

İstatistik-II Dersi

11.Bölüm: Regresyon Analizi

ÇOKLU DOĞRUSAL REGRESYON ANALİZİ

Basit doğrusal regresyon modelinde, bir bağımsız değişken ve bir bağımlı değişken yer almaktaydı. Çoklu doğrusal regresyon modelinde ise bir bağımlı değişken ve birden fazla bağımsız değişken yer alır. Model şu şekilde kurulur:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon_i$$

Modelde β_0 ; tüm bağımsız değişkenlerin 0 değerini alması durumundaki sabit etkiyi, β_1 ; bağımsız değişken X_1 'deki bir birimlik değişimin bağımlı değişken üzerindeki kısmi etkisini, β_2 ; bağımsız değişken X_2 'deki bir birimlik değişimin bağımlı değişken üzerindeki kısmi etkisini, β_k ; bağımsız değişken X_k 'deki bir birimlik değişimin bağımlı değişken üzerindeki kısmi etkisini ve ε_i ise hata terimini ifade etmektedir.

Çoklu doğrusal regresyonda EKKY

Çoklu regresyon modellerinde, bağımsız değişken sayısı arttıkça kullanılacak formüllerin hesaplaması biraz daha zorlaşmaktadır. Fakat bağımsız değişken sayısı ne olursa olsun kullanılacak formüller tek bağımsız değişkenli EKKY nin genelleştirilmiş halidir.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ çoklu doğrusal regresyon modelindeki parametre tahminlerini elde etmek için gerekli formüller şunlardır:

$$\begin{aligned} Y &= n\hat{\beta}_0 + \hat{\beta}_1 \sum X_1 + \hat{\beta}_2 \sum X_2 \\ YX_1 &= \hat{\beta}_0 \sum X_1 + \hat{\beta}_1 \sum X_1^2 + \hat{\beta}_2 \sum X_1 X_2 \\ YX_2 &= \hat{\beta}_0 \sum X_2 + \hat{\beta}_1 \sum X_1 X_2 + \hat{\beta}_2 \sum X_2^2 \end{aligned}$$

Bu üç denklem birlikte çözülerek parametrelerin tahmin değerleri bulunur.

Çoklu doğrusal regresyonda Ortalamadan Farklar Yöntemi

Parametre tahminlerini elde etmek için gerekli formüller şu şekilde olacaktır:

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 \\ \hat{\beta}_1 &= \frac{(\sum yx_1)(\sum x_2^2) - (\sum yx_2)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2} \\ \hat{\beta}_2 &= \frac{(\sum yx_2)(\sum x_1^2) - (\sum yx_1)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2} \end{aligned}$$

Burada $x_1 = (X_1 - \bar{X}_1)$, $x_2 = (X_2 - \bar{X}_2)$ ve $y = (Y - \bar{Y})$ olarak tanımlanmaktadır.

Örnek: 6 aylık döneme ait ÜFE, Dolar kuru ve beklenti anketi verileri tabloda verilmiştir. Dolar kuru ve beklenti anketinin ÜFE'yi etkilediği bilinmektedir. Buna göre doğrusal ilişkiyi gösteren doğrusal regresyon denklemini EKKY ve ortalamadan farklar yöntemi ile bulunuz.

	ÜFE	DOLAR	BEKLENTİ
OCAK	5,6	2,2	5,7
ŞUBAT	5,8	2,3	5,7
MART	5,7	2,2	5,5
NİSAN	5	2,3	5,2
MAYIS	6,4	2,4	6,2
HAZİRAN	5,8	2,4	6

Örnek: Aylık araba satış miktarlarını, arabalara uygulanan vergi miktarları ve benzin fiyatının etkilediği düşünülmektedir. Aşağıda verilen veriler yardımıyla regresyon denklemini tahmin ediniz.

	Satılan araba	Vergi miktarı	Benzin fiyatı
Ocak	160	13	2,5
Şubat	152	13	2,6
Mart	140	14	2,6
Nisan	160	12	2,4
Mayıs	130	13	2,6
Haziran	122	13	2,8
Temmuz	126	13	2,7
Ağustos	130	12	2,8
Eylül	120	15	3
Ekim	120	15	3,1
Kasım	118	16	3
Aralık	111	16	3,3

Çoklu regresyon modelinin ve katsayıların anlamlılığı

Modelde amaç;

- Bağımsız değişkenler yardımı ile bağımlı değişkeni kestirmek,
- Bağımsız değişkenlerden hangisi-hangilerinin bağımlı değişkeni daha çok etkilediğini bulmak ve aralarındaki karmaşık yapıyı tanımlamaktır. Model anlamlılığı için basit doğrusal regresyondaki gibi ANOVA tablosu ve F testi kullanılır. Regresyon için k serbestlik derecesi, artıklar için n-k-1 serbestlik derecesi ve genel değişim için ise n-1 serbestlik derecesidir.

$$F_0 > F_{\alpha,k,n-k-1}$$

ise H_0 red edilir.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : En az bir \beta_j \neq 0$$

H₀: Kurulan regresyon modeli anlamsızdır

H₁: Kurulan regresyon modeli anlamlıdır.

$$RKT = \sum \hat{y}_i^2 = \hat{\beta}_1 \sum y_i x_1 + \hat{\beta}_2 \sum y_i x_2 + \dots + \hat{\beta}_k \sum y_i x_k$$

$$GKT = \sum y_i^2$$

$$HKT = GKT - RKT$$

Değişimin Kaynağı	Kareler toplamaları	Serbestlik Derecesi	Kareler ortalaması	F testi
Regresyona bağlı değişim	RKT	k	RKO=RKT/k	F=RKO/HKO
Hatalara bağlı değişim	HKT	n-k-1	HKO=HKT/n-k-1	
Genel değişim	GKT	n-1		

Çoklu Regresyon Analizi Varsayımları

- a) Bağımlı değişken bir tesadüfi bir değişkendir ve normal dağılım göstermektedir.
- b) Tahmin hataları tesadüfidir, birbiri ile ilişki göstermezler ($Cov(\varepsilon_i, \varepsilon_j) = 0$) ve ortalaması sıfır ($E(\varepsilon) = 0$) ve varyansı $V(\varepsilon) = \sigma^2 I$ şeklindedir. Bu matris varyans kovaryans matrisidir. Bu matrisin köşegen elemanları σ^2 olup diğer elemanları sıfırdır.
- c) Bağımsız değişkenler rassal değişkenler değildir. Yani bağımsız değişkenlerin aldıkları değerler sabit veya önceden belirlenmiş yada isteğe bağlı olarak seçilmiş değerler olduğu varsayılır.
- d) Hatalar birbirinden bağımsızdır (otokorelasyon yoktur.)
- e) Her bağımsız değişkenin değerlerine ait olan bağımlı değişken değerlerinin alt setleri varyansları birbirine eşittir (esit varyanslılık=homoscedasticity).
- f) Bağımsız değişkenler arasında basit doğrusal ilişkilerin olmaması gerekir. Bağımsız değişkenler arasındaki basit doğrusal korelasyon katsayılarının 0 veya 0'a çok yakın olması şartı şeklinde de açıklanabilen bu varsayıma, istatistikte "**Çoklu Doğrusal Bağlantı**" (**Multicollinearity**) olmama durumu adı verilmektedir.
- g) Bağımsız değişkenler ile hata terimi arasında ilişki yoktur.
 $Cov(\varepsilon_i, X_{ji}) = 0$, $j=1,2,\dots,p$ ve $i=1,2,\dots,n$
- h) Hipotez testlerini yapabilmek ve güven aralıklarını oluşturabilmek için hataların normal dağılıma sahip olduğu varsayılır.

Çoklu regresyon modeli için standart hata ve varyans formülleri

Parametre tahminleri için varyans ve standart sapma formüllerinde de tahminin standart hatası

kullanılmaktadır: $\hat{\sigma} = \sqrt{\frac{\sum \varepsilon_i^2}{n-k}}$ Buna göre formüller şu şekilde olacaktır:

$$Var(\hat{\beta}_0) = \left[\frac{1}{n} + \frac{\bar{X}_1^2 \sum x_2^2 + \bar{X}_2^2 \sum x_1^2 - 2\bar{X}_1\bar{X}_2 \sum x_1x_2}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2} \right] \hat{\sigma}^2$$

$$s(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)}$$

$$Var(\hat{\beta}_1) = \left[\frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2} \right] \hat{\sigma}^2$$

$$s(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$$

$$Var(\hat{\beta}_2) = \left[\frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2} \right] \hat{\sigma}^2$$

$$s(\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_2)}$$

Çoklu regresyon modeli parametreleri için hipotez testleri ve test istatistikleri

Tahmin edilen regresyon parametresinin anlamlılığı test edilerek, yapılan tahminlerin güvenilirliği de araştırılabilir. Regresyon modelinde her parametre tahmini için de hipotez testi kurulabilir. Hipotezler ve test istatistiği:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$t_h = \hat{\beta}_i / s(\hat{\beta}_i)$$

$t_h > t_{n-k; \alpha/2}$ ise hipotez red edilir.

Çoklu regresyon modelinde belirtme katsayısı (Determinasyon Katsayısı= R^2)

Basit doğrusal regresyonda olduğu gibi çoklu doğrusal regresyonda da belirtme katsayısı regresyon kareler toplamının, genel kareler toplamına oranı ile bulunur. k adet bağımsız değişkene sahip model için formüller:

$$R^2 = \frac{\hat{\beta}_1 \sum y_i x_{i1} + \hat{\beta}_2 \sum y_i x_{i2} + \dots + \hat{\beta}_k \sum y_i x_{ik}}{\sum y_i^2}$$

İki bağımsız değişken için belirtme katsayısı formülü

$$R^2 = \frac{RKT}{GKT} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_1 \sum y x_1 + \hat{\beta}_2 \sum y x_2}{\sum y_i^2}$$

Burada $x_1 = (X_1 - \bar{X}_1)$, $x_2 = (X_2 - \bar{X}_2)$ ve $y = (Y - \bar{Y})$ olarak tanımlanmaktadır.

Çoklu doğrusal regresyonda düzeltilmiş belirtme katsayısı da kullanılır.

$$R_{adj}^2 = 1 - \frac{\frac{RKT}{(n-k)}}{\frac{GKT}{(n-1)}}$$

Örnek: Verilen değerler için çoklu regresyon modelini kurarak modelin anlamlılığını, parametrelerin anlamlılıklarını tartışınız. Belirtme katsayısını ve düzeltilmiş belirtme katsayısını bulunuz.

X_1	X_2	Y
2	1	12
5	2	16
8	1	13
9	3	14
8	4	15
7	5	18

Örnek: 3 bağımsız değişkeni olan ve 12 örneklemden oluşturulmuş bir çoklu regresyon modelinden üretilen ANOVA tablosu verilmiştir. Tablodaki harflerin yerine gelecek sayıları bulunuz ve yorumlamaları yapınız. (d.f=serbestlik derecesi, SS =Kareler toplamı, MS=Kareler ortalaması)

Kaynak	d.f	S.S.	M.S.	F
Regresyon	a	d	f	96
Artık	b	e	12	
Toplam	c			

Kaynaklar:

1. Erilli, N.A. (2018). İstatistik-2. Ankara: Seçkin Yayıncılık.
2. Ankara Üniversitesi açık ders notları .Prof.Dr. Olcay Arslan

https://acikders.ankara.edu.tr/pluginfile.php/73932/mod_resource/content/1/Unite_10.pdf