



# LOJİSTİK REGRESYON YÖNTEMİ İLE AYVALIĞI TURİZM AMAÇLI TERCİH ETMEDE ÖNEMLİ RİSK FAKTÖRLERİNİN BELİRLENMESİ

DETERMINING THE IMPORTANT RISK FACTORS IN PREFERRING AYVALIK FOR  
TOURISTIC PURPOSE USING THE METHOD OF LOGISTIC REGRESYON

Orhan ŞAHİN<sup>1</sup>

## Öz

Bu çalışmanın amacı, ikili sonuç değişkeni ile hem sürekli hem de kesikli değişkenlerden oluşan bağımsız değişkenler arasındaki ilişkiyi ortaya koyabilen lojistik regresyon analizinin incelenmesidir.

Lojistik regresyon analizine bir uygulama göstermek amacıyla Ayvalık'ı turizm amacıyla tercih etmede önemli risk faktörlerini belirlemek için bir anket uygulaması yapılmıştır.

Lojistik değişkene dahil edilecek bağımsız değişkenler adımsa teknikler, geriye dönük eleme yöntemi ile belirlendikten sonra, çok değişkenli modele dahil edilen her bir değişkenin önemliliği gösterilmiştir.

**Anahtar Kelimeler:** Lojistik regresyon analizi, odds oranı, turizm

## Abstract

Purpose of this study is to survey the logistic regression analyse which reveals the relation between independent variables that make up of either constant or discrete variables and dichotomous outcome variable.

To carry out an application for logistic regression analyse a survey application has been made which states the important risk factors on preferring Ayvalık for tourism.

Independent variables step techniques that will be included logistic variables after being determined by backwards elimination method the importance of every single variable which included in multivariate model is pointed out.

**Keywords:** Logistic Regression Analysis, Odds ratio, tourism

---

<sup>1</sup> Yrd.Doç.Dr., Balıkesir Üniversitesi, ayvalık MYO, [osahin@balikesir.edu.tr](mailto:osahin@balikesir.edu.tr)

## 1.GİRİŞ

Turizm faaliyetleri dünya ekonomisinde giderek artan oranlı bir gelişim göstermektedir. Birçok ülkede turizm, GSMH ile istihdamın önemli bir bölümünü meydana getirmektedir. Buna ilave olarak turizmden elde edilen uluslararası döviz gelirleri ülkelerin mevcut bütçe açıklarını finanse etmeye yardımcı olmaktadır. Bölgesel düzeyde ise turizm işsizlik sorununun çözümüne yardım etmekte ve tarım sektöründen dolayı kaybedilen rekabet avantajının yeniden kazanılmasına yarar sağlamaktadır. Bununla birlikte, turizm ekonomisi birçok ülkedeki ulaşım, ticaret, inşaat, konaklama, yiyecek-içecek sektörleri ve diğer hizmet sektörleri ile birlikte doğrudan ya da dolaylı bir yayılma etkisi göstermektedir (Proença ve Soukiazı 2008). Turizm gelirinin zengin ülkelerden daha fakir ülkelere, gelişmişlerden gelişmekte ve az gelişmiş ülkelere doğru dağılımına katkı yaparak ülkeler arasında yakınsamaya da neden olmaktadır. Böylece turizm bölgesel gelişmeye de yardımcı olmakta ve bölgesel ekonomik farklılıkların azaltılmasına olanak sağlamaktadır. Dolayısıyla sektöre yapılacak ulusal/uluslararası yatırımlar, kısa sürelerde etkisini göstererek geri kalmış bölgelerin ilerlemesine olumlu yönde etki edebilecektir (Proença ve Soukiazı 2008).

Bölgesel gelişmeye önemli katkıları olan turizm faaliyetlerinin üzerine etkili olacak risk faktörlerini belirlemek üzere Ülkemizin bir turizm destinasyonu olan Ayvalık bölgesinde lojistik regresyon analizi yapılmıştır.

Gözlemleri verilerin yapısında bulunan olası gruplara atamak için kullanılan yöntemlerden birisi Lojistik Regresyon Analizidir. Lojistik modelin biyolojik deneylerin analizi için kullanımı ilk olarak Berkson (1944) tarafından önerilmiş, Cox (1970) bu modeli gözden geçirerek çeşitli uygulamalarını yapmış, özet gelişmeler ise ilk Andersson (1979, 1983) tarafından verilmiştir. Ayrıca verilerin lojistik modele uyumu ile ilgili birçok çalışmalar da yapılmıştır. Bunlar arasında Aranda-Ordaz (1981) ve Johnson(1985) tarafından yapılan çalışmalar en önemlileridir. Pregibon (1981) iki grup lojistik modelde etkin (influential), aykırı (outlier) gözlemleri ve belirleme ölçütlerini (diagnostic), Lesaffre (1986), Lesaffre ve Albert (1989) ise çoklu grup lojistik modellerde etkin ve aykırı gözlemlerle belirleme ölçütlerini incelemişlerdir. Lojistik regresyon modellerinin yaygın bir şekilde kullanılabilir hale gelmesi, katsayı tahmin yöntemlerinin geliştirilmesi ve lojistik regresyon modellerinin daha ayrıntılı incelenmesine sebep olmuştur. Cornfield (1962), lojistik regresyondaki katsayı tahmin işlemlerinde diskriminant fonksiyonu yaklaşımını ilk kez kullanarak popüler hale getirmiştir. Lee (1984) basit dönüşümlü (cross-over) deneme planları için linear lojistik modeller üzerinde durmuştur. Bonney (1987) lojistik regresyon modelinin kullanımı ve geliştirilmesi üzerinde çalışmıştır. Robert ve ark. (1987) lojistik regresyonda standart Ki-kare, olabilirlik oran (G2), "pseudo" en çok olabilirlik tahminleri, uyum mükemmelliği ve hipotez testleri üzerine çalışmalar yapmışlardır. Duffy (1990) lojistik regresyonda hata terimlerinin dağılışı ve parametre değerlerinin gerçek değerlere yaklaşımını incelemiştir. Başarır (1990) klinik verilerde çok değişkenli lojistik regresyon analizi ve ayrımsama sorunu üzerinde çalışmıştır. Hsu ve Leonard (1995) lojistik regresyon fonksiyonlarında Bayes tahminlerinin elde edilmesi işlemleri üzerine çalışmışlar ve lojistik regresyonda Monte Carlo dönüşümünün kullanılabilirliğini göstermişlerdir. Cox ve ark. (1998) kardiyovasküler hastalıklar ve hipertansiyon arasındaki ilişkiyi incelemişlerdir. Lojistik regresyon modelleri, son yıllarda biyoloji, tıp, ekonomi, tarım ve veterinerlik ve taşıma sahalarında yaygın olarak kullanılmaktadır. Gardside ve Glueck (1995) insanlarda beslenme şekli, sigara ve alkol kullanımı, fiziksel aktivite gibi risk faktörlerinin kalp hastalığı üzerindeki etkilerini incelemiştir. Kloiber ve ark (1996), Peoples ve ark. (1991), Buescher ve ark. (1993)

kadınlarda düşük doğum ağırlığını etkileyen risk faktörlerini; Santos ve ark. (1998) kafein tüketimi ve düşük doğum ağırlığı arasındaki ilişkiyi; Sable ve Herman (1997) erken doğum ve düşük doğum ağırlığı arasındaki ilişkiyi incelemişlerdir.

Türkiye de de bu konuda çeşitli alanlarda çalışmalar yapılmıştır. Bunlardan bazıları ise şunlardır: Akkaya ve Pazarlıoğlu (1998) lojistik regresyon modellerinin ekonomi alanında kullanımını örneklerle incelemişlerdir. Kartal vd.(2004) Lojistik Regresyon analizinin dış hekimliğinde bir uygulamasını göstermişlerdir. Ünsal ve Güler (2005) Türk bankacılık sektörünü lojistik regresyon analiziyle incelerken, Tatlıdil, Başarır ve Hökmen (1990) ülkelerin sosyo ekonomik gelişmişliklerine göre sınıflandırılmasına ilişkin çalışma yapmışlardır. Vupa ve Çelikoğlu (2006) akciğer kanseri hastalar için lojistik regresyon modeli önermişlerdir. Ayrıca, Aktaş ve Yılmaz (2001) LPG kullanan özel araç sürücülerinin sınıflandırılmasını, Çolak ve Özdamar (2004) ölümle sonuçlanan trafik kazalarında risk faktörlerini, lojistik regresyon analiziyle incelemişlerdir. Büyükerişler ve Özlem GÜZEL(2014) Profesyonel Turist Rehberleri ve Alman Turistler üzerinde lojistik regresyon incelemesi yapmışlardır. Oktay ve Orçanlı (2014) Atatürk Üniversitesinde İnternet Bankacılığının Kullanımını Etkileyen Faktörlerin Belirlenmesi amacıyla lojistik regresyon kullanmışlardır.

Çeşitli varsayım bozulmaları olduğunda Lojistik Regresyon Analizi, diskriminant analizi ve çapraz tablo uygulamalarına alternatif olarak uygulanmaktadır. Kullanım nedeni olarak en temel yaklaşım doğrusal regresyon analizinde yapılabilir; bağımlı değişken 0 ve 1 gibi ikili (binary) ya da ikiden çok kategori içeren kesikli değişken olduğunda normallik varsayımı bozulmakta ve doğrusal regresyon analizi uygulanamamaktadır. Lojistik regresyonu doğrusal regresyondan ayıran en belirgin özellik ise lojistik regresyonda sonuç değişkenin ikili veya çoklu olmasıdır. Lojistik regresyon ve doğrusal regresyon arasındaki bu fark hem parametrik model seçimine, hem de varsayımlara yansımaktadır. Lojistik regresyonda da, doğrusal regresyon analizinde olduğu gibi bazı değişken değerlerine dayanarak tahmin yapılmaya çalışılır. Ancak bu iki yöntem arasında üç önemli fark vardır ( Elhan, A.H. 1997: 4): 1. Doğrusal regresyon analizinde tahmin edilecek olan bağımlı değişken sürekli iken, Lojistik Regresyon Analizinde bağımlı değişken kesikli bir değer almaktadır. 2. Doğrusal regresyon analizinde bağımlı değişkenin değeri, Lojistik Regresyon Analizinde ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir. 3. Doğrusal regresyon analizinde bağımsız değişkenin çoklu normal dağılım göstermesi şartı aranırken, Lojistik Regresyon Analizinde böyle bir şart yoktur.

Lojistik regresyon modelleri, zayıf ölçekle ölçülmüş değişkenler arasındaki ilişkinin şeklini ortaya koyan modellerdir. (Oktay ve Üstün,2001:347)

Bu modellerde genellikle incelenen olayın gerçekleşmesi 1, gerçekleşmemesi ise 0 ile ifade edilir. Bundan dolayı ikili kategorili lojistik regresyonda genellikle riskli durumların:1, risksiz durumların:0 ile kodlanması yararlı olur. Yani bazı alanlarda olgunun ortaya çıkması:1, çıkmaması:0 ile tanımlanabilmektedir. (Alpar,2011:616)

Lojistik model, ortaya çıkacak riski 0 ile 1 arasında herhangi bir değer olarak tahmin etmeye yarar. Başka bir deyişle 1'in üstünde veya 0'ın altında bir risk olmaz. Bu durum her model için her zaman doğru olmamaktadır. (Hosmer-Lemeshov,180:1043-1069).

## 2.Lojistik Regresyon Modeli

Bağımlı değişken nominal ölçekli bir değişken olduğunda En Küçük Kareler Tekniği ile elde edilen tahminler yetersiz kalmaktadır. Diğer bir anlatımla, tahmin edilen varyanslar artık minimum değildir. Çünkü EKK tekniği bağımlı değişkenin normal dağılıma uyduğunu varsaymaktadır. Bağımlı değişken nominal ölçekli olduğunda ise, bu varsayım sağlanamamaktadır. Bağımlı bir modelde, bağımlı değişken nominal ölçekli ise, EKK tekniğine alternatif olarak kullanılabilen teknikler arasında Lojistik Regresyon Analizi yer almaktadır (Albayrak, 2006: 439).

Lojistik Regresyon Analizi, temelde Regresyon Analizi olmakla birlikte bir ayırıcı analiz tekniği olma özelliğini de taşımaktadır. Bu kapsamda, Regresyon Analizinden üç önemli farklılığı vardır. (Coşkun vd, 2004: 43). - Regresyon Analizinde bağımlı değişken sayısal iken Lojistik Regresyon Analizi'nde kesikli bir değerdir. - Regresyon Analizinde bağımlı değişkenin değeri, Lojistik Regresyon 'da ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı kestirilir. - Regresyon Analizi'nde bağımsız değişkenlerin çoklu normal dağılım göstermesi koşulu aranırken, Lojistik Regresyon 'un uygulanabilmesi için bağımsız değişkenlerin dağılımına ilişkin hiçbir koşul gerekmemektedir.

İki Değişkenli (Binay) Lojistik Regresyon Analizinin temel amacı, diğer regresyon yöntemlerinde olduğu gibi bağımsız değişkenler ile bağımlı değişken arasındaki nedensellik ilişkisini incelemektir. Başka bir deyişle, amaç en az değişkeni kullanarak en iyi uyumu sağlayacak sonuç değişkeni ve açıklayıcı değişkenler arasındaki ilişkiyi tanımlayan kabul edilebilir modeli kurmaktır. (Ulupınar, 2007: 39).

İki Değişkenli Binary Lojistik Regresyon Analizinin temeli olasılık oranına dayanır. Olasılık oranı, bir olayın gerçekleşmesi olasılığı ile söz konusu olayın gerçekleşmemesi olasılığını karşılaştırır. İki Değişkenli Binary Lojistik Regresyon Analizi olasılık oranının doğal logaritması alınarak elde edilir. Olasılık oranının doğal logaritması alınarak elde edilen İki Değişkenli Binary Lojistik Regresyon Modelinin parametrelerini tahmin ederken en yüksek olabilirlik ve Wald istatistikleri yaygın olarak kullanılmaktadır. (Berenson ve Levine, 1996: 837)

En çok olabilirlik yönteminde, gözlenen veri setini elde etme olasılığını en büyük yapacak şekilde bilinmeyen parametreler için değerler üretilir. Bu yöntemi uygulayabilmek için önce olabilirlik fonksiyonu olarak adlandırılan bir fonksiyonun oluşturulması gerekir. Bu fonksiyon, gözlenen verinin olasılığını bilinmeyen parametrelerin bir fonksiyonu olarak belirtir. Bu fonksiyonu en büyük yapan değerler, bilinmeyen parametrelerin en çok olabilirlik kestiricileridir. Yani en çok olabilirlik yönteminde, bir olayın olması olasılığı en çok yapılmaya çalışılır. (Alpar, 2011)

Regresyon katsayılarının önemli olup olmadığını test etmede kullanılacak ikinci test Wald testidir. Wald testine ait test istatistiğinin dağılımı standart normal dağılıma yaklaşır. Her değişken için listedeki standart hatalar kullanılarak Z testi yapılır. Wald testi, örnek hacminin büyük olması durumunda anlam kazanır. (Buse, 1982: 153) Eğitim parametresinin en yüksek ihtimal tahmincisi standart hatasının tahmini değeri ile mukayese edilir.  $0 < 1 = \beta$  iken test istatistiğinin dağılımı standart normal dağılıma uygundur. Bu teste ait test istatistiği,

$$W = \frac{\hat{B}_1}{SE(\hat{B}_1)} \text{ Formülü ile elde edilir.}$$

Lojistik regresyon modelleri zayıf ölçekle ölçülmüş değişkenler arasındaki ilişkinin şeklini ortaya koyan modellerdir. Yapılan birçok çalışmada bağımlı değişken sadece iki sonuca sahiptir. Genellikle üzerinde durulan olayın gerçekleşmesi “1”, gerçekleşmemesi ise “0” ile gösterilir

### 3.Lojistik Regresyon fonksiyonu;

$$\pi(x) = \frac{\exp(B_0 + B_1X)}{1 + \exp(B_0 + B_1X)} \text{ şeklindedir.}$$

Bu ifadenin bir diğer şekli ise;

$$\pi(x) = [1 + \exp(-B_0 - B_1X)]^{-1} \text{ olarak yazılabilir. (Agresti,1996:103)}$$

$E(y/x)$  değeri şartlı ortalama olarak bilinir.

$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = B_0 + B_1X$  eşitliğine logit transformasyon adı verilir, transformasyon değişkeni  $g(x)$ , modeldeki parametreler ile doğrusaldır, süreklidir ve  $-\infty, +\infty$  aralığında değişen değerler alır.  $\pi(x)$  arttıkça  $g(x)$ 'te artar ve eğer  $\pi(x) < 0,5$  ise  $g(x)$  negatif,  $\pi(x) > 0,5$  ise  $g(x)$  pozitif değerler alır. (Hosmer ve Lemeshov,1989:307)

### 4.Lojistik Regresyon Parametrelerinin Önem testi.

Lojistik regresyonda katsayılar kestirildikten sonra regresyon parametrelerinin önem testi lojistik regresyonda gözlenen ve beklenen değerlerin karşılaştırılması log olabilirlik fonksiyonu ile yapılmaktadır.

$D = -2 \ln \left[ \frac{\text{Şu andaki modelin olabilirliği}}{\text{Doymuş modelin olabilirliği}} \right]$  Bu eşitlik “likelihood ratio” olarak adlandırılır. Bağımsız bir değişkenin önemine karar vermek için denklemden o bağımsız değişkenin bulunduğu ve bulunmadığı durumlardaki  $D$  değerleri,  $G$  istatistiği kullanarak karşılaştırılır.  $G$  istatistiği  $p$  serbestlik derecesiyle ki-kare dağılımı gösterecektir.

$G = D$  (Değişkensiz model için)  $-D$  (Değişkenli model için)

$$G = -2 \ln \left[ \frac{\text{Değişkensiz modelin olabilirliği}}{\text{Değişkenli modelin olabilirliği}} \right]$$

Katsayıların önem kontrolü yapıldıktan sonra katsayıların yorumlanması odds oranları kullanılarak yapılmaktadır. (Christensen, R. “Log-Linear Models.”Springer Verlag,1990)

### 5.Lojistik Regresyon Katsayılarının Yorumlanması

Lojistik cevap fonksiyonunda tahmin edilen bir regresyon katsayısının yorumlanması lineer regresyon modelindeki kadar kolay değildir.  $B_1$  Katsayısı yorumlanırken  $X$ 'deki bir birimlik artış için  $\frac{\pi}{1-\pi}$  odds tahmini ile  $\exp(bi)$  çarpılarak elde edilen lojistik cevap fonksiyonundan yararlanır. Doğrusal modelde  $B_1$  bağımsız değişken  $X$ 'deki 1 birim değişimin bağımlı değişkende ne kadarlık bir değişime neden olduğunu gösterirken, lojistik modelde  $X$ 'deki 1 birim değişimin lojitte ne kadar değişime neden olduğunu gösterir.

### 6.Materyal ve metod:

Bu çalışmada, materyal olarak Ayvalık ilçesinde 2016 yılında yapılan anket çalışması sonucu elde edilen veriler kullanılmıştır. Ayvalık turizmini etkileyen risk faktörleri belirlenmiş,

buna uygun olarak bir anket formu geliştirilmiştir. Anket 300 kişiye uygulanmış, elde edilen veriler analizde kullanılmıştır. Bağımlı değişken olarak “Turizm amaçlı Ayvalık’ı tercih eder misiniz?” sorusu olup “Hayır” ve “Evet” şeklindedir ve iki şıklı olup “0” hayır ve “1” evet olacak şekilde kodlanmıştır. Ayvalık turizmini etkileyen faktörlerin hangi düzeyde etkilediğinin bilinmesi gerekmektedir.

### 7. Modelin Oluşturulması

Uygulamada sadece bir tane en iyi model yerine çok sayıda iyi modelin olması olasıdır ve istatistiksel olarak algoritmalarından hiçbirisi en iyi modeli sağlamayı garanti edememektedir. Bu nedenle, olası modellerin geniş bir aralıkta değerlendirilmesi istenir (Alpar, 2011). Model seçim stratejisi çalışmanın amacına bağlı olarak düzenlenebilir. Bazı uygulamalarda, elde edilen pek çok değişkenden hangilerinin sonuç değişkenini etkilediği belirlenmek istenebilir (Alpar, 2011). Buradaki amaç hangi değişken ya da değişkenlerin bağımlı değişkeni etkilediğini araştırmaktır. Çalışmamızda kullanılan yöntem geriye dönük adımsa yöntem olup değişkenlerin çıkarılmasında kullanılan  $\alpha$  anlamlılık düzeyi 0,1’dir. Modeli genel olarak test etmede ise kullanılacak  $\alpha$  anlamlılık düzeyi ise 0,05’dir.

**Tablo 1. Geriye Dönük Eleme (Olabilirlik Oranı) Yöntemi İle Elde Edilen Model ve Adım Ki-Kare Sonuçları**

		Ki-Kare	sd	p
Adım 1	Step	52,577	19	,000
	Block	52,577	19	,000
	Model	52,577	19	,000
Adım 2	Step	-,001	1	,973
	Block	52,575	18	,000
	Model	52,575	18	,000
Adım 14	Step	-1,813	1	,178
	Block	41,868	6	,000
	Model	41,868	6	,000

Katsayıların yorumunu katsayıların sifıra eşit olup olmadığı şeklinde hipotezler kurmak suretiyle test edebiliriz. Model ki-kare ile ilgili hipotez testleri;

H0: Model anlamsızdır. ( $B_1 = B_2 = \dots = B_k = 0$ )

H1: Model anlamlıdır. ( $B_1 \neq B_2 \neq \dots \neq B_k \neq 0$ ) şeklinde kurulur. Tablo 1’de her adımda modelin anlamlılığını test eden Omnibus testleri ve ayrıca Tablo 2’de verilere uygulanan geriye dönük adımsa yöntemin 14. adımında elde edilen Wald istatistiği değerleri görülmektedir. Omnibus testlerinde geçen model ki-kare istatistiği, lojistik regresyon modelini genel olarak test etmektedir. Bu istatistik, sabit terimin dışındaki tüm logit katsayılarının sifıra eşit olup olmadığını sınamaktadır. Model Ki-Kare istatistiği, incelenen modelin parametre sayısı ile yalnız sabit terimli modelin parametreleri arasındaki farka eşit bir serbestlik derecesi ile Ki-Kare dağılımına

uymaktadır. Katsayılar incelendiğinde Wald istatistiğine ait p değerleri ve Tablo 2’de Omnibus Testindeki her adımda elde edilen modellerin parametrelerinin model Ki-kare değerleri anlamlılık düzeyimiz olan 0,05’den küçük olmasından dolayı her iki test istatistiğine göre kestirilen katsayıların anlamlı olduğuna karar veririz. (Model ki-kare: 41,868 sd:6 p=0.000 ve 3. Adımda kestirilen katsayıların Wald değerleri, sd.leri ve p değerleri Tablo 3’de yer almaktadır.) Geriye dönük adımsa yönteminde değişkenin modelden çıkartıldığında modelin her aşamada anlamlılığını test eden Omnibus Testleri Tablo 2’de gösterilmiştir. Değişkenler modelden çıkartılırken Wald İstatistiği değerleri kullanılmıştır. Wald testi, eğim parametresi  $\beta_1$ ’in en çok olabilirlik tahmini ile bu tahminin standart hatasını karşılaştırmaya dayanmaktadır.  $B_1$ ’nin standart hatası, kovaryans matrisindeki köşegen elemanlarının kareköklerinin alınmasıyla elde edilmektedir.

Eğim parametresini gösteren  $H_1=0$  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 olduğundan, Wald istatistiği aşağıdaki biçimde de ifade edilebilir (Alpar, 2011: 626).

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

Geriyeye dönük adımsa yönteminin başlangıç aşamasında, tüm değişkenler modele alınır ve bağımlı değişkenle ilişkisi en az olan değişkenden başlanarak değişkenler modelden çıkartılır. Geriyeye dönük adımsa eleme yönteminde değişkenlerin modelden çıkarılmasında kullanılacak hipotez ve karşı hipotez aşağıda sunulmuştur. Adımsal ki-kare ile ilgili hipotez testleri; ,

$H_0$ :Modelden Çıkarılan Değişkenin Modele Katkısı Yoktur.

$H_1$ :Modelden Çıkarılan Değişkenin Modele Katkısı Vardır.’ şeklinde kurulur.

Değişkenlerin modele alınma sıraları bloklar halinde tanımlandığından model ve blok ki-kare değerleri aynıdır. Tüm değişkenlerin modele alındığı ilk adımda, adım, blok ve model testi sonuçları aynıdır. Bu sonuç, her üç testte modelin anlamlı olduğunu göstermektedir. ( $\chi^2=52,540$ ;  $p=0,000$ )

Birinci adımda, modele katkısı en az olan yaş değişkeni modelden çıkartıldığında model ki-kare 52,575 olarak elde edilmektedir. Adım ki-kare değerleri ( $\chi^2 = 52,575-52,577=0,001$ );  $p=0,973$ )değişkeninin çıkartılmasının modelde anlamlı bir değişime neden olmadığını belirtmektedir. Diğer adımlarda da benzer işlemlere devam edilerek en son model elde edilir.

İkinci ve diğer aşamalarda modele katkısı en az olan değişken modelden çıkartıldığında, on dördüncü aşamada sonuç modeli elde edilmiştir. Değişkenlerin modelden çıkarılmasında %95 Güven aralığı bir değerini içeren ve Wald değeri ikiden küçük değerli değişken çıkarılmak suretiyle devam edilmiştir. Elde edilen son model anlamlıdır. Katsayıların yorumunu katsayıların sıfıra eşit olup olmadığı şeklinde hipotezler kurulduğunda ise müteakip cümlelerde olduğu gibi ifade edebiliriz. Model ki-kare ile ilgili hipotez testleri; ,

$H_0$ : Model anlamsızdır.( $B_1 = B_2 = \dots = B_k = 0$ )

$H_1$ : Model anlamlıdır.( $B_1 \neq B_2 \neq \dots \neq B_k \neq 0$ ) şeklinde kurulur.

Tablo 2’de her adımda modelin anlamlılığını test eden Omnibus testleri ve ayrıca Tablo3’de verilere uygulanan geriye dönük adımsa yöntemin 3. adımında elde edilen Wald istatistiği değerleri görülmektedir. Omnibus testlerinde geçen model ki-kare istatistiği, lojistik regresyon

modelini genel olarak test etmektedir. Bağımsız değişkenlerden hiçbirinin bağımlı üstünlük oranıyla anlamlı doğrusal bir ilişki göstermediğini ileri süren sıfır hipotezini test etmektedir. Bir başka deyişle bu istatistik, sabit terimin dışındaki tüm logit katsayılarının sıfıra eşit olup olmadığını sınıamaktadır. Model Ki-Kare istatistiği, incelenen modelin parametre sayısı ile yalnız sabit terimli modelin parametreleri arasındaki farka eşit bir serbestlik derecesi ile Ki-Kare dağılımına uymaktadır. Katsayılar incelendiğinde Wald istatistiğine ait p değerleri ve Tablo 2’de Omnibus Testindeki her adımda elde edilen modellerin parametrelerinin model Ki-kare değerleri anlamlılık düzeyimiz olan 0,05’den küçük olmasından dolayı her iki test istatistiğine göre kestirilen katsayıların anlamlı olduğuna karar veririz. (Model ki-kare: 46,378 sd:8 p=0.000 ve 3 Adımda kestirilen katsayıların Wald değerleri, sd.leri ve p değerleri Tablo 2’de yer almaktadır.). Başlangıç ve sonuç modeli içeren tablo aşağıdaki gibidir.

**Tablo 2. Geriye Dönük Eleme Yöntemi Sonuçları**

	$b_j$	$S(b_j)$	Wald	sd	P	Odds Oranı Exp( $b_j$ )	%95 Güven Aralığı		
							Alt	Üst	
<b>Adım 1</b>	Cinsiyet	1,511	,778	3,767	1	,052	4,530	,985	20,082
	Medeni Hal	-,877	,943	,866	1	,352	,416	,066	2,639
	Yaş	,005	,385	,000	1	,990	1,005	,472	2,139
	Aile Reis	-,059	,192	,095	1	,757	,942	,646	1,374
	Çocuk Say	,102	,342	,090	1	,765	1,108	,567	2,165
	Eş Çalışma	-2,115	,843	6,291	1	,012	,121	,023	,630
	Aile Reis Eğit	-,345	,411	,704	1	,401	,708	,317	1,585
	Aylık Gelir	,483	,517	,870	1	,351	1,620	,588	4,467
	Yıllık Tatil Harcamaları	,395	,456	,748	1	,387	1,484	,607	3,629
	Konaklama Süresi	-,620	,342	3,293	1	,070	,538	,275	1,051
	Hangi Tür Konaklama Tercih	,669	,281	5,673	1	,017	1,953	1,126	3,387
	Ayvalıka Hangi Yolla Geldiniz	-,425	,569	,558	1	,455	,654	,215	1,993
	Ayvalıkta Konaklama Fiyatı	,568	,639	,789	1	,375	1,764	,504	6,178
	Ayvalık Tercih Nedenleri	,130	,189	,470	1	,493	1,139	,786	1,650
	Ayvalık Tercih Etmede Bilgi Kaynağı	,720	,349	4,259	1	,039	2,055	1,037	4,074
	Ayvalık Markamıdır	,907	,379	5,729	1	,017	2,476	1,178	5,201
	Ayvalığı Tercih Markamı	,809	,929	,757	1	,384	2,245	,363	13,88
	Ayvalık Ulaşım	,333	,423	,621	1	,431	1,395	,609	3,197
	Konaklama Tercihiniz	-,313	,152	4,252	1	,039	,731	,543	,985
	Constant	-,867	4,256	,041	1	,839	,420		
<b>Adım 14</b>	Cinsiyet	1,035	,613	2,854	1	,091	2,816	,847	9,362
	Eş Çalışma	-1,561	,643	5,892	1	,015	,210	,059	,740
	Hangi Tür Konaklama Tercih	,543	,221	6,012	1	,014	1,721	1,115	2,655
	Ayvalık Tercih Etmede Bilgi Kaynağı	,795	,278	8,216	1	,004	2,215	1,286	3,816
	Ayvalık Markamıdır	,964	,313	9,458	1	,002	2,622	1,419	4,848
	Konaklama Tercihiniz	-,290	,108	7,245	1	,007	,749	,606	,924
	Constant	,028	1,681	,000	1	,987	1,029		



Son aşamada, modelde cinsiyet, Eş Çalışma Durumu, Konaklama Tercihi, Ayvalığı Tercih Etmede Bilgi Kaynağı, Ayvalık Markamıdır ve Konaklama Tercihiniz değişkenleri kalmıştır. Sonuç model için lojistik regresyon modeli aşağıdaki gibi yazılacaktır.

$$\pi(x) P_1(\text{Tercih etme} / x) = \frac{1}{1 + e^{-(0,028 + 1,035 \text{ Cins.} + 0,543 \text{ Kon.Ter.} + 0,795 \text{ Bil.Kay.} + 0,964 \text{ Ay.Marka.} - 0,290 \text{ Kon.Türü})}}$$

Geriye yönelik adımsa yöntemle elde edilen sonuç modelde altı değişkenin etkisi anlamlı bulunmuştur. Bu değişkenler ile oluşturulan model için SPSS yazılımından elde edilen model ki-kare ve  $R^2$  değerleri aşağıda verilmiştir.

**Tablo 3. Modelin Özeti**

Adım	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	78,041	,161	,455
2	78,042	,161	,455
3	78,136	,160	,455
4	78,243	,160	,454
5	78,836	,159	,449
6	79,342	,157	,445
7	80,119	,155	,439
8	80,735	,153	,434
9	81,465	,151	,428
10	82,314	,149	,421
11	83,236	,146	,414
12	84,480	,143	,404
13	86,936	,135	,384
14	88,749	,130	,369

Doğrusal Regresyon Analizindeki bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi ölçen  $R^2$  istatistiğine benzeyen bir istatistik lojistik regresyon analizinde bulunmamaktadır. Bu nedenle, regresyon analizindeki  $R^2$  değeri ile lojistik regresyon analizindeki  $R^2$  değerini karşılaştırmak uygun değildir. Bununla birlikte literatürde lojistik regresyon analizi için birkaç  $R^2$  istatistiğine yer verilmektedir.

McFaden  $R^2$ , Cox-Snell ve Nagelkerke  $R^2$  istatistikleri en sık kullanılan istatistiklerdir. Bu istatistikler genel olarak küçük çıkma eğiliminde olduklarından farklı modellerin performansını değerlendirmek için kullanılmaları önerilmektedir. Bu nedenle bazı yazarlar sonuçlar sunulurken  $R^2$  istatistiklerini verilmesini önermemektedir. (Alpar, 2011: 643) Uygulamamızda Tablo 4'de, Cox-Snell  $R^2$  istatistiği 14. adımda (son modelde) yaklaşık olarak % 13 olarak elde edilmiştir. Bu istatistik bağımlı değişken ile bağımsız değişkenler arasında yaklaşık % 13'lük bir ilişkinin olduğunu göstermektedir. (Kalaycı, 2010: 293) Nagelkerke  $R^2$  istatistiği ise Cox ve Snell  $R^2$  istatistiğinin 0-1 aralığında değerler almasını sağlamak amacıyla geliştirilmiştir (Kalaycı, 2010: 293). Nagelkerke  $R^2$  istatistiği Tablo 4'de son adımda yaklaşık olarak % 36,9 olarak elde edilmiştir. Bu istatistik bağımlı değişken ile bağımsız değişkenler arasında yaklaşık % 36,9'luk

bir ilişkinin olduğunu göstermektedir. Çeşitli kaynaklarda,  $R^2$  istatistiklerinin 0,20 ile 0,40 arasında çıkması yeterli olduğu açıklanmaktadır. Bu nedenle modelin uygunluğunun değerlendirmesinde elde edilecek olan değerler bu değerler arasında olmasını bekleriz. Sonuç olarak, elde edilen değerler bu değerler arasında olması nedeniyle kurulan modelin uygun olduğunu söyleyebiliriz. (Alpar, 2011: 666) Ayrıca  $-2\text{LogL}$ 'nin her adımda model için en faydasız değişkenin çıkartılmasına karşılık artması modelin ayrı bir bakış açısından anlamlı olduğunu gösterir.

Modelin uyum iyiliğini değerlendirmek için Hosmer-Lemeshow testide uygulanır. Bu istatistiğe yönelik hipotezimiz ise,

**Tablo4. Hosmer and Lemeshow Test**

Adım	Ki-Kare	sd	p
1	1,352	8	,995
14	3,349	8	,911

H0: Parametreler belirleyicilik açısından iyi bir ayrımcılığa sahiptir.

H1: Parametreler belirleyicilik açısından iyi bir ayrımcılığa sahip değildir.' şeklindedir.

Model  $Ki\text{-Kare} = 3,349 < \chi^2(0,1;8) = 13,362$  olduğundan dolayı H0 kabul edilmiş ve Lojistik Regresyon Modeli'nin turizm için Ayvalığı tercih edip etmeyecekleri gruplara ayırmada yeterli bir model olduğu bulunmuştur. Buradan, model uyumunun oldukça iyi olduğu ve parametreler belirleyicilik açısından iyi bir ayrımcılığa sahip olduğu sonucuna varılabilir.

Modelin uyumunu değerlendirmek için bir diğer yaklaşımda doğru sınıflandırma oranlarının incelenmesidir. Bu amaçla bireylerin gerçekte oldukları sınıflar ile lojistik regresyon modeli yardımıyla kestirilen sınıfların çapraz tablosu elde edilir.

**Tablo 5. Çoklu Lojistik Regresyon Modeli İçin Doğru Sınıflama Tablosu**

Gözlenen		Kestirilen		
		Turizm İçin Ayvalık'ı Tercih Edermisiniz		Doğru Sınıflandırma Oranı
		Hayır	Evet	
Turizm İçin Ayvalığı Tercih Edermisiniz	Hayır	3	14	17,6 (3/17)
	Evet	2	281	99,3 (281/283)
Toplam				94,7 (284/300)

Modelin bireyleri doğru sınıflandırma oranı %94,7dir. Turizm için Ayvalığı tercih edeceklerden hayır diyenlerin %17,6 sı hayır, evetlerin %99,3 evet (duyarlılık) şeklinde sınıflanmıştır.

## 9.SONUÇ

Çalışmamızda, bağımlı değişkenin iki düzeyli, bağımsız değişkenler arasında da kategorik değişkenin olduğu durumlarda, gözlemlerin gruplara atanmasında bir ayırimsama modeli olarak kullanılan ve son yıllarda diskrimant analizine alternatif olarak geniş bir uygulama alanı bulan Lojistik Regresyon Analizi incelenmiştir. Çok değişik alanlarda uygulama şansına sahip olan ve anket türü verilerin analizinde çok kullanılan lojistik regresyon konusu ile ilgili bu çalışmada amaç, turizm amaçlı Ayvalığı tercih etmede etkili olan faktörleri ortaya koymak amacıyla 300 anketlik araştırma yapılmıştır. Bağımlı değişkenin iki düzeyli olması ile ilgili bir uygulama yapılmıştır. Turizm için Ayvalığı tercih etmede etkili olacak faktörlerin belirlenmesi amacıyla yapılan bu uygulamada bağımlı değişken iki düzeyli olup tercih etmem 0, tercih ederim 1 olarak alınmıştır. Ayvalığı turizm amacıyla tercih etmede etkili olacağı düşünülen 18 değişken belirlenmiş ve bir lojistik regresyon modeli oluşturulmuştur. Ele alınan bütün bağımsız değişkenlerin anlamlı olup olmadıklarını sınamak amacıyla adımsa teknikler kullanılmıştır. Geriye Dönük Eleme Yöntemi ile yapılan uygulama sonucunda, ele alınan değişkenlerden önemli olanlarından altı tanesinin modelde yer aldığı görülmüştür. Modele alınan değişkenlerin anlamlı olup olmadığına bir göstergesi olarak kabul edilen Wald ölçütüne bakıldığında, modele alınan değişkenler için Wald değerinin 2'den büyük olduğu görülmektedir.

Modelin genel olarak doğru sınıflandırma oranı %94,7 'dir. Bu sonuçlara göre kurulan modelin geçerli ve kullanılabilir bir model olduğu görülmektedir. Turizm için Ayvalığı tercih etmede etkili olabilecek 18 risk faktörü olan değişken incelenmiştir. Tablo 2'de görüldüğü gibi, bu değişkenlerden sadece Cinsiyet, Eş Çalışma Durumu, Ayvalık Markamıdır, Konaklama Şekli, Konaklama Türü ve Ayvalığı Tercih Etmede Bilgi Kaynağı değişkenleri ile en uygun model oluşturulmuştur. Tahmin edilen lojistik regresyon modelinden çıkartılacak yorumlar, modeldeki risk faktörleri için tahmin edilen odds oranları ile yapılır. Risk faktörleri için odds oranları, tahmin edilen katsayıların üsteli alınarak elde edilir. Tablo 2'de Exp(B) sütununda odds oranları verilmiştir. Elde edilen odds oranı değeri 1'den küçük olduğunda risk azalmakta, 1'e eşit olduğunda risk aynı olmakta ve 1'den büyük olduğunda ise risk artmaktadır. CINS değişkenine ait odds oranı 2,816 olarak bulunmuştur. Bu katsayısı, Ayvalığı tercih etmede bayan birine göre erkek bireyin tercih etme riski 2,816 yüksek olduğunu gösterir. Aynı şekilde Eş Çalışma durumuna göre tek çalışan bireye göre her ikiside çalışıyor olanların tercih etme 0,21 kat daha fazla olduğunu gösterir. Ayvalık marka diyenlerin tercih etme riski değildir diyenlere karşılık 2,622 kat fazla olduğunu gösterir.

Yoğun bir rekabetin yaşandığı hizmet sektörünün içinde yer alan turizm işletmelerinin rekabette ileride olabilmesi için tüketici tercihlerini etkileyen faktörlerden hangisinin daha çok önemli olduğunu bilmeleri işletmelere rekabet avantajı sağlayacaktır.

**EK 1.** Ayvalık turizmini etkileyen risk faktörleri olabilecek değişkenler aşağıda verilmiştir.

- 1: Cinsiyet : Kesikli Değişkendir : 1:Erkek 2:Bayan
- 2: Medeni Hal : Kesikli Değişkendir : 1:Evli 2:Bekar
- 3: Yaş : Sürekli Değişkendir.
- 4: Aile Reisinin Mesleği : Kesikli Değişkendir.
- 5:Ailedeki Çocuk Sayısı : Kesikli Değişkendir.

- 6: Eşlerin Çalışma Durumu : Kesikli Değişkendir : 1:Her ikisi de çalışıyor. 2:Birdisi çalışıyor.
- 7: Aile Reisinin Eğitimi : Kesikli Değişkendir.
- 8: Aylık Gelir : Kesikli Değişkendir.
- 9: Yıllık Tatil Harcaması : Kesikli Değişkendir.
- 10: Konaklama Süresi : Kesikli Değişkendir.
- 11: Hangi tür konaklamayı tercih edersiniz? : Kesikli Değişkendir.
- 12: Ayvalık'a hangi yolla geldiniz? : Kesikli Değişkendir : 1:Bireysel 2:Seyahat Acentası 3:Diğer
- 13: Ayvalık'ta Konaklama Fiyatları : Kesikli Değişkendir : 1:Ucuz 2:Normal 3:Pahalı
- 14: Ayvalık'ı tercih nedeniniz nedir? : Kesikli Değişkendir.
- 15: Ayvalık'ta tatil yapma tercihinize etkili olan bilgi kaynağı nedir? : Kesikli Değişkendir.
- 16: Ayvalık turizmde marka mıdır? : Kesikli Değişkendir : 0:Hayır 1:Evet
- 17: Ayvalık'ı tercih etmeniz bir marka olmasından mıdır? : Kesikli Değişkendir : 0:Hayır 1:Evet
- 18: Ayvalık'a ulaşımında hangi araçları kullanmaktasınız? : Kesikli Değişkendir.
- 19: Hangi tür konaklamayı tercih ediyorsunuz? : Kesikli Değişkendir.

## KAYNAKÇA

- Akaya, Ş., M.V. Pazarlıoğlu (1998), Ekonometri . Erkan Matbaacılık, İzmir,
- Albayrak, A.S. (2006). Uygulamalı çok değişkenli İstatistik Teknikleri, 1. Baskı, Asil Yayıncılık, Ankara.
- Agresti, A. (1990) Analysis of Ordinal Categorical Data, John Wiley and Sons, New York
- Akgül, A. ve O. Çevik, (2003), İstatistiksel Analiz Teknikleri-SPSS'te İşletme Yönetimi Uygulamaları, Emek Ofset Ltd. Şti Ankara.
- Aktaş, C. ve Yılmaz V. (2001), Eskişehir de Lpg Kullanan Özel Araç Sürücülerinin Sınıflandırılmasında Lojistik Regresyon Analizi, İstanbul Kentiçi Ulaşım Sempozyumu, İstanbul, 251-256.
- Alpar Reha. (2011), Uygulamalı Çok Değişkenli İstatistiksel Yöntemler, 3. Baskı, Ankara, Detay Yayınları, S.616
- Anderson, J. A. (1979), Multivariate Logistic Compounds Biometrika, 66: 17–191. Anderson, J. A. (1983), Robust İ nference Using Logistic Models Bulletin of international Statistical Institute, 48: 35-53.
- Berenson, M. L. ve Levine, D. M. (1996). Basic Business Statistics: Concepts and Applications, Sixth Edition, Prentice-Hall International, 837-838.
- Başarır, G. (1990), Çok Değişkenli Verilerde Ayrısama Sorunu ve Lojistik Regresyon Analizi . (Uygulamalı istatistik doktora tezi.) H.Ü., 1-36, Ankara.

- Bircan, H.,(2004),”Lojistik Regresyon Analizi : Tıp Verileri Üzerine Bir Uygulama,”Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,2,185-208.
- Cox, D. R. – Snell, E. S.1989, Analysis of Binary Data, London.
- Coşkun, S.ve diğerleri (2004), “Lojistik Regresyon Analizinin İncelenmesi ve Diş Hekimliğinde Bir Uygulaması,” Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi, Cilt:7, Sayı:1,42-50.
- Çolak, E. ve Özdamar, K. (2004), Ölümle Sonuçlanan Trafik Kazalarında Risk Faktörlerinin Koşullu ve Sınırlandırılm Lojistik Regresyon Yöntemleri ile incelenmesi, OGU T p Fak. Dergisi, 26, 1, 7-14.
- Büyükışler, Didar vd. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi Y.2014, C.19, S.1, s.125-138
- Duffy, D.E. (1990). On Continuity-corrected Residuals in Logistic Regression Biometrika, 77: 287-293.
- Elhan, A.H. (1997), Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması. (Biyostatistik Yüksek Lisans Tezi, A.Ü.,4-29,Ankara.
- Hsu, J.S., Leonard, T. (1995), Hierarchical Bayesian Semiparametric Procedures for Logistic Regression . Biometrika, 84: 85-93.
- Hosmer,D.W.-Lemeshow,S. (1980) Goodness of fit Tests for the multiple Logistic Regression Model,Communications in Statistics,Seri A9.1043-1069.
- Hosmer, D.W.-Lemeshow, S. (1989) Applied Logistic Regression,John Wiley&Sons.
- Johnson, W. (1985), Influence Measures for Logistic Regression, Another Point of View, Biometrika, 72, 1, 59–65 .
- Kalaycı, Ş. (2010). SPSS Uygulamalı Çok Değişkenli İstatistiksel Teknikleri, 5.Baskı, Asil Yayın Dağıtım, Ankara.
- Kloiber, L.L. Winn, N.J., Shaffer, S.G., Hassanein, R.S. (1996), Late Hyponatremia in very Low Birth Weight Infants: Incidence and Associated Risk Factors . Journal of the American Dietetic Association, 96: 880-884..
- Lee, C.T. (1984), Logistic Models for Cross-over Designs. Biometrika, 71: 216-217.
- Oktay E., Ü.Özen ve Ö. Alkan (2011).İnternet Bankacılığını Tercih Etkili Olan Faktörlerin Analizi. Atatürk Üniversitesi Örneği, Proceeding of The 12 th International Symposium on Econometrics Statistics and Operations Research, p.72-77, 26-29 May 2011, Denizli.
- Oktay, E. Orçanlı, Kenan. Atatürk Üniversitesinde İnternet Bankacılığının Kullanımını Etkileyen Faktörlerin Belirlenmesi ,Uşak Üniversitesi Sosyal Bilimler Dergisi 2014, 7/2
- Özdamar.K.,(2002),Paket Programlar ile İstatistiksel Veri Analizi I, Kaan Kitabevi, Eskişehir.
- Ünsal, A. ve Güler, H. (2005), Türk Bankacılık Sektörünün Lojistik Regresyon Ve Diskriminant Analizi ile incelenmesi , VII. Ulusal Ekonometri ve İstatistik Sempozyumu,İstanbul Ün.
- Ulupınar, S. D. (2007). 2001 Kriz Dönemi, Öncesi ve Sonrasında Türk Ticari Bankalarının Karlılıklarının Lojistik Regresyon Analizi ile İncelenmesi, İstatistik Bilim Dalı Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul.

Vupa, Ö. ve Çelikođlu, C. (2006), Model Building in Logistic Regression Models About Lung Cancer Data.,Anadolu Ü. Bilim ve Teknoloji Dergisi, cilt: 7, 1, 127-141.

Tatlıdil, H. (1992),Uygulamalı Çok Deđişkenli İstatistiksel Analiz,Engin Yayınları,Ankara.

Tatlıdil, H., Başarrı, G. ve V. Hökmen (1990), Ülkelerin Sosyo Ekonomik Gelişmişliklerine Göre Kümelenmesine ve Sıralanmasına Yeni Yaklaşımlar, Planlama Dergisi, 26, 103-120.

<https://tr.scribd.com/doc/47455636/Chapter-24-Logistic-regression>