

PAPER DETAILS

TITLE: LOJISTIK REGRESYON ANALIZI YARDIMIYLA SUÇLU PROFILININ BELIRLENMESI

AUTHORS: Ayse OGUZLAR

PAGES: 21-35

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/30026>

LOJİSTİK REGRESYON ANALİZİ YADIMIYLA SUÇLU PROFİLİNİN BELİRLENMESİ

Ayşe OĞUZLAR^(*)

Özet: Lojistik regresyon analizi bağımlı değişkenin kategorik bir değişken olması durumunda çoklu regresyon modelinin bir türü olarak düşünülebilir ve sosyal bilimlerle ilgili araştırmalarda sıkça kullanılır. Bu çalışmada Bursa Emniyet Müdürlüğü ile ortak yapılan bir proje kapsamında, ahlak, yankesicilik ve narkotik bürolarına ilişkin verilerden yararlanılarak, suçluların profilinin belirlenmesinde lojistik regresyon analizinden yararlanılmıştır. Öncelikle sözü edilen üç büro için, en iyi model bulunarak daha sonra modelin uyum iyiliği değerlendirilmiştir. Ardından bahis (odds) oranları yorumlanarak ilgili bürolara ait suçlu profili belirlenmeye çalışılmıştır.

Anahtar Kelimeler: Lojistik regresyon analizi, Odds oranı, Olabilirlik oran testi.

Abstract: Logistic regression analysis can be considered as a kind of multi-regression model and it is commonly used at researches on social sciences. In this study with cooperation of police headquarters of Bursa, Logistic regression analysis is used for definition of profiles of criminals based on the records of moral, burglary and narcotic bureaus. Firstly the best model was defined for the subject bureaus mentioned above and then the goodness of fit was evaluated. As a result criminal profiles of each bureaus were defined according to evaluation of odds ratios.

Keywords: Logistic regression analysis, Odds ratio, Likelihood ratio test.

I. Giriş

Lojistik regresyon analizi veya kısaca logit modeller, sosyal ve biyolojik bilimlerde sıkça kullanılmaktadır. Sosyal bilimlerdeki araştırmaların çoğunda, bağımlı değişkenin iki mümkün değerden birine sahip olabileceği varsayılmaktadır. Örneğin; bir genç yüksek okuldaki mezun olmuş olabilir veya olmayabilir, işçi çalışıyor olabilir veya çalışmıyor olabilir ya da bir hasta tedaviye cevap verebilir veya vermeyebilir. İki mümkün ve farklı değer içeren bu tür verilere iki değerli (binary) veriler denilmektedir. İki değerli veya ikili değişkenler literatürde (0;1) değişkenleri olarak da adlandırılmaktadır. İki değerli değişkenlerle çalışan bir araştırmacının hedefi, bağımsız değişkenlerin koşullu bir kümesine bağımlı olarak, başarı veya başarısızlık olasılığının kestirilmesidir (Powers ve Xie, 2000:41).

Lojistik regresyon analizi, bağımlı veya çıktı değişkeninin kategorik olduğu durumlarda çoklu regresyon analizi tekniğinin bir uzantısıdır. Pratikte de kategorik çıktı değişkenini içeren durumlara oldukça sık rastlanılmaktadır.

^(*) Yrd. Doç. Dr. Uludağ Üniversitesi İİBF Ekonometri Bölümü

Bağımlı değişkenin kategorik olması durumunda veri analizi için diğer mümkün teknikler olarak diskriminant analizi, probit analizi, logaritmik doğrusal regresyon ve lojistik regresyon sayılabilir (<http://www.ex.ac.uk/~SEGLEa/multivar2/disclogi.html>). Sözü edilen bu tekniklerin uygulanabilirliği de farklı durumların sağlanmasını gerektirmektedir: Örneğin, logaritmik doğrusal regresyon tüm bağımsız veya regresör değişkenlerin kategorik olmasını gerektirirken, diskriminant analizi tüm bağımsız değişkenlerin sürekli olmasını öngörmektedir. Lojistik regresyon analizi ise, kategorik ve sayısal bağımsız değişkenlerin varlığı durumunda daha az varsayım gerektirmektedir. Lojistik regresyon, kategorik bir bağımlı değişkenin ön kestirimi amacıyla bağlı olarak diskriminant analizine benzerdir ve daha az varsayımı gerektirir. Diğer yandan diskriminant analizinin gerektirdiği varsayımların sağlanması durumunda da lojistik regresyon uygulanabilir (Akgül ve Çevik, 2003:390). Ayrıca teorik alt yapı olarak regresyon denkleminin bulunması ve katsayılarının yorumu anlamında doğrusal regresyon ile benzerlikler göstermektedir.

Lojistik dönüşüm, başarı veya başarısızlığın bahis (odds) oranlarının doğal logaritmasıdır. Başarı durum olasılığının (p) lojistik dönüşümü aşağıdaki biçimde gösterilebilir (Hosmer ve Lemeshow; 2000:6):

$$\log \text{it}(p_i) = \log \left(\frac{p_i}{1 - p_i} \right) \quad (1)$$

(1) nolu denklem, genelleştirilmiş doğrusal model çerçevesinde bir bağıntı (link) fonksiyonu olarak ele alındığında ve x'ler bağımsız değişkenleri göstermek üzere, aşağıdaki logit model elde edilir:

$$\text{Log} \left(\frac{p_i}{1 - p_i} \right) = Z_i = \sum_{k=0}^p \beta_k x_{ik} \quad (2)$$

(1) nolu denklemde yer alan $\log \text{it}(p_i)$, p olasılığının lojistik dönüşümünü ifade etmektedir. p 0' a yaklaştığında, $\log \text{it}(p_i) \rightarrow -\infty$ a, p 1' e yaklaştığında ise $\log \text{it}(p_i) \rightarrow +\infty$ a yaklaşmaktadır. p_i , bağımlı değişkenin 1 değerini alma olasılığını, $1 - p_i$, bağımlı değişkenin 0 değerini alma olasılığını göstermek üzere p_i ' nin $1 - p_i$ ' ye oranı bahis oranı (odds ratio) olarak adlandırılır ve (3) numaralı eşitlikteki gibi elde edilir:

$$\frac{p_i}{1 - p_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} \quad (3)$$

Bahis oranı 1' e yakın çıkan değişkenler Y' nin değişimine önemli bir katkısı olmayan değişkenlerdir. Bu tür değişkenlerin katsayıları anlamlı bulunmaz ise, ilgili değişkenin önemli bir etmen olmadığı sonucu çıkarılabilir. Katsayının anlamlı olması koşulu ile, 1' den büyük bulunan bahis oranı değeri, ilgili değişkeninin önemli bir etken olduğunu gösterir. 0' a yakın çıkan bahis oranı değerleri ise, katsayının anlamlı olması şartıyla, değişkeninin önemli bir etmen olduğunu fakat Y' nin düşük değerler almasına neden olduğu negatif bir etki sağladığı söylenebilir (Özdamar, 1999:487). Lojistik regresyon, bağımlı değişkeninin logit değişkenine dönüşümünün ardından en çok olabilirlik tahminini (maximum likelihood estimation) kullanır (<http://www2.chass.ncsu.edu/garson/pa765/logistic.htm>). Lojistik regresyon analizinde katsayıların tahmin edilmesinin ardından, uygun bulunan veya uydurulan modelin içerdiği değişkenlerin anlamlılığı değerlendirilmektedir. Lojistik regresyon modelinde değişkenlerin anlamlı olup olmadıklarını sınavan ve yaygın olarak kullanılan üç test mevcuttur. Bunlar aşağıda sırasıyla kısaca açıklaması yapılan olabilirlik oran testi (likelihood ratio test), Wald testi (Wald test) ve skor testi (score test) dir.

II. Parametrelerin Anlamlılığının Testi

Olabilirlik fonksiyonunun kullanılmasıyla önkeşirilen ve gözlenen modelin kıyaslanmasında aşağıda gösterilen ve sapma (deviance) olarak adlandırılan eşitlikten yararlanılmaktadır:

$$D = -2 \log \left[\frac{\text{kestirilen modelin olabilirliği}}{\text{doymuş modelin olabilirliği}} \right] \quad (4)$$

Olabilirlik fonksiyonunun yazılması ile,

$$D = \sum_{i=1}^n d_i^2 = -2 \sum_{i=1}^n (y_i \log(\hat{P}_i / y_i) + (1 - y_i) \log((1 - \hat{P}_i) / (1 - y_i))) \quad (5)$$

biçimine dönüşen sapma ölçütü, p modelde yer alan parametre sayısını göstermek üzere (n-p) serbestlik dereceli ki-kare tablo değeri ile kıyaslanmaktadır (Tatlıdil, 1996:299). Lojistik model için sapma, doğrusal regresyon analizinde kullanılan SSE (hataların kareleri toplamı) ile aynı role sahiptir. Aynı zamanda bu eşitlik hipotez testi amaçlı olarak da kullanılabilir. Bu hipotez testi olabilirlik oran testi olarak adlandırılmaktadır. Modelde yer alan parametrelerin anlamlı olup olmadıklarının belirlenmesinde, bağımsız değişkeni içeren modelin sapması bağımsız değişkeni içermeyen modelin sapması ile karşılaştırılmaktadır

(Anderson, 1990:278). D değeriindeki bu değişim G istatistiği olarak adlandırılmaktadır. G istatistiği aşağıdaki biçimde tanımlanabilir:

$$G = D(\text{değişkeni içermeyen model için}) - D(\text{değişkeni içeren model için})$$

$$G = -2 \ln [\text{değişkensiz olasılık} / \text{değişkenli olasılık}]$$

G istatistiği, tüm katsayılarının testinde kullanılabileceği gibi eğim parametresi $H_0: \beta_1=0$ hipotezinin testinde 1 serbestlik dereceli bir ki-kare dağılımına sahiptir.

Wald testi, eğim parametresi β_1 ' in en çok olasılık tahmini ile bu tahminin standart hatasını karşılaştırmaya dayanmaktadır. $\hat{\beta}_1$ ' nin standart hatası, kovaryans matrisindeki köşegen elemanlarının kareköklerinin alınmasıyla elde edilmektedir. Lojistik regresyon modeli için bu test istatistiği aşağıdaki biçimde tanımlanabilir:

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \quad (6)$$

Eğim parametresini gösteren $\beta_1=0$ hipotezi için W istatistiği standart normal dağılım göstermektedir. Wald testinin başka bir açıdan da değerlendirmek mümkündür. Normal rassal bir değişkenin karesinin alınması 1 serbestlik dereceli bir ki-kare rassal değişkenine eşit olacağından, Wald istatistiği aşağıdaki biçimde de ifade edilebilir:

$$W^2 = \left(\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \right)^2 \quad (7)$$

Parametrelerin anlamlılığının testinde kullanılan bir başka test ise skor testidir. Bu test en olasılık denklemlerinin türevlerinin koşullu dağılımlarına dayanmaktadır ve aşağıdaki biçimde formüle edilebilir:

$$ST = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (8)$$

Bu istatistik eğim parametresini gösteren $\beta_1=0$ hipotezi için standart normal dağılım göstermektedir.

III. Modelin Oluşturulması

Model oluşturmadaki amaç, en az bağımsız değişken yardımıyla bağımlı değişkendeki değişimi en fazla biçimde açıklamaktır. Eğer fazla sayıda

değişken modele dahil edilirse standart hata tahminleri büyüyecektir. Aynı zamanda fazla sayıda bağımsız değişken ile modelin kurulması ve geliştirilmesi süreçleri daha karmaşık bir hal alacaktır. Lojistik regresyon modelinde değişken seçimi için çok farklı yöntemler mevcuttur. Bunlar tek değişkenli ve çok değişkenli analiz olmak üzere iki temel ayrıma göre yapılır. Çok değişkenli analiz iki yöntemi içermektedir. Bunlar adımsal (stepwise) yöntem ve en iyi alt setler yöntemidir. En iyi alt setler yöntemi lojistik regresyon analizinde pek kullanılmamaktadır (Costanza vd., 1992:775). Adımsal yöntem ise ileri yönlü seçim (forward selection) ve geriye doğru eleme (backward elimination) olmak üzere ikiye ayrılmaktadır (Lee ve Koval, 1997:566). Hata teriminin normal dağılım göstermesinden dolayı yeni değişkenin modele katkısının önemi için Doğrusal regresyon analizinde F testinden yararlanılmasına karşılık, lojistik regresyon modelinde hata terimleri binom dağılımı göstereceğinden daha önce sözü edilen G istatistiği ile karar verilir.

İleri yönlü seçim tekniğinde, değişkenler modele, yeni bir değişken uyumu arttırmayınca ya da değin eklenirler. Her bir adımda, uyumu en fazla arttıracak değişken seçilir. Geriye doğru eleme tekniğinde ise süreç en karmaşık tüm değişkenlerin modelde yer almasıyla başlar ve değişkenler ardıl olarak modelden çıkarılır. Her adımda, modele etkisi en az olan değişken modelden dışlanarak en uygun model bulunmaya çalışılır.

Olabilirlik oran testi, modelde yer alan bağımsız değişkenler için p katsayılarının tümünün anlamlılığının test edilmesinde de kullanılabilir. Bu testin G istatistiğine dayandığından daha önce söz edilmişti. Modelin anlamlı olup olmadığının belirlenmesinde, değişken(ler)siz modelin log-olabilirliği değişken(ler) ile modelin log-olabilirliği ile kıyaslanmalıdır. Bu durum için G istatistiği aşağıdaki biçimde hesaplanır:

$$G = -2 \ln [\text{değişken(ler)siz olabilirlik} / \text{değişken(ler)li olabilirlik}]$$

Modelin anlamlılığının kontrol edilmesinde hipotez takımları;

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{En az biri 0'dan farklı}$$

şeklinde dir. G istatistiği ($\gamma_2 - \gamma_1$) serbestlik dereceli ki-kare dağılımı göstermektedir. Serbestlik derecesi hesaplanırken kullanılacak γ_2 değeri tam modeldeki değişken sayısına 1 eklenerek ve γ_1 değeri ise düşürülmüş modeldeki değişken sayısına 1 eklenerek bulunacaktır. Eğer yukarıda verilen sıfır hipotezi red edilirse modelin anlamlı olduğu, eğer sıfır hipotezi red edilemez ise düşürülmüş modelin tam model kadar iyi olduğu sonucuna varılır.

IV. Modelin Uyum İyiliği

Modelin uyum iyiliği, bağımlı değişkeni açıklamak için oluşturulan en iyi modelin etkinliğinin bir ölçüsünü bize göstermektedir. Modelin uyum

iyiliğinin ölçüsü olarak bu başlık altında Hosmer-Lemeshow testi ve sınıflandırma tabloları ele alınacaktır.

Hosmer-Lemeshow testinde tahmin edilen olasılık değerleri gruplandırılmaktadır. Örneğin; tahmin edilen n olasılık değeri olduğu düşünüldüğünde, ilk sütun en küçük tahmin edilen olasılık değerini, n . sütun ise en büyük tahmin edilen olasılık değerini göstermiş olsun. Bu küçükten büyüğe doğru sıralandırılmış n tane tahmini olasılık değerinin gruplandırılmasında, iki farklı gruplama stratejisi izlenebilir: (1) Tahmini olasılık değerlerinin persantillerine dayalı olarak tablolaştırılması, (2) Tahmini olasılık değerlerinin kesin değerlere uygun olarak tablolaştırılması.

Hosmer-Lemeshow uyum iyiliği istatistiği $\hat{C}$, gözlenen ve teorik frekanslardan oluşan $2 \times g$ tablosu için χ^2 değeri olarak bulunur ve aşağıdaki biçimde hesaplanır:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n_k^* \bar{\pi}_k)^2}{n_k^* \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (9)$$

Formülde yer alan; O_k ; gözlenen frekanslardır ve $O_k = \sum_{j=1}^{n_k^*} y_j$ olarak bulunur.

$\bar{\pi}_k$; tahmin edilen olasılık değerlerinin ortalamasıdır ve $\bar{\pi}_k = \sum_{j=1}^{n_k^*} \frac{m_j \hat{\pi}_j}{n_k^*}$

şeklinde hesaplanır.

Sınıflandırma Tabloları (classification tables), uyum iyiliği için kullanılan bir başka ölçüttür. Bu tablo, bağımlı değişkenin çapraz sınıflandırılması ile elde edilir. Sınıflandırma tablosunda, bağımlı değişkeninin gözlenen ve kestirilen lojistik olasılıklarından türetilen (0 veya 1) değerleri yer almaktadır. Türetilen bağımlı değişken değerlerinin elde edilmesinde bir kesim değerinin (c) tanımlanması zorunludur. Kestirilen olasılık değeri c' yi aştığında türetilen bağımlı değişken 1, aksi durumda 0 değerini alacaktır. Kesim değeri için en çok kullanılan değer 0,5' dir. Ele alınan kestirim değeri 0,5 değerini geçtiğinde 1, aksi durumda 0 grubuna atama yapılmaktadır. Bu şekilde atama yapmanın bazı dezavantajlarının olduğu da bir gerçektir. Örneğin kestirim değeri olarak 0,52 ile 0,48 arasında neredeyse bir fark olmadığı halde, 0,5 kesim değerine göre atama yapıldığında, neredeyse birbirine eşit olan bu iki değer farklı gruplara atanabilecektir. Diskriminant analizinde olduğu gibi amaç sınıflandırma olduğunda sınıflandırma tablolarının kullanımı daha anlamlı olacaktır (Hosmer ve Lemeshow, 2000:156).

V. Uygulama

Bursa emniyet müdürlüğünden alınan suçlu veri tabanı kullanılarak lojistik regresyon analizi yapılmaya çalışılmıştır. Analizlerin yapılacağı bürolar ilgi çekici olması açısından, ağır suçlar, ahlak ve yankesicilik bürolarıdır. Veri tabanı bir bütün olarak ele alınarak hangi büroya ilişkin veriler analiz ediyorsa, o büronun büro değişkeninin değeri 1, diğer tüm değerler 0 olarak kodlanmıştır. Örneğin veri tabanında ağır suçlar bürosunun verileri analiz ediyorsa, ağır suçlara ilişkin büro değişkeninin değeri 1, diğerleri 0 olarak ele alınmıştır. Büro değişkeni dışında analize dahil edilen değişkenler aşağıdaki şekilde verilmiştir:

Bağımlı Değişken:

BURO

Bağımsız Değişkenler:

Suçun İşlendiği Olay Saati (OLAYSAAT):

1. 00.00-05.59
2. 06.00-11.59
3. 12.00-17.59
4. 18.00-23.59

Suçu İşleyen Bireyin Cinsiyeti (CİNSİYET):

1. Erkek
2. Kadın

Suçu İşleyen Bireyin Yaşı (YAŞ)

Suçu İşleyen Bireyin Doğum Yeri (DOGYER):

1. Akdeniz Bölgesi
2. Doğu Anadolu Bölgesi
3. Ege Bölgesi
4. Güney Doğu Anadolu Bölgesi
5. İç Anadolu Bölgesi
6. Karadeniz Bölgesi
7. Marmara Bölgesi
8. Göçmen (Yurt Dışı)

Suçu İşleyen Bireyin Öğrenim Durumu (ÖĞRENİM):

1. Okur Yazar Değil
2. Okur Yazar
3. İlkokul Mezunu
4. Ortaokul mezunu
5. Lise Mezunu
6. Meslek Yüksekokulu Mezunu
7. Üniversite Mezunu

Suçu İşleyen Bireyin Mesleği (MESLEK):

1. Boşta Gezer
2. Öğrenci
3. Çiftçi
4. Ev Hanımı
5. İşçi (Kamu veya Özel)
6. Kamu Personeli (Devlet Memuru)
7. Serbest Meslek (Avukat, Mühendis, Doktor, Muhasebeci, Teknisyen,.....)
8. Esnaf ve Sanatkar
9. Yönetici
10. Emekli

Bursa Emniyet Müdürlüğüne ait polis suç veri tabanına dahil 13 bürodan ahlak, yankesicilik ve narkotik büroları birbirlerinden farklı özelliklere sahip olduğundan dolayı keyfi olarak seçilmiştir.

Öncelikle ahlak bürosuna ilişkin olarak yukarıda sözü edilen altı model ele alınarak, SPSS’ de geriye doğru olabilirlik oranı seçim kriteri (Backward LR) kullanılarak ele alınan altı değişken içerisinde önemli olanları belirlenmeye çalışılmıştır. Elde edilen sonuçlar Tablo 1’ deki gibidir.

Tablo 1: *Ahlak Bürosuna İlişkin Geriye Doğru Olabilirlik Oranı Değişim Tablosu*

	Değişkenler	Model Olabilirlik Oranı	Olabilirlik Oran Değerindeki Değişim	S.d.	Değişimin Anlamlılık Düzeyi
Adım 1	OLAYSAAT	-24603,095	3369,960	4	,000
	CINSİYET	-22923,987	11,743	1	,001
	YAS	-23744,545	1652,859	1	,000
	DOGYER	-23049,604	262,977	8	,000
	MESLEK	-23077,028	317,826	10	,000
	OGRENİM	-23082,273	328,315	7	,000

Tablo 1’ e bakıldığında ele alınan tüm değişkenlerin önemli bulunduğu ve tümünün modele dahil edildiğini görmekteyiz. Altı bağımsız değişkenden oluşan modelin uyum iyiliğini gösteren Hosmer ve Lemeshow test sonucunun ($\chi^2 = 293,817$, anlamlılık düzeyi 0,000) anlamlı bulunduğu anlaşılmaktadır. Modelin verilere uygun olmasının anlaşılmasının ardından sınıflandırma tablosu elde edilmiştir. Ahlak bürosu verilerine ve altı değişken ele alınarak bulunan modelin doğru sınıflandırma oranı %94,2’ dir. Bu değer modelin çok iyi bir biçimde doğru sınıflandırma yaptığını göstermektedir. Modelin uygun olduğunun belirlenmesinin ardından denklemde yer alan değişken kategorilerine ilişkin bahis oranlarının yorumlanması amacıyla Tablo 2’ den yararlanılmıştır.

Tablo 2: Ahlak Bürosuna İlişkin Denklemden Yer Alan Değişkenler ve Kategorilerine İlişkin Özet Değerler

	Değişkenler	B	S.E.	Wald	S.d.	Anlamlılık Düzeyi	Bahis Oranı
Adım 1	OLAYSAAT			2333,629	4	,000	
	OLAYSAAT(1)	-2,189	,108	409,347	1	,000	,112
	OLAYSAAT(2)	-,407	,029	195,529	1	,000	,666
	OLAYSAAT(3)	-2,457	,085	843,132	1	,000	,086
	OLAYSAAT(4)	-1,705	,047	1344,252	1	,000	,182
	CINSİYET(1)	,175	,052	11,334	1	,001	1,191
	YAS	,045	,001	1678,334	1	,000	1,046
	DOGYER			211,128	8	,000	
	DOGYER(1)	-,836	,173	23,347	1	,000	,433
	DOGYER(2)	-1,609	,192	70,196	1	,000	,200
	DOGYER(3)	-,578	,145	15,916	1	,000	,561
	DOGYER(4)	-,763	,178	18,273	1	,000	,466
	DOGYER(5)	-,735	,157	21,983	1	,000	,480
	DOGYER(6)	-,293	,150	3,816	1	,051	,746
	DOGYER(7)	-,306	,144	4,522	1	,033	,736
	DOGYER(8)	-,285	,140	4,121	1	,042	,752
	OGRENİM			305,279	7	,000	
	OGRENİM(1)	1,078	,285	14,296	1	,000	2,940
	OGRENİM(2)	1,674	,283	34,885	1	,000	5,335
	OGRENİM(3)	1,606	,284	32,013	1	,000	4,982
	OGRENİM(4)	1,280	,284	20,327	1	,000	3,598
	OGRENİM(5)	1,265	,290	19,071	1	,000	3,543
	OGRENİM(6)	1,334	,289	21,317	1	,000	3,797
	OGRENİM(7)	1,084	,335	10,444	1	,001	2,956
	MESLEK			288,266	10	,000	
	MESLEK(1)	,598	,110	29,688	1	,000	1,819
	MESLEK(2)	,222	,111	4,023	1	,045	1,249
	MESLEK(3)	-1,831	,684	7,162	1	,007	,160
	MESLEK(4)	,022	,189	,013	1	,909	1,022
	MESLEK(5)	,011	,166	,004	1	,947	1,011
	MESLEK(6)	,552	,114	23,372	1	,000	1,737
	MESLEK(7)	,142	,177	,640	1	,424	1,152
	MESLEK(8)	,680	,109	39,195	1	,000	1,974
	MESLEK(9)	,884	,117	57,154	1	,000	2,420
	MESLEK(10)	1,328	,769	2,985	1	,084	3,774
	Sabit	-5,512	,340	263,138	1	,000	,004

Öncelikle Tablo 2’ de yer alan anlamlılık düzeylerine bağlı olarak ilgili değişken kategorisinin anlamlı olup olmadığına bakılarak anlamlı olmayan kategoriler için bahis oranları yorumlanmamıştır. Ardından bahis oranları arasından suçlu profilini belirlemeye yarayacak olanlar (1 ve daha büyük olanlar) yorumlanmaya çalışılmıştır. Erkek olan kişilerin ahlak suçu işleme oranı ahlak suçu işlemeyenlere oranla 1,191 kat daha fazladır. Ayrıca yaş değişkeni ele alındığında yaştaki bir birimlik artış ahlak suçu işleme riskini

%4,6 oranında arttırmaktadır. Ahlak suçu işleme oranı ahlak suçu işlemeyenlere göre okur yazar olmayanlarda 2.940 kez, okur yazar olanlarda 5,335, ilkokul mezunu olanlarda 4,982, ortaokul mezunlarında 3,598, lise mezunlarında 3,543, meslek yüksek okulu mezunlarında 3,797 ve üniversite mezunlarında 2,956 kez daha fazladır. Boşta gezenlerin ahlak suçu işleme oranı, ahlak suçu işlemeyenlere oranla 1,819 kez, öğrencilerin 1,249 kez, kamu personeli olanların 1,737 kez, esnaf ve sanatkarların 1,974 kez, yönetici olanların 2,420 kez daha fazladır.

Yankesicilik bürosuna ilişkin geriye doğru olabilirlik oranı değişim tablosu aşağıda görüldüğü şekildedir:

Tablo 3: *Yankesicilik Bürosuna İlişkin Geriye Doğru Olabilirlik Oranı Değişim Tablosu*

	Değişkenler	Model Olabilirlik Oranı	Olabilirlik Oran Değerindeki Değişim	S.d.	Değişimin Anlamlılık Düzeyi
Adım 1	OLAYSAAT	-16141,545	2686,776	4	,000
	CINSİYET	-14800,483	4,652	1	,031
	YAS	-14815,283	34,251	1	,000
	DOGYER	-21091,519	12586,723	8	,000
	OGRENİM	-15163,229	730,145	7	,000
	MESLEK	-14943,834	291,353	10	,000

Tablo 3 incelendiğinde ele alınan tüm değişkenlerin önemli bulunduğu ve modele dahil edildiği anlaşılmaktadır. Ayrıca ahlak bürosunda olduğu gibi altı değişkenden oluşan modelin uyum iyiliğini gösteren Hosmer ve Lemeshow test sonucunun da ($\chi^2 = 35,529$, anlamlılık düzeyi 0,000) anlamlı bulunduğu anlaşılmaktadır. Modelin verilere uygun olduğunun anlaşılmasının ardından yankesicilik bürosuna ilişkin sınıflandırma tablosu elde edilmiştir. İlgili tablodan yankesicilik bürosu için bulunan doğru sınıflandırma oranının %95,9 olduğu anlaşılmaktadır. Bu değer modelin çok iyi doğru sınıflandırma yüzdesine sahip olduğunu göstermektedir. Modelin uygun olduğunun anlaşılmasının ardından bahis oranlarının yorumlanması amacıyla aşağıda gösterilen Tablo 4' den yararlanılmıştır.

Tablo 4: Yankesicilik Bürosuna İlişkin Denklemden Yer Alan Değişkenler ve Kategorilerine İlişkin Özet Değerler

	Değişkenler	B	S.E.	Wald	S.d.	Anlamlılık Düzeyi	Bahis Oranı
Adım 1	OLAYSAAT			1996,864	4	,000	
	OLAYSAAT(1)	,546	,073	56,272	1	,000	1,726
	OLAYSAAT(2)	-1,451	,071	417,719	1	,000	,234
	OLAYSAAT(3)	,776	,051	233,280	1	,000	2,172
	OLAYSAAT(4)	1,187	,040	877,443	1	,000	3,277
	CINSİYET(1)	-,109	,050	4,703	1	,030	,897
	YAS	-,009	,002	33,410	1	,000	,991
	DOGYER			9257,092	8	,000	
	DOGYER(1)	-1,522	,169	81,117	1	,000	,218
	DOGYER(2)	2,172	,117	344,299	1	,000	8,780
	DOGYER(3)	-,696	,117	35,527	1	,000	,498
	DOGYER(4)	-1,215	,155	61,124	1	,000	,297
	DOGYER(5)	-,271	,121	5,043	1	,025	,762
	DOGYER(6)	-,657	,124	28,206	1	,000	,518
	DOGYER(7)	-1,270	,121	110,960	1	,000	,281
	DOGYER(8)	-3,045	,122	621,940	1	,000	,048
	OGRENİM			627,601	7	,000	
	OGRENİM(1)	1,670	,274	37,209	1	,000	5,313
	OGRENİM(2)	2,710	,270	100,650	1	,000	15,034
	OGRENİM(3)	2,689	,271	98,559	1	,000	14,713
	OGRENİM(4)	2,187	,271	65,058	1	,000	8,908
	OGRENİM(5)	1,779	,283	39,414	1	,000	5,921
	OGRENİM(6)	1,731	,284	37,075	1	,000	5,649
	OGRENİM(7)	1,063	,421	6,388	1	,011	2,896
	MESLEK			251,186	10	,000	
	MESLEK(1)	1,063	,316	11,309	1	,001	2,895
	MESLEK(2)	1,382	,316	19,157	1	,000	3,984
	MESLEK(3)	,339	,398	,723	1	,395	1,403
	MESLEK(4)	1,513	,399	14,379	1	,000	4,541
	MESLEK(5)	1,178	,331	12,681	1	,000	3,247
	MESLEK(6)	,427	,322	1,751	1	,186	1,532
	MESLEK(7)	-,596	,448	1,765	1	,184	,551
	MESLEK(8)	1,122	,316	12,630	1	,000	3,070
	MESLEK(9)	,780	,327	5,693	1	,017	2,181
	MESLEK(10)	-3,456	10,188	,115	1	,734	,032
	Sabit	-5,406	,440	151,250	1	,000	,004

Öncelikle Tablo 2 için yapıldığı gibi anlamlılık (Sig) değerlerine bakılarak ilgili değişken kategorisinin anlamlı olup olmadığına bakılacaktır. Anlamlı olanların belirlenmesinin ardından bahis oranları arasında suçlu profilini belirlemeye yarayacak olanlar (1' den daha büyük olanlar) yorumlanmaya çalışacaktır. Yankesicilik suçunu 00.00-05.59 arasında işleyenlerin oranı işlemeyenlere oranla 1,726 kat, 12.00-17.59 arasında 2,172 kat, 18.00-23.59 arasında 3,277 kat oranla daha fazladır. Doğu Anadolu Bölgesi

doğumlu olup, yankesicilik suçunu işleyenlerin oranı işlemeyenlere oranla 8,780 kat daha fazladır. Yankesicilik suçu işleyen ve okur yazar olmayanların oranı bu suçu işlemeyenlere oranla 5,313 kat, okur yazar olanların oranı 15,034 kat, ilkokul mezunu olanların oranı 14,713 kat, ortaokul mezunu olanların oranı 8,908 kat, lise mezunu olanların oranı 5,921 kat, meslek yüksek okulu mezunu olanların oranı 5,649 kat ve üniversite mezunu olanların oranı ise 2,896 kat daha fazladır. Boşta gezen ve yankesicilik suçu işleyenlerin oranı, bu suçu işlemeyenlere oranla 2,895 kat, yankesicilik suçu işleyip öğrenci olanların oranı bu suçu işlemeyenlere oranla 3,984 kat, ev hanımı olanlarda 4,541 kat, işçi olanlarda 3,247 kat, esnaf ve sanatkar olanlarda 3,070 kat ve yönetici olanlarda 2,181 kat daha fazladır.

Narkotik bürosuna ilişkin geriye doğru olabilirlik oranı değişim tablosu ise aşağıda görüldüğü şekildedir:

Tablo 5: *Narkotik Bürosuna İlişkin Geriye Doğru Olabilirlik Oranı Değişim Tablosu*

	Değişkenler	Model Olabilirlik Oranı	Olabilirlik Oran Değerindeki Değişim	S.d.	Değişimin Anlamlılık Düzeyi
Adım 2	OLAYSAAT	-6357,921	470,929	4	,000
	YAS	-6177,657	110,399	1	,000
	OGRENİM	-6234,551	224,189	7	,000
	MESLEK	-6205,224	165,535	10	,000
	DOGYER	-8406,079	4567,244	8	,000

Tablo 5 incelendiğinde ikinci adımda cinsiyet değişkeninin modelden çıkarılmasına karar verildiği (anlamlılık düzeyi 0,477 olduğundan dolayı) anlaşılmaktadır. Cinsiyet değişkeni modelden dışlandığında, geriye kalan beş değişkenden oluşan modelin uyum iyiliğini gösteren Hosmer ve Lemeshow test sonucunun ($\chi^2 = 22,398$, anlamlılık düzeyi 0,004) anlamlı bulunduğu anlaşılmaktadır. Bu testin ardından narkotik bürosu verileri için sınıflandırma tablosu elde edilmiştir. Narkotik bürosu için bulunan doğru sınıflandırma oranının %98,6 olduğu anlaşılmaktadır. Bu değer, modelin mükemmel bir doğru sınıflandırma yüzdesine sahip olduğunu göstermektedir. Model uyumunun belirlenmesinin ardından bahis oranlarının yorumlanması amacıyla aşağıdaki Tablo 7' den yararlanılmıştır.

Tablo 6: Narkotik Bürosuna İlişkin Denklemden Yer Alan Değişkenler ve Kategorilerine İlişkin Özet Değerler

	Değişkenler	B	S.E.	Wald	S.d.	Anlamlılık Düzeyi	Bahis Oranı
Adım 2	OLAYSAAT			66,681	4	,000	
	OLAYSAAT(1)	-44,350	5,498	65,078	1	,000	,000
	OLAYSAAT(2)	-,008	,068	,013	1	,908	,992
	OLAYSAAT(3)	-,117	,090	1,691	1	,193	,890
	OLAYSAAT(4)	,012	,070	,029	1	,866	1,012
	YAS	,022	,002	92,809	1	,000	1,022
	DOGYER			2969,417	8	,000	
	DOGYER(1)	2,504	,219	131,150	1	,000	12,237
	DOGYER(2)	-,462	,244	3,583	1	,058	,630
	DOGYER(3)	,190	,219	,758	1	,384	1,210
	DOGYER(4)	1,013	,229	19,517	1	,000	2,754
	DOGYER(5)	-,250	,233	1,156	1	,282	,778
	DOGYER(6)	-,858	,247	12,100	1	,001	,424
	DOGYER(7)	-1,937	,253	58,724	1	,000	,144
	DOGYER(8)	-3,512	,252	193,718	1	,000	,030
	OGRENİM			203,184	7	,000	
	OGRENİM(1)	,650	,289	5,057	1	,025	1,915
	OGRENİM(2)	-,196	,291	,454	1	,500	,822
	OGRENİM(3)	,942	,287	10,775	1	,001	2,566
	OGRENİM(4)	,581	,288	4,061	1	,044	1,788
	OGRENİM(5)	,186	,321	,337	1	,561	1,205
	OGRENİM(6)	,189	,323	,340	1	,560	1,207
	OGRENİM(7)	-,407	,584	,485	1	,486	,666
	MESLEK			112,500	10	,000	
	MESLEK(1)	7,525	10,654	,499	1	,480	1853,602
	MESLEK(2)	7,451	10,654	,489	1	,484	1721,623
	MESLEK(3)	6,020	10,670	,318	1	,573	411,762
	MESLEK(4)	8,509	10,658	,638	1	,425	4961,299
	MESLEK(5)	5,993	10,658	,316	1	,574	400,780
	MESLEK(6)	7,026	10,655	,435	1	,510	1125,109
	MESLEK(7)	6,787	10,660	,405	1	,524	886,244
	MESLEK(8)	7,872	10,654	,546	1	,460	2622,362
	MESLEK(9)	7,547	10,655	,502	1	,479	1894,795
	MESLEK(10)	-,543	70,644	,000	1	,994	,581
	Sabit	-12,265	10,661	1,324	1	,250	,000

Diğer iki büroda yapıldığı gibi bu büroda da öncelikle anlamlı olan değişken kategorileri belirlenerek, bu kategorilere ilişkin 1' den büyük çıkan bahis oranları yorumlanmaya çalışılmıştır. Yaş değişkeni ele alındığında, yaşın bir birim artmasının narkotik suçlarından birini işleme riskini %2,2 oranında arttırdığı anlaşılmaktadır. Akdeniz Bölgesi doğumlu olup, narkotik suçlarını işleyenlerin oranı işlemeyenlere oranla 12,237 kat ve Güney Doğu Anadolu Bölgesi doğumlu olanların 2,754 kat daha fazladır. Narkotik suçları işleyen ve okur yazar olmayanların oranı bu suçu işlemeyenlere oranla 1,915 kat, ilköğretim

mezunu olanların 2,566 kat ve ortaokul mezunu olanların 1,788 kat daha fazladır. Meslek grubu kategorilerinin bahis oranlarının aşırı yüksek oluşu dikkat çekici bir durum oluşturmaktadır fakat kategoriler anlamlı olmadığından dolayı yorum yapılmamıştır.

VI. Sonuç

Bursa Emniyet Müdürlüğü ile yapılan bir proje kapsamında elde edilen verilerden hareketle, ahlak, yankesicilik ve narkotik büroları için, kişilerin ilgili büroya ait suç işlemesinde etkili olan faktörler belirlenmeye çalışılmıştır. Bu yapılırken öncelikle önemli olan değişkenlerin belirlenmesinde geriye doğru seçim kriteri kullanılmıştır. Bu bağlamda; ahlak ve yankesicilik bürolarına ilişkin veriler ele alındığında, modele dahil edilen altı değişkenin tümünün ve narkotik bürosuna ilişkin sonuçlar incelendiğinde ise cinsiyet değişkeni dışında beş değişkeninin anlamlı olduğu sonucuna varılmıştır.

Ahlak bürosuna ilişkin değişken kategorileri arasında, anlamlı ve bahis oranı 1' den büyük olanlara bakıldığında, ahlak suçu işleyenlerin en belirgin özelliği olarak öğrenim durumu değişkeninin kategorileri göze çarpmaktadır. Öğrenim durumu değişkenine ait kategorilerin tümünün bahis oranının 1' den büyük olduğu görülmektedir. Dolayısı ile ahlak suçunu işlemede bireyin öğrenim durumu belirleyicidir. Özellikle okur yazar ve ilkokul mezunu olanların bu suçları işlemeye meyilli oldukları söylenebilir. Ayrıca mesleği emekli olanlar da ahlak suçu işlemeye eğilimli olan diğer bir grubu oluşturmaktadır.

Yankesicilik bürosuna ait elde edilen bahis oranlarına bakıldığında da, yine öğrenim durumu değişkeninin yankesicilik suçlarını işleme üzerinde etkin olduğu söylenebilir. Bu büroya ait öğrenim durumu değişkeninin tüm kategorilerinin bahis oranları 1' den büyük çıkmıştır. Özellikle okur yazar veya ilkokul mezunu olanların bu suçu işlemeye daha eğilimli oldukları söylenebilir. Meslek değişkeninin kategorilerine bakıldığında da, bahis oranı 1' den büyük olanlara rastlanılmaktadır. Yine bu değişkeninin kategorilerden biri olan ev hanımlarının bu suçu daha fazla işlemeye meyilli oldukları söylenebilir. Bir başka farklı durum da, bu büroya ait doğum yeri değişkeninin kategorileri arasından yalnızca Doğu Anadolu bölgesi doğumluların bahis oranı 1' den büyük bulunmuştur. Doğu Anadolu bölgesi doğumlu olanların bu suçları işlemeye daha eğilimli oldukları söylenebilir.

Son olarak narkotik bürosu için elde edilen sonuçlara bakıldığında, yine ağırlıklı olarak öğrenim durumu değişkeninin bu suçları işlemede belirleyici olduğunu söylemek mümkündür. Özellikle ilkokul mezunu olan veya okur yazar olmayanlar bu suçları işlemeye daha meyillidir. Meslek değişkeninin kategorileri için bulunan bahis oranları çok yüksek çıkmasına karşılık, anlamsız olduklarından yorumlanmamıştır. Doğum yeri değişkeninin kategorilerine bakıldığında da, en fazla bahis oranı değerine sahip Akdeniz bölgesi doğumluların öne çıktığını görmekteyiz. Dolayısıyla Akdeniz bölgesi doğumlu

olanların bu tür suçları işlemeye eğilimli olacağını söylemek yanlış olmayacaktır.

Kaynaklar

- Akgül A. ve Çevik O., (2003), İstatistiksel Analiz Teknikleri “SPSS’ te İşletme Yönetimi Uygulamaları”, Emek Ofset, Ankara.
- Anderson Erling B. (1990), *The Statistical Analysis of Categorical Data*, Springer Verlag, Berlin.
- Costanza M.E., Staddat A.M., Gaw V. ve Zapple J.G. (1992), “The Risk Factors of Age and Family History and Relationship To Screening Mammography Utilization”, *Journal of The American Statistical Association*, 40, 776.
- Hosmer David W. ve Lemeshow Stanley (2000), *Applied Logistic Regression*, John Wiley and Sons, New York.
- Lee K. ve Koval J.J. (1997), “Determination of The Best Significance Level in Forward Stepwise Logistic Regression”, *Communication in Statistics*, 26(B), 566.
- Logistic regression and discriminant analysis, <http://www.ex.ac.uk/~SEGLea/multvar2/disclogi.html> >(24.09.2004).
- Logistic Regression, <http://www2.chass.ncsu.edu/garson/pa765/logistic.htm> > (18.09.2004).
- Özdamar K. (1999), *Paket Programlar İle İstatistiksel Veri Analizi 1*, Kaan Kitapevi, Eskişehir.
- Powers Daniel A. ve Xie Yu (2000), *Statistical Methods for Categorical Data Analysis*, Academic Press, ABD.
- Tatlıdil H. (1996), *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Cem Web Ofset, Ankara.