T.C. Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:6, Sayı:2, Yıl:1991

REGRESYON ANALİZİNDE DEĞİŞKENLER ÜZERİNE UYGULANAN TRANSFORMASYONLAR

Ali Şen (*) Cenk Özler (**)

ÖZET

Bu makalede regresyon problemlerinde bağımlı ve bağımsız değişkenler üzerine uygulanan bazı transformasyon yöntemleri tanıtılmıştır. Ayrıca transformasyon parametresinin nasıl seçilecegi üzerinde durulmuştur.

1- GİRİŞ

Regresyon analizi iki veya daha çok sayıdaki değişkenler arasındaki ilişkiyi göstermek amacıyla kullanılan bir tekniktir.

Değişkenler arasındaki ilişkiler üçe ayrılırlar:

1) Deterministik ilişkiler,

2) Yarı deterministik ilişkiler,

3) Stokastik ilişkiler.

Deterministik ilişkiler, değişkenler arasında kesin bir ilişkiyi gösterdiklerinden regresyon analizi kapsamında değildir. Bu tür ilişkilere örnek olarak fizik olaylarını gösterebiliriz. Yarı deterministik ve stokastik ilişkiler regresyon analizi kapsamına girerler.

$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$, matris notasyonunda doğrusal regresyon modelini gösterir. $\underline{Y}$ responsları, $\underline{X}$ açıklayıcı (bağımsız) değerleri, $\underline{\beta}$ parametreleri, $\underline{\varepsilon}$ hata terimlerini gösterir. Ayrıca,

$$\underbrace{\underline{Y}_{(nx1)}}_{\text{Toplam değişkenlik}} = \underbrace{\underline{X}_{(nxp)}\,\underline{\beta}_{(px1)}}_{\text{Açıklanan değişkenlik}} + \underbrace{\underline{\varepsilon}_{(nx1)}}_{\text{Açıklanamayan değişkenlik}}$$

olarak yazılabilir

(*) D.E.Ü. İ.İ.B.F. Ekonometri Bölümü Yardımcı Doçenti

(**) D.E.Ü. İ.İ.B.F. Ekonometri Bölümü Araştırma Görevlisi f(x,y), x ve y değişkenlerinin birleşik fonksiyonu olsun. x'in verilen bir değeri için y'nin beklenen değerini hesaplarsak

$$E(y / X) = \frac{\int_{-\infty}^{\infty} y \, f(X, y) \, dy}{\int_{-\infty}^{\infty} f(X, y) \, dy}$$ olur [6].

$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$ modelinde $\underline{Y}$'nin beklenen değerini hesaplarsak sonuçta $E(\underline{Y}) = \underline{X}\,\underline{\beta}$ bulunur.

Klasik En Küçük Kareler (KEKK) yöntemiyle tahmin edilen doğrusal regresyon modeli bazı varsayımlara dayanmaktadır. Bu varsayımlar:

1) $\varepsilon_i \sim N(0, \sigma^2)$

2) σ^2 her gözlem için sabittir.

3) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad , \quad i \neq j$

4) $Cov(x_i, x_j) = 0 \quad , \quad i \neq j$

5) $Cov(\varepsilon_i, x_j) = 0$

Gauss-Markov teoremine göre,hataların ortalaması sıfır ve varyansının sabit olması halinde, EKK tahminleyicileri BLU [(best, linear,unbiased)=(en iyi, doğrusal, sapmasız)] özelliklerini taşırlar. EKK tahminleyicilerinin bütün tahminleyiciler içerisindeki yeri aşağıdaki şemayla gösterilebilir:

Tüm tahminleyiciler

Doğrusal tahminleyiciler

Sapmasız tahminleyiciler

• EKK tahminleyicileri (minimum varyanslı)

Doğrusal tahminleyiciler, bütün tahminleyicilerin alt kümesidir. Aynı şekilde sapmasız tahminleyicilerde doğrusal tahminleyicilerin bir alt kümesidir. EKK tahminleyicileri ise sapmasız tahminleyiciler içinde tek bir noktadır. Ayrıca EKK tahminleyicileri minimum varyanslıdırlar.

2- REGRESYON DEĞİŞKENLERİ ÜZERİNE UYGULANAN TRANSFORMASYONLAR

Transformasyon, gözlemin taşıdığı özelliği değiştirmeden onu başka bir biçime sokmak anlamına gelmektedir.

Günümüzde istatistikte transformasyonlar iki amaç için kullanılmaktadır [8] :

1) Teorik amaçlar veya genel uygunluk için yaklaşımlar sağlamak

2) Veri analizine yardımcı olmak.

Tukey (1957)' ye göre aşağıdaki şartlar sağlandığında verilerin analizi daha kolay yapılır [8]:

(1) Etkiler eklenebilir olmalı

(2) Hataların değişkenliği sabit olmalı

(3) Hataların dağılımı simetrik ve normal olmalıdır.

Box ve Cox (1964)'un şartları ise şunlardır [2]:

(1) E(y)'ler basit yapıda olmalı

(2) Hataların değişkenliği sabit olmalı

(3) Hataların dağılımı normal olmalı

(4) Gözlemler birbirinden bağımsız olmalıdır.

Doğrusal regresyon modeli bu şartları sağlamadığı takdirde transformasyonlara başvurulur.

Şimdi bazı transformasyon ailelerini tanıyalım:

Tukey (1957), pozitif $y > 0$ responsları için kuvvet transformasyonları ailesini kullanmıştır [8].

$$
y^{(\lambda)} = \begin{cases} y^{\lambda} & , \quad \lambda \neq 0 \\ \log(y) & , \quad \lambda = 0 \end{cases} \qquad (1)
$$

Box ve Cox (1964), pozitif y > 0 responsları için daha önce Tukey (1957) tarafından kullanılan monotonik kuvvet transformasyonları ailesinin genelleştirilmesi üzerinde çalıştılar [2].

$$
y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & , \quad \lambda \neq 0 \\ \log(y) & , \quad \lambda = 0 \end{cases} \qquad (2)
$$

Buradaki λ, transformasyonu tanımlayan bir parametredir. Bu transformasyon aileleri log, karekök, ters transformasyonlar gibi özel durumları kapsamaktadır. Ayrıca $\lambda = 0$ olduğunda süreklidir. $y^{(\lambda)}$, $\lambda \geq 1$ olduğu durumlarda dışbükey, $\lambda < 1$ olduğu durumlarda içbükeydir. Bu formüller y'nin eğrisel ve pozitif olduğu durumlarda, kullanışlı ve duyarlıdır. Varyans analizinde elde edilen sonuçlar (1)'de (2)'de yapılacak bir transformasyon ile değişmeyeceğinden bu iki fonksiyon birbiriyle eşdeğerdedir.

Negatif responsların varolması durumunda, genişletilmiş kuvvet transformasyonları ailesi bize yardımcı olacaktır [2].

$$
y^{(\lambda)} = \begin{cases} \dfrac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & , \quad \lambda_1 \neq 0 \\ \log(y + \lambda_2) & , \quad \lambda_1 = 0 \end{cases} \qquad (3)
$$

Burada da $y + \lambda_2 > 0$ ya da $y > -\lambda_2$ olmak zorundadır. Bazı durumlarda λ_2'nin yerine uygun değerini koyarak değiştirilmiş $y + \lambda_2$ responslarını (1) ve (2) de kullanabiliriz.

John ve Draper (1980) ise modulus transformasyonlarını geliştirmişlerdir [3].

$$y^{(\lambda)} = \begin{cases} \text{sign}(y)\left[\dfrac{(|y|+1)^{\lambda} - 1}{\lambda}\right] & , \ \lambda \neq 0 \\ \text{sign}(y)\left[\log(|y|+1)\right] & , \ \lambda = 0 \end{cases} \tag{4}$$

Bu transformasyon ailesi monotoniktir. $\lambda = 0$ olduğu durumlarda süreklidir ve negatif responslar için uygun bir transformasyon ailesidir. Bütün responslar pozitif olduğunda modulus ailesi, özel bir durum olarak kuvvet ailesine benzeyecektir.

[0, b] gibi bir aralıkta sınırlandırılmış $y > 0$ responsları için Mosteller ve Tukey (1977) ve Atkinson (1982) tarafından katlı-kuvvet transformasyonları geliştirilmiştir [3].

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - (b-y)^{\lambda}}{\lambda} & , \ \lambda \neq 0 \\ \log(y/(b-y)) & , \ \lambda = 0 \end{cases} \tag{5}$$

Eğer responslar 0 ve b'nin civarında toplanmışsa bu aile kuvvet ailesine yaklaşacaktır.

Yukarıda ele alınan tüm transformasyonlar dönme (rolation), öteleme ve genleşmenin kombinasyonu olan konform (transformasyondan önce üç noktadan geçen açıları transformasyondan sonra da koruyan) transformasyonlar olup, doğrusal transformasyonlardır [7].

Bağımsız değişkenler üzerine yapılan transformasyonlardan biri olan Ridge Regresyon'u inceleyecek olursak:

Ridge tahminleyicileri çoklu doğrusal bağlantı durumlarında EKK'ya alternatif olarak Hoerl ve Kennart(1970) tarafından kullanılan bir sapmalı tahminleyiciler sınıfıdır.

Standart doğrusal regresyon modeli

$$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$$

$\underline{\beta}$ ' nın EKK tahminleyicileri

$$\hat{\underline{\beta}}_{LS} = (\underline{X}^T\underline{X})^{-1}\,\underline{X}^T\underline{Y}$$

$\underline{\beta}$' nın ridge tahminleyicileri ise

$\hat{\underline{\beta}}(\lambda) = (\underline{X}^T\underline{X} + n\lambda I)^{-1}\underline{X}^T\underline{Y}$, $\lambda \geq 0$ şeklindedir.

Ridge tahminleyicileri taşıdıkları bazı özellikler nedeniyle EKK tahminleyicilerine alternatif olmaktadır. Mesela ridge tahminleyicilerinin hata kareler toplamı EKK tahminleyicilerinkinden daha küçük olabilmektedir.

$$E \,\|\, \underline{\beta} - \underline{\beta}_n(\lambda) \,\|^2 < E \,\|\, \underline{\beta} - \underline{\beta}_n(0) \,\|^2$$

Bu eşitsizlik bazı λ değerleri için ridge tahminleyicilerinin EKK tahminleyicilerinden daha küçük kareler ortalamasına sahip olabileceğini göstermektedir [4].

3- TRANSFORMASYON PARAMETRESİNİN SEÇİMİ:

Box ve Cox (1964) λ 'nın seçimi için En Yüksek Olabilirlik (Maximum-Likelihood) ve Bayesian metodları üzerinde durmuşlardır.

Tüm λ değerleri için $y^{(\lambda)}$, y'nin monotonik bir fonksiyonu olarak kabul edilir. Ayrıca, bilinmeyen λ parametresi için transformasyona tabi tutulmuş responsların oluşturduğu vektör $\underline{Y}^{(\lambda)} = (y_i^{(\lambda)})$ aşağıdaki gibi yazılabilir.

$$\underline{Y}^{(\lambda)} = \underline{X}\,\underline{\beta} + \underline{\varepsilon} \qquad (1)$$

Burada $\underline{X}$ bağımsız değişkendir. $\underline{\varepsilon}$ ise hata terimidir. $\underline{\varepsilon}$, diğer değişkenlerden bağımsızdır ve yaklaşık olarak sıfır ortalamaya ve σ^2 varyansa sahip normal dağılım gösterir. $Y^{(\lambda)}$'nın beklenen değeri ise;

$$E(\underline{Y}^{(\lambda)}) = \underline{X}\,\underline{\beta} \qquad (2)$$

olur. En Yüksek Olabilirlik metoduyla λ' yı tahminlemek istersek:

Transformasyona tabi tutulmamış gözlemlerin olasılık sıklık fonksiyonları aşağıdaki gibidir:

$$J(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}(\underline{Y}^{(\lambda)}-\underline{X}\,\underline{\beta})^T(\underline{Y}^{(\lambda)}-\underline{X}\,\underline{\beta})\right\} \qquad (3)$$

Burada J, transformasyonun Jakobyen'idir ve

$$J(\lambda : Y) = \prod_{i=1}^{r}\left|\frac{dy_i^{(\lambda)}}{dy_i}\right|$$ şeklinde gösterilir.

En Yüksek Olabilirlik tahminleri iki aşamada bulunur. İlk olarak verilen bir λ için Olabilirlik fonksiyonu, bir tamsayı dışında EKK problemi için yoğunluk fonksiyonu olarak ele alınır. Bu durumda $\underline{\beta}$'nın En Yüksek Olabilirlik tahminleri $y^{(\lambda)}$ fonksiyonu için EKK tahminleridir. Buradaki σ, λ'nın varyansı olan $\hat{\sigma}^2(\lambda)$ dir.

$$\hat{\sigma}^2(\lambda) = (\underline{Y}^{(\lambda)})^T\,\underline{X}_r\,\underline{Y}^{(\lambda)}/n = S(\lambda)/n \qquad (4)$$

$$\underline{X}_r = I - \underline{X}(\underline{X}^T\underline{X})^{-1}X^T \qquad (5)$$

$S(\lambda)$ ise $y(\lambda)$ 'nın varyans analizindeki hata kareler toplamıdır.

Sabit bir λ için maksimize edilmiş log Olabilirlik

$$L_{max}(\lambda) = -\frac{1}{2} n \log \hat{\sigma}^2(\lambda) + \log J(\lambda : \underline{Y}) \qquad (6)$$

olacaktır. Sabit terim gözönüne alınmamıştır.

$Z^{(\lambda)} = Y^{(\lambda)} / J^{1/n}$ normalize transformasyonu kullanılarak

$$L_{max}(\lambda) = -\frac{1}{2} n \log \hat{\sigma}^2 (\lambda : \underline{Z})$$

elde edilir.

λ'nın En Yüksek Olabilirlik tahmini (6) veya (7)' nin

$\frac{dL_{max}(\lambda)}{d\lambda} = 0$ şeklinde maximize edilmesiyle elde edilir.

Bayesian düşünceye göre;

$v_r = n - \text{rank}\, \underline{X}$ ve sonra

$$S^2(\lambda) = \frac{\underline{Y}^{(\lambda)T} \underline{X}_r \underline{Y}^{(\lambda)}}{v_r} = \frac{S(\lambda)}{v_r}$$

olur. Sonuçta

$$L_b(\lambda) = -\frac{1}{2} v_r \log(S(\lambda : \underline{Z}) / v_r) \qquad (8)$$

bulunur.

Hem $L_{max}(\lambda)$ hem de $L_b(\lambda)$, $S(\lambda : \underline{Z})$' nin monotonik fonksiyonlarıdır. Genelde bu fonksiyonlar birbirleriyle eşdeğerdir [2].

4- SONUÇ VE ÖNERİLER

$y > 0$ koşuluyla bağımlı değişkenin transformasyonunda, En Yüksek Olabilirlik metoduyla tahminlenecek olan λ parametresinin alacağı değerlere göre transformasyonlar aşağıdaki gibi olurlar:

$\lambda = -2$ => $y^{(\lambda)} = y^{-2}$

$\lambda = -1$ => $y^{(\lambda)} = y^{-1}$

$\lambda = -0.5$ => $y^{(\lambda)} =$

$\lambda = 0$ => $y^{(\lambda)} = \log y$

$\lambda = 0.5$ => $y^{(\lambda)} =$

$\lambda = 2$ => $y^{(\lambda)} = y^{2}$

$\lambda = 3$ => $y^{(\lambda)} = y^{3}$

.

.

.

.

Ancak, En Yüksek Olabilirlik metoduyla tahminlenen λ parametresi tam olarak bu değerden birine eşit çıkmayabilir. Mesela $\hat{\beta} = 2.13$ olabilir. Böyle bir sayıyı hesaplamalara katmak bazen zahmetli olabilir.

Uygulanan transformasyonların ne derece başarılı sonuçlar sağladığı ise her gözlem seti için elde edilen vektörünün ortalama karesel hataları ile ölçümlenebilir. Bu konuda açık olan bir çalışma alanı olarak; λ belli iken transformasyon sonrası parametre ve gözlem tahminlerinin asimptotik özellklerinin incelenmesini gösterebiliriz.

SUMMARY

In this article, some transformations applied to dependent and independent variables in regression problems are ccnsidered. It is also considered how the transformation parameter can be selected.

KAYNAKLAR

BARLETT, M.S.(1947) The use of transformations. Biometrics, 3, 39-52

Box, G.E.P and D.R. Cox (1964) An analysis of transformations

J. Roy. Statist. Soc., 18-26, 211-243

COOK, R.D. and Weisberg, (S. 1983a) Residuals and Influence in Regression. London: Chapman and Hall

ERCİL, A. (1984) Cross-Validated ridge regression: theory and simulations ISA. Vol 15: Part 5, 1767-1773

ERGEN, M.Ö. (1978) Transformasyonlar Yönteminin Esasları. Uygulamalı İstatistik Cilt:1, Sayı:1, 51-59

KENDALL M.G. and Stuart A. (1973) The Advanced Theory of Statistics. Griffin: London

KREYSZİG E. (1966) Advanced Engineering Mathematics. John Wiley and Sons Inc. NewYork, London, Sydney

TUKEY. J.W (1957) On the comparative anatomy of transformations Ann. Math. Statist., 28, 602-632