

## ÇOK DEĞİŞKENLİ REGRESYON ANALİZLERİNDE SÜREKLİ BAĞIMSIZ DEĞİŞKENLER ARASINDAKİ ETKİLEŞİMİN İNCELENMESİ

Doç. Dr. Mustafa Baloğlu\*  
Gaziosmanpaşa Üniversitesi - Tokat

### ÖZET

Çok değişkenli regresyon analizi sosyal bilimler araştırmalarında kullanılan genel güçlü ve esnek bir istatistiksel analiz türüdür. Bununla birlikte, çok değişkenli regresyon analizinin en çok ihtiyaç duyulan fakat en az kullanılan özelliklerinden biri etkileşim değişkenlerinin analize entegre edilmesidir. Çalışma bu bağlamda bazı genel metotları açıklamakta ve etkileşim değişkenlerinin çok değişkenli regresyon analizlerine entegre edilmesini, taranmasını, test edilmesini içermektedir.

**Anahtar Sözcükler:** Çok Değişkenli Regresyon Analizi, Etkileşim Değişkeni

### Abstract

Multiple regression analysis is a powerful and flexible statistical technique used in social sciences. It can use single or multiple predictors that may be of continuous, discrete, and/or both levels. One of the most necessary, yet rarely used features of multiple regression analysis is the integration of interaction terms. This study described some general ways of including, detecting, and testing interaction effects of continuous variables in multiple regression analysis.

**Key words:** Multiple Regression Analysis, Interaction Variable in Multiple Regression

\* Makale hakkında yazışma adresi: Doç. Dr. Mustafa Baloğlu, Gaziosmanpaşa Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü, Tokat 60110; Tel: 356 252 1616-3366; Fax: 356 212 1748; e-mail: baloglu@gop.edu.tr veya baloglu@hotmail.com

### Giriş

Basit regresyon analizinin amacı bir kategorik veya sürekli bağımsız değişken (X) yoluyla bir sürekli bağımlı değişkenin (Y) yordanmasıdır. Fakat, araştırma alanı ne olursa olsun, sosyal bilimlerde tek bir değişkenin herhangi bir sosyal yapıyı tam anlamıyla açıklaması mümkün değildir. Bunun sebebi ise sosyal yapıların oldukça karmaşık ve birbiri ile örüntülü olmasıdır. Dolayısıyla iki veya daha çok kategorik ve/veya sürekli değişkenin yordayıcı olarak kullanılabilmesi çok değişkenli regresyon analizi sosyal bilimlerde araştırmalarında daha çok tercih edilmektedir. Genel olarak çok değişkenli regresyon formülü

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e_i$$

#### Formül 1.

şeklinde ifade edilebilir. Bu formülde  $\beta_0$  regresyon kesişimi,  $\beta_1, \dots, \beta_n$  regresyon katsayıları,  $e_i$  ise regresyonun denkleminin hatasıdır.  $X$ 'ler ise her bir yordayıcı değişkendir (bağımsız değişken).  $Y_i$ 'in yordanmasında  $\hat{Y}$  kullanılır ve yordayıcıların hata karelerinin toplamı,  $\sum (Y_i - \hat{Y})^2$ , her zaman olabilecek en düşük değerdedir. Böylece  $Y_i$

$$Y_i = b_0 + b_1 X_1 + \hat{b}_2 X_2 + \dots + b_n X_n$$

#### Formül 2.

şeklinde de ifade edilebilir. Ek olarak,  $\beta_0$  ve  $\beta_1$ 'in parametre tahminleri değişkenlerin ortalama ve standart sapmaları kullanılarak aşağıdaki gibi hesaplanabilir.

$$b_0 = \bar{Y} - b_1 \bar{X} \text{ ve } b_1 = r_{xy} \frac{s_y}{s_x} = \frac{\text{COV}_{xy}}{s_x^2}$$

#### Formül 3.

Formül 1 deki gibi bir çoklu regresyon modelinin en büyük eksikliği, modelin parametrelerinin yalnızca doğrusal eklemeli olmasıdır. Yani,  $Y$ 'nin  $X_1$  (bir bağımsız değişken) üzerinden yordanması,  $Y$ 'nin  $X_2$  (bir başka bağımsız değişken) ile yordanmasından bağımsız olmasıdır. Bir başka deyişle modeldeki değişkenler arasındaki parametrelerin etkileşimi yok (yada sıfır) farz edilmektedir. Bu tür bir model

doğrusal olmayan veya birbirleriyle etkileşim halinde olan değişkenleri kapsamaktadır. Araştırmacılar ise bu tür doğrusal modellerin sosyal bilimlerdeki araştırmalardaki kompleks yapıları açıklamada yetersiz kaldıklarını savunmaktadırlar (Aiken & West, 1996). Bu çalışmanın amacı etkileşim değişkenlerinin çoklu regresyon modellerine nasıl entegre edilebileceğini ve bu tür etkileşimlerin istatistiksel etkilerinin nasıl test edilebileceğini açıklamaktır.

Araştırmacılar çok uzun zamanlardan beri değişkenler arasındaki etkileşimi incelemektedirler, çünkü doğada varolan şeylerin tek başına ve diğerlerinden tamamen bağımsız olduğunu varsaymak pek de akılcı değildir. Bu özellikle de sosyal bilimlerde daha belirgin bir özelliktir. Araştırmacıların değişkenler arasındaki ilişkiyi incelemekte kullandıkları analiz türü faktöryel varyans analizidir. Bir literatür taramasında, etkileşim halinde olan sürekli bağımsız değişkenleri içeren 148 yayınlanmış makale incelenmiş ve bu çalışmaların % 77'sinin etkileşim değişkenli çoklu regresyon analizi kullanması gerekirken, sürekli değişkenleri yapay sürekli değişkene dönüştürmek yoluyla etkileşimsiz varyans analizini kullandıkları görülmüştür (Aiken & West, 1996).

Etkileşim değişkenlerinin çoklu regresyon analizlerinde sıklıkla kullanılmamasının en önemli nedenlerinden birisi çoklu regresyon analizine etkileşim değişkeni entegre etme aşamalarını açıklayan çalışmaların azlığıdır. Özellikle ülkemizde bu alandaki çalışmalar çok yetersizdir. Halbuki teorik olarak bu alandaki çalışmalar yıllar öncesine dayanmaktadır (Aiken & West, 1996). Örneğin, Saunders (1955), ilk olarak etkileşim değişkenlerinin etkisinin çoklu regresyon analizinde test edilmesi ile ilgili metodlar geliştirmiş ve bu yeni yöntemi "ortalayıcı çoklu regresyon (moderated multiple regression)" olarak adlandırmıştır. İlerleyen yıllarda Cohen (1978) konu hakkındaki yöntemleri geliştirmiştir. Aiken ve West (1996) çoklu regresyon analizlerinde etkileşim etkilerini test etmek üzere oldukça detaylı adımlar önermiştir.

Etkileşim değişkeninin etkisini test edebilmek için sürekli değişkenleri daha düşük ölçme düzeylerine indirgemek iki önemli kısıtlamaya sebep olmaktadır. İlk olarak, bir sürekli değişken daha düşük ölçme düzeyine indirildiğinde taşıdığı bilgi miktarı azalmaktadır. Ayrıca, bu çeşit bir değişkenin kullanıldığı analizlerde, testin gerçekte varolan anlamlı bir etkiyi yakalama yeteneği (yani testin gücü) de azalmaktadır ve dolayısıyla tip II tür hata yapma olasılığı artmaktadır. Bu gibi nedenlerden dolayı sürekli değişkenleri kategorik düzeye düşürmek yerine, sürekli değişkenlerin etkileşiminin incelenebileceği çoklu regresyon analizini kullanmak daha uygun olacaktır.

### Çoklu Regresyon Analizinde Etkileşim Değişkeni

İki değişken arasındaki etkileşim bu iki değişkenin çarpımı şeklinde ifade edilebilir. Fakat, bu şekilde bir ifade, ölçümlerdeki düşük güvenilirliğe bağlı olarak regresyon katsayısındaki önyargıyı üstel olarak yükseltmektedir. Aiken ve West (1996) bu durumun engellenmesi için değişkenlerin çarpımı alınmadan önce değişkenlerin merkezileştirilmesini önermişlerdir. Yani değerlerin kayma skorları şekline dönüştürülmesi ve daha sonra çarpımlarının alınmasıdır.

Etkileşim ifadesini de içeren bir genel çoklu regresyon modeli

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \dots + \beta_n X_n + e_i$$

#### Formül 4.

şeklinde yazılabilir. Burada  $X_1 X_2$  birinci ve ikinci değişken arasındaki etkileşimi ifade etmektedir. Formülde  $X_1 X_2$  çarpımı şeklindeki ifade,  $X_1$ 'in Y regresyonunun  $X_2$ 'nin alacağı spesifik değerlere ve  $X_2$ 'nin Y regresyonunun  $X_1$ 'in alacağı spesifik değerlere bağlı olduğunu belirtmektedir. Bu bağımlı ilişkiden dolayı bu tür bir etki şartlı etki olarak da adlandırılmaktadır.

Değişik  $X_2$  değerleri için  $X_1$ 'in Y regresyonunda eğimi test etmek için Formül 4. şu şekilde yeniden yazılabilir:

$$\hat{Y} = b_0 + (b_1 + b_3 X_2) X_1 + b_2 X_2$$

#### Formül 5.

Formüldeki  $b_1 + b_3 X_2$  regresyon denkleminin eğimidir. Denklem değişik  $X_2$  değerleri için çözümlenerek her bir değer için farklı bir regresyon denklemi elde edilebilir.  $X_2$  için hangi değerlerin seçileceği, araştırma sorusuna ve/veya araştırma konusundaki geçmiş literatür verilerine bağlıdır. Cohen ve Cohen (1983) daha genel bir öneri getirerek, ortalamadan bir standart sapma üstü ve bir standart sapma altının  $X_2$  değerleri olarak seçilebileceğini belirtmektedirler. Örneğin,  $X_1$ 'in Y regresyon denkleminde  $X_2$  için ( $\bar{X}_{X_2} = 0$  ve  $SS_{X_2} = 1$ ) ortalamadan bir standart sapma üstü ve bir standart sapma altı değerleri denklemleri şöyle yazılabilir:

$$\hat{Y}_{üst} = b_0 + (b_1 + b_3)X_1 + b_2 \text{ ve } \hat{Y}_{alt} = b_0 + (b_1 - b_3)X_1 - b_2$$

**Formül 6.**

Elde edilen denklemler daha sonra grafik üzerine aktarılarak etkileşim etkisi görsel olarak da incelenebilir.

$X_2$  için farklı değerler kullanılarak değişik regresyon denklemleri bulunduğu sonra bu regresyon denklemlerinin her birinde eğimin istatistiksel olarak sıfırdan farklı olup olmadığı  $t$ -testi analizi ile incelenebilir. İlk olarak, varyans-kovaryans matrisi kullanılarak her bir eğimin standart hatası formül 7 ile bulunur.

$$s = \sqrt{s_{11} + 2X_2s_{13} + X_2^2s_{33}}$$

**Formül 7.**

Formüldeki  $s_{11}$  ve  $s_{33}$   $b_1$  ve  $b_3$ 'ün varyanslarıdır ve  $s_{13}$  ise  $b_1$  ile  $b_3$  arasındaki kovaryanstır. Örneğin, ortalamanın bir standart sapma altı ve bir standart sapma üstü için eğimin standart hatası ( $\bar{X}_{X_2} = 0$  ve  $SS_{X_2} = 1$  olduğu durumda) sırasıyla  $s_{ust} = \sqrt{s_{11} + 2s_{13} + s_{33}}$  ve  $s_{alt} = \sqrt{s_{11} - 2s_{13} + s_{33}}$  olarak bulunur.

Daha sonra da eğim(ler)in istatistiksel olarak sıfırdan farklı olup olmadığı Formül 8 ile test edilebilir.

$$t = \frac{b_1 + b_3 X_2}{\sqrt{s_{11} + 2X_2s_{13} + X_2^2s_{33}}}$$

**Formül 8.**

Bulunan  $t$  değeri  $n-k-1$  serbestlik derecesinde test edilir ( $n$  = örneklem sayısı;  $k$  = bağımsız değişken sayısı). Bizim örneğimiz için istatistiksel manidarlık  $X_2$ 'nin ortalamasının standart sapma altı için  $t_{alt} = \frac{b_1 - b_3}{\sqrt{s_{11} - 2s_{13} + s_{33}}}$  ve bir standart sapma

üstü için  $t_{ust} = \frac{b_1 + b_3}{\sqrt{s_{11} + 2s_{13} + s_{33}}}$  ile test edilebilir.

Benzer şekilde, değişik  $X_1$  değerleri için  $Y$ 'nin  $X_2$  üzerindeki regresyonu formül 9 yardımıyla hesaplanabilir.

$$\hat{Y} = b_0 + b_1 X_1 + (b_2 + b_3 X_1) X_2$$

**Formül 9.**

Formüldeki  $b_2 + b_3 X_1$  değişik  $X_1$  değerleri için Y'nin  $X_2$  üzerindeki eğimidir. Değişik  $X_1$  değerleri için farklı çoklu regresyon denklemleri aynı şekilde bulunabilir ve  $t = \frac{b_2 + b_3 X_1}{\sqrt{s_{22} + 2X_1 s_{23} + X_1^2 b_{33}}}$  t-testi ile n-k-1 serbestlik derecesinde test edilebilir.

**Kaynakça**

- Aiken, L. S., & West, S. G. (1996). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Cohen, J. (1978). Partialled products are interactions; partial powers are curve components. *Psychological Bulletin*, 85, 858-866.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analyses for the behavioral sciences*. (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Saunders, D. R. (1955). The "moderator variable" as a useful tool in prediction. *Proceedings of the 1954 invitational Conference on Testing Problems* (pp. 54-58). Princeton, NJ: Educational Testing Service.