kategoriye sahip olduğunda $E(Y|x)$ için bir model ortaya koymada pek çok kümülatif dağılımdan yararlanılabilir. Matematiksel açıdan son derece esnek ve kolay olması, ayrıca bilimsel olarak anlamlı yorumlara götürmesi nedeniyle lojistik dağılım tercih edilir (Hosmer&Lemeshov, 2000).

Lojistik dağılım kullanıldığında, x verildiğinde Y'nin koşullu ortalaması $\pi(x)=E(Y|x)$ şeklinde tanımlanır. Ayrıca $\pi(x)=P(Y=1|x)$ ve $1-\pi(x)=P(Y=0|x)$ olasılıkları da yazılabilir. Bu durumda lojistik regresyon modelinin özel formu aşağıdaki gibidir:

$$\pi(x)=E(Y|x)=\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \tag{2}$$

Bağımsız değişken 0 ve 1 şeklinde kodlandığında aşağıdaki tablo oluşturulabilir.

Tablo 1: Bağımsız Değişken İkili Kodlandığında Lojistik Regresyon Modelinin Değerleri

Bağımlı Değişken (Y)	Bağımsız Değişken (X)	
	x=1	x=0
y=1	$\pi(1)=\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0)=\frac{e^{\beta_0}}{1+e^{\beta_0}}$
y=0	$1-\pi(1)=\frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0)=\frac{1}{1+e^{\beta_0}}$
Toplam	1,0	1,0

Tablo 1'de bağımlı değişkenin bahsi (odds) x=1 olduğunda $\frac{\pi(1)}{1-\pi(1)}$, x=0 olduğunda ise $\frac{\pi(0)}{1-\pi(0)}$ şeklinde tanımlanır. Bu bahis değerlerinin oranına ise bahis oranı (odds ratio) adı verilir ve

$OR=\frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$ şeklinde gösterilir (Hosmer&Lemeshov, 2000).

(2) nolu eşitliğe lojit dönüşüm uygulanırsa;

$$g(x)=\ln\left[\frac{\pi(x)}{1-\pi(x)}\right]=\ln\left[\frac{P(Y=1|x)}{P(Y=0|x)}\right]=\beta_0+\beta_1 x \tag{3}$$

eşitliği elde edilir (Anderson, 1997). Log-bahis (bahislerin logaritması) olarak adlandırılan, lojit dönüşüm ya da kısaca lojit, doğrusal regresyon modellerinin pek çok özelliğine sahiptir. Örneğin, lojistik regresyonda bağımlı değişken ile bağımsız değişken arasında doğrusal bir ilişki yok iken, bu değişkenlerin lojit değerleri arasında doğrusal bir ilişki vardır (Garson, 1999).

Eşitlik (3)'de yer alan β_1 katsayısı; x bir birim değiştiğinde, log-bahis oranının nasıl değiştiğini, β_0 katsayısı ise, x sıfır olduğunda log-bahis oranının nasıl değiştiğini ifade eder (Everit, 1992; Gujarati, Çev: 1999).

Bu bilgiler ışığında olasılık, bahis (odds) ve log-bahis (lojit) aralıkları aşağıdaki tabloda özetlenmiştir.

Tablo 2: Olasılık, Bahis ve Log-Bahis (Lojit) Aralıkları

Tanım	Gösterim	Minimum	Maksimum
Olasılık	$\pi(x)$	0	1
Bahis (odds)	$\pi(x)/[1-\pi(x)]$	0	$+\infty$
Log-Bahis (lojit)	$g(x)$	$-\infty$	$+\infty$

3. LOJİSTİK REGRESYON ANALİZİNDE PARAMETRE TAHMİNLERİ VE İSTATİSTİKSEL ANLAMLILIĞIN SINANMASI

Lojistik regresyon analizinde parametreler en çok olabilirlik yöntemi kullanılarak tahmin edilir. (x_i, y_i) gözlem çiftinin en çok olabilirlik fonksiyonuna katkısını ifade etmek için aşağıdaki fonksiyon kullanılır (Hosmer&Lemeshov, 2000).

$$\xi(x_i) = \pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i} \tag{4}$$

Gözlem sonuçlarının birbirinden bağımsız olduğu varsayımından hareketle, β'nın olabilirlik fonksiyonu:

$$l(\beta) = \prod_{i=1}^{n} \xi(x_i) \tag{5}$$

şeklinde ifade edilir. En çok olabilirlik tahminleri $l(\beta)$'yi maksimum yapan β hesaplanarak bulunabilir. Fakat olabilirlik fonksiyonunu maksimum yapan β değerini bulmadan önce kolay işlem yapabilmek için 5 numaralı eşitliğin logaritması alınır.

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1 - y_i)\ln[1 - \pi(x_i)]\} \tag{6}$$

$L(\beta)$'yi maksimum yapan değeri bulmak için β_0 ve β_1'e göre $L(\beta)$'nin türevi alınıp eşitlikler sıfıra eşitlenir. Bu durumda sabit (β_0) için $\sum_{i=1}^{n}[y_i - \pi(x_i)] = 0$ ve eğim (β_1) için $\sum_{i=1}^{n} x_i[y_i - \pi(x_i)] = 0$ eşitliklerine ulaşılır (Rush Sloan, http://www.trinity.edu...).

Olabilirlik Oran Testleri: Lojistik regresyon analizinde regresyon katsayılarının anlamlılığını test etmek için olabilirlik oran istatistiğinden yararlanılır. Bu istatistik;

D=-2ln[Tahmin edilen modelin olabilirliği / Doymuş modelin olabilirliği] (7)

şeklindedir. Burada doymuş model; tüm parametreleri ve tüm etkileşim etkilerini içeren model olarak tanımlanır.

Lojistik regresyon modelinde yer alan bir bağımsız değişkenin anlamlılığını ölçmek için, değişkenin modelde olması durumundaki ve modelden çıkarılması durumundaki *D* değerine bakılır ve aşağıdaki istatistik elde edilir.

$$G = D_{(\text{değişkeni içermeyen model})} - D_{(\text{değişkeni içeren model})} \tag{8}$$

(8) nolu eşitlik düzenlendiğinde;

G=-2ln[Değişken içermeyen modelin olabilirliği / Değişken içeren modelin olabilirliği] (9)

şeklinde elde edilir (Hosmer&Lemeshov, 2000). *J* bağımlı değişkenin kategori sayısını ve *I* da tahmin edilen parametre sayısını göstermek üzere bu istatistik (*J*-1)(*I*-1) serbestlik derecesi ile bir χ^2 dağılımına sahiptir (Agresti, 1990). Ayrıca katsayıların anlamlılığı için kurulan hipotez çifti aşağıdaki gibidir (Sharma, 1996).

$$H_0 : \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad H_1 : \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Wald İstatistiği: Lojistik regresyon modelinde yer alan her bir bağımsız değişkenin istatistiksel anlamlılığının belirlenmesinde "t" değerlerinin karelerinden oluşan Wald istatistiği kullanılır. 1 serbestlik derecesi ile asimptotik χ^2 dağılımına sahip olan Wald istatistiği aşağıdaki gibi hesaplanır (Sharma, 1996).

$$W_j = \left[\hat{\beta}_j / \hat{SE}(\hat{\beta}_j) \right]^2 \tag{10}$$

Modelin Genel Anlamlılığının (Uyum İyiliğinin) Sınanması: Lojistik regresyon analizinde modelin genel anlamlılığının sınanması için Pearson χ^2 ve Sapma istatistiğinden yararlanılır.

Pearson

$$\chi^2 : r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j\hat{\pi}_j)}{\sqrt{m_j\hat{\pi}_j(1-\hat{\pi}_j)}} \Rightarrow \chi^2 = \sum_{j=1}^{J} r(y_j, \hat{\pi}_j)^2 \tag{11}$$

Sapma:

$$d(y_j, \hat{\pi}_j) = \pm\left\{2\left[y_j \ln\left(\frac{y_j}{m_j\hat{\pi}_j}\right) + (m_j - y_j)\ln\left(\frac{(m_j - y_j)}{m_j(1-\hat{\pi}_j)}\right)\right]\right\}^{1/2} \tag{12}$$

formülde yer alan "+" veya "-" işareti $(y_j - m_j\hat{\pi}_j)$'nin işareti ile aynıdır. Burada m_j; j kategorisine ait toplam değeri vermektedir. Buradan hareketle;

$$D = \sum_{j=1}^{J} d(y_j, \hat{\pi}_j)^2 \tag{13}$$

şeklinde elde edilir (Hosmer&Lemeshov, 2000). Ayrıca uyum iyiliğini ölçen her iki istatistik de (doymuş modelin parametre sayısı – tahmin edilen modelin parametre sayısı) şeklinde hesaplanan serbestlik derecesi ile χ^2 dağılımına sahiptir.

Pseudo R^2: Doğrusal regresyonda bulunan R^2 ile aynı anlama gelen ve 0 ile 1 arasında yer alan bu istatistik sadece sabit içeren modelin olabilirliğinden bağımsız değişkenlerin tümünün yer aldığı modelin olabilirliğinden farkın alınıp sadece sabit içeren modelin olabilirliğine bölünmesiyle elde edilir (Hair&Anderson&Tatham, 1998).

$$Pseudo\ R^2 = \frac{-2LL_{(\text{sabit terimli model})} - (-2LL_{(\text{doymuş model})})}{-2LL_{(\text{sabit terimli model})}} \tag{14}$$

Sınıflama Tabloları: Daha önce de belirtildiği gibi Lojistik regresyon analizi, gözlem sonuçlarının sınıflandırılmasında veya gruplandırılmasında kullanılmaktadır. Gözlem sonuçlarının ele alınan gruplardan birine atanmasında $\hat{\pi}_i = \left[1 \big/ \left(1 + e^{-(x_i'\beta)}\right) \right]$ olasılıkları bulunur. Bunun yanı sıra, eğer gözlem sonuçlarından elde edilen $\hat{\pi}_i$ değeri 0,50'den küçük ise Y=0 grubuna, $\hat{\pi}_i$ değeri 0,50'den büyük ise diğer gruba atanır. Elde edilen sınıflandırma tablosunda köşegen değerler model tarafından doğru sınıflandırılan birimlerin sayısını vermektedir (Özdinç, 1999).

4. MULTİNOMİNAL LOJİSTİK REGRESYON ANALİZİ

Multinominal lojistik regresyon modellerinde bağımlı değişkenin sınıflayıcı ölçme düzeyinde ölçülmüş ve en az üç kategoriye sahip olması gerektiğine daha önce değinmiştik. Bağımlı değişken üç kategoriye sahip olduğunda, diğer bir deyişle, Y'nin kategorileri 0, 1 ve 2 biçiminde kodlandığında, iki lojistik model kurulur. Bu modeller Y=1'e karşı Y=0 ve Y=2'ye karşı Y=0 şeklindedir. Burada Y=0 referans kategorisi iken Y=2'ye karşı Y=1'i karşılaştıran lojistik fonksiyon, yukarıda tanımlı iki karşılaştırmaya ilişkin lojistik fonksiyonların farklarından elde edilmektedir (Tatlıdil, 1996; Hosmer&Lemeshov, 2000). Bu fonksiyonlar,

$$\begin{aligned} g_1(x) &= \ln\left[\frac{P(Y=1|\mathbf{x})}{P(Y=0|\mathbf{x})}\right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p = \mathbf{x}'\boldsymbol{\beta}_1 \\ g_2(x) &= \ln\left[\frac{P(Y=2|\mathbf{x})}{P(Y=0|\mathbf{x})}\right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots + \beta_{2p}x_p = \mathbf{x}'\boldsymbol{\beta}_2 \end{aligned} \tag{15}$$

şeklinde tanımlanmaktadır. Bağımlı değişken kategorilerinin her birinin koşullu olasılıklarının bağımsız değişken vektörünü verdiği aşağıda gösterilmiştir (Powers&Xie, 2000).

$$
\begin{aligned}
P(Y=0|\mathbf{x}) &= \frac{1}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}, \\
P(Y=1|\mathbf{x}) &= \frac{e^{g_1(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}, \\
P(Y=2|\mathbf{x}) &= \frac{e^{g_2(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}
\end{aligned}
\tag{16}
$$

Bu durumda koşullu olasılıkların genel ifadesi;

$$
P(Y=j|\mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{\sum_{k=0}^{2} e^{g_k(\mathbf{x})}}
\tag{17}
$$

şeklinde gösterilir.

5. MULTİNOMİNAL LOJİSTİK REGRESYON ANALİZİNDE PARAMETRE TAHMİNLERİ

Multinominal lojistik regresyonda olabilirlik fonksiyonu oluşturulurken bir gözlemin grup üyeliğini belirlemede iki değer alan üç değişkenden yararlanılır. Bu değişkenler, $\sum_{j=0}^{2} Y_j = 1$ olmak üzere, "Y=0 ise Y_0=1, Y_1=0, Y_2=0", "Y=1 ise Y_0=0, Y_1=1, Y_2=0" ve "Y=2 ise Y_0=0, Y_1=0, Y_2=1" biçiminde kodlanır. Gözlem sonuçları birbirinden bağımsız olarak elde edilmiş n hacimlik örneklem için koşullu olabilirlik fonksiyonu,

$$
l(\beta) = \prod_{i=1}^{n} \left[\pi_0(\mathbf{x}_i)^{y_{0i}} \pi_1(\mathbf{x}_i)^{y_{1i}} \pi_2(\mathbf{x}_i)^{y_{2i}} \right]
\tag{18}
$$

şeklindedir. Bu eşitliğin logaritması alınarak olabilirlik fonksiyonuna ulaşılır.

$$
L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} y_{1i} g_1(\mathbf{x}_i) + y_{2i} g_2(\mathbf{x}_i) - \ln\left(1 + e^{g_1(\mathbf{x}_i)} + e^{g_2(\mathbf{x}_i)}\right)
\tag{19}
$$

$L(\beta)$'nın birinci kısmi türevleri alınıp eşitlikler sıfıra eşitlenerek olabilirlik denklemleri bulunur. $\pi_{ji} = \pi_j(\mathbf{x}_i)$ olmak üzere bu denklemlerin genel gösterimi aşağıdaki gibidir (Hosmer&Lemeshov, 2000).

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{n} x_{ki}\left(y_{ji} - \pi_{ji}\right) \tag{20}$$

Bununla birlikte, olabilirlik fonksiyonunun ikinci kısmi türev matrisi en çok olabilirlik tahmincisinin kovaryans matrisinin tahmincisi ve bilgi matrisini bulmak için kullanılır. Lojistik regresyon analizinde katsayıların anlamlılığı ve uyum iyiliği için kullanılan ölçütlerin tümü multinominal lojistik regresyon analizinde de kullanılır.

6. UYGULAMA

Yukarıda teorik yapısı ortaya konulan lojistik regresyon analizi Devlet İstatistik Enstitüsü'nden alınan Ekim 1999 Hanehalkı İşgücü Anketi veri grubuna uygulanmıştır. Uygulamanın amacı, ücretsiz çalışan ile genel anlamda gelir karşılığı çalışan işgücünü çeşitli bağımsız değişkenler ile karşılaştırarak mevcut durumu ortaya koymaktır. Bu nedenle, istihdamda olanlar ve yeni çalışabilir nüfus tanımına göre (DİE) 15-64 yaş arası veriler analize dahil edilmiştir. İlgili veri grubunda bağımlı değişken olarak "çalıştığı yerde işteki durumu", bağımsız değişkenler olarak cinsiyet, medeni durum, işyerinde kaç kişinin çalıştığı, eğitim durumu ve nüfus grubu gibi değişkenler alınarak analize tabi tutulmuştur. Ele alınan bağımlı değişken parantez içinde DİE tanımları ve anahtar kelimeler olmak üzere;

1- Düzenli Gelir, Düzenli Çalışan (düzenli ücretli-maaşlı) (***DG-DÇ***),

2- Düzensiz Gelir, Düzensiz Çalışan (yevmiyeli-mevsimlik, arizi, geçici, ücretli ev işçisi) (***DSG-DSÇ***),

3- Düzensiz Gelir, Düzenli Çalışan (işveren, kendi hesabına çalışan) (***DSG-DÇ***) ve

4- Ücretsiz Çalışan (ücretsiz aile işçisi) (***ÜSÇ***)

şeklinde dört grupta incelenmiştir. Kısıtlar dikkate alındığında çalışma 20.123 kişi üzerinden yürütülmüştür. Değişkenlere ait frekanslar EK-1'de sunulmuştur.

Yukarıdaki amacı gerçekleştirmek için, model tahmin edilmiş, öncelikle tahmin edilen modelin uyum bilgisi ile modelde kullanılan her bir bağımsız değişkenin istatistiksel anlamlılığını veren tablolara yer verilmiştir (Tablo 3-4).

Tablo 3: Model Uyum Bilgisi, Uyum İyiliği ve Pseudo R^2

Model Uyum Bilgisi					*Uyum İyiliği*				*Pseudo R^2*	
Model	**-2LL**	χ^2	**s.d.**	**p**		χ^2	**s.d.**	**p**	Cox and Snell	,646
Sabit Terimli	25414,967				Pearson	11284,637	1785	,000	Nagelkerke	,705
Doymuş	4498,467	20916,501	48	,000	Sapma	2929,463	1785	,000	McFadden	,419

Model uyum bilgisi ile Pearson ve Sapma değerlerine bakıldığında kurulan modelin istatistiksel olarak anlamlı olduğu görülmektedir. Bunun yanı sıra doğrusal regresyonda bulunan R^2 ile aynı anlama gelen Pseudo R^2 değerleri ise oldukça yüksek çıkmıştır (bkz.Tablo 3).

Tablo 4: Olabilirlik Oran Testleri

Etki	**-2LL**	χ^2	**s.d.**	**p**
Sabit	4498,467	,000	0	,
Cinsiyet	6733,214	2234,748	3	,000
Medeni Durum	4792,216	293,749	9	,000
Çalışan Kişi Sayısı	12050,291	7551,824	6	,000
Eğitim Durumu	5300,404	801,937	12	,000
Nüfus Grubu	6246,361	1747,894	3	,000
Yaş	5596,966	1098,499	15	,000

Daha önce de belirtildiği gibi Lojistik regresyon modelinde yer alan bir bağımsız değişkenin anlamlılığını ölçmek için olabilirlik oran testinden yararlanılır. Tablo 4 incelendiğinde tüm bağımsız değişkenlerin istatistiksel olarak anlamlı olduğu görülmektedir.

Modele ait genel istatistikler incelendikten sonra, parametre tahminleri (β), Wald istatistiklerine bağlı anlamlılık düzeyleri (p) ve daha önce belirtilen bahis oranlarının (B.Oranı) yer aldığı Tablo 5 düzenlenmiştir.

Tablo 5'te 1. Model incelendiğinde; düzenli gelire sahip, düzenli çalışan işgücü ücretsiz çalışan işgücüne göre erkeklerde kadınlara nazaran 7,27 kat daha yüksektir. Yine aynı grup için 30-39 yaş grubunda 2,29, 40-49 yaş grubunda 2,32 ve 50-59 yaş grubunda ise 1,71 kat 60 yaş üstüne nazaran daha yüksek olduğu görülmektedir.

Nüfus grubu dikkate alındığında, ücretsiz çalışanlara göre düzenli gelire sahip düzenli çalışan işgücünde, kent nüfus grubunda olanlara kıyasla kır nüfus grubunda olanlar daha düşüktür. Eğitim durumu dikkate alındığında ise düzenli gelire sahip, düzenli çalışan işgücü ücretsiz çalışan işgücüne kıyasla üniversite ve üstü eğitim alanlar diğer eğitim gruplarına göre daha yüksek çıkmıştır. DG-DÇ'lar ÜSÇ'la karşılaştırıldığında, medeni durum değişkeninde eşi ölmüşlere kıyasla bekar ve evli olan işgücü daha düşüktür.

2.Model ele alındığında, ücretsiz çalışan (ÜSÇ) işgücüne göre düzensiz gelire sahip, düzensiz çalışan (DSG-DSÇ) işgücü, erkeklerde bu kez kadınlara nazaran 11,846 kat; 30-39 yaş grubunda 2,56 kat, 40-49 yaş grubunda 2,31 kat ve 50-59 yaş grubunda ise 1,92 kat 60 yaş üstüne nazaran daha yüksektir. Yine 1. modeldeki gibi kent nüfus grubunda olanlara göre kır nüfus grubunda olanlar; eşi ölmüşlere kıyasla bekar ve evli olan işgücü daha düşük çıkmıştır.

Düzenli gelire sahip, düzensiz çalışan (DG-DSÇ) işgücünün ücretsiz çalışan (ÜSÇ) işgücüne göre kıyaslandığı üçüncü modelde ise erkek işgücünün istihdamda olması kadınlara nazaran 24,58 kat daha fazla olduğu anlaşılmaktadır. Bir diğer ifade ile, ücretsiz çalışan işgücünde kadınlar daha yoğundur.

Tablo 5: İşteki Durum Bağımlı Değişkeni İçin Lojistik Regresyon Analizi (Referans Kategorisi: Ücretsiz Çalışan)

	Kategoriler	**1. Model (Y=1'e karşı Y=0)**			**2. Model (Y=2'e karşı Y=0)**			**3. Model (Y=3'e karşı Y=0)**		
		β	**p**	**B.Oranı**	β	**p**	**B.Oranı**	β	**p**	**B.Oranı**
Sabit		8,555	*,000*		3,354	*,000*		5,508	*,000*	
Cinsiyet [1]	Erkek	1,984	*,000*	7,272	2,472	*,000*	11,846	3,202	*,000*	24,579
Medeni Durum [2]	H.Evlenmedi	-1,945	*,000*	,143	-2,564	*,000*	7,700E-02	-3,321	*,000*	3,612E-02
	Evli	-1,392	*,000*	,249	-2,079	*,000*	,125	-2,038	*,000*	,130
	Boşandı	,371	,465	1,449	-,660	,187	,517	-,597	,192	,550
Çalışan Kişi Sayısı [3]	-10	-5,792	*,000*	3,051E-03	-3,084	*,000*	4,576E-02	-,920	*,015*	,399
	10-24	-2,948	*,000*	5,246E-02	-1,971	*,000*	,139	-,921	*,022*	,398
Eğitim Durumu [4]	O-Y Bilmiyor	-3,131	*,000*	4,366E-02	,819	*,011*	2,267	-2,351	*,000*	9,526E-02
	İlkokul	-1,588	*,000*	,204	1,039	*,001*	2,826	-1,463	*,000*	,232
	Ortaokul	-1,659	*,000*	,190	,353	,272	1,423	-1,496	*,000*	,224
	Lise	-,977	*,000*	,376	,372	,242	1,451	-1,009	*,000*	,365
Nüfus Grubu [5]	Kır	-2,874	*,000*	5,646E-02	-2,195	*,000*	,111	-1,736	*,000*	,176
Yaş [6]	15-19	-1,694E-02	,941	,983	,119	,587	1,126	-3,862	*,000*	2,102E-02
	20-29	1,700E-02	,935	1,017	3,943E-02	,839	1,040	-2,343	*,000*	9,608E-02
	30-39	,829	*,000*	2,292	,942	*,000*	2,564	-,855	*,000*	,425
	40-49	,843	*,000*	2,324	,836	*,000*	2,307	-,484	*,001*	,616
	50-59	,535	*,016*	1,708	,654	*,001*	1,923	-,211	,161	,809

Referans kategorileri: [1] Kadın; [2] Eşi öldü; [3] 25+; [4] Üniversite+; [5] Kent; [6] 60+

Tablo 6: Sınıflama Tablosu

Gözlenen	Tahmin Edilen				
	DG-DÇ	**DSG-DSÇ**	**DSG-DÇ**	**ÜSÇ**	**%**
DG-DÇ	7299	262	1372	295	79,1
DSG-DSÇ	525	207	1313	296	8,8
DSG-DÇ	542	93	4801	296	83,8
ÜSÇ	302	91	334	2095	74,2
%	43,1	3,2	38,9	14,8	**71,6**

Kurulan model sonucunda ele alınan gözlem sonuçlarının doğru sınıflandırılması yaklaşık %72'dir. Bu durum sosyal bilimler için iyi bir sonuç olarak değerlendirilebilir.

7. SONUÇ

İşteki durumuna göre işgücünü lojistik regresyon analizine tabi tuttuğumuzda öncelikle ücretsiz çalışan işgücünde kadınların daha yoğun olduğu, medeni durum açısından değerlendirildiğinde ise düzenli gelire sahip işgücünde eşi ölmüş olanların daha fazla yer aldığı görülmektedir. Nitekim teorik olarak da ücretsiz aile işçilerinin genelde aile fertlerinin işlerini yaptıkları bilinmekte, dolayısıyla yaşamını idame ettirmek zorunda kalan eşi ölenlerin gelir karşılığı çalıştıkları anlaşılmaktadır.

Ücretli çalışan kesimin tümünün sermaye sahibi olan kesime nazaran en üretken yaş olan 30-59 yaş arasında çok daha yoğun olduğu ortaya çıkmıştır. Tüm modellerde ücretsiz çalışan işgücünde kır nüfus grubunda olanlar daha fazladır. Düzenli gelire sahip ve kendi hesabına çalışan işgücünde üniversite ve üstü eğitim alanlar diğer eğitim gruplarına göre daha yüksek çıkmıştır.

KAYNAKÇA

Agresti Alan, (1990); **Categorical Data Analysis**, New York: John Wiley & Sons.

Andersen Erling, (1997); **Introduction to the Statistical Analysis of Categorical Data**, Springer-Verlag Berlin.

Dayton Mitchell, (1992); Logistic Regression Analysis: http://www.education.umd.edu/EDMS/LRA/LRA.pdf

Everitt Brian, (1996); **Making Sense of Statistics in Psychology**, New York: Oxford Uni. Press.

Everitt, B.; Dunn, G.(1992); **Applied Multivariate Data Analysis**, New York: Oxford Uni. Press.

Garson D., (1999); Logistic Regression: http://www2.chass.ncsu.edu/garson/pa765/logistic.htm

Gujarati Damodar, (1999); **Temel Ekonometri**, Çev: Ü. Şenesen, G. G. Şenesen, İstanbul: Literatür Yayıncılık.

Hair J., Anderson R., Tatham R., Black W., (1998); **Multivariate Data Analysis**, USA: Prentice-Hall, Inc.

Hosmer D., Lemeshov S., (2000), **Applied Logistic Regression**, New York: John Wiley & Sons.

Johnson Dallas, (1998); **Applied Multivariate Methods For Data Analysis**, USA: Brooks/Cole Publishing Company.

King G., Zeng L., (2000); Logistic Regression in Rare Events Data: http://gking.harvard.edu/files/os.pdf

Horton J. N., Laird M. N., (2001); Maksimum Likelihood Analysis of Logistic Regression Model with Incomplete Covariate Data and Auxiliary Information: http://www.biostat.harvard.edu/~horton/biometrics.pdf

Özdamar Kazım, (1999); **Paket Programlar İle İstatistiksel Veri Analiz 1**, 2b, Eskişehir: Kaan Kitabevi.

Özdinç Özer, (1999); **Derecelendirme Sürecinde Ekonometrik Bir Değerlendirme**, Ankara: SPK Yayın No:130.

Powers D., Xie Y., (2000); **Statistical Methods For Categorical Data Analysis**, London: Academic Press.

Rush Sloan, Logistic Regression: The Standard Method of Analysis in Medical Research: http://trinity.edu/departments/mathematics/studpapers/s3.pdf

Sharma Subhash, (1996); **Applied Multivariate Technique**, New York: John Wiley & Sons.

Tatlıdil Hüseyin, (1996); **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Ankara: Cem Web Ofset Ltd. Şti.

EK-1 Yapılan Analizlere Ait Kategorik Değişkenlerin Frekans Dağılımı

İstihdamda Olanlar İçin		
Değişkenler	**Kategoriler**	**N**
Çalıştığı işteki durumu	Düzenli Gelir, Düzenli Çalışan	9228
	Düzensiz Gelir, Düzensiz Çalışan	2341
	Düzensiz Gelir, Düzenli Çalışan	5732
	Ücretsiz Çalışan	2822
Cinsiyet	Erkek	15512
	Kadın	4611
Medeni durum	Hiç Evlenmedi	4363
	Evli	15288
	Boşandı	206
	Eşi Öldü	266
Işyerinde çalışan kişi sayısı	10' dan az	12416
	10-24 kişi	2301
	25' den fazla	5406
Eğitim durumu	O-Y Bilmiyor	1410
	İlkokul	10484
	Ortaokul	2205
	Lise	3776
	Üniversite +	2248
Nüfus grubu	Kır	6153
	Kent	13970
Yaş kategorileri	15-19	1788
	20-29	5009
	30-39	6062
	40-49	4428
	50-59	1821
	60+	1015
Toplam		20123