GÜFBED/*GUSTIJ* (2021) 11 (4): 1145-1153
DOI: 10.17714/gumusfenbil.863338 Araştırma Makalesi / *Research Article*

Genelleştirilmiş lineer modellerde kısmi ve augmented kısmi artıklar ve grafikleri

Partial and augmented partial residuals and plots in generalized linear models

Esin AVCI[*1,a]
[1]*Giresun Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 28200, Merkez, Giresun*

• Geliş tarihi / *Received*: 18.01.2021 • Düzeltilerek geliş tarihi / *Received in revised form*: 14.07.2021 • Kabul tarihi / *Accepted*: 30.07.2021

Öz
Genelleştirilmiş lineer modeller; fen, mühendislik ve sosyal alanlarda sıkça karşılaşılan bağımlı değişkenin kesikli veya sürekli dağılıma sahip olması durumunun, bağımsız değişken(ler)le modellenmesine olanak veren yöntemlerden biridir. Bu çalışmada, bağımsız değişkenlerin modelde yer alma biçimlerinin belirlenmesinde tanısal grafiklerden yararlanılmıştır. Bu amaçla kısmi ve augmented kısmi artıklar ile bu artıklara ait grafikler tanıtılarak, modellemede sıkça kullanılan artık grafikleri ile karşılaştırılmıştır. Uygulama olarak, 1965-1998 yılları arasındaki grev sayılarının sendikalaşma oranı, ücret oranı, işyeri ve kapanan işyeri sayılarıyla olan ilişkisi analiz edilmiş ve en uygun model seçilmeye çalışılmıştır. Analiz sonucunda; kısmi artık ve augmented kısmi artık grafiklerinin, ele alınan tüm bağımsız değişkenlerin modelde yer alma biçimini daha belirgin biçimde gösterdiği sonucuna varılmıştır.

Anahtar kelimeler: Artık, Augmented kısmi artık, Genelleştirilmiş lineer modeller, Grev, Kısmi artık.

Abstract
Generalized linear models; It is one of the methods that allows the modeling of the dependent variable having discrete or continuous distribution, which is frequently encountered in science, engineering and social fields, with independent variable(s). In this study, diagnostic graphics were used to determine the form of the independent variables were included in the model. For this purpose, partial and augmented partial residuals and graphs of these residuals were introduced and compared with the residual graphs that are frequently used in modeling. As an application, the relationship between the number of strikes between 1965 and 1998 with the unionization rate, wage rate, number of workplaces and closed workplaces was analyzed and the most suitable model was chosen. It is concluded that partial residual and augmented partial residual graphs show the functions of all the independent variables are included in the model more accurately.

Keywords*: Residual, Augmented partial residual, Generalized linear models, Strikes, Partial residual*

*[a] Esin AVCI; esinavci@hotmail.com, Tel: (0454) 310 53 63, orcid.org/0000-0002-9173-0142

1. Giriş

Lineer ve lineer olmayan regresyon modellerinde istatistiksel analiz yöntemleri bağımlı değişkenin normal dağıldığı varsayımına dayanmaktadır. Bağımlı değişkenin normal dağıldığı durumların yanı sıra sürekli olmadığı durumları analiz etmek için geliştirilmiş modeller de bulunmaktadır. Örneğin tedaviye katılma durumunun modellemesinde bağımlı değişken (katılma=1 ve katılmama=0) değerlerini almaktadır. Bazı durumlarda da belirli bir zaman aralığında bir olayın kaç kez tekrarlandığı; örneğin bir günde meydana gelen ölümle sonuçlanan trafik veya iş kazaları, bir yıl ya da bir günde meydana gelen deprem sayısı veya bir saat içinde bankaya uğrayan müşteri sayısı ilgi konusu olabilmektedir. Bu durumlarda bağımlı değişken kesiklidir. Normal dağılıma sahip olmayan, sürekli değerler alan uygulama alanları da vardır. Bu tür verilerin analizine olanak veren geliştirilmiş modellerden biri Genelleştirilmiş Lineer Modellerdir (GLM).

Genelleştirilmiş lineer modeller, bağımlı değişken dağılımının üstel dağılım ailesi üyesi olduğu varsayımına dayanmaktadır. Üstel dağılım ailesi; binom, negatif binom, Poisson, normal, gamma ve üstel gibi daha genel ve uygulama alanı daha kapsamlı olan dağılımları içermektedir. Üstel dağılım ailesinde varyans ortalamanın bir fonksiyonudur. Bu durum değişen varyansa yol açmaktadır. Bu nedenle etkin bir parametre tahmini için en küçük kareler tahmin yöntemi yerine ağırlıklandırılmış en küçük kareler, Newton-Raphson ve Fisher-Scoring algoritmalarından yararlanılmaktadır (Myres vd., 2001).

Genelleştirilmiş lineer modeller, bağımlı değişkenin normal dağılım göstermediği durumlar için veri dönüştürme yöntemine alternatif olarak kullanılabilmektedir. Bu modellerden elde edilen sonuçlar, ters dönüşüme gerek olmaması ve daha dar güven aralığa sahip olması nedeniyle veri dönüşümüne göre daha doğru sonuçlar verdiği saptanmıştır (Lewis vd., 2001; Mccullagh ve Nelder, 1989).

Genelleştirilmiş lineer modeller ilk defa Nelder ve Wadderburn (1972) tarafından ele alınmıştır. Daha sonraki yıllarda McCullagh ve Nelder (1989), Aitken M. (1989), Lindsey (1997), Uusipaikka (2000), McCulloch ve Searle (2001), Myers vd. (2001), Dobson (2002) genelleştirilmiş lineer modeller teorisi hakkında çalışmaya devam etmişlerdir.

Genelleştirilmiş lineer modellerde, lineer ve lineer olmayan regresyon modellerinde kullanılan grafiksel ve istatistiksel yöntemlerin benzerleri bağımsız değişken(ler)le bağımlı değişken arasındaki ilişkiyi veren bağıntı fonksiyonunu oluşturmak için kullanılmaktadır.

Bu çalışmada, genelleştirilmiş lineer modeller için çeşitli artıklar ve bu artıklardan yararlanan tanısal grafikler verilmiştir. Bu amaçla, TÜİK ve Çalışma ve Sosyal güvenlik Bakanlığından derlenen 1965-1998 yılları arasında Türkiye de yapılan grevler ve bu grevlerin belirleyicileri ile olan ilişkisi sözü edilen yöntemlerle analiz edilmiştir. Veriler analiz edilirken grafiksel ve istatistiksel yöntemlerin uygulanmasında S-plus istatistik paket programı kullanılmıştır.

2. Gereç ve yöntem

Lineer modeller genelleştirilmiş lineer modellerin bir uzantısıdır. Dolayısıyla lineer modellerden yola çıkarak genelleştirilmiş lineer modeller tanımlanacaktır.

Lineer modellerde Y bağımlı değişken vektörü, μ ortalama ve σ^2 varyans ile normal dağıldığı varsayılmaktadır. Bağımlı değişken ortalaması μ ile bağımsız değişkenler $(X_1, X_2, \dots, X_p)$ arasındaki ilişki aşağıdaki gibi olmaktadır.

$$E(Y) = \mu = \sum_{j=1}^{p} x_j \beta_j$$
$$i=1,2,\dots,n \text{ ve } j=1,2,\dots,p \tag{1}$$

Burada β'lar, değeri genelde bilinmeyen ve veriden tahmin edilmesi gereken değerlerdir. Matris gösterimi ile (μ $n \times 1$, X $n \times p$ ve β $p \times 1$ boyutludur)

$$E(Y) = \mu = X\beta \tag{2}$$

Burada X model matrisidir ve β parametre vektörüdür. Model matrisi, modelde olması düşünülen tüm bağımsız değişkenleri içeren matristir.

Genelleştirilmiş lineer modellerde bağımlı değişken dağılımı, birçok dağılımı içeren üstel dağılım ailesine sahiptir. Üstel dağılım ailesinin yoğunluk fonksiyonu aşağıdaki biçimdedir.

$$f(y; \theta, \phi) = exp\left\{\frac{[y\theta - b(\theta)]}{a(\phi)} + c(y, \phi)\right\} \tag{3}$$

Burada $a(.), b(.)$ ve $c(.)$ özel fonksiyonlardır. θ, merkez ölçüsü parametresi ve ϕ, genellikle ölçek parametresi olarak adlandırılır.

Bağımlı değişkene ait ortalama değerinin (μ) bağımsız değişkenlerle ilişkisi bağıntı fonksiyonu ile belirlenir. Bağıntı fonksiyonunu η ile gösterilirse;

$$\eta = g(\mu) = X\beta \tag{4}$$

olur. Burada β, bilinmeyen ve verilerden tahmin edilmesi gereken parametrelerdir. X ise model matrisidir (McCullagh ve Nelder, 1989).

Bağımlı değişkenin dağılımı ile bağıntı fonksiyonları birbirinden bağımsız değildir. Örneğin; binom dağılımında başarı ihtimal değerleri $0 \leq p \leq 1$ aralığında değer alması gerektiğinden, seçilen bağıntı fonksiyonunun bu temel kuralı bozmayacak şekilde seçilmesi gerekmektedir. Ayrıca, seçilen bağıntı fonksiyonları türevi alınabilen ve monoton olmalıdır (Myres vd., 2001). Bağıntı fonksiyonu, lineer modellerde ortalamaya eşit olmaktadır. Lineer modellerde olduğu gibi, bağıntı fonksiyonunun bağımlı değişkenin ortalamasına (beklenen değerine) eşit olduğu durumlarda bağıntı fonksiyonuna "*Özdeş*" denir. Bağımsız değişkenler aralığı oldukça dar olduğunda ve modelde lineer olmama problemi dışında herhangi bir sorun olmadığında, bağımlı değişken dağılımına bakılmaksızın özdeş bağıntı fonksiyonu çok iyi sonuçlar verebilmektedir. θ parametresine eşit olan bağıntı fonksiyonuna "*Kanoniksel*" adı verilmektedir (McCullagh ve Nelder, 1989).

Bağıntı fonksiyonunu doğruluğunu grafiksel olarak kontrol etmenin en kolay yolu; bağımlı değişken değerlerini $\hat{\eta}_i$ ön tahmin değerlerine karşı çizmektir. Grafik düz bir doğru biçimi gösteriyorsa, bu durum bağıntı fonksiyonun doğru tanımlandığını; eğrisel bir biçim gösteriyorsa, bağıntı fonksiyonunun doğru tanımlanmadığını göstermektedir. İkili (Binary) değer alan bağımlı değişkenler için bu grafik bilgi verici değildir.

Genelleştirilmiş lineer modellerde lineer modellerde olduğu gibi, tahmin edilen modelin veriye uyumu ile varsa etkin gözlemlerin ortaya çıkarılması ve bağımsız değişken(ler)in modelde yer alma fonksiyonlarının tespit edilmesi gibi tanısal amaçlar için artıkların ve bu artıklara ilişkin grafiklerin elde edilmesi gerekmektedir.

2.1. *Genelleştirilmiş lineer modellerde artıklar ve grafikleri*

2.1.1. Genel artıklar ve grafikleri

Genel artık olarak adlandırılan artıklar; lineer modellerden bilinen bağımlı değişken değerinin tahmin değerinden farkına oldukça benzemektedir. Ancak genelleştirilmiş lineer modellerde, bağımlı değişken ortalaması ile bağımsız değişkenler bağıntı fonksiyonu yardımıyla ifade edildiğinden, genel artıklar;

$$g(\hat{\mu}_i)(y_i - \hat{\mu}_i) = (\frac{d\eta_i}{d\hat{\mu}_i})(y_i - \hat{\mu}_i) \tag{5}$$

şeklinde elde edilir. Bu artıklar, S-plus istatistik paket programında "working residual" olarak adlandırılmaktadır.

Artık değerinin dikey, ilgilenilen bağımsız değişkenin yatay eksene çizilmesiyle elde edilen tanısal grafiğe "*Artıkların Bağımsız Değişkene Karşı Grafiği*" denilmektedir. Grafik sıfır ortalamalı ve sabit varyanslı bir doğru biçimi gösteriyorsa; ilgilenilen bağımsız değişkenin modele lineer biçimde eklenmesi; eğrisellik durumunda ise, oluşan biçime göre $x^2, \sqrt{x}, log x$ vb. gibi lineer olmayan biçimde eklenmesi gerektiği vurgulanmaktadır.

2.1.2. Kısmı artık ve grafiği

Lineer regresyon için kısmı artık grafiği 1972'de Larsen ve McCleary tarafından tanımlanmıştır. Genelleştirilmiş lineer modeller için kısmı artık grafiği McCullagh ve Nelder (1989) tarafından verilmiş ve bu grafik bağımsız değişkenlerin fonksiyonel yapılarının kontrolü için kullanılmıştır. Landwehr vd. (1984) lojistik regresyonla ağırlıklandırılmış lineer regresyon arasında ilişki kurarak lojistik regresyon için kısmı artık grafiklerini oluşturmuştur. O'Hara Hines ve Carter (1993), kısmı artık grafiklerinin genelleştirilmiş lineer modellerde etkili gözlemlerin değerlendirmesi üzerine çalışmışlardır.

Kısmı artık grafiği, ilgilenilen bağımsız değişkenin lineer olmayan, monoton bir fonksiyonla açıklanmasının iyi olup olmadığını incelemeye yarayan bir tanısal grafiktir.

Bağımlı değişkenin $p \times 1$ ve $q \times 1$ boyutuna sahip $x_{11}, \ldots, x_{1n}$ ve $x_{21}, \ldots, x_{2n}$ bağımsız değişken vektörleri ile açıkladığı varsayılsın. Bu durumda bağımlı değişkene ait ortalama; x_{1i} ve x_{2i} değişkenleri için,

$$g(\mu_i) = \eta_i = \beta_1 x_{1i} + h(x_{1i}) \quad i=1,\dots,n \tag{6}$$

şeklinde kurulan bağıntı fonksiyonu ile açıklansın. Burada β_0 bilinmeyen bir parametre, β_1 $p \times 1$ boyutlu bilinmeyen parametre vektörü ve $h(.)$ise x_{2i} bağımsız değişkenlerinin bir fonksiyonu, g ise bilinen bağıntı fonksiyonudur.

Uygulamada $h(.)$ genellikle bilinmediğinden; doğal bir yaklaşım olarak veriler $g(\hat{\mu}_i) = \hat{\eta}_i$ bağıntı fonksiyonu ve (3) ile verilen olasılık yoğunluk fonksiyonu için tahmin edilmektedir. Burada $\hat{\eta}_i$ lineer ön tahmini;

$$\hat{\eta}_i = \alpha_0 + \alpha_1' x_{1i} + \alpha_2' x_{2i} \quad i=1,\dots,n \tag{7}$$

ile verilir. Burada α_0 bilinmeyen skalar parametre ve α_1 ve α_2 sırasıyla $1 \times p$ ve $1 \times q$ boyutlu bilinmeyen parametre vektörleridir. $\hat{\mu}_i$, (3) olasılık yoğunluk fonksiyonuna ve (7) bağıntı fonksiyonuna sahip y_i'ler için ortalama fonksiyonudur.

(7)'de verilen bağıntı fonksiyonu yeniden x_2 değişkeninin herhangi bir fonksiyonu biçiminde düzenlendiğinde,

$$\hat{\eta}_i = b_0 + b_1' x_{1i} + b_2' l(x_{2i}) \quad i=1,\dots,n \tag{8}$$

elde edilir. x_2 bağımsız değişkeni ile ilgilenildiğini varsayalım. x_2 değişkeni için kısmı artık değeri,

$$\hat{p}_{rt} = \left(\frac{d\eta}{d\hat{\mu}}\right)(y - \hat{\mu}) + b_2' x_2 \tag{9}$$

ile elde edilir. Bu değerin x_2 değişkenine karşı çizimi ile *“**Kısmi Artık Grafiği**”* elde edilir.

$l(x_2) = x_2$ olduğunda kısmı kısmi artık grafiği durumları incelendiğinde verilerden elde edilen regresyon tahmini;

$$\hat{\eta} = b_0 + b_1' x_1 + b_2' x_2 \tag{10}$$

olur. Kısmi artıkların ikinci Taylor seri açılımı ile elde edilen yaklaşık değer;

$$[(\beta_0 - b_0) + (\beta_1 - b_1)' x_1 + h(x_2) - \frac{1}{2}(\eta - \hat{\eta})^2 \frac{\left(\frac{d^2\eta}{d\mu^2}\right)}{\left(\frac{d\eta}{d\mu}\right)^2} + e]$$
$$= [(\beta_0 - b_0) + (\beta_1 - b_1)' x_1 + h(x_2) + \hat{Q}(X) + e] \tag{11}$$

olarak elde edilir. Burada $\hat{Q}(X)$ değeri g bağıntı fonksiyonundaki lineer olmamayı veren ölçüttür. Eğer b_0 ve b_1sırasıyla $\beta_0$ve β_1 değerlerine yeterince yakın ve $|\hat{Q}(X)|$ yeterince küçük ise; o zaman kısmı artık grafiği daha iyi sonuç vermektedir. Dolayısıyla kısmi artık grafiğinin davranışı b_0, b_1 ve $|\hat{Q}(X)|$ 'e bağlıdır.

Kısmi artık grafiğinden daha ileri seviyedeki bilgileri anlamak için kitleye ait kısmi artık değerleri elde edilmelidir.

$$p_{rt} = (y - \tilde{\mu})\left(\frac{d\eta}{d\tilde{\mu}}\right) + \beta_2' x_2 \approx (\beta_0 - \gamma_0) + (\beta_1 - \gamma_1)' x_1 + Q(X) + h(x_2) + \varepsilon \tag{12}$$

Burada $\varepsilon = (y - \tilde{\mu})\left(\frac{d\eta}{d\tilde{\mu}}\right)$, $Q(X) = -(\eta - \tilde{\eta})^2 G(\tilde{\mu})$ ve $E(\varepsilon|x_1) = 0$ olduğundan $\tilde{\eta} = E(\varepsilon|x_2) = E\{E(\varepsilon|x_2)|x_2\} = 0$ olur.

Artık değerindeki $(\beta_0 - \gamma_0)$ terimi, uygulamada çok önemli olmayan h fonksiyonunun eklenmiş sabitle açıklanan kısmıdır.

$(\beta_1 - \gamma_1)' x_1$ ve $Q(X)$ terimleri değişime ve yanlılığa potansiyel katkıları vardır. Yanlılığa olan potansiyel katkıları;

$$E(p_{rt}|x_2) \approx (\beta_0 - \gamma_0) + (\beta_1 - \gamma_1)' E(x_1|x_2) + E(Q(X)|x_2) + h(x_2) \tag{13}$$

ile görülebilir.

Eğer x_1 ve x_2 bağımsız ise; $E(X) = 0$ olduğundan $E(x_1|x_2) = 0$ olur. Dolayısıyla yanlılık sadece $Q(X)$'e yani bağıntı fonksiyonundaki lineer olmamaya bağılıdır. Eğer bağıntı fonksiyonu ortalamaya göre lineerse; $Q(X) = 0$ olur. Çünkü; $(\eta - \hat{\eta}) = 0$ olur. Uygulamada ortalama değerler uç değerlerden uzak değerler aldığı sürece *Q*-yanlılığı artsa bile problem oluşturmaz (Cook ve Croos-Dabrera, 1998).

β_1'in γ_1'e eşit olabilmesi için $E(x_1|x_2)$ koşullu beklenen değerinin x_2'ye göre lineer olması gerekmektedir. β_1, $\gamma_1\gamma_1$'e eşit olduğunda ise; $(\beta_1-\gamma_1)'E(x_1|x_2)$ teriminden kaynaklanabilecek yanlılık ortadan kalkmaktadır. Böylece kısmi artık grafiği, ilgilenilen değişkenin varsayılan modelde doğru biçimde yer almasını sağlamaktadır.

2.1.3. Agumented kısmi artık ve grafiği

Mallows (1986), yılında bağımsız değişkenlerin birbiriyle ilişkili olması nedeniyle bazen kısmi artıkların lineer olmayan fonksiyon etrafında daha düzgün serpilme gösterdiğini belirtmiştir. Cook (1993), augmented kısmi artıklarının $E(x_1|x_2)$ koşullu beklenen değerinin x_2'ye göre daha lineer olmasında esneklik sağladığını ve özellikle ilgilenilen değişken 2. dereceden bir fonksiyona sahip olduğunda etkin olduğunu belirtmiştir. Genelleştirilmiş lineer modeller için augmented kısmi artık;

$$\hat{p}_{aug}=\left(\frac{d\eta}{d\hat{\mu}}\right)(y-\hat{\mu})+b_2'x_2+b_3'x_2^2$$

i=1,…,n (14)

ile tanımlanır. Bu artık değerlerinin x_2'ye karşı çizimiyle "***Augmented Kısmi Artık Grafiği***" elde edilmektedir (Cai ve Tsai, 1999).

3. Bulgular

Bu çalışmada, 1965-1998 yılları arasında Türkiye'de yapılan grevler ve bu grevlerin belirleyicileri Poisson regresyon metodu ile incelenmiştir. Grevlerin belirleyicileri olarak, sendikalaşma oranı, işyeri sayısı, ücret oranları ve kapanan işyeri sayıları kullanılmıştır. Kullanılan veriler TÜİK kaynakları ve Çalışma ve Sosyal güvenlik Bakanlığından derlenmiştir. 1980-1983 yılları arasında askeri yönetim olduğunda grev görülmemektedir. Dolayısıyla bu yıllara ait veriler çıkarılmıştır.

Kolmogorov-Simirnov test istatistiği ile 1965-1998 yılları arasındaki grev sayılarının 106 ortalamalı Poisson dağılımlına sahip olduğu sonucuna varılmıştır (D= 0.5668, p>0.05).

Bağımlı değişkenin gözlemleri arasında ilişki olup olmadığı, Durbin-Watson test istatistiği ile kontrol edilebilir.

$H_0: \rho = 0$ (otokorelasyon yoktur)

test istatistiği sonucunda; d=1.97 elde edilmiştir. Bu değer $d_u = 1.74$ kritik değerinden büyük olduğundan bağımlı değişken gözlemleri arasında otokorelasyon (ilişki) olmadığı sonucuna varılmıştır.

Başlangıç olarak modeli;

$$g(\hat{\mu})=\hat{\eta}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2+\hat{\beta}_3x_3+\hat{\beta}_4x_4 \quad (15)$$

şeklinde tahmin edelim. Burada;

y: 1965-1998 yılları arsındaki grev sayıları.
x_1: 1965-1998 yılları arasındaki sendikalaşma oranı.
x_2: 1965-1998 yılları arasındaki işyeri sayısı.
x_3: 1965-1998 yılları arasındaki ücret oranları.
x_4: 1965-1998 yılları arasındaki kapanan işyeri sayısıdır.

Tahmin edilen parametrelerin anlamlılıkları Wald testi ile incelenebilir. Wald istatistiğine göre, modelde diğer bağımsız değişkenler olduğunda x_1 değişkeninin anlamlılığı için (parametre tahmin değerleri ve standart hataları Tablo 1'den alınmıştır);

$H_0: \beta_1 = 0$ Hipotezine göre test istatistiği aşağıdaki gibi hesaplanmıştır.

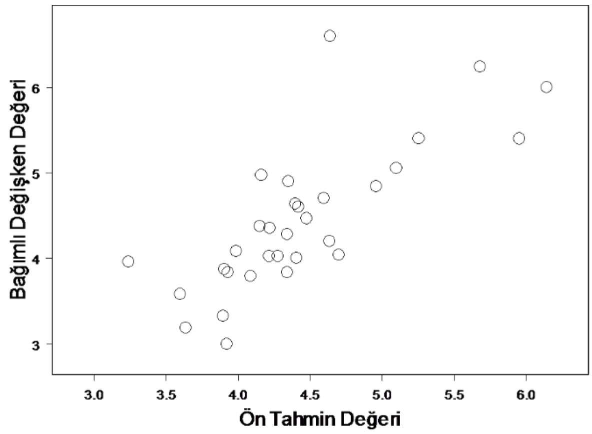
$z_1^2=\left(\frac{0.1449}{0.0427}\right)^2=11.5155$. Kritik değer olan $\chi^2_{1;0.05}=3.841$ den büyük olduğundan katsayı önemli bulunur. Dolayısıyla $x_1$değişkeni model için önemlidir denebilir. Diğer parametreler için benzer şekilde hipotez testi yapılabilir.

Tablo 1. Katsayıların anlamlılık test değerleri

Katsayılar	Parametre tahmin değerleri	Parametre tahminine ait standart hata	z değeri	p değeri
$\hat{\beta}_0$	2.0237	1.0274	1.9697	0.0244*
$\hat{\beta}_1$	0.1449	0.0427	3.3970	0.0003*
$\hat{\beta}_2$	-0.0002	0.00006	-2.6999	0.0034*
$\hat{\beta}_3$	0.0353	0.0065	5.4240	0.0000*
$\hat{\beta}_4$	-0.0008	0.0003	-2.5140	0.0059*
Uyum İyiliği Testi	$\chi^2=1702.26$			0.0000*

Tablo 1'den ele alınan tüm bağımsız değişkenlerin bağımlı değişkeni açıklamada önemli olduğu görülmektedir $(z_j^2 > \chi^2_{1;0.05} = 3.841)$. Uyum iyiliği testinden tahmin edilen modelin istatistiksel olarak anlamlı olduğu görülür ($p<0.05$)

Bağıntı fonksiyonu ile ilgili hataların ortaya çıkarılmasında tanısal bir grafik olan bağımlı değişken değerleri $\hat{\eta}_i$ ön tahmin değerlerine karşı çizildiğinde Şekil 1'deki grafik elde edilmiştir.

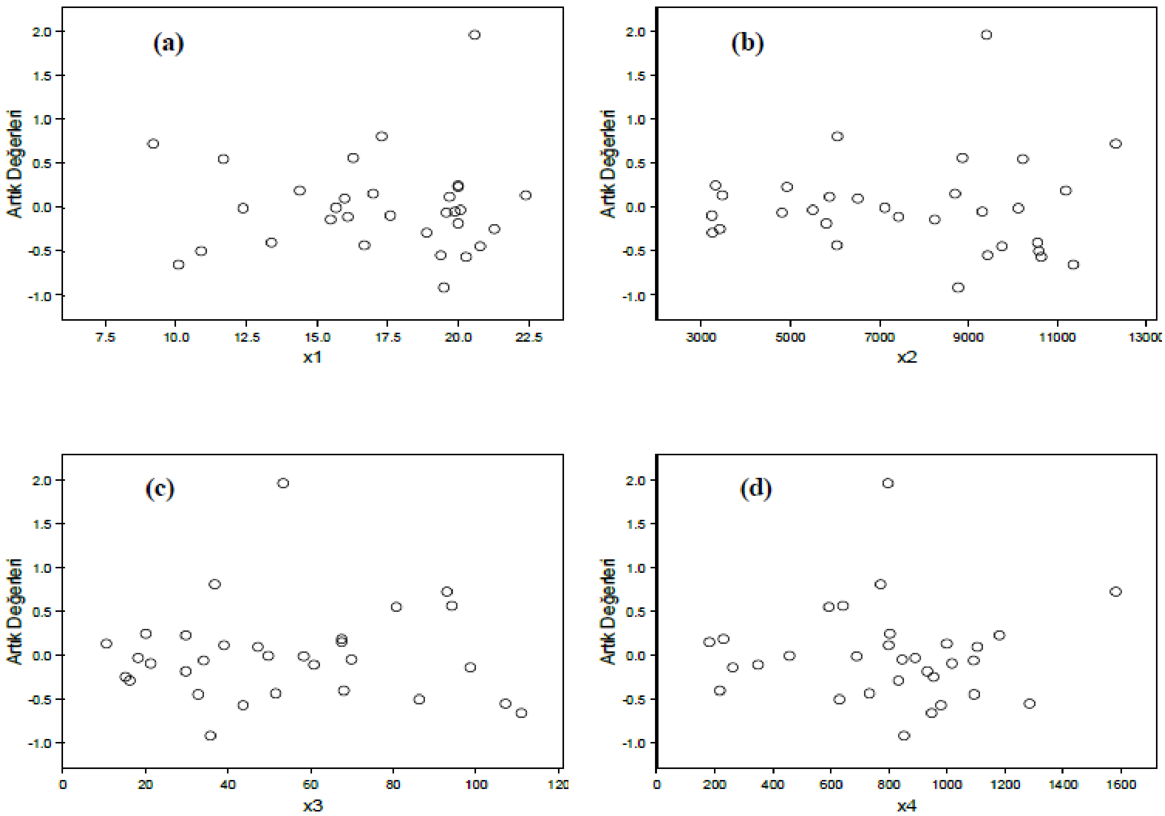


Şekil 1. Bağımlı değişken değerlerinin ön tahmin değerlerine karşı grafiği

Şekil 1'den görüldüğü gibi düz bir doğru biçimi gösterdiğinden bağıntı fonksiyonun doğru seçilmiş olabileceği söylenebilir.

Ele alınan bağımsız değişkenlerin modelde yer alma fonksiyonunun belirlenmesi için sırasıyla; artık kısmi artık ve augmented kısmi artıklardan elde edilen tanısal grafikler incelenmiştir.

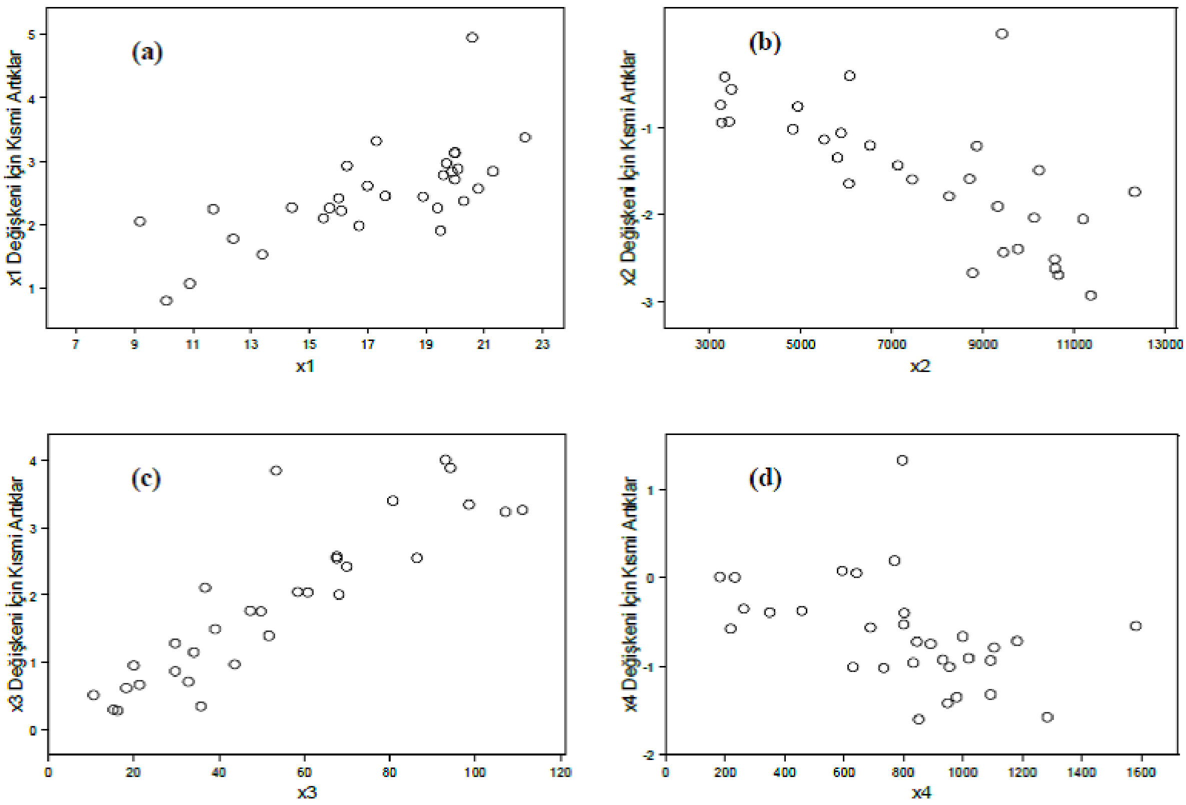
$\left(\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i}\right)(y_i - \hat{\mu}_i)$'den elde edilen artık değeri sırasıyla her bir bağımsız değişkene karşı çizildiğinde aşağıdaki grafikler oluşmaktadır (Şekil 2).



Şekil 2. Artıkların bağımsız değişkenlere karşı grafiği

Şekil 2'den (a), (b), (c) ve (d) grafikleri ele alınan dört bağımsız değişkenin modelde net bir biçimde lineer olarak yer almasını göstermemektedir.

Kısmi artık değerleri; $g(\hat{\mu}) = \hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4$ denkleminden ilgili parametre değerleri hesaplandıktan sonra x_1 değişkeni için $pre = \left(\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i}\right)(y_i - \hat{\mu}_i) + \hat{\beta}_1 x_1$ eşitliği ile ve x_1 değişkeni için $pre = \left(\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i}\right)(y_i - \hat{\mu}_i) + \hat{\beta}_2 x_2$ ile hesaplanmaktadır. Diğer değişkenler için benzer yolla hesaplanabilir. β tahmin değerleri Tablo 1'de verilmiştir. Kısmi artık grafikleri, elde edilen kısmi artık değerlerinin ilgili değişkenlere karşı çizimi ile elde edilmiştir (Şekil 3).



Şekil 3. Kısmi artıkların bağımsız değişkenler için grafiği

Kısmi artık grafikleri modele alınmış olan bağımsız değişkenlerin modelde nasıl bir şekilde yer alması gerektiğini belirten tanısal bir grafik olduğundan, Şekil 3 (a) ve (c) x_1 ve x_3 değişkenlerinin pozitif yönde lineer, Şekil 3 (b) ve (d) ise; x_2 ve x_4 bağımsız değişkenlerinin negatif yönde lineer bir biçimde modele alınması gerektiğini belirtmektedir.

Son olarak, augmented kısmi artık grafiklerinin elde edilmesi için tahmin edilen modele ilgilenilen bağımsız değişkenin karesel terimi eklenir. Bu modelden önce artıklar elde edilir sonra bu artık değerlerine ilgilenilen değişkenin lineer ve karesel terimi eklenerek augmented kısmi artıklar elde edilir. x_1 değişkeni için augumented kısmi artık hesaplanma adımları şu şekildedir:

Önce model,

$$\hat{\eta} = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \hat{\gamma}_3 x_3 + \hat{\gamma}_4 x_4 + \hat{\gamma}_5 x_1^5 \tag{16}$$

olarak tahmin edilir. Buradan elde edilen $\left(\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i}\right)(y_i - \hat{\mu}_i)$ artıkları x_1 değişkeninin lineer ve karesel terimi toplanarak augmented kısmi artıklar elde edilir.

$$p_{aug} = \left(\frac{\partial \hat{\eta}}{\partial \hat{\mu}}\right)(y - \hat{\mu}) + \hat{\gamma}_1 x_1 + \hat{\gamma}_5 x_1^5 \tag{17}$$

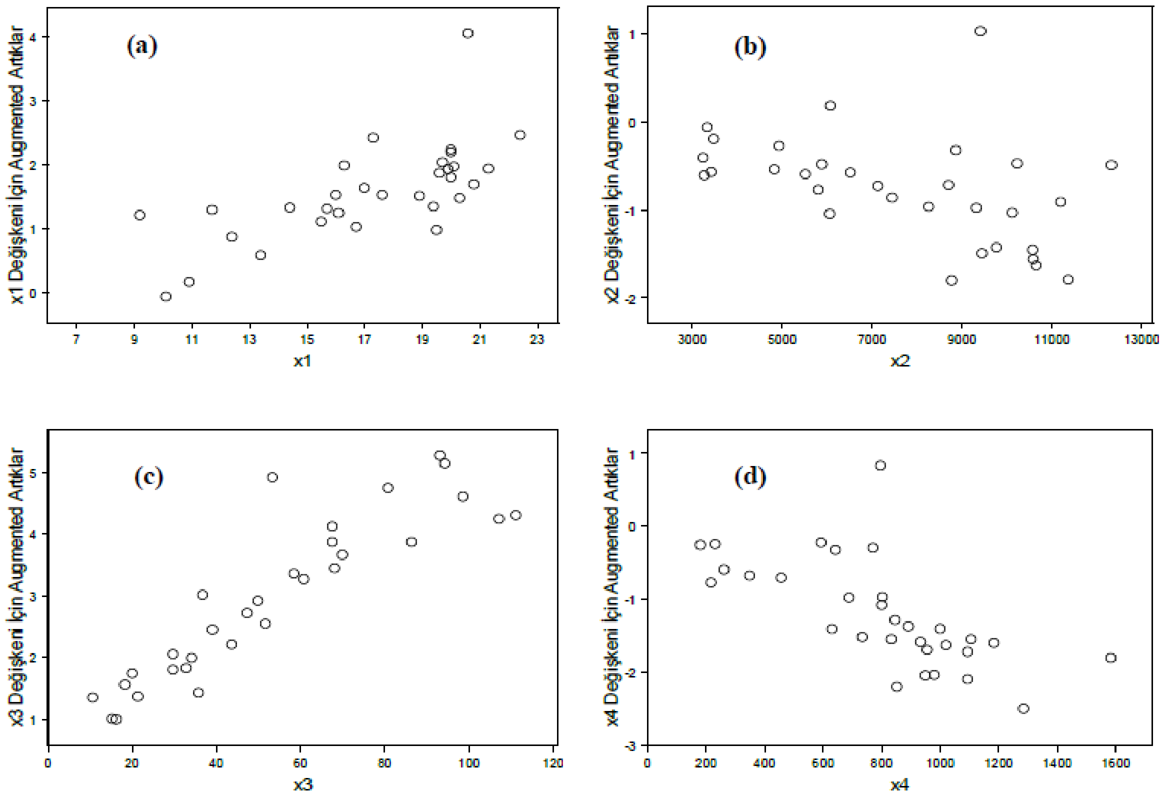
$\hat{\gamma}_1$ve $\hat{\gamma}_5$ katsayıları sırasıyla; 0.0234 ve 0.0038 dir. Diğer değişkenler için benzer biçimde elde edildiğinde $\hat{\gamma}_1$ ve $\hat{\gamma}_5$ katsayı değerleri aşağıdaki tabloda verilmiştir.

Tablo 2. $\hat{\gamma}_1$ve $\hat{\gamma}_5$ katsayı değerleri

Değişken	$\hat{\gamma}_1$	$\hat{\gamma}_5$
x_1	0.0234	0.0038
x_2	-0.0001	0
x_3	0.0761	-0.0003
x_4	-0.0015	0

Augmented kısmi artık grafikleri, elde edilen augmented kısmi artık değerlerinin ilgili değişkenlere karşı çizimi ile elde edilmiştir (Şekil 4).

Augmented kısmi artık grafikleri ilgili bağımsız değişken modele özellikle karesel bir biçimde dahil edilmesi gerektiğinde güçlü tanısal bir grafiktir.



Şekil 4. Augmented kısmi artıkların bağımsız değişkenler için grafiği

Şekil 4 (a) ve (c) x_1 ve x_3 değişkeninin modele pozitif yönde lineer biçimde, Şekil 4.(b) ve (d) ise; x_2 ve x_4 değişkeninin negatif yönde lineer biçimde katılması gerektiğini göstermektedir.

4. Sonuçlar

Genelleştirilmiş lineer modeller, bağımlı değişkenin normal dağılımın yanı sıra binom, negatif binom, Poisson, gamma ve üstel gibi daha genel ve uygulama alanı daha kapsamlı olan dağılımları içermektedir. Bağımlı değişken ortalaması ile bağımsız değişkenler bağıntı fonksiyonu yardımıyla ilişkilendirilmektedir.

Lineer modellerde olduğu gibi genelleştirilmiş lineer modellerde de bağımsız değişken(ler)in modelde yer alma fonksiyonlarının tespit edilmesinde artıklara ve bu artıklara ilişkin grafiklere gereksinim duyulmaktadır. Bu amaçla, yaygın olarak kullanılmayan kısmi artık ve augmented kısmi artık ve grafikleri ele alınmıştır.

Bu çalışmada, genelleştirilmiş lineer modellerin üstel dağılım aile üyelerinden olan Poisson modeli grev sayılarının sendikalaşma oranı, işyeri sayısı, ücret oranları ve kapanan işyeri sayısı ile olan ilişkisinin belirlenmesi için artık grafiklerinden ve istatistiksel analiz sonuçlarından yararlanılmıştır. Analiz sonucunda; kısmi artık ve augmented kısmi artık grafiklerinin yaygın olarak kullanılan artık grafiklerine göre daha belirgin bir biçimde grev sayılarının, sendikalaşma oranı ve ücret oranları ile pozitif, işyeri ve kapanan işyeri sayısı ile negatif yönde lineer bir ilişkisi olduğunu göstermiştir.

Sendikalaşma oranının grev sayılarına olan marjinal etkisi $e^{\hat{\beta}_1 x_1}$ ile belirlenmiştir. Buna göre sendikalaşma oranı %20 olduğunda ortalama grev sayıları %18 oranında artış göstermiştir.

Ücret oranları arttıkça grev sayılarında artış görülmesi, grevin temel nedenlerinden biri olan ücretlere enflasyon oranında zam yapılmaması olarak gösterilebilir. Örneğin 1990 yılında %60.4 olan enflasyon oranına %50 ücret artışı verildiğinde ortalama grev sayılarında %6 oranında bir artış görülmüştür.

Endüstri ilişkilerinin ayrılmaz bir parçası olan grev sayıları artan işyeri sayısı ile azalma göstermektedir. Örneğin işyeri sayısı 3000 olduğunda ortalama grev sayıları %45 oranında azalma göstermiştir. Bu durum işsizlik sorununun varlığına bağlanabilmektedir. Benzer biçimde kapanan işyeri sayısı 1000 olduğunda ortalama grev sayıları %55 oranında azalma göstermiştir.

Kaynaklar

Aitken, M., Anderson, D., Francis, B. and Hinde, J. (1989). *Statistical modelling in GLIM*. Oxford: Clarendon Press.

Cai, Z. and Tsai, C. (1999). Diagnostics for nonlinearity in generalized linear models. *Computational Statistics&Data Analysis*, *29*, 445-469. https://doi.org/10.1016/S0167-9473(98)00079-6

Cook, R.D. (1993). Exploring partial residual plots. *Technometrics*, *35*, 351-362.

Cook, R.D. and Croos-Dabrera, R. (1998). Partial residual plots in generalized linear models. *American Statistical Association*, *93*(442), 730-739. https://doi.org/10.1080/01621459.1998.10473725

Dobson, A.J. (2002). *An iintroduction to generalized linear models* (2nd ed). London: Chapman and Hall.

Landwehr, J.M., Pregibon, D. and Shoemaker, A.C. (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, *79*(385), 61-71.

Larsen, W.A. and McCleary, S.J. (1972). The use of partial residual plot in regression analysis. *Technometrics*, *14*, 781-790.

Lewis, S.L., Montgomery, D.C. and Myers, R.H. (2001). Examples of designed experiments with nonnormal responses. *Journal of Quality Technology*, *33*, 265-278. https://doi.org/10.1080/00224065.2001.11980078

Lindsey, J.K. (1997). *Applying generalized linear model*. New York: Springer Verlag.

Mallows, C.L. (1986). Augmented partial residuals. *Technometrics*, *28*, 313-319.

McCullagh, C.E. and Searle, S.R. (2001). *Generalized linear and mixed effects*. America: John Wiley&Sons, Inc.

McCullagh, P. and Nelder, J.A. (1989). *Background an outline of generalized linear models. İçinde: Generalized linear models* (2nd ed.). London: Chapman and Hall.

Myers, R.H., Montogomery, D.C. and Vining G.G. (2001). *Generalized linear models with applications in engineering and science*. John Wiley and Sons. Inc.

Nelder, A.J. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, *135*, 370-384.

O'Hara Hinest, R.J. and Carter, E.M. (1993). Improved added variable and partial residual plots for the detection of influential observations in generalized linear models. *Applied Statistics*, *42*(1), 3-20. https://doi.org/10.2307/2347405

Uusipaikka, E. (2000). *Confidence intervals in generalized regressions models*. Boca Raton: CRC Press.