


Original article

Çoklu Doğrusal Regresyon ve Genelleştirilmiş Toplamsal Modellerin Açıklanan Varyans Sonuçlarının Karşılaştırılması ¹

Comparison of Explained Variance Results of Multiple Linear Regression and Generalized Additive Models

Hikmet Şevgin  *

Department of Measurement and Evaluation, Faculty of Educational, University of Muş Alparslan, Muş, Turkey

Özet

Bu çalışmanın amacı aynı veri setinde Çoklu Doğrusal Regresyon analizi ile Genelleştirilmiş Toplamsal Modellerin açıklanan varyans değeri açısından karşılaştırmaktır. Doğrusal modeller oldukça basit olmasına rağmen, bazı gerçek hayattaki örneklerin doğrusal olmaması nedeniyle, doğrusal olmayan regresyon etkilerini tanımlamak ve karakterize etmek için daha esnek bir istatistiksel yöntem olan Genelleştirilmiş Toplamsal Modeller alternatif olarak kullanılabilir. Bu bağlamda regresyon yöntemlerinin karşılaştırılarak değerlendirilmesi ve sonuçlarının karşılaştırılması önemli görülmektedir. Çalışmaya ait veri seti PISA 2015 çalışması içinden Türk öğrencilere ait okuma becerileri puanları ile öğrencilere ait çeşitli demografik değişkenlerden oluşmaktadır. Örneklemde yer alan öğrenci sayısı 5895 olup kayıp verilerin dikkate alınmaması sonrası çalışma 5089 öğrenciye ait veri üzerinde sürdürülmüştür. Gerekli varsayımlar sınanması sonrası yapılan analizlerde; Çoklu Doğrusal Regresyon analizi toplam varyansın %23,8'ini açıklarken Genelleştirilmiş Toplamsal Modeller %25,33'ünü açıklamıştır. Bağımsız değişkenlerin bağımlı değişkende açıkladıkları varyans miktarı bakımından Genelleştirilmiş Toplamsal Modellerin daha yüksek olduğu sonucu elde edilmiştir. Genelleştirilmiş toplamsal modeller Çoklu Doğrusal Regresyon analiz yöntemine göre açıklanan varyans açısından daha yüksek elde edilmesinde Genelleştirilmiş toplamsal modellerin doğrusal tahmin edicileri olarak düzleştirici fonksiyonları kullanmasının etkili olduğu söylenebilir. Genelleştirilmiş toplamsal modeller, varsayımların sağlanamadığı veriler için Çoklu doğrusal regresyon analiz yöntemine alternatif olarak, neden sonuç temelli çalışmalarda kullanılabilirliği düşünülmektedir. Bu doğrultuda Genelleştirilmiş Toplamsal Modellerin Eğitim Bilimleri çalışmalarında kullanılması önerilmektedir.

Anahtar Kelimeler: Genelleştirilmiş Toplamsal Modeller, Çoklu Doğrusal Regresyon, Açıklanan Varyans.

Abstract

The aim of this study is to compare Multiple Linear Regression Analysis and Generalized Additive Models in terms of explained variance value in the same data set. Although linear models are fairly simple, Generalized Additive Models, a more flexible statistical method, can be used alternatively to describe and characterize nonlinear regression effects, as some real-life examples are not linear. In this context, it is important to compare the regression methods and to compare the results. The data set of the study consists of the reading skills scores of Turkish students from the PISA 2015 study and of various demographic variables of the students. The number of students in the sample is 5895 and the study was carried out on the data of 5089 students after the

* Corresponding author:

Şevgin Hikmet is an Dr. Research Assistant in the Department of Measurement and Evaluation at Muş Alparslan University in Muş, Turkey. Her research interests include the Measurement and Evaluation, Educational Data Mining, Scale Development, Power Analysis. He has lived, in Muş, Turkey.
Email: hikmetsevgin@gmail.com

¹ Bu çalışma "6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresin"de sunulan sözlü bildirinin genişletilmiş halidir.

missing data were not taken into account. In the analysis made after the necessary assumptions had been tested; while Multiple Linear Regression analysis explained 23.8% of the total variance, Generalized Additive Models explained 25.33%. It has been concluded that Generalized Additive Models are higher in terms of the amount of variance explained by the independent variables in the dependent variable. It can be said that generalized additive models use flattening functions as linear estimators of generalized additive models to obtain higher variance explained according to the Multiple Linear Regression analysis method. Generalized additive models are thought to be used in cause and effect based studies as an alternative to the multiple linear regression analysis methods for data where assumptions cannot be satisfied. Accordingly, it is recommended to use Generalized Additive Models in Educational Sciences studies.

Keywords: Generalized Additive Models, Multiple Linear Regression, Explained Variance.

Received: 05 September 2020 * **Accepted:** 10 March 2021 * **DOI:** <https://doi.org/10.29329/ijiape.2021.333.10>

GİRİŞ

Regresyon analizi iki ya da daha fazla değişken arasındaki ilişkiyi modellemek ve incelemek, bir matematiksel eşitlik ile açıklamak için kullanılan istatistiksel bir analiz yöntemidir (Montgomery, Peck ve Vining, 2013). Regresyon analizi bağımlı ve bağımsız değişkenler arasındaki ilişkileri neden sonuç temelinde ele alan analizler arasında yer almaktadır. Normallik, varyansların homojenliği, doğrusallık gibi bazı varsayımların sağlanması durumunda parametrik, bu varsayımların sağlanmadığı durumlarda ise parametrik olmayan ya da yarı parametrik regresyon yöntemlerinden biri kullanılabilir (Kayri, Çokluk, Şevgin, Elkonca ve Ceyhan, 2012; Shavelson, 2016). Regresyon analizinde değişkenler arasındaki ilişki doğrusal ise doğrusal regresyon analizi, değil ise doğrusal-olmayan (eğrisel) regresyon analizi olarak isimlendirilir (Büyüköztürk, Çokluk ve Köklü, 2011). Çoklu doğrusal regresyon (ÇDR) analizi doğrusal regresyon analizlerinden biridir. Burada ilk olarak ÇDR analiz yöntemi hakkında ve daha sonra Genelleştirilmiş toplamsal modeller (GAM) hakkında bilgiler sunulmuştur.

Bağımsız değişken sayısı iki veya daha fazla olduğunda bağımlı değişken ile bağımsız değişkenler arasındaki ilişki ÇDR analizi ile incelenmektedir. ÇDR’de bağımlı değişken (Y), bir grup bağımsız değişken (X_1, X_2, \dots, X_p) tarafından kestirilir ve bu ilişkiye ait regresyon denklemi aşağıdaki gibi ifade edilir;

$$E(Y|X_1, X_2, \dots, X_p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (1)$$

Burada $i = (1, 2, \dots, p)$ ’dir ve β_0 Y’nin kesişimini (bütün X değerlerinin 0 olduğu durumda Y’nin değeri) ifade eder. $\beta_1, \beta_2, \dots, \beta_p$ regresyon sırasında her bir bağımsız değişkene atanan katsayılarıdır. X_1, X_2, \dots, X_p ’ler ise bağımsız değişkenleri ifade eder. ÇDR analizinde standart (Enter), aşamalı ve hiyerarşik yaklaşım olmak üzere yaygın olarak üç model kullanılmaktadır (Sümbüloğlu ve Akdağ, 2007; Büyüköztürk, 2011; Özdamar, 2013).

ÇDR analiz yönteminin uygulanabilmesi için veri setinin bazı varsayımları sağlaması gerekir (Büyüköztürk, 2011; Özdamar, 2013; Shavelson, 2016). Bunlar;

- Bağımlı değişken normal dağılım sergilemelidir (Normallik).
- Bağımlı değişkenin varyansı bağımsız değişkenin her türlü olası kombinasyonunda eşittir (varyansların homojenliği).
- Bütün bağımsız değişkenler sabit tutulduğunda, bağımlı değişken ile her bir bağımsız değişken arasındaki ilişki doğrusaldır (Doğrusallık).
- Bağımsız değişkenler hatasız ölçülmüş olmalı ve aralarında çoklu doğrusal bağımlılık (Multicollinearity) bulunmamalıdır.

ÇDR analizi öncesi veri seti üzerinde yukarıda belirtilen varsayımların irdelenmesi gerekir. Genelleştirilmiş toplamsal modeller (GAM) ilgili varsayımların gerçekleşmesine ihtiyaç duymaz (Wood, 2006; Hastie, Tibshirani ve Friedman, 2009).

GAM analiz yöntemi 1990 yılında Hastie ve Tibshirani tarafından geliştirilen istatistiksel bir yöntemdir (Hastie ve Tibshirani, 1990). Bağımlı değişkenlerin normallik varsayımını sağlamadığı durumlarda Genelleştirilmiş Doğrusal Model (GLM) sıkça kullanılan bir yaklaşımdır (Kan, Kılınç ve Çavuş, 2017). GAM analiz yöntemi GLM ailesinin parametrik olmayan bir uzantısıdır. Yani GAM, GLM'nin değişime uğramış bir versiyonudur. GAM'a düzeltirme fonksiyonu eklenmesiyle bu değişim sağlanmış olmaktadır. Yani GAM'la GLM arasındaki tek fark GAM'ın doğrusal tahmin edici olarak bilinmeyen düzeltirici fonksiyonları kullanmasıdır (Şenel, Cengiz, Savaş ve Terzi, 2009).

Doğrusal ve doğrusal olmayan regresyon modellerinde istatistiksel analiz yöntemleri bağımlı değişkenin normal dağılım varsayımı üzerinde şekillenmektedir (Savaş ve Cengiz, 2009). Doğrusal regresyon analizinde istatistiksel analiz yöntemleri bağımlı değişkenin normal dağıldığı varsayımına dayanmaktadır. Hem bağımlı değişkenin normal dağılım göstermediği hem de bağımlı değişkenle bağımsız değişken arasındaki ilişkinin parametrik (doğrusal olmadığı) olmadığı durumlarda, GAM modellemeye esneklik getirir (Şenel, Cengiz, Savaş ve Terzi, 2009; Finch vd., 2011). Örneğin, normal dağılım bağımlı değişkenlerin kategorik olduğu durumlarda uygun değildir veya bağımlı değişkenle bağımsız değişken arasında her zaman bilinen bir matematiksel fonksiyon (doğrusal, polinom ya da logaritmik gibi) olmayabilir ya da ilişki kolayca tanımlanmayabilir. Bu gibi durumlarda GAM var olan ilişkileri belirleyebilmek için tanımlı bir fonksiyonla, parametrik olmayan bir düzeltiriciyi yer değiştirerek bu varsayımı genelleştirir (Hastie ve Tibshirani, 1990).

$$E(Y|X_1, X_2, \dots, X_p) = s_0 + s_1(X_1) + s_2(X_2) + \dots + s_p(X_p) \quad (2)$$

Burada $i = (1, 2, \dots, p)$ 'dir ve $s_i(X)$, düzeltirme fonksiyonlarıdır. X_i 'ler ise bağımsız değişkenleri ifade eder.

$$E(Y|X_1, X_2, \dots, X_p) = s_0 + \sum_{i=1}^p s_i(X_i) \quad (3)$$

GAM analiz yöntemi, ÇDR analiz yöntemine göre bazı avantajlar taşımaktadır. ÇDR analiz yönteminde normallik, doğrusallık ve homojenlik gibi varsayımların sağlanması gerekirken, GAM analiz yönteminde bu tip istatistiksel varsayımlara ihtiyaç duyulmamaktadır (Hastie ve Tibshirani, 1990). Diğer bir deyişle bağımlı değişken normallik varsayımını sağlasın ya da sağlamasın, veri seti GAM analiz yöntemi ile incelenebilmektedir. Yine ÇDR analiz yönteminde bağımsız değişkenlerin sürekli olması gerekir. Şayet kategorik değişkenlerin bağımlı değişken üzerinde etkileri incelenecekse “dummy” değişken olarak tanımlanarak analize dâhil edilmeleri gerekir (Büyüköztürk, 2011). GAM analiz yöntemi ise, bağımsız değişkenlerin ister sürekli ister kategorik yapıya sahip olabileceği durumlar için kullanılabilen bir yöntemdir (Hastie, Tibshirani ve Friedman, 2009).

Doğrusal modeller oldukça basit olmasına rağmen, bazı gerçek hayattaki örneklerin doğrusal olmaması nedeniyle, doğrusal olmayan regresyon etkilerini tanımlamak ve karakterize etmek için daha esnek bir istatistiksel yöntem olan GAM analiz yöntemi alternatif olarak kullanılabilir (Hastie, Tibshirani ve Friedman, 2009). Bu bağlamda regresyon yöntemlerinin karşılaştırılarak değerlendirilmesi ve sonuçlarının karşılaştırılması önemli görülmektedir. Bu çalışmada regresyon analizi yöntemlerinden Çoklu Doğrusal Regresyon yöntemi (ÇDR) ile GAM analiz yöntemine ait açıklanan varyans değeri (R^2) karşılaştırılarak PISA 2015 veri seti arasından Türk öğrencilere ait veriler üzerinde uygulama yapılmış ve bu iki regresyon yönteminin açıklanan varyans (R^2) performansını göstermek amaçlanmıştır. Bu amaç doğrultusunda çalışmada aşağıdaki soruya cevap aranmıştır:

PISA 2015 uygulamasında Türk öğrencilerin Okuma Becerilerini etkilediği düşünülen değişkenlerin Çoklu Doğrusal Regresyon ve GAM analiz yöntemleriyle kestirilmesi ile elde edilen sonuçların açıklanan varyans (R^2) açısından karşılaştırılması sonucu bir fark var mıdır?

YÖNTEM

Araştırmanın Modeli

Bu araştırma genel tarama yöntemlerinden betimsel nitelikte ilişkisel tarama modelinde bir araştırmadır. Tarama modelleri, geçmişte ya da halen var olan bir durumu olduğu şekliyle betimlemeyi amaçlayan araştırma yaklaşımlarıdır. Genel tarama modeli, çok sayıda elemandan oluşan bir evrende, evren hakkında genel bir yargıya varmak amacıyla evrenin tümü ya da evrenden alınacak bir grup üzerinde yapılan tarama düzenlemeleridir. Araştırmaya konu olan durum, kendi koşulları içinde ve olduğu gibi tanımlanmaya çalışılır (Karasar, 2009).

Çalışma Grubu

Bu çalışmada kullanılan veri seti, PISA 2015 çalışması içinden Türk öğrencilere ait okuma becerileri puanları ile öğrencilere ait çeşitli demografik değişkenlerden oluşmaktadır. Çalışma grubunda kayıp veri miktarı % 5'ten fazla olduğu için kayıp verileri veri setinden çıkarma işlemi uygulanmıştır. Çalışma grubunda yer alan 5895 öğrenci sayısı, eksik veriye sahip öğrencilerin çıkarılması sonrası 5089 olarak elde edilmiştir.

İşlem-Verilerin Analizi

Çalışmada kullanılan verilerin yapısı itibarıyla, kategorik yapıya sahip bağımsız değişkenler dummy kodlanarak sürekli hale getirilmiştir. Araştırmada ÇDR analizinin gerektirdiği varsayımların sınanması adına normallik, doğrusallık ve homojenlik için Kolmogorov-Smirnov testi, histogram grafiği ve standardize edilmiş artık değerler ile standardize edilmiş yordanan değerlerin yer aldığı saçılım grafikleri incelenmiştir. Daha sonra ÇDR analiz yöntemi ve GAM analiz yöntemi kullanılmıştır. Analizler için SPSS ve STATISTICA programlarından yararlanılmıştır.

BULGULAR

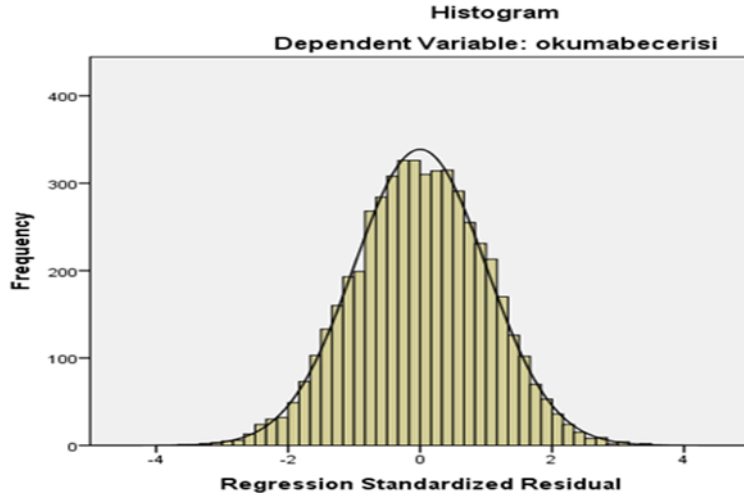
ÇDR analiz yönteminin varsayımlarının sınanması adına bağımlı değişkenin normal dağılım gösterip göstermediği Kolmogorov-Smirnov testi ile test edilmiş ve histogram Grafiği ile de gösterilmiştir. Tablo 1'de görüldüğü üzere Kolmogorov-Smirnov testi sonucu $p > 0,05$ büyük olduğu için bağımlı değişkenin normal dağılım gösterdiği varsayılır.

Tablo1. Normallik Testi

	Kolmogorov-Smirnov		
	Statistic	df	p
Okuma Becerisi	0,013	5089	0,064

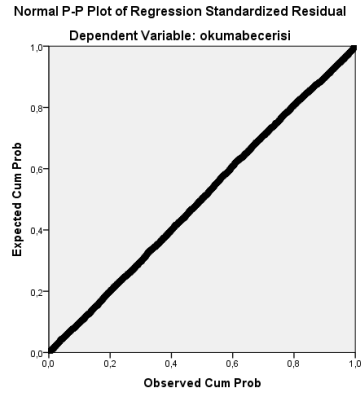
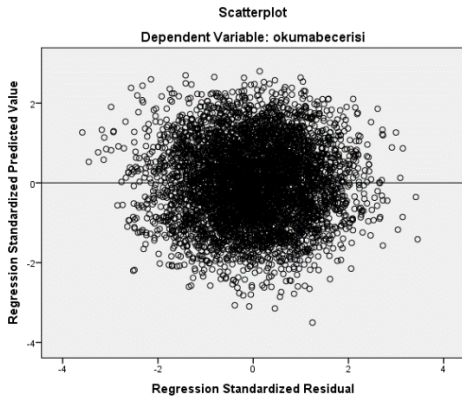
* $p > 0,05$

Yine standardize edilmiş bağımlı değişkenimize ait histogram ve normal dağılım eğrisi şekil 1'de görüldüğü üzere normal bir dağılım gösterdiği söylenebilir.



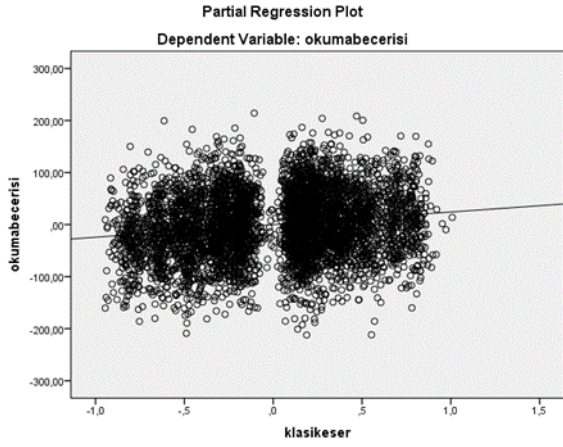
Şekil 1. Normal Dağılım Eğrisini gösteren Histogram Grafiği

Değişkenlerin çok değişkenli normallik ve doğrusallık gösterip göstermediği standartlaştırılmış hata değerlerine ilişkin grafikler ile incelenmiş şekil 2 ile 15 arasında gösterilmiştir.

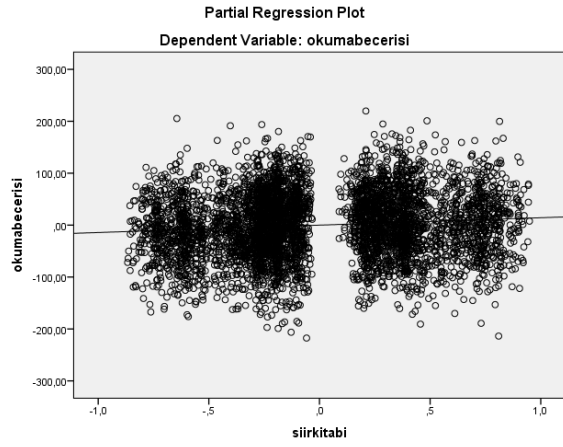


Şekil 2. Okuma Becerisi Saçılım Grafiği Şekil 3. Standartlaştırılmış Regresyon Doğrusu

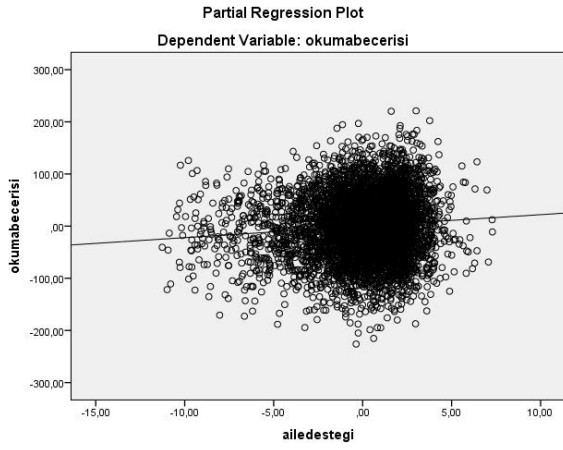
Şekil 2 ve 3 incelendiği zaman, standardize edilmiş artık değerler ile standardize edilmiş yordanan değerler için oluşturulan saçılım grafiği doğrusal bir ilişkiyi göstermektedir. Noktalar bir eksen etrafında toplanma eğilimindedir.



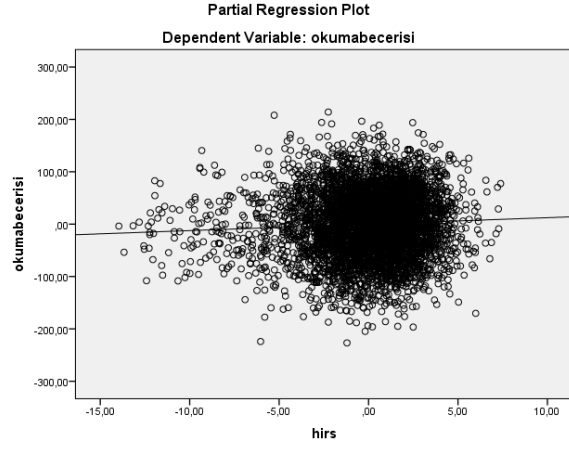
Şekil 4. Klasik eser saçılım grafiği



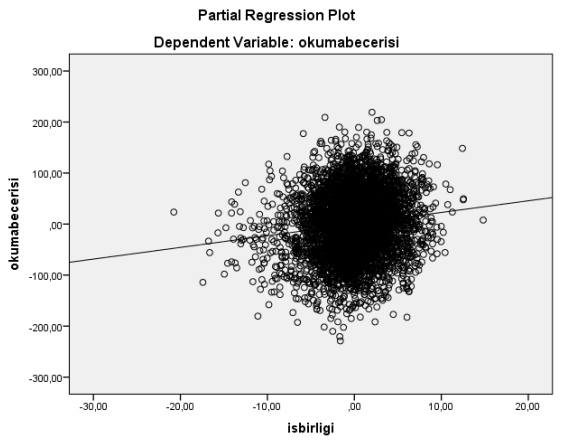
Şekil 5. Şiir kitabı saçılım grafiği



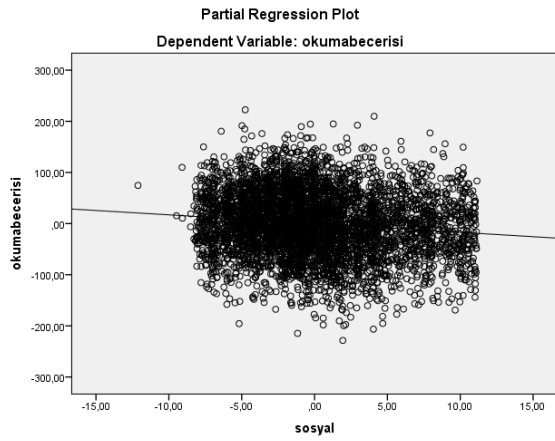
Şekil 6. Aile desteği saçılım grafiği



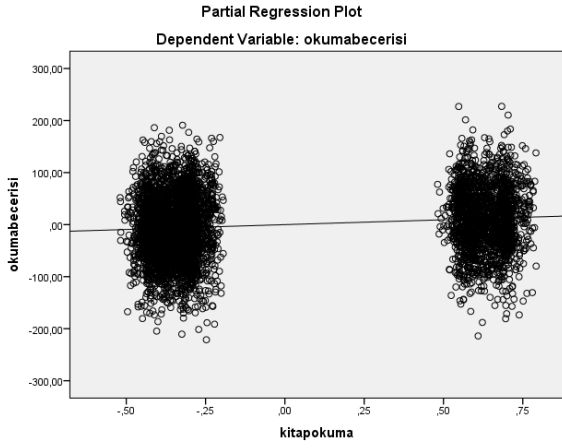
Şekil 7. Hırs saçılım grafiği



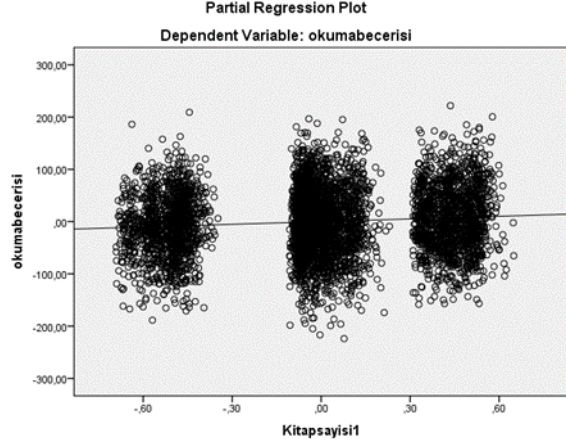
Şekil 8. İşbirliği saçılım grafiği



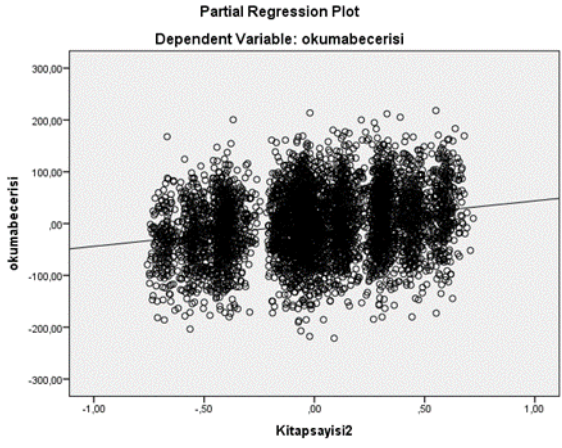
Şekil 9. Sosyal saçılım grafiği



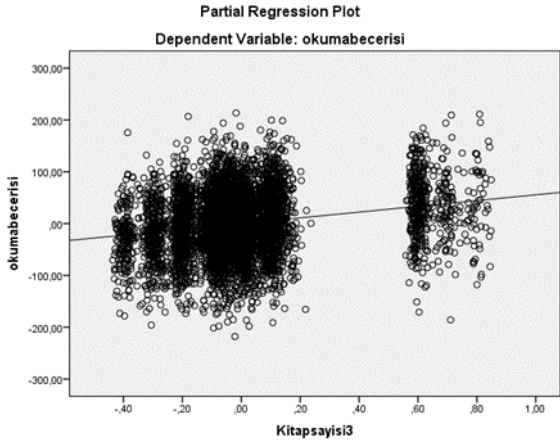
Şekil 10. Kitap Okuma saçılım grafiği



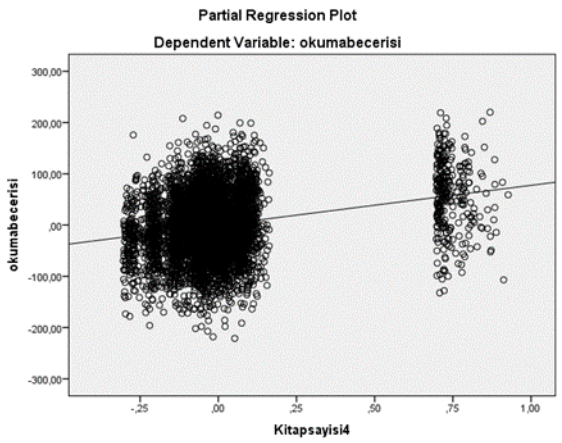
Şekil 11. Kitap Sayısı 1 saçılım grafiği



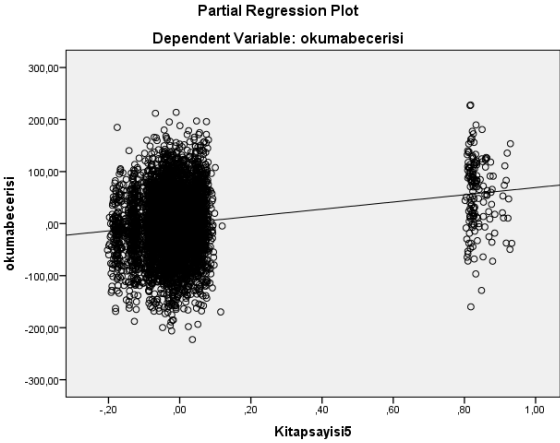
Şekil 12. Kitap Sayısı 2 saçılım grafiği



Şekil 13. Kitap Sayısı 3 saçılım grafiği



Şekil 14. Kitap Sayısı 4 saçılım grafiği



Şekil 15. Kitap Sayısı 5 saçılım grafiği

Şekil 4 ve 15 arasındaki grafikler incelendiğinde bağımsız değişkenlerin bağımlı değişkenle aralarındaki kısmi ilişkilerini temel alan dağılım grafikleri görülmektedir. Yukarıdaki grafiklerde de görüldüğü üzere klasik eser, şiir kitabı, aile desteği, hırs, işbirliği, kitap okuma ve kitap sayısı arasında

pozitif yönde doğrusal bir ilişkinin olduğu görülmektedir. Sosyal değişken ile okuma becerisi arasında ise negatif yönde doğrusal bir ilişkinin olduğu görülmektedir.

Bağımsız değişkenlerin kendi arasında yüksek düzeyde korelasyona sahip olması çoklu bağlantı problemi olarak karşımıza çıkmaktadır. Tablo 2’de değişkenlere ait VIF (variance inflation factor) ve Tolerance değerleri verilmektedir.

Tablo 2. Bağımsız değişkenlere ilişkin VIF-Tolerance Değerleri

Model (Enter)	Beta	Sig.	Tolerance	VIF
Klasikeser	,168	,000	,751	1,331
Şiir kitabı	,098	,000	,773	1,293
Ailedestegi	,085	,000	,838	1,194
Hırs	,052	,000	,803	1,245
İşbirliği	,125	,000	,815	1,227
Sosyal	-,108	,000	,992	1,008
Kitap okuma	,124	,000	,982	1,018
Kitapsayısı1	,107	,000	,622	1,609
Kitapsayısı2	,276	,000	,576	1,737
Kitapsayısı3	,227	,000	,715	1,398
Kitapsayısı4	,248	,000	,789	1,268
Kitapsayısı5	,167	,000	,865	1,156

Tolerance değeri 0,20’den daha düşük ve VIF değerlerinin 10’dan daha yüksek çıkması durumunda çoklu bağlantı probleminin olduğuna işaret eder (Büyüköztürk, 2011; Özdamar, 2013). Çoklu bağlantı problemi modelin güvenilirliğini düşürmektedir. Tablo 2’de elde ettiğimiz Tolerance ve VIF değerleri çoklu bağlantı probleminin olmadığına işaret etmektedir.

ÇDR analizimiz için gerekli olan varsayımların sınanması sonrası bağımlı değişkenle bağımsız değişken arasındaki ilişkinin miktarı ve açıklanan varyansı Tablo 3’te verilmektedir.

Tablo 3. Çoklu Doğrusal Regresyon Açıklanan Varyans Değeri

Model	R	R ²	Adjusted R Square
Enter Modeli	0,489	0,239	0,238

R² bağımsız değişkenlerin bağımlı değişkendeki değişimin yüzde kaçını açıkladığını gösteren bir indekstir. Toplam varyansın %23,8’nin bağımsız değişkenler tarafından açıklandığını ifade edebiliriz. GAM analiz yönteminde bağımsız değişkenlerin bağımlı değişken üzerindeki varyansın yüzde kaçını açıkladığı Tablo 4’de verilmektedir.

Tablo 4. Genelleştirilmiş Toplamsal Modelle Açıklanan Varyans Değeri

Response: Okuma Becerisi				
Distribution: Normal; link function: Identity				
	Num. of obs	Num. of smoths	R	R square *100%
Statistic	5089	10	0,503	25,335

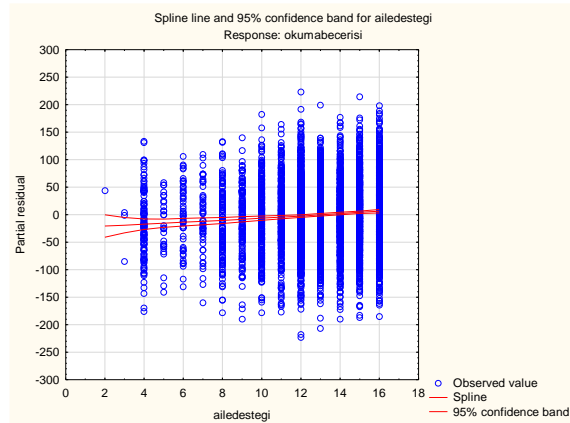
Toplam varyansın % 25,33'nün bağımsız değişkenler tarafından açıklandığını ifade edebiliriz. İki analiz arasında ortaya çıkan bu fark GAM analiz yönteminin doğrusal tahmin edici olarak düzleştirici fonksiyonları kullanması ile oluşmaktadır. ÇDR analiz yöntemi en küçük kareler fonksiyonu kullanılarak regresyon doğrusu çizilmektedir. En küçük kareler fonksiyonu tüm noktaların regresyon doğrusuna olan uzaklıklarının kareler ortalamasını en aza indirir. GAM analiz yöntemi ise tüm veri setini ele alıp düzleştirici fonksiyonları kullanarak küçük doğrusal parçacıkların toplamsal halini ortaya koymaktadır. Böylelikle düzgünleştirme uzanımları (smoothing spline) çizilmektedir. Tablo 5'de GAM analiz yönteminde non-linear özellik gösteren değişkenlere ait p anlamlılık değerleri sunulmaktadır.

Tablo 5. Genelleştirilmiş Toplamsal Modellerde Sürekli Verilere ait Non-Linear Değerler

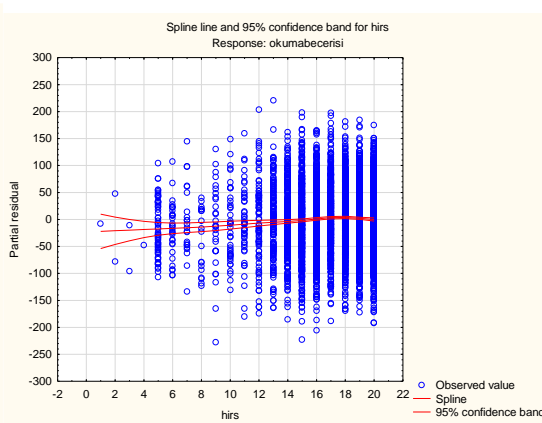
Non-Linear Bağımsız Değişkenler	Non-Linear p-value
Aile Desteği	0,937127
Hırs	0,006456
İşbirliği	0,000000
Sosyal	0,000000

*P < 0,05

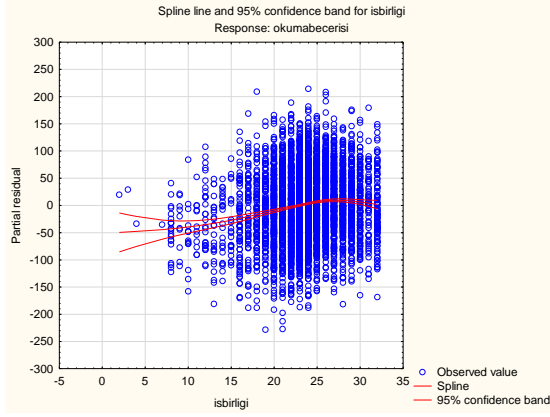
GAM analiz yönteminde kategorik yapıda bağımsız değişkenler herhangi bir dönüşüme uğramadan analiz sürecine dâhil edilebilmektedir. Tablo 5 incelendiğinde de sürekli yapıya sahip değişkenlere ilişkin p anlamlılık değerleri sunulmaktadır. Hırs, işbirliği ve sosyal değişkenleri ($p < 0,05$) non-linear bir yapı sergiledikleri, aile desteği değişkeninin ise lineer bir yapı sergilediği izlenmektedir. Bu değişkenlere ait grafikler şekil 16-17-18-19 da sunulmaktadır.



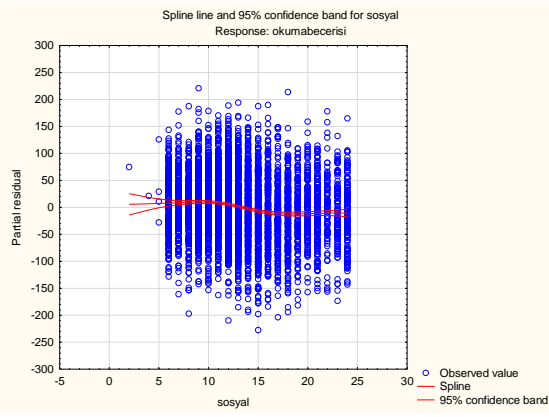
Şekil 16. Aile Desteği saçılım grafiği



Şekil 17. Hırs saçılım grafiği



Şekil 18. İşbirliği saçılım grafiği



Şekil 19. Sosyal saçılım grafiği

Şekil 16'da aile desteği değişkenine ait gözlenen değerler % 95 güven aralığında düzeltirici bir spline (uzanım) ile doğrusal bir yapı sergilemektedir. Şekil 17-18-19'da sırasıyla hırs, işbirliği ve sosyal değişkenlerine ait gözlenen değerlerin % 95 güven aralığında düzeltirici bir spline ile kısmi düzeltirmelerini göstermektedir.

TARTIŞMA ve SONUÇ

ÇDR analiz yöntemi ile GAM analiz yöntemi arka planda kullandıkları algoritmalarından dolayı R^2 açısından birbirinden farklı sonuçlar vermiştir. Bağımsız değişkenlerin bağımlı değişkende açıkladıkları varyans miktarı bakımından GAM analiz yönteminin açıkladığı varyans miktarının (% 25,3) ÇDR analiz yönteminden (% 23,8) daha yüksek olduğu görülmektedir. Araştırmada kullanılan veri setinin yapısı itibariyle gerek bağımlı değişkenin normal dağılım sergilemesi gerekse bağımsız değişkenlerin bağımlı değişken ile uyumlu bir yapıda olması bu farkın düşük kalmasında önemli olduğu düşünülmektedir. Finch ve diğerleri (2011) okul öncesi çocuklar üzerinde yaptıkları çalışmada GAM analiz yönteminin (% 25) ÇDR analiz yöntemine (% 22) göre daha yüksek açıklanan varyansa sahip olduğunu belirtmiştir. Benzer şekilde Rawal ve diğerleri (2019) yaptıkları çalışmada açıklanan varyansı ÇDR analiz yöntemine (% 45) göre GAM analiz yöntemi için (% 58) daha yüksek elde etmiştir. GAM analiz yönteminin ÇDR analiz yöntemine göre açıklanan varyans açısından daha yüksek elde edilmesinde GAM analiz yönteminin doğrusal tahmin edicileri olarak düzeltirici fonksiyonları kullanmasının etkili olduğu söylenebilir.

GAM analiz yöntemi normallik doğrusallık ve homojenlik gibi varsayımlara gereksinim duymayan güçlü robust bir yöntemdir (Hastie, Tibshirani ve Friedman, 2009). GAM analiz yöntemi bu tür verilerin incelenebilmesine olanak tanımasından ötürü Eğitim Bilimleri çalışmalarında kullanılması önerilmektedir. Bu anlamda GAM, varsayımların sağlanamadığı veriler için ÇDR analiz yöntemine alternatif olarak, neden sonuç temelli çalışmalarda kullanılabilceği düşünülmektedir.

KAYNAKÇA

- Büyüköztürk, Ş. (2011). *Sosyal Bilimler İçin Veri Analizi El Kitabı* (14. Baskı). Ankara: Pegem Yayıncılık.
- Büyüköztürk, Ş., Çokluk, Ö., ve Köklü, N. (2011). *Sosyal Bilimler için İstatistik* (8. Baskı). Ankara: Pegem Yayıncılık.
- Finch, W. H., Chang, M., Davis, A. S., Holden, J. E., Rothlisberg, B. A. ve McIntosh, D. E. (2011). The prediction of intelligence in preschool children using alternative models to regression. *Behavior research methods*, 43(4) 942–952.
- Kan Kılınç, B. ve Çavuş, M. (2017). Comparative simulation study for model adequacy with binary response variable under multicollinearity – nonparametric approaches. *SAÜ Fen Bilimleri Enstitüsü Dergisi*, 21 (2), 169-177.
- Karasar, N. (2009). *Bilimsel Araştırma Yöntemi* (20. Baskı). Ankara: Nobel Akademi Yayınları.
- Kayri, M., Çokluk, Ö., Şevgin, H., Elkonca, F. ve Ceyhan, G. (2012). Öğrencilerin Akademik Erteleme Eğilimlerinin Ordinal Lojistik Regresyon ile İncelenmesi. *Abant İzzet Baysal Üniversitesi Eğitimde ve Psikolojide Ölçme ve Değerlendirme III. Ulusal Kongresi*, 19-21 Eylül, Bolu.
- Hastie, T. J. ve Tibshirani, R. J. (1990). *Generalized Additive Models* (First Edition). London: Chapman and Hall Press.
- Hastie, T., Tibshirani, R. ve Friedman J. (2009). *The Elements of Statistical Learning Data Mining, Inference, and Prediction* (Second Edition). Germany: Springer.
- Montgomery, D. C., Peck, E. ve Vining, G. G. (2013). *Linear Regression Analysis* (5. Edition). Çev. Edt. Erar M. A., Ankara: Nobel Yayınları.
- Rawal, A., Chakraborty, S., Li, B., Lewis, K., Godoy, M., Paulette, L. ve Weindorf D. C. (2019). Determination of base saturation percentage in agricultural soils via portable X-ray fluorescence spectrometer. *Geoderma*, 338, 375-382.
- Savaş, N. ve Cengiz, M. A. (2009). Yaşam Memnuniyetini Etkileyen Faktörlerin Belirlenmesinde Genelleştirilmiş Toplamsal Modellerin Kullanımı. *EÜFBED - Fen Bilimleri Enstitüsü Dergisi Cilt-Sayı: 2-1*.
- Shavelson, R. J. (2016). *Statistics for Social Sciences* (3. Edition). Çev. Edt. Güler N., Ankara: Pegem Yayıncılık.
- Sümbüloğlu, K. ve Akdağ, B. (2007). *Regresyon Yöntemleri ve Korelasyon Analizi*. Ankara: Hatiboğlu Yayıncılık.
- Şenel, T., Cengiz, M. A., Savaş, N. ve Terzi, Y. (2009). Çoklu Doğrusal Regresyonda Model Seçiminde Genelleştirilmiş Toplamsal Modellerin Kullanımı. *EÜFBED-Fen Bilimleri Enstitüsü Dergisi, Cilt 2, Sayı 2*.
- Tabachnick, B. G. ve Fidell, L. S. (2015). *Using Multivariate Statistics* (6. Edition). Çev. Edt. Baloğlu M., Ankara: Nobel Yayınları.
- Özdamar, K. (2013). *Paket programlar ile istatistiksel veri analizi: MINITAB 16-IBM SPSS 21*. Nisan Kitabevi.
- Wood, S. N. (2006). *Generalized Additive Models: an Introduction With R*. USA: CRC Publications.