6.ÇOKLU DOĞRUSAL REGRESYON

6.1. Çoklu Doğrusal Regresyon Kavramı

Basit doğrusal regresyonda, tek bağımsız değişken ile bağımlı değişken arasındaki ilişki açıklanmaktadır. Eğer bağımsız değişken sayısı birden çok ise regresyon incelemesine çoklu regresyon denir. İktisadi olaylar bazen tek bağımsız değişken ile açıklanabilse de, birden fazla bağımsız değişkenin etkisinin de incelenmesi yararlı olabilecektir.

n sayıda açıklayıcı değişkeni bulunan bir çoklu doğrusal model $Y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$ şeklinde gösterilir. Bu modelde değişkenlere ait katsayıların herbiri o değişkene ait kısmi eğimdir. Diğer değişkenlerin sabit tutulması durumunda geçerlidir. Örneğin Y'nin X_2 değişkenine göre kısmi eğimi β_2'dir.

Basit doğrusal regresyon modelini açıklarken iki örnek esas alınmıştı. Burada tekrar o örneklere dönelim. Örneğin, şekerpancarı verimi ile kullanılan gübre miktarı arasındaki ilişkiyi açıklamak için oluşturulan doğrusal model, basit doğrusal regresyon modelidir. Diğer bir örnek, bireylerin tüketim harcamalarında görülen toplam değişimin yüzde ne kadarı bireylerin gelirleri ile açıklanabileceğidir (Friedman, 1957).

Bu ilişkileri, matematiksel olarak y=f(x) bağıntısı ile göstermek mümkündür. Ancak, bununla birlikte gözlenemeyen ya da bilinemeyen etkenlerin varlığı değişimin açıklanamayan kısmın oluşturacaktır. Değişimin iki ana kaynağı,

Toplam değişim = Açıklanabilen değişim + Açıklanamayan değişim şeklinde yazılabilir (Tarı, 2006).

Bu şekilde tek bir bağımsız değişken ile açıklayamadığımız bağımlı değişkeni çoklu bağlanım ile incelemek mümkündür. Örneğimizde; şekerpancarı verimi ile kullanılan gübre miktarı arasındaki ilişkiyi açıklamak için oluşturulan doğrusal modelde, buğday verimini etkileyen yağmur, sıcaklık, işgücü vb. gibi diğer etmenlerin de dikkate alınması gerekebilir.

Diğer örneğimizde ise, bireylerin tüketim harcamalarında görülen toplam değişimin yüzde ne kadarı bireylerin gelirleri ile açıklanabilir şeklinde idi. Oysa iktisat teorisine göre, gelir dışında diğer faktörlerin de tüketim harcamalarını etkileyeceği ortaya konmuştur. Fiyatlar, zevk ve alışkanlıklar, zaman gibi değişkenleri modele dahil ederek tüketim harcamalarındaki değişimi daha iyi açıklamak mümkün olabilmektedir.

Bu örneğe göre, basit doğrusal regresyon modeline fiyatları ilave edelim. Tüketim bağımlı değişken olduğu için Y ile, gelir bağımsız değişken X_1 ve fiyatlar bağımsız değişken X_2 ile ifade edilsin.

Aşağıdaki soruları inceleyelim:

1. İki etken birlikte tüketimi ne kadar açıklayabilmektedir?

2. Hep aynı miktarda gelir için (gelir miktarı sabit tutulsun) fiyatın etkisi nedir? Tüketimi ne kadar açıklamaktadır?

3. Aynı fiyat için gelirin tek başına etkisi nedir? Tüketimi ne kadar açıklamaktadır?

Bu sorular, çoklu doğrusal regresyon modelinin incelenmesi sayesinde cevaplandırılabilecektir.

Örneğe ilişkin, iki bağımsız değişkenli anakütle çoklu doğrusal regresyon modelimiz, $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$ şeklinde ifade edilecektir. İktisat teorisine göre, tüketim harcamalarındaki değişmeler, gelir ve fiyat tarafından açıklanmaktadır. Bu modelde, β_0 modelin bilinmeyeni ve sabit terimdir. β_1 ve β_2 ise modelin bilinmeyen parametreleridir. ε_i, modelin hata terimi ve toplam değişim içerisindeki açıklanamayan değişimi ifade etmektedir.

Sonuç olarak, çoklu regresyon şu amaçlar için yapılır:

1. Bağımlı değişken Y'deki değişimi açıklamak,

2. Diğer etkenler ya da değişkenler veri kalmak kaydı ile, bağımlı değişkene etkilerin kestiriminin yapılması,

3. Bağımlı değişken ya da etkilenen değişkene ilişkin ortalama y_i değerlerinin hesaplanması.

6.2. Çoklu Doğrusal Regresyon Denklemi

Anakütle çoklu doğrusal regresyon denklemini bir i'inci gözlem için aşağıdaki gibi,

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i$ yazılabilir.

Ayrıca değişkenler türünden ise,

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$ şeklinde yazmak mümkündür.

Örneklem regresyon denklemi ise,

$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + ... + \hat{\beta}_k X_{ik} + e_i$ olarak ifade edilecektir.

Burada $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ parametreleri $\beta_1, \beta_2, ..., \beta_k$ 'nın tahmin edicileridir.

6.3. Temel Varsayımlar

Basit doğrusal regresyonda olduğu gibi, çoklu doğrusal regresyonda da elde edilen sonuçların geçerli olması temel varsayımların geçerliliği ile ilgilidir. Basit doğrusal regresyon varsayımları, çoklu regresyon için de geçerli olmakla birlikte, ilave olarak iki varsayım daha mevcuttur. Buna göre, çoklu doğrusal regresyonun varsayımları aşağıdaki gibi verilmiştir:

a) Normallik varsayımı: $\varepsilon_i \sim N(0,\sigma^2), i = 1,2,...,n$ için. Hata terimleri normal dağılım özelliği gösterir.

b) Sıfır ortalama: $E(\varepsilon_i) = 0$, i=1,2,…,n. Hata terimlerinin beklenen değeri sıfırdır.

c) Sabit varyans: $V(\varepsilon_i) = \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2 = \sigma^2$’dir. Regresyon hata terimlerinin varyansı sabittir. Eğer veriler bu varsayımı sağlamıyorsa ortaya değişen varyanslılık sorunu çıkar.

d) Otokorelasyon olmaması: $Cov(\varepsilon_i,\varepsilon_j) = E[(\varepsilon_i - E(\varepsilon_i)][(\varepsilon_j - E(\varepsilon_j)] = 0, \quad i \neq j$. Hata terimleri arasında ilişki olması durumu otokorelasyon (ardışık bağımlılık, özilişki) adını alır. Ancak bu varsayıma göre, bir hata terimi kendinden önce ya da sonra yer alan hata terimlerini etkilememelidir. ε_i ’lerin kestirimi olan e_i’lerin birbirleri ile ilişkili olmamaları istenir.

e) Bağımsız değişkenin tesadüfî değişken olmaması: $E(x_i\varepsilon_i) = 0, \; i = 1,2,....., n$. X’ler raslantı değişkeni olduğunda X ve ε değişkenlerinin bağımsız olmaları gereklidir.

f) Çoklu bağlantı olmaması: Bir bağımsız değişken diğer değişkenler tarafından açıklanabiliyor ise, bu değişkenler arasında doğrusal bağımlılık mevcuttur. Bağımsız değişkenlerin ikişer ikişer aralarında yüksek derecede ilişki bulunması halidir. Bu durumda X matrisinin rankı k+1’den küçük olacak, dolayısıyla $(X'X)^{-1}$ matrisinin tersi alınamayacaktır. Uygulamalarda bağımsız değişkenler arasında az ilişki bulunabilmekte ancak yüksek derecede ilişki olması durumunda ise çoklu bağlantı ya da çoklu doğrusal bağlılık ortaya çıkmaktadır (Güriş ve Çağlayan, 2010) (Dikmen, 2012).

6.4. Parametrelerin Tahmini

Basit doğrusal regresyon modelinde olduğu gibi, çoklu doğrusal regresyon modelinin parametreleri de aynı yöntemle tahmin edilebilmektedir. Ancak, çoklu doğrusal regresyon modelinde k sayıda bağımsız değişken ve bağımlı değişkenin tahmin edilmesi söz konusu olduğunda işlemler biraz daha karmaşık hale gelebilmektedir (Dikmen, 2012).