

**YAŞAM MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN  
BELİRLENMESİNDE GENELLEŞTİRİLMİŞ TOPLAMSAL  
MODELLERİN KULLANIMI****THE USAGE OF GENERALIZED ADDITIVE MODELS FOR  
DETERMINING FACTORS EFFECTED ON LIFE SATISFACTION****Nurettin SAVAŞ<sup>1\*</sup> ve Mehmet Ali CENGİZ<sup>2</sup>**<sup>1</sup>*Erzincan Üniversitesi, Fen-Edebiyat Fakültesi, Matematik Bölümü, ERZİNCAN*<sup>2</sup>*Ondokuz Mayıs Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü,  
SAMSUN***Geliş Tarihi:** 25 Mart 2009**Kabul Edilme Tarihi:** 6 Mayıs 2009**ÖZET**

Biyolojik, ekonomik veya teknolojik verilerin değerlendirilmesinde özellikler arası ilişkiler incelenirken en çok kullanılan istatistiksel tekniklerden biri de çoklu regresyon analizidir. Varsayımları sağlandığında oldukça güçlü bir parametrik değerlendirme yöntemidir. Bu varsayımlardan biri de açıklayıcı değişkenlerle bağımlı değişkenler arasındaki ilişkilerin tanımlı bir matematiksel fonksiyon olmasıdır (doğrusal, polinomial veya üstel gibi). Bununla beraber birçok uygulamada böyle bir matematiksel fonksiyonun tanımlanması çok zor hatta imkânsız olabilmektedir. Genelleştirilmiş Toplamsal Modeller (GAM), böyle klasik bir fonksiyonu tanımlama yerine var olan ilişkileri belirleyen modele parametrik olmayan bir düzleştirici terim ekleyerek bu varsayımın etkisini ortadan kaldırmaya çalışır. Bu çalışmada yaşam memnuniyet derecelerine ilişkin bir zeka testi merkezinden elde edilen anket verilerine GAM uygulanmıştır.

**Anahtar Kelimeler:** Doğrusal Modeller, Toplamsal Modeller, Genelleştirilmiş Doğrusal Modeller, Genelleştirilmiş Toplamsal Modeller.

**ABSTRACT**

Multiple regression analysis is one of the most widely used statistical techniques to evaluate the relationships among one response and several predictors obtained from biological, economical and technical data sets. It is a powerful tool when the assumptions are met, including that the relationships between the predictors and the response are well described with a defined mathematical function (e.g., straight-line, polynomial, or exponential). In many applications, however, the reliance on a defined mathematical function is limited. Many phenomena do not have a

\* Sorumlu yazar: [nsavas@erzincan.edu.tr](mailto:nsavas@erzincan.edu.tr)

relationship which can be easily defined. Generalized Additive Models (GAM) enable us to relax this assumption by replacing a defined function with a non-parametric smoother to uncover existing relationships. Smoothing is a method which will highlight a trend by separating it from variability due to noise. In this study, GAMs are applied to real IQ test data regarding the life satisfaction level obtained from a survey.

**Key Words:** Linear Models, Additive Models, Generalized Linear Models, Generalized Additive Models.

## 1.GİRİŞ

Veri analizinde kullanılan ve en çok bilinen istatistik değerlendirme araçlarından biri, doğrusal regresyon modelidir. En basit bir durum, bir bağımlı değişken (olan  $Y$ ) ve bir açıklayıcı değişken ( $X$ )'in  $n$  tane ölçüme sahip olmasıdır. Bu amaç doğrultusunda  $Y$ 'nin ortalaması  $X$ 'in bir doğrusal fonksiyonu olarak ifade edilebilir.

Doğrusal ve doğrusal olmayan regresyon modellerinde istatistiksel analiz yöntemleri bağımlı değişkenin normal dağıldığı varsayımına dayanmaktadır. Bağımlı değişkenin sürekli olmadığı durumları analiz etmek için de genelleştirilmiş modellerde bulunmaktadır. Örneğin; bir hastanın tedaviye verdiği cevap (sonuç) değişkeni 1 ve 0 değerlerini alabilir. Bir başka durumda belli bir zaman aralığında bir olayın kaç kez tekrarlandığı ilgi konusu olabilir. Bu duruma bir günde meydana gelen ölümle sonuçlanan trafik veya iş kazaları, bir yıl ya da bir günde meydana gelen deprem sayısı örnek olarak verilebilir. Bu durumlarda bağımlı değişken sürekli değildir. Ayrıca değişkenleri sürekli olup da normal dağılım göstermeyen verilerde söz konusu olabilir. Bu tür verilerin analizine imkan verecek geliştirilmiş modeller Genelleştirilmiş Doğrusal Modellerdir (GLM).

GLM, ilk kez Nelder ve Wedderburn (1972) tarafından ileri sürülmüş ve çok geniş uygulama alanlarında kullanılmıştır. Bu alanlardaki ilk detaylı kitaplar McCullagh ve Nelder (1989), Aitken et al. (1989) ve Dobson (1990) tarafından yazılmıştır ve daha sonraki yıllarda sayısız çalışmalar yapılmıştır. Cengiz (1997) genelleştirilmiş doğrusal modellerin genel bir özetini vermektedir.

Modeldeki açıklayıcı değişkenlerin sayısı çok olduğunda parametrik olmayan metot iyi performans göstermez. Açıklayıcı değişkenin çokluğu tahminlerin varyansını artırır. Açıklayıcı değişken sayısı arttırıldığında boyut artacağından varyansın hızlı bir şekilde artması problemi “boyut problemi” olarak bilinir. Bir başka problem ilişkinin yorumlanmasında ortaya çıkar. Bu zorlukların aşılması noktasında Stone (1985) toplamsal modelleri önermektedir. Bu modeller çok değişkenli regresyon fonksiyonları için bir toplamsallık sağlamaktadır. Toplamsal yaklaşımın iki türlü faydası vardır. Birincisi; toplamsal terimin her biri bir tek değişkenli düzleştirici kullanılarak tahmin edildiği için “boyut problemi” ortaya çıkmaz. İkincisi; her bir terimin tahmini bağımlı değişkenin bağımsız değişkenlerle nasıl değiştiğinin açıklanmasıdır.

Toplamsal Modellerin her türlü bağımlı değişken için genelleştirilmiş hali Hastie & Tibirshani (1990) ve Hastie (1991) tarafından önerilmektedir. Bu durum Genelleştirilmiş Toplamsal Modeller (GAM) olarak adlandırılır. Bu modeller, bilinen GLM'nin Toplamsal Modellere modifiye edilmiş halidir. Bağımlı değişkenin ortalamasının bir toplamsal tahmin ediciye bağlı olduğu bu modellerde doğrusal olmayan bir link fonksiyonu kullanılır. GLM'de olduğu gibi GAM, bağımlı değişkenin dağılımının üstel dağılımlar ailesinden olmasını ister. GAM'la GLM arasındaki tek fark GAM'ın doğrusal tahmin edici olarak bilinmeyen düzleştirici fonksiyonları kullanmasıdır. Normal dağılımlı bağımlı değişken için Toplamsal Modellere, parametrik olmayan lojistik modeller ve parametrik olmayan log-doğrusal modeller örnek olarak verilebilir.

## 2. GENELLEŞTİRİLMİŞ TOPLAMSAL MODELLER

Regresyon modelleri farklı değişkenlerin etkileşimli davranışını anlamak için tahmin etme, tahmin, sınıflandırma kuralları ve veri analitik araçlarını sağlayarak pek çok uygulama alanında kullanılmasında önemli bir role sahiptir. Dikkat edilecek olursa basit olmasına rağmen doğrusal modeller, sık sık gerçek yaşam etkilerinde genellikle doğrusal olmadığı için başarısızlığa neden olur. Daha esnek istatistiksel modeller doğrusal olmayan regresyon etkilerini tanımlamak için kullanılabilir. Bu modeller; genelleştirilmiş toplamsal modeller olarak adlandırılırlar. Örneğin tıp

araştırmalarında sıkça kullanılan istatistiksel binary veriler için bir lojistik regresyon modelidir. Burada binary verilerin ortalaması olan  $\mu = P(y=1)$ 'nin tahmin edicilerle bir regresyon modeli kullanarak ortaya konulan ilişkisi ve logit bağlantı fonksiyonu,

$$\log \left[ \frac{\mu}{1-\mu} \right] = \alpha + \beta_1 x_1 + \dots + \beta_p x_p \quad (1)$$

dir.

Toplamsal doğrusal regresyon modeli daha genel bir fonksiyonel form olan bu model,

$$\log \left[ \frac{\mu}{1-\mu} \right] = \alpha + f_1(x_1) + \dots + f_p(x_p) \quad (2)$$

her bir doğrusal terimle yer değişir. Burada her bir  $f_j$ , parametrik olmayan bir fonksiyondur.  $f_j$  fonksiyonları için parametrik olmayan formun daha esnek bir model oluştururken toplamsallık kalıcıdır ve daha önceki durumlarda olduğu gibi modeli yorumlamayı sağlar.

Toplamsal lojistik regresyon modeli, geliştirilmiş toplamsal modellerin bir örneğidir. Genellikle bir  $y$  bağımlı değişkenin ortalaması  $\mu$ ,  $g$  bağlantı fonksiyonu ile tahmin edicilerin toplamsal fonksiyonu ile ilişkilidir.  $g$  link fonksiyonu;

$$g = \alpha + f_1(x_1) + \dots + f_p(x_p) \quad (3)$$

dir.

Diğer klasik bağlantı fonksiyonları ve ilişkili geliştirilmiş toplamsal modeller aşağıda verilmiştir.

- $g(\mu) = \mu = \sum_j f_j(x_j)$ , sıradan doğrusal modelin toplamsal genişletilmesi
- $g(\mu) = \log(\mu) = \sum_j f_j(x_j)$ , sayısal veriler (Poisson) için bir logaritmik toplamsal modeldir.

Bu üç ifade gamma ve negatif binom dağılımını içeren üssel ailesinin örneklem modellerinden türetilebilir. Bu aileler

genelleştirilmiş toplamsal modellere uygun bir şekilde genişletilmiş genelleştirilmiş doğrusal modeller diye iyi bilinen sınıfı türetirler (McCullagh ve Nelder, 1989).

$f_j$  fonksiyonları bir algoritmanın temel yapı bloklarını oluşturan bir serpilme diyagramını kullanarak daha esnek bir şekilde tahmin edilir. Tahmin edilen  $\hat{f}_j(x_j)$  fonksiyonu  $x_j$ 'nin etkisinde muhtemel doğrusal olmayan durumları ortaya çıkarabilir. Amaç genellikle doğrusal ve bazı değişkenlerin kesikli faktörleri olduğunda diğer parametrik oluşumlarla da kolaylıkla karşılaştırılabilir olduğunu göstermektir. Doğrusal olmayan terimler ana etkilerle sınıflandırılmaz. Ya iki veya daha fazla değişkende doğrusal olmayan bileşenlerine sahip olunabilir veya kesikli bir faktörün her bir seviyesi için eğrilerin ayrılabilir durumda olmasıdır. Sonuçta aşağıdakilerden herhangi biri olarak sınıflandırılabilir.

- $g(\mu) = X' \beta + \alpha_k + f(z)$ : Yarı parametrik bir model; burada  $X$  doğrusal olarak modellenecek tahmin edicilerin bir faktörü,  $\alpha_k$ , kesikli bir faktörün  $k$ -ncı derecesi için etki ve tahmin edici  $Z$ 'nin etkisi parametrik olmayan şekilde modellenir.
- $g(\mu) = f(x) + g_k(z)$ : Burada  $k$  bir faktörün seviyesi ve sonuçta  $k$  ve  $Z$ 'nin bir etkisi için bir etkileşim ifadesi yaratır.
- $g(\mu) = f(x) + g(z, w)$ : Burada  $g$  iki değişkendeki parametrik olmayan bir fonksiyondur.

Genel doğrusal olmayan modeller aşağıdaki gibi gösterilir.

$$Y = f(X_1, X_2, \dots, X_p) + \varepsilon \quad (4)$$

Burada  $X_1, X_2, \dots, X_p$ 'ler açıklayıcı değişken,  $\varepsilon$  rasgele hatalar ve  $f: R^p \rightarrow R$  bir keyfi olarak seçilen düzgün bir fonksiyondur. Ancak doğrusal olmayan model  $p > 2$  için genellikle uyumsuzluk gösterir. Ayrıca pek çok açıklayıcı değişken olduğunda genel doğrusal olmayan regresyonu yorumlamak kolay değildir. Bir toplamsal model doğrusal modellerinin özel bir durumu olan çoklu regresyon modellerinin bir genelleştirilmesini temsil eder. Toplamsal

modelde;  $f_j$ 'yi tahmin eden  $X_j$ 'nin keyfi olarak seçilen fonksiyonunun terimleri olan  $f_j(X_j)$  doğrusal denklemdeki  $\beta_j X_j$ 'nin yer değiştirmesidir. Diğer bir ifadeyle her bir açıklayıcı değişken için tek bir katsayı yerine, spesifik olmayan bir fonksiyonun her bir tahmin edicisi için tahmin edilebilir. Bu amaçla bağımlı değişken değerlerinin daha iyi bir tahminini elde etmek için toplamsal model kullanılabilir. Bağımlı değişken  $Y$ , bağımsız değişkenleri  $X_1, X_2, \dots, X_p$ 'ye kadar olan bir toplamsal model,

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \varepsilon \quad (5)$$

formuna sahiptir. Burada hatalar  $X_j$ 'den bağımsız,  $E(\varepsilon) = 0$  ve  $Var(\varepsilon) = \sigma^2$ 'dir. Toplam ifadesi, yeni tahmin ediciler  $X_3 = X_1 \cdot X_2$  olarak tanımlanırsa  $X_1 X_2$  gibi bir etkileşim, modele eklenmediği terimlerin toplamını gösteren modeli kastetmektedir. Daha da genel olarak  $f(\dots)$  iki değişkenli düzeltilmiş durumda tahmin edilen spesifik bir fonksiyon olmayan  $f(X_1, X_2)$  gibi terimleri elde etmesi sağlanabilir.

$$E(Y|X_1, X_2, \dots, X_p) = \sum_{j=1}^p f_j(X_j) \quad (6)$$

olduğu için  $E(Y)$  üzerine  $X_j$ 'nin etkisi diğer kovaryantlarının (açıklayıcı değişkenlerinin) her bir değeri için aynıdır.

Kabul edelim ki  $Y$  bir bağımlı değişken ve  $X_1, X_2, \dots, X_p$  açıklayıcı değişkenler olsun. Standart doğrusal regresyon modeli;

$$E(Y|X_1, \dots, X_p) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (7)$$

koşullu beklentisi için doğrusal bir form önerir.  $\beta_1, \beta_2, \dots, \beta_p$ 'nin tahminleri en küçük kareler metodu ile elde edilir. Toplamsal Modeller;

$$E(Y|X_1, \dots, X_p) = s_0 + s_1(X_1) + \dots + s_p(X_p) \quad (8)$$

de verilen koşullu beklentinin modellenmesi ile doğrusal modelleri genelleştirir. Burada  $s_i(X_i)$ ,  $i = 1, \dots, p$  düzleştirme fonksiyonlarıdır.

Tahmin ediciler olması için düzleştirme fonksiyonları  $s_i$ 'lerin  $Es_j(X_j) = 0$  gibi standartlaştırılmış koşulları sağlaması gerekir.

Bu fonksiyonlar parametrik bir form sağlamaz. Dolayısıyla parametrik olmayan yollarla tahmin edilmesi gerekir. Çoğu istatistiksel veri analizinde klasik doğrusal modeller ve toplamsal modeller kullanılıyorken onların uygun olmadığı problem türleri de vardır.

Örneğin normal dağılım kategorik bağımlı değişkenlerin modellenmesi için uygun değildir. GAM bu zorlukları aşmada bize yardımcı olur. Normal dağılım gibi diğer birçok dağılımlara da toplamsal modeller uygulanabilir. Böylece GAM daha geniş aralıklı veri aralıklı analizine uygulanabilir (SAS Institute Inc., 1999 ve 2004).

GLM'e benzer olarak GAM'da bir tesadüfi bileşen, bir toplamsal bileşen ve bir link fonksiyonu içerir.  $Y$  bağımlı değişkeni,

$$f_Y(y; \theta; \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (9)$$

tesadüfi bileşeni ile verilen üstel dağılımlar ailesine sahip olduğu varsayılır.  $\theta$  doğal (normal) parametre ve  $\phi$  ölçek parametresidir. Bağımlı değişken  $\mu$ 'nün ortalaması  $X_1, X_2, \dots, X_p$  açıklayıcı değişkenlere bir  $g$  link fonksiyonu ile ilişkilendirilir.

$$\eta = s_0 + \sum_{i=1}^p s_i(X_i) \quad (10)$$

ifadesi toplamsal bileşeni tanımlar. Burada  $s_1(\cdot), \dots, s_p(\cdot)$  düzgün fonksiyonları gösterirken  $\mu$  ve  $\eta$  arasındaki ilişki  $g(\mu) = \eta$  ile gösterilir.

GAM ve GLM benzer durumlara uygulanabilir. Ancak farklı analitik amaçlara hizmet ederler. GLM; modelin parametreleri için çıkarım ve tahmine vurgu yaparken GAM; veriyi parametrik olmayan bir şekilde açıklama üzerine vurgu yapar. GAM veriyi açıklamada ve açıklayıcı değişkenlerle bağımlı değişken arasındaki ilişkiyi görüntülemeye daha uygundur.

### 3. UYGULAMA

Kullanılan veri, zeka testleri yapılan bir merkezde belirli bir aralıkla merkeze gelen 118 kişiye uygulanan anket çalışmasından elde edilmiştir. Anket çalışmasından çıkarılan temel değişkenler; kişinin hayat memnuniyet derecesi(0=düşük derecede memnuniyet, 1=yüksek derecede memnuniyet), yaş(yıl olarak), zeka testi sonucu, eğitim(yıl olarak), aylık gelir(TL cinsinden), ağırlığı(Kg) gibi sürekli değişkenler alınmıştır. Bu veri; ilk olarak açıklayıcı değişkenler olan yaş, zeka, eğitim, gelir ve ağırlık değişkenlerinin memnuniyet derecesi üzerine etkilerini incelemek amacı ile binary lojistik regresyona uygulanmıştır.

Bu amaçla SAS programı içerisinde PROC GENMOD kullanılmıştır. GENMOD işlemi değişkenlere Genelleştirilmiş Doğrusal Modeller uygular. Parametre vektörünün tahmini için en çok olasılık tahmin yöntemini kullanır. Genelde bu tahminler için açık çözüm formları yoktur. Dolayısıyla nümerik çözüm için yöntemler kullanılır.

Bu analizin amacı memnuniyet derecesi ile açıklayıcı değişkenler arasındaki ilişkiyi ortaya koymaktır. Verilerin tanımlandığı DATA adımı aşağıda Şekil 1.'de verilmiştir. İlk olarak sadece açıklayıcı değişkenlerin ana etkilerinin alındığı Genelleştirilmiş Doğrusal Modeller kullanılmıştır. Bu modelleme GENMOD1 olarak isimlendirilmiştir.

```

title 'Comparing PROC GAM with PROC GENMOD';
data memnun;
input yas zeka memnun egitim gelir agirlik @@;
datalines;
34 88 0 5 867 55
55 83 0 8 777 67
...
40 99 1 11 2370 76
33 105 1 11 990 77
;
proc genmod;
model memnun = yas zeka egitim gelir agirlik
/ link=logit dist=binomial;
run;

```

Şekil 1. Verilerin tanımlanması



Burada link fonksiyonu olarak logit ve bağımlı değişken olan memnuniyet derecesinin dağılımı olarak binomial dağılımı alınmıştır. Toplam zeka testi yaptıran 118 kişiden 66'sı hayattan memnuniyetinin düşük derecede olduğunu söylerken 52'si hayattan memnuniyetinin yüksek derecede olduğunu ifade etmiştir. Açıklayıcı değişken etkilerinin GENMOD analizi Tablo1.'de verilmiştir.

**Tablo 1.** Parametre tahminleri analizi tablosu

Parametre tahminleri analizi							
Parametre	SD	Tahmin	S. Hata	Wald 95% Güven aralığı		Ki-kare	p
sabit	1	34.7232	7.4116	20.1968	49.2496	21.95	<.0001
yaş	1	-0.0298	0.0283	-0.0853	0.0257	1.11	0.2925
zeka	1	-0.3807	0.0801	-0.5378	-0.2236	22.56	<.0001*
eğitim	1	-0.1826	0.0891	-0.3571	-0.008	4.2	0.0404*
gelir	1	-0.0018	0.0007	-0.0032	-0.0005	6.93	0.0085*
ağırlık	1	0.0645	0.0218	0.0218	0.1072	8.76	0.0031*

Bu sonuçlara dayalı olarak zeka, eğitim, gelir ve ağırlık değişkenlerinin anlamlı faktörler olduğunu, yaşın anlamlı olmadığını söyleyebilir. GENMOD işlemi bağımlı değişken memnuniyet ile açıklayıcı değişkenler zeka, eğitim, gelir ve ağırlık arasında kesin bir doğrusal ilişkinin olduğunu varsayar. Bu ilişkiyi daha az sınırlayıcı hale getirmek için ilk olarak PROC GAM süreci kullanılacaktır. PROC GAM süreci parametrik olmayan bir yöntemle açıklayıcı değişkenlerin bir bağımlı değişkeni nasıl açıkladığını ortaya koyan bir yöntemdir.

Uygulama aşamasını sadece Kübik düzleştirici zinciri ile sınırlandırıldığından dolayı modelin SAS makro yazılımı Şekil 2.'de verilmiştir. MODEL ifadesinde bir toplamsal model yazılmıştır ve her bir açıklayıcı değişken için bir tek değişkenli düzleştirici zincir uygulanmıştır. Dolayısıyla her bir açıklayıcı değişken 3 serbestlik derecesine sahip bir kübik zincir kullanılarak fit edilmiştir. Bu 3 serbestlik derecesinin bir tanesi uyumun doğrusal kısmı için alınırken kalan ikisi uyumun doğrusal olmayan zincir kısmı için alınmıştır.

```

title 'Comparing PROC GAM with PROC GENMOD';
proc gam data=memnun;
  model memnun=spline(yas,df=3) spline(zeka,df=3)
  spline(egitim,df=3)
  sline(gelir,df=3) spline(agirlik,df=3)
  /dist= binomial;
  output out=estimate p;
run;

```

Şekil 2. SAS makrosu

GAM modeli uygulandıktan sonra düzleştirici parametre tahminleri ve bunlara karşılık gelen genelleştirilmiş çapraz geçerlilik değerleri Tablo 2.'de ve sapma analizi Tablo 3 'te verilmiştir.

Tablo 2. Düzleştirme bileşenleri için uyum tablosu

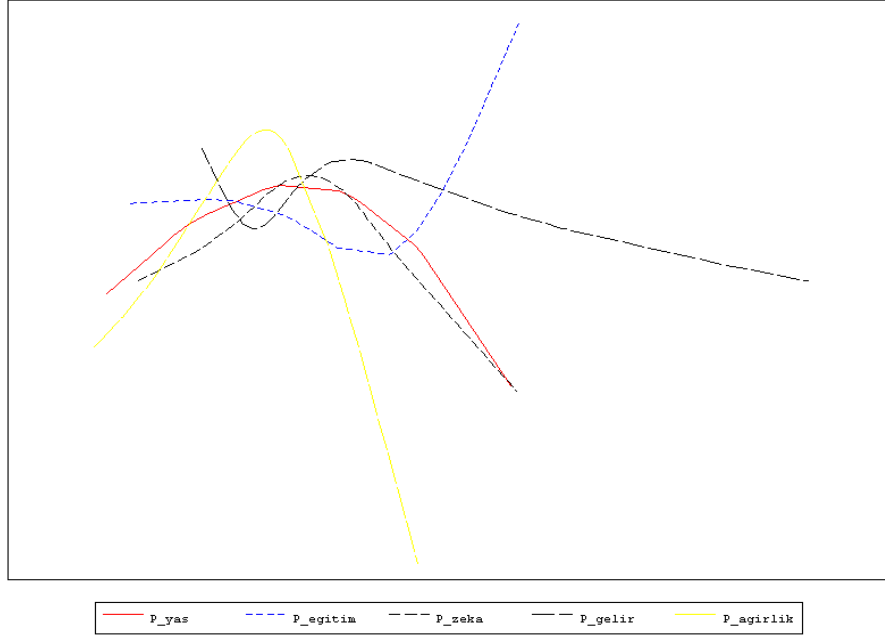
Düzleştirme model analizi			
Düzleştirme bileşenleri için uyum çizelgesi			
Bileşen	Düzleştirme Parametresi	SD	GCV
Zincir(yaş)	0.999959	3	1184.824166
Zincir (zeka)	0.997288	3	14.572082
Zincir (eğitim)	0.983906	3	8.890708
Zincir (gelir)	0.999999	3.000002	29599
Zincir (ağırlık)	0.999999	3	4284.869413

Tablo 3. Düzleştirme model analizi tablosu

Düzleştirme model analizi (Sapma analizi)				
Kaynak	SD	Kareler Toplamı	Ki-kare	p
Zincir (yaş)	3	1.851377	7.4813	0.048
Zincir (zeka)	3	2.303436	9.3081	0.0255
Zincir (eğitim)	3	3.442165	13.9097	0.003
Zincir gelir	3	1.427178	5.7672	0.1235
Zincir(ağırlık)	3	22.508637	90.9565	<.0001

GAM sonuçlarının en önemli kısmı sapma analizi tablosudur. Modeldeki her bir düzleştirici etki için tablo full model ve ilgilenilen değişken olmaksızın elde edilen model arasındaki sapmayı karşılaştıran bir Ki-kare test değeri verir. Tablo 3.'te sapma analizi sonuçları gösterir ki % 5 anlamlılık seviyesinde yaşın, zekanın, eğitimin ve ağırlığın etkisi anlamlıdır. Oysa gelirin etkisi anlamsız olarak ortaya çıkmıştır.

Görüldüğü gibi PROC GAM ve PROC GENMOD süreçleri farklı sonuçlar ortaya koymuştur. PROC GENMOD'da sadece yaş anlamsızken PROC GAM da yaş anlamlı olurken ağırlık anlamsız hale gelmiştir. Bu farklılıkları incelemek için bağımlı değişken memnuniyete karşı açıklayıcı değişkenler yaş, zeka, eğitim ve ağırlığın grafiği çizilebilir. Bu grafik Şekil 3. 'de verilmiştir.



Şekil 3. Yaş, zeka, eğitim ve ağırlığın grafiği

Bu grafiklerden, yaş, eğitim, zeka, ağırlığa karşılık gelen kısmi tahminlerin kuadratik bir biçime sahip olduğu gözlemlenirken gelire karşılık gelen kısmi tahminlerin daha karışık bir biçime sahip olduğu anlaşılmaktadır. Yapılan aykırı gözlem incelemelerinden de

benzer grafikler ortaya çıkmaktadır. Dolayısıyla bu grafiklerden çıkan sonuçlarla Tablo 3.'teki sapma analizi sonuçları ile örtüşmektedir.

Bu verilerin GENMOD kullanılarak yapılan ilk analizi ile GAM analizi arasındaki en önemli farklar GENMOD1'de yaş anlamsızken GAM'da anlamlı olması ve gelir GENMOD1'de anlamlı iken GAM'da anlamsız olmasıdır. Bunun temel nedeni Şekil 1.'deki grafikten ve Tablo 3.'ten anlaşılacağı gibi yaşın etkisi kuadratik iken GENMOD1'de doğrusal alınması ve gelirin etkisi tam olarak kuadratik değil iken GAM'da kuadratik alınmasıdır.

O halde GAM'ın sonuçlarından hareketle söylenebilir ki yaş, zeka ve ağırlık etkileri kuadratik alınabilirken gelirin etkisi kuadratik olmayacaktır. Bu sonuçları göz önüne alarak yaş, zeka ve ağırlık etkilerini kuadratik, gelirin etkisini doğrusal olarak tekrar GENMOD işlemi uygulanabilir. GENMOD2 olarak isimlendirilen bu modelin makrosu Şekil 4. 'deki gibidir.

```

title 'Comparing PROC GAM with PROC GENMOD';
proc genmod data=memnun ;
  model memnun = yas yas*yas
  zeka zeka*zeka
  egitim egitim*egitim
  gelir
  agirlik agirlik*agirlik
  /link=logit dist=binomial;
  contrast 'yas' yas , yas*yas 1;
  contrast 'zeka' zeka 1, zeka*zeka 1;
  contrast 'egitim' egitim 1, egitim*egitim 1;
  contrast 'agirlik' agirlik 1, agirlik*agirlik 1;
run;

```

Şekil 4. Genmod2 SAS makrosu

Bazı kuadratik formları da içeren yeni GENMOD modelinin uyum iyiliği kriterlerine bakıldığında ilk GENMOD modelindeki uyum iyiliği kriter değerlerinden çok daha küçük sapma değerleri

elde edilmiştir. Her iki model için uyum iyiliği belirleme için kullanılan kriterlerin karşılaştırılması Tablo 4.'de verilmiştir.

**Tablo 4.** Uyum iyiliği kriterlerinin karşılaştırılması

Kriter	GENMOD1	GENMOD2
	Değer	Değer
Sapma	69.0492	37.4314
Ölçeklenmiş sapma	69.0492	37.4314
Pearson Ki-kare	101.2149	53.2505
Ölçeklenmiş Pearson Ki-Kare	101.2149	53.2505
Log Olabilirlik	-34.5246	-18.7157

Tablo 4.'den görüldüğü gibi GENMOD1 olarak isimlendirilen sadece ana etkileri alındığı ilk modelin sapma değerleri GENMOD2 olarak adlandırılan ana etkilerin bazı karesel formların alındığı ikinci modelin sapma değerlerinden oldukça yüksektir. GENMOD2 modelinin parametre tahminleri analizi Tablo 5.'te verilmiştir.

**Tablo 5.** Parametre tahminleri analizi tablosu

Parametre tahminleri analizi							
Parametre	SD	Tahmin	Standart Hata	Wald 95% Güven Aralığı		Ki-Kare	p
sabit	1	260.6662	127.002	11.7469	509.5855	4.21	0.0401
yaş	1	-0.7122	0.3083	-1.5124	0.088	3.94	0.0411
yaş*yaş	1	0.0069	0.0013	-0.0016	0.0154	4.56	0.0409
Zeka	1	-3.7792	0.6341	-8.9419	1.3834	4.96	0.0014
zeka*zeka	1	0.0185	0.0043	-0.0096	0.0466	4.66	0.0475
eğitim	1	1.729	1.0029	-0.2367	3.6946	2.97	0.0447
eğitim* eğitim	1	-0.0898	0.0454	-0.1789	-0.0007	3.9	0.0482
gelir	1	-0.0025	0.0006	-0.0055	0.0006	2.56	0.0095
ağırlık	1	-1.7578	0.7311	-3.1908	-0.3248	5.78	0.0162
Ağırlık* ağırlık	1	0.0128	0.0052	0.0027	0.0229	6.16	0.0131

Tablo 5.'te görüldüğü gibi bütün parametreler %5 anlamlılık seviyesinde anlamlıdır. Oysa gerek GENMOD1 gerekse de GAM modelinde bazı değişkenler anlamsızdır. Ana etkiler için kontrast test sonuçları Tablo 6. 'da verilmiştir.

**Tablo 6.** Kontrast sonuçları tablosu

Sabit	Kontrast sonuçları			
	SD	Ki-kare	Pr > Ki-kare	Tip
Yaş	2	9.84	0.044	LR
zeka	2	32.68	<.0001	LR
eğitim	2	10.86	0.0044	LR
ağırlık	2	29.46	<.0001	LR

Tablo 6.'dan anlaşılacağı gibi kontrast test sonuçlarından yaş, zeka, eğitim ve ağırlık anlamlı çıkmıştır.

#### 4. SONUÇ VE ÖNERİLER

Genelleştirilmiş toplamsal modeller (GAM), bağımlı değişken ile açıklayıcı değişkenler arasında var olan ilişkinin matematiksel formunun ortaya konmasına yardımcı olur. Genelleştirilmiş doğrusal modeller ismiyle kullanılan doğrusal modellerin üst çatısı olan yapı bu ilişkilerin parametrik olduğunu var sayar. Oysa uygulamada çoğu zaman araştırma gereği hissedilmeyen bu ilişkinin yapısı kesin varsayımlar altında hatalı modellenebilir. GAM bu ilişkileri tanımlı bir fonksiyonun yerine parametrik olmayan bir düzleştirici koyarak kesin varsayımları hafifletmeye imkân kılar.

Bu çalışmada açıklayıcı değişkenle bağımlı değişkenin aralarındaki ilişkinin doğrusal olmadığı halde GLM ile doğrusal varsayıldığı durumda ilgilenilen değişken %5 anlamlılık seviyesinde anlamsız iken GAM kullanılarak aradaki ilişkinin doğrusal değil kuadratik bir forma daha uygun olduğu hem istatistiksel test olarak hem de görsel grafiksel olarak görülmüştür.

Daha açık ifade edilirse, yaş değişkeninin memnuniyet üzerine etkisi doğrusal olmadığı halde doğrusal varsayılan GLM'de anlamsız çıkarken aralarındaki ilişkinin kübik olduğu GAM ile gösterilmiştir. Daha sonra kübik olarak GLM'de modellendiğinde

anlamalı çıkmıştır. Bunun terside gerçekleşmiştir. Gelir değişkenin memnuniyet derecesi üzerine etkisi kübik olmadığı GAM ile anlaşılmiş ve daha sonraki modele sadece doğrusal etki göz önüne alınarak katılmıştır.

Sonuç olarak, açıklayıcı değişkenlerin memnuniyet derecesi üzerine etkilerinin yapısı GAM ile incelendikten sonra tekrar GLM uygulandığında model karşılaştırma kriterlerine göre sapmalar Tablo 4.'te görüldüğü gibi büyük oranda azalmıştır.

#### KAYNAKLAR

- Aitken, M., Anderson, D., Francis, B.,Hinde, J. (1989). *Statistical modelling in GLIM.*, Clarendon Press, 374s, Oxford.
- Cengiz, M.A. (1997). Bivariate logistic regression analysis. Technical report, the University of Salford, MCS-97-11.
- Dobson, A.J. (1990). *An introduction to generalized linear models*, 2<sup>nd</sup> ed., Chapman and Hall, London.
- Hastie, T. J. (1991). Generalized additive models, in statistical models (Editörler: J. M., Chambers, T. J. Hastie), Wadsworth & Brooks Cole Advanced Books & Software, 249-307, Pacific Grove, California.
- Hastie, T. J., Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall, 335s, New York.
- Nelder, J.A., Wedderburn, R.W.M. (1972). Generalized linear models, *Journal of Royal Statist. Soc.*, 135(A), 370-384.
- McCullagh, P., Nelder, J.A. (1989). *Generalized linear models*. Second Edition, Chapman and Hall, London.
- SAS Institute Inc. (1999). SAS procedures guide, Version 8, Cary, NC: SAS Ins. Inc. SAS Institute Inc. (2004). SAS/STAT 9.1 User's guide, ary, NC: SAS Institute Inc.
- Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, 13, 689-705.

\*\*\*\*