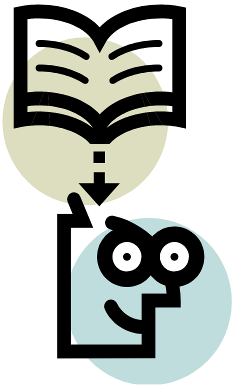

REGRESYON ANALİZİ

13.1 Giriş

İstatistik biliminin en önemli konularından birisini regresyon analizi oluşturmaktadır. Regresyon analizi, araştırma, matematik, finans, ekonomi, tıp gibi bilim alanlarında yoğun olarak kullanılmaktadır. Regresyon analizinin temelinde; gözlenen bir olayın değerlendirilirken, hangi olayların etkisi içinde olduğunun araştırılması yatmaktadır. Bu olaylar bir veya birden çok olacağı gibi dolaylı veya direkt etkileniyor da olabilirler.

Regresyon analizi yapılırken, gözlem değerlerinin ve etkilenilen olayların bir matematiksel gösterimle yani bir fonksiyon yardımıyla ifadesi gerekmektedir. Kurulan bu modele regresyon modeli denilmektedir.

Regresyon analizi incelenirken, genellikle konusunu oluşturan, etkilendiği olaylara değişkenler adı verilir bu değişkenlerin yer alacağı matematiksel model incelenir.

Değişken, belirli bir zaman aralığı gözönüne alınıp, o zaman aralığında bir kütleyi oluşturan belli birimdeki olayları içeren örneklerdir. Sayılabilir veya ölçülebilir nitelikte olmalıdır.

Bir hissenin fiyatını bir değişken alırsak, ona dolaylı olarak veya direkt etkili bir veya birden çok değişken alabiliriz (Örneğin: Faiz oranları, enflasyon, ekonomik, politik, finansal olaylar vs.). Sadece faiz oranlarının etkisi ile ilgileniyorsak, tek değişkenli bir matematiksel model, faiz oranları ile birlikte enflasyon oranı ile de ilgileniyorsak, iki değişkenli bir matematiksel modelden söz ediyoruzdur. Faiz oranları hisse senedinin fiyatını direkt etkileyen bir unsur olmadığı halde faiz oranlarının yükseldiği durumda hisse senedinin fiyatının düşüyor olmasının gözlemlenmesi bir etkileşim olduğunun göstergesidir[1].

Öncelikle Regresyon modelinin kullanılması, ilgilenilen olayla ilgili olarak, bir sebep-sonuç ilişkisi bulunması gerekmektedir. Örneğin 2000-2008 yılları arasındaki hisse senedi fiyatlarını incelersek, seçilen zaman aralığında bir matematiksel model kurma gereği vardır ve bu modelde bir sebep, sonuç ilişkisi aranmaktadır. Sebep, hisse senedinin fiyatını yükselten veya düşüren unsurlardır. Faiz oranları, ekonomik nedenler, enflasyon oranları vs. olarak incelenebilir. Sonuç ise hisse senedinin fiyatının değişmesidir.

Sebep-sonuç ilişkisi, regresyon modeli kurulurken, bağımlı ve bağımsız değişkenler olarak anlatılmaktadır. Yukarıdaki hisse senedi fiyatı sonuç olan bağımlı değişken, faiz oranları, ekonomik nedenler, enflasyon oranları vs. sebep olan bağımsız değişkenlerdir.

Regresyon analizi yapılırken kurulan matematiksel modelde yer alan değişkenler bir bağımlı değişken ve bir veya birden çok bağımsız değişkenden oluşmaktadır.

Tek değişkenli regresyon analizi bir bağımlı değişken ve bir bağımsız değişken arasındaki ilişkiyi inceler. Tek değişkenli regresyon analizi ile bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkiyi temsil eden bir doğrunun denklemi formüle edilir. İçinde bir adet bağımlı değişken ve birden fazla bağımsız değişkenin bulunduğu regresyon modelleri çok değişkenli regresyon analizi olarak bilinir.

13.2 Regresyon Denklemleri

İki olayın değişkenleri X ve Y olmak üzere en uygun regresyon denklemini bulmak için X ve Y değerlerinin grafik üzerinde sırasıyla apsis ve ordinat olarak düşünülmesi faydalı olcaktır. Dağılma grafiğindeki bu noktalar belirli

[1] http://analiz.ibsyazilim.com/egitim/regresyon.html

bir seyir gösterdikleri takdirde Y=f(X) fonksiyonu regresyon denklemi olacaktır.

Bir başka değişle, iki değişken arasında var olan ilişkinin en uygun bir şekilde hangi matematik fonksiyonla ifade edilebileceğinin aratırılması sonucunda elde edilen denklem regresyon denklemidir. Bu denklem değişkenlerden birinin bilinmesi durumunda diğeri için tahmin yapmanıza imkan verir.

Dağılma grafiğinin noktalarının seyri regresyon denkleminin tipi için önemlidir. Noktalar bir doğru etrafında dağılmış ise iki değişken arasında doğrusal bir ilişki düşünülebilir. Eğer dağılışta bir bükülme noktası varsa bu takdirde iki değişken arasında 2.dereceden bir regresyon denklemi bağlantısı olduğu düşünülebilir. Genelleştirecek olursak;

- Hiç bükülme noktasu yok ise regresyon denklemi doğrusaldır.

$$Y = a + bX$$

- Bir bükülme varsa regresyon denklemi 2.derecedendir.

$$Y = a + bX + cX^2$$

- İki bükülme varsa regresyon denklemi 3.derecedendir.

$$Y = a + bX + cX^2 + dX^3$$

- Üç bükülme varsa regresyon denklemi 4.derecedendir.

$$Y = a + bX + cX^2 + dX^3 + eX^4$$

Yukarıdaki denklemlerin yanısıra aşağıdaki tipde de regresyon denklemleri oluşturulabilir:

$$Y = \frac{1}{a + bX}$$

$$Y = ab^X yada\ logY = loga + Xlogb$$

$$Y = aX^b yada\ logY = loga + blogX$$

13.2.1 Normal Denklemleri

İstatistik olaylar arasındaki ilişkinin en iyi şekilde belirtilebilmesi için, noktalar arasından geçirilecek olan doğru ya da eğrinin bu noktalara olan uzaklıları toplamı minimum olmalıdır. Bu nedenle a,b,c gibi paremetrelerin bu şarta uygun olarak belirlenmesi gerekmektedir.

İstatistiki olaylar arasındaki ilişki doğrusal ise ilgili normal denklemleri şu şekildedir:

$Y = a + bX$ doğru denklemi için

$$\sum Y_i = na + b\sum X_i$$

$$\sum Y_i X_i = a\sum X_i + b\sum X_i^2$$

İstatistiki olaylar arasındaki ilişki 2.dereceden ise ilgili normal denklemleri şu şekildedir:

$Y = a + bX + cX^2$ denklemi için

$$\sum Y_i = na + b\sum X_i + c\sum X_i^2$$

$$\sum Y_i X_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3$$

$$\sum Y_i X_i^2 = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4$$

$X - \bar{X} = x \; ve \; Y - \bar{Y} = y$ olmak üzere

$b_{yx} = \frac{\sum xy}{\sum x^2}$ ve $b_{xy} = \frac{\sum xy}{\sum y^2}$ ifadelerine de sırasıyla y'nin x'e göre ve x'in y'ye göre regresyon katsayıları adı verilir.

13.2.2 Tahminin Standart Hatası

X'in belirli bir değerine karşılık gelen Y değerine Y' teorik değeri diyelim. Gerçekleşen değerler ile teorik değerler toplamı birbirine eşit olmakla beraber değerler tek tek karşılaştırıldığında az ya da çok bir fark olacaktır. Bu farklar elde edilmiş olan regresyon doğrusunun gerçek değerlerden mininum uzaklıklarına eşittir. Bu değer yapılan herhangi bir tahminin hatasını gösterir. Formülü şu şekildedir:

$S_{yx} = \sqrt{\frac{\sum(Y-Y')^2}{n}}$ bu formül Y'nin X'e göre standart hatasıdır.Benzer şekilde X'in Y'ye göre standart hatası da $S_{xy} = \sqrt{\frac{\sum(X-X')^2}{n}}$ biçiminde hesaplanabilir. Burada genellikle $S_{yx} \neq S_{xy}$ dir.

13.3 Korelasyon

Herhangi iki olay arasında pozitif ya da negatif bir ilişki söz konusudur. Bu ilişki tam ilişkinin bir yüzdesi olarak belirtilir. İki ya da daha fazla değişkenin arasındaki ilişkinin yönünün ve derecesinin araştırılması korelasyon analizinin konusudur. Korelasyon katsayısı da bu ilişkinin derecesinin tayininde kullanılan bir ölçüdür.

Aşağıdaki formül yardımıyla korelasyon katsayısı hesaplanabilir:

$$r = \frac{\sum[(X-\bar{X})(Y-\bar{Y})]}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}}$$

Burada $X - \bar{X} = x \; ve \; Y - \bar{Y} = y$ alınırsa

$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$ elde edilir. Denklem kesrinin paydası n/n ile çarpılıp gerekli işlemler yapılırsa denklemin son hali aşağıdaki gibi olur:

$$r = \frac{\sum xy}{n\sigma_x\sigma_y.}$$

Korelasyon katsayısının değeri -1 ile +1 arasındadır. +1 olduğunda değişkenler arasında ikişki tam ve pozitifdir denir. -1 için de tersi söz konusudur. Korelasyon değerinin ∓ 1 e yakın olması ilişkinin kuvvetli olduğuna, 0'a yakın olması ise ilişkinin zayıf olduğuna işaret eder.

Korelasyon katsayısı regresyon katsayılarının aritmetik ortalamasıdır.

$$r = \mp\sqrt{b_{xy} b_{yx}}$$

Regresyon katsayıları pozitif ise korelasyon katsayısı da pozitif, Regresyon katsayıları negatif ise korelasyon katsayısı da negatifdir.

Son olarak tahminin standart hatasının formülle hesaplanışında kullanılan formülü verelim:

$S_{yx} = \sqrt{\frac{\sum y^2 - b_{yx}\sum xy}{n}}$ ve $S_{xy} = \sqrt{\frac{\sum x^2 - b_{xy}\sum xy}{n}}$ biçimindedir.

13.4 Örnek Çalışma

Aşağıdaki tabloda 5 ailenin çocuk sayıları ve anne yaşları gösterilmektedir.

Aile No	Çocuk sayıları(X)	Anne yaşları(Y)
1	2	25
2	1	20
3	5	35
4	4	45
5	3	25
TOPLAM	15	150

a) Y'nin X'e göre regresyon doğrusunu bulalım.

$Y = a + bX$ denkleminde a ve b katsayılarını bulmak yeterlidir.

$\sum Y_i = na + b\sum X_i \quad \sum Y_i X_i = a\sum X_i + b\sum X_i^2$ normal denklemlerini kuralım.

Aile No	Çocuk sayıları(X)	Anne yaşları(Y)	X^2	XY
1	2	25	4	50
2	1	20	1	20
3	5	35	25	175
4	4	45	16	180
5	3	25	9	75
TOPLAM	15	150	55	500

$$\sum \mathbf{Y_i} = \mathbf{na} + \mathbf{b}\sum \mathbf{X_i} \Rightarrow \mathbf{150 = 5a + 15b}$$

$$\sum \mathbf{Y_i X_i} = \mathbf{a}\sum \mathbf{X_i} + \mathbf{b}\sum \mathbf{X_i^2} \Rightarrow \mathbf{500 = 15a + 55b}$$

Denklemlerin ortak çözülmesiyle a=15 ve b=5 bulunur. Buna göre regresyon doğru denklemi $Y = 15 + 5X$ olur.

b)Anne yaşları ile ilgili teorik değerleri elde edilen regresyon doğrusu yardımıyla tespit edelim.

$$X = 2 \text{ için } Y' = 15 + 5.2 = 25$$

$$X = 1 \text{ için } Y' = 15 + 5.1 = 20$$

$$X = 5 \text{ için } Y' = 15 + 5.5 = 40$$

$$X = 4 \text{ için } Y' = 15 + 5.4 = 35$$

$$X = 3 \text{ için } Y' = 15 + 5.3 = 30$$

c) Teorik değerlere dayanarak tahminin standart hatasını bulalım.

Anne yaşları(Y)	Y’	Y-Y’	(Y-Y’)^2
25	25	0	0
20	20	0	0
35	40	-5	25
45	35	10	100
25	30	-5	25
150	150	0	150

$$S_{yx} = \sqrt{\frac{\sum(Y-Y')^2}{n}} = \sqrt{\frac{150}{5}} = 5.47$$

d) Tahminin standart hatasını formül yardımıyla bulalım.

$$S_{yx} = \sqrt{\frac{\sum y^2 - b_{yx}\sum xy}{n}}$$

Çocuk sayıları(X)	Anne yaşları(Y)	X-Xort=x	Y-Yort=y	xy	y^2
2	25	-1	-5	5	25
1	20	-2	-10	20	100
5	35	2	5	10	25
4	45	1	15	15	225
3	25	0	-5	0	25
15	150	0	0	50	400
Xort=3 ve Yort=30 dir.					

$$b_{yx} = \frac{\sum xy}{\sum x^2} = \frac{50}{10} = 5, \qquad S_{yx} = \sqrt{\frac{400-5.50}{5}} = 5.47$$

e) Son olarak r korelasyon katsayısını bulalım.

$b_{yx} = \frac{\sum xy}{\sum x^2} = \frac{50}{10} = 5$ ve $b_{xy} = \frac{\sum xy}{\sum y^2} = \frac{50}{400} = 0,125$

$r = \mp\sqrt{b_{xy} b_{yx}} = \mp\sqrt{0.125 * 5} = 0.79$

Kaynaklar

1.M.,Akar, S.Şahinler, İstatistik, Ç.Ü.Ziraat Fakültesi ,Genel Yayın no:4,Adana,1997.

2. F.,İkiz, H.Püskülcü, Ş.Eren,İstatistiğe Giriş, EÜ Basımevi,İzmir,1996.

3. Ö.,Serper, Uygulamalı İstatistik, Ezgi Kitapevi, Bursa, 2000.

4. Y.,Özkan, Uygulamalı İstatistik I, Alfa Yayınları, İstanbul,1999.

5.N.,Çömlekçi,İstatistik,Bilim Teknik Yayınevi, Eskişehir,1984.