ÇOKLU REGRESYON MODELİ, ANOVA TABLOSU, MATRİSLERLE REGRESYON ÇÖZÜMLEMESİ,REGRES-YON KATSAYILARININ YORUMU

ÇOKLU REGRESYON MODELİ

- Ekonomi ve işletmecilik alanlarında herhangi bir bağımlı değişkeni tek bir bağımsız değişken ile açıklamak mümkün değildir. Ekonomik modeller, genellikle birden fazla sebebin sonucudurlar. Çok fazla sayıda değişken bir araya gelerek bir diğer değişkeni etkileyebilmektedirler. Bu değişkenler aynı zamanda kendi aralarında da birbirlerini etkileyebilmektedir. Bu sebeple, bu tür birden fazla değişkenin kullanılması gereken durumlarda tekli regresyon analizi yapılması mümkün değildir. Birden fazla bağımsız değişken kullanılarak yapılan regresyon analizine "**<u>çoklu regresyon analizi</u>** (multiple regression analysis)" adı verilmektedir.

.....

- X_i'ler bağımsız değişkenleri ve Y de bağımlı değişkeni göstermek üzere en genel çoklu regresyon denklemi;

$$\mathbf{Y=a_0+a_1X_1+a_2X_2+\ldots+a_kX_k+e_i}$$
$$\mathbf{=a_0+\Sigma a_rX_r+e_i}$$

- Çoklu regresyon modelleri de EKK kullanılarak çözülebilir. Tekli regresyonda olduğu gibi tahmini denklem kurularak diğer hesaplamalar yapılır.

İki bağımsız değişkenli modelin EKK ile çözümü:

- Üzerinde hesaplama yapacağımız model iki bağımsız değişken(X_2 ve X_3) ile bir bağımlı değişken (Y) içeren

$$\mathbf{Y=a+bX_2+cX_3+e_i}$$

modeli olacaktır. Bu regresyon denklemine ait **tahmin modeli:**

$$\hat{Y} = \hat{a} + \hat{b}X_2 + \hat{c}X_3$$

Burada e_i **hata terimi:**

$$e_i = Y - \hat{Y}$$

Katsayıların hesaplanışı:

- Çoklu regresyon modelinde de tıpkı tekli modelde olduğu gibi katsayılar hesaplanırken bağımsız değişkenlerin ortalamadan sapmaları kullanılmaktadır. Aşağıda sırası ile b,c ve a katsayılarının nasıl tahmin edileceğine ait formüller verilecektir. Formüller için kullanılacak x_i ve y değerlerinin eşiti olan ifadeler yazılmıştır.(i=1,2,3)

$$x_i = X_i - \overline{X}_i$$

$$y = Y - \overline{Y}$$

$$\hat{b} = \frac{\left(\sum y.x_2\right)\left(\sum x_3^2\right) - \left(\sum y.x_3\right)\left(\sum x_2.x_3\right)}{\left(\sum x_2^2\right)\left(\sum x_3^2\right) - \left(\sum x_2.x_3\right)^2}$$

$$\hat{c} = \frac{\left(\sum y.x_3\right)\left(\sum x_2^2\right) - \left(\sum y.x_2\right)\left(\sum x_2.x_3\right)}{\left(\sum x_2^2\right)\left(\sum x_3^2\right) - \left(\sum x_2 x_3\right)^2}$$

$$\hat{a} = \overline{Y} - \hat{b}\overline{X}_2 - \hat{c}\overline{X}_3$$

…..

- Regresyon katsayıları hesaplanıp regresyon tahmin modeli kurulduktan sonra belirlilik katsayısı olan R^2 hesaplanır. Bu sayede katsayıların anlamlılığı, modelin uygunluğu gözlemlenecektir. Genel çoklu regresyon modeli için R^2 hesabı; (b,c,..,z katsayılar x_i ler de tanımlanan değerlerdir)

$$R^2 = \frac{\hat{b}.\sum y.x_2 + \hat{c}.\sum y.x_3 + \dots\dots\dots + \hat{z}.\sum y.x_k}{\sum y^2}$$

Düzeltilmiş R^2:

- R^2 belirlilik katsayısı çoklu modellerde genellikle yeterli değildir. Çünkü çoklu regresyon modelleri için denkleme yeni değişken ilave edilmesi durumunda R^2 değeri genellikle artmaktadır. Bu yüzden anlamlı bir test yapabilmek için çoklu modellerde düzeltilmiş R^2 hesaplanmalıdır.($\overline{R^2}$)

n:gözlem sayısı k:modeldeki değişken sayısı(bağımsız değişken+bağımlı değişken)

.....

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-k}$$

- Tekli regresyon modellerinde olduğu gibi belirlilik katsayısı 1'e ne kadar yakın ise mevcut olan model o kadar uygundur(anlamlıdır).

.....

- Modelde tahmin edilen katsayıların güvenilirliği standart hata ve varyansın küçüklüğüne bakılarak test edilir. Bu bize tahmin değerlerinin gerçek değerlere uygunluğu için kısmen bir oran vermektedir.
- Regresyon modelindeki bağımsız değişkenlerin katsayıları modelin durumu, anlamlılığı, gücü hakkında bilgi verdiği halde bağımlı ve bağımsız değişkenler arasındaki ilişkinin yönünü ve kuvvetini göstermemektedir. Bu nedenle korelasyon analizi ile bağımlı ve bağımsız değişken veya değişkenler arasındaki ilişkiyi ölçeriz.
- Görüldüğü gibi eklenen yeni değişkenlere ilişkin ufak uyarlamalar ile çoklu regresyon modelleri de tekli regresyon modellerine benzer işlemler ile yorumlanabilmektedir.

ANOVA TABLOSU

- Regresyon modeli için hesaplamalar yapılarak tahmin değerleri bulunduktan sonra **anova tablosu** adı verilen bir tablo hazırlanır.

SST:kareler toplamı → $SST = \sum (Y_i - \overline{Y})^2$

SSE:artıkların kareleri toplamı → $SSE = \sum (Y_i - \hat{Y}_i)^2$

SSR:tahminlerin kareleri toplamı →

$$SSR = \sum (\hat{Y}_i - \overline{Y})^2 = \sum \hat{Y}_i^2 - n\overline{Y}^2$$

.....

Kaynak (sov)	*Serbestlik derecesi(sd veya df)*	*SS*	*MS*	*F*
Model (regresyon)	2-1=1	SSR	MSR= SSR/2-1	MSR/MSE
Artık	n-2	SSE	MSE= SSE/n-2	
Toplam	n-1	SST		

.....

- Anova tablosu içerisindeki F istatistiği ile model parametrelerinin (katsayıların) anlı olup olmadığı test edilmektedir.
- Ayrıca anova tablosu kullanılarak belirlilik katsayısı da hesaplanabilir: R^2=SSR/SST
- R^2 değeri sayesinde Y bağımlı değişkeninin değerleri arasındaki varyasyonun model tarafından ne oranda açıklandığı gözlemlenebilir.

Matrisler İle Regresyon Çözümlemesi:

- Regresyon denklemini matris hesaplamaları ile de bulmak mümkündür. Bunun için eldeki verileri matris olarak ifade etmemiz gereklidir.
- Regresyon modelimiz $Y_i=a_0+a_1X_i+e_i$ (i=1,2,…,n) olsun. Buradaki Y_i ve X_i değerleri sırasıyla veri setinde her bir gözleme karşılık gelen değerledir. O halde her bir i için elimizde aşağıdaki denklem sistemi mevcuttur:

$$Y_1=a_0+a_1X_1+e_1$$
$$Y_2=a_0+a_1X_2+e_2$$
$$\ldots$$
$$\ldots$$
$$Y_n=a_0+a_1X_n+e_n$$

…..

- Bu denklem sisteminin matris olarak ifade edecek olursak:

$$\underline{Y} = \underline{\beta} X + \underline{e}$$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ \dots \\ Y_n \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \dots & \dots \\ \dots & \dots \\ 1 & X_n \end{bmatrix} \quad \underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ \dots \\ e_n \end{bmatrix}$$

Çözüm:

- $\underline{\beta}$, $\underline{\hat{\beta}}$ için EKK tahmin edicisi ise bunun için çözüm;

$$\underline{\hat{\beta}} = \left(X^{'}X\right)^{-1} X^{'} \underline{Y}$$

- Matris çarpımının yapılması ile 2 x 1 tipinde bir matris bulunur. Bu matrisin birinci satırı a_0 katsayısı için, ikinci satırı ise a_1 katsayısı için bir tahmin olup regresyon tahmin modelinde aranan katsayılardır. Bunları yerine yazarak tahmin modeline ulaşılır.

.....

- Ayrıca tahmin modeli kurulduktan sonra Y değerleri için aranan tahmin sonuçları ise matris yoluyla yandaki şekilde hesaplanabilir:
- Bu işlemler sırasında tahminler için yapılan hata ise;

$$\underline{\hat{Y}} = X \underline{\hat{\beta}}$$

$$\underline{\hat{e}} = \underline{Y} - \underline{\hat{Y}}$$

Regresyon Katsayılarının Yorumlanması:

- Tahmin edilen katsayıların yorumu için **değişkenlerin birimi** ve **regresyon denkleminin yapısı** önemlidir.

- **1. Değişkenleri Mutlak Sayılarla Ölçülen Doğrusal Denklemler:** denklem formu $Y=a_0+a_1X_1+\ldots+a_kX_k+e$ şeklindedir. Burada a_0 sabit terim, a_i ler katsayılar, Y bağımlı değişken, X_i ler bağımsız değişkenler, e ise hata terimini göstermektedir. (i=1,2,…,k)

.....

- **Sabit terim:** bağımsız değişkenlerin hepsi birden 0 iken (X_i=0) bağımlı değişken Y' nin alacağı değerdir.
- **Katsayılar:** a_j katsayısı diğer bağımsız değişkenler sabit iken X_j deki bir birimlik değişme Y' yi a_j birim kadar değiştirmektedir.

- **örneğin;** K malına olan talep modeli tahmin edilmiş ve sonuç $Q_t=10-0.5P_t+0.7Y_t$ olarak bulunmuştur.(P_t: fiyat, Y_t:gelir, Q_t:talep) (ölçü birimi milyon TL)
  - a_0=10:K malının fiyatı ve gelir sıfır iken malın talebi 10 milyon TL olacaktır.
  - a_1=-0.5: bu dönemin geliri sabit iken K malının fiyatındaki 1milyon TL'lik artış malın talebini 0.5milyon TL azaltacaktır.
  - a_2=0.7: K malının fiyatı sabit iken bu dönemin gelirindeki 1 milyon TL'lik artış malın talebini 0.7 milyon TL artırmaktadır.

.....

- **2. Değişkenleri % ile İfade Edilen Denklemler**:denklem formu $Y=a_0+a_1X_1+\ldots+a_kX_k+e$ şeklindedir.
  - **Sabit terim:** açıklayıcı değişkenlerdeki değişim % 0 iken açıklanan değişkenin % kaç olduğunu gösterir.
  - **Katsayılar:** diğer açıklayıcı değişkenlerdeki % değişim sabit iken (yokken) X_j değişkenindeki %1lik değişim Y değişkenini % a_j kadar değiştirmektedir.
  - **Örneğin:**E döviz kuru, M para arzı, P tüketici fiyatındaki % değişim, r faiz oranı ve e hata terimini göstermek üzere ilgili regresyon tahmin modeli şöyledir: $\mathbf{E_t=0.9+0.2M_t-0.4r_t+0.8P_t}$

….

– **Sabit terim**: a_0=0.9:diğer tüm faktörler (M, r, P) sıfır iken döviz kurundaki değişim %0.9 olacaktır.
– **Katsayılar**: a_1=0.2:faiz oranı ve fiyatlarda % değişim yokken para arzındaki %1lik artış döviz kurunu %0.2 artıracaktır.

a_2= -0.4 : ???

a_3= 0.8 : ???

ÖRNEK UYGULAMA

Aşağıdaki tabloda X_i'ler babaların, Y_i'ler ise erkek çoçukların boy uzunluklarını cm. cinsinden göstermektedir.

i	X_i	Y_i
1	163	165
2	164	167
3	170	169
4	172	170
5	165	164
6	167	168
7	168	171
8	166	163

Soru:

a) Regresyon denklemini yazın.
b) a_0,a_1,Y ve e değerlerinin tahminlerinin hesaplayın.
c) 169cm boy uzunluğundaki bir babanın oğlunun boy uzunluğunu tahmin edin(kestirin).
d) Anova tablosunu oluşturun ve model için belirlilik katsayısını hesaplayın.

Çözüm: (a)

- $Y_i=a_0+a_1X_i+e_i$ n=8 ve i=1,2,...,8

$$\left.\begin{array}{l} Y_1=a_0+a_1X_1+e_1 \\ \dots\dots \\ Y_8=a_0+a_1X_8+e_8 \end{array}\right\} \Rightarrow$$

$$\underline{Y}=\begin{bmatrix}165\\167\\\dots\\163\end{bmatrix} \quad \underline{\beta}=\begin{bmatrix}a_0\\a_1\end{bmatrix} \quad X=\begin{bmatrix}1 & 163\\1 & 164\\\dots & \dots\\1 & 166\end{bmatrix} \quad \underline{e}=\begin{bmatrix}e_1\\e_2\\\dots\\e_8\end{bmatrix}$$

$$\underline{Y}=\underline{\beta}X+\underline{e}$$

Prof.Dr.A.KARACABEY
Doç.Dr.F.GÖKGÖZ

25

Çözüm: (b)

$$\underline{\hat{\beta}}=\left(X'X\right)^{-1}X'\underline{Y} \Rightarrow$$

$$X'X=\begin{bmatrix}1 & 1 & \dots & 1\\X_1 & X_2 & \dots & X_8\end{bmatrix}\begin{bmatrix}1 & X_1\\1 & X_2\\\dots & \dots\\1 & X_8\end{bmatrix}=\begin{bmatrix}8 & \sum X_i\\\sum X_i & \sum X_i^2\end{bmatrix}$$

$$\left(X'X\right)^{-1}=\frac{1}{8\sum X_i^2-\left(\sum X_i\right)^2}\begin{bmatrix}\sum X_i^2 & -\sum X_i\\-\sum X_i & 8\end{bmatrix}=\begin{bmatrix}429.3699 & -2.5723\\-2.5723 & 0.0154\end{bmatrix}$$

$$X'Y=\begin{bmatrix}1 & 1 & \dots & 1\\X_1 & X_2 & \dots & X_8\end{bmatrix}\begin{bmatrix}Y_1\\Y_2\\\dots\\Y_8\end{bmatrix}=\begin{bmatrix}\sum Y_i\\\sum X_iY_i\end{bmatrix}=\begin{bmatrix}1337\\223155\end{bmatrix}$$

Prof.Dr.A.KARACABEY
Doç.Dr.F.GÖKGÖZ

26

....

$$\hat{\underline{\beta}} = \left(X'X\right)^{-1} X' \underline{Y} = \begin{bmatrix} 56.1965 \\ 0.6647 \end{bmatrix}$$

$$\hat{Y}_i = \hat{a}_0 + \hat{a}_1 X_i = 56.1965 + 0.6647\, X_i$$

$$\hat{\underline{Y}} = X \hat{\underline{\beta}} = \begin{bmatrix} 1 & 163 \\ \dots & \dots \\ 1 & 166 \end{bmatrix} \begin{bmatrix} 56.1965 \\ 0.6647 \end{bmatrix} = \begin{bmatrix} 167.54 \\ \dots \\ 166.55 \end{bmatrix}$$

$$\hat{e} = \underline{Y} - \hat{\underline{Y}} = \begin{bmatrix} 0.43 \\ \dots \\ -3.55 \end{bmatrix} \Rightarrow \sum \hat{e}_i = -0.02 \cong 0$$

ÇÖZÜM: (c-d)

- X= 169 cm ise modelde yerine yazılırsa

$$\hat{Y} = 56.1965 + 0.6647X \cong 168.53cm.$$

- İlgili değerleri bulup anova tablosunu oluşturursak:

$$\overline{Y} = \frac{1}{n}\sum Y_i = \frac{1}{8}1337 = 167.125$$

$$SST = \sum Y_i^2 - n\overline{Y}^2 \cong 59$$

$$SSE = \sum \hat{e}_i^2 = 30.21$$

$$SSR = \sum \hat{Y}_i^2 - n\overline{Y}^2 = SST - SSE = 28.79$$

Anova tablosu:

Kaynak	Sd	SS	MS	F
Model	2-1=1	28.79	28.79/1	28.79/5.035
Artık	8-2=6	30.21	30.21/6	
Toplam	8-1=7	59		

- R^2=SSR/SST=0.48796 o halde Y' ler içindeki değişimin yaklaşık olarak %49'u model tarafından açıklanmaktadır.
- Babaların boyu 1cm arttığı zaman çocukların boyu da 0.6647cm artmaktadır. Ayrıca babaların boyu sıfır iken çocukların boyu 56.1965cm olabilmektedir.

BAŞARILAR