

REGRESYON ANALİZİ

Regresyon analizi, **aralarında sebep-sonuç ilişkisi bulunan** iki veya daha fazla değişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili tahminler (estimation) ya da kestirimler (prediction) yapabilmek amacıyla yapılır. Doğada birçok olayda sebep sonuç ilişkisine rastlamak mümkündür.

Örnek: **Sebe**

Sonuc

Gelir

Harcama

Yaş

Boy

Gübre

Verim

Yem miktarı

Süt miktarı

Çalışma süresi

Alınan not

Bu analiz tekniğinde iki (basit regresyon) veya daha fazla değişken (çoklu regresyon) arasındaki ilişki açıklamak için matematiksel bir model kullanılır ve bu model **regresyon modeli** olarak adlandırılır. Aşağıda basit regresyon analizi anlatılmıştır.

Basit regresyon modeli

$$Y = \alpha + \beta X + \epsilon$$

şeklinde bir bağımlı ve bir de bağımsız değişken içeren bir modeldir. Burada

Y; bağımlı (sonuç) değişken olup belli bir hataya sahip olduğu varsayılır.

X; bağımsız (sebe) değişkeni olup hatasız ölçüldüğü varsayılır.

α ; sabit olup $X=0$ olduğunda Y 'nin aldığı değerdir.

β ise regresyon katsayısı olup, X 'in kendi birimi cinsinden 1 birim değişmesine karşılık Y 'de kendi birimi cinsinden meydana gelecek değişme miktarını ifade eder.

ϵ ; tesadüfi hata terimi olup ortalaması sıfır varyansı σ^2 olan normal dağılım gösterdiği varsayılır. Bu varsayım parametre tahminleri için değil katsayıların önem kontrolleri için gereklidir.

Parametrelerin (Katsayıların) Tahmini

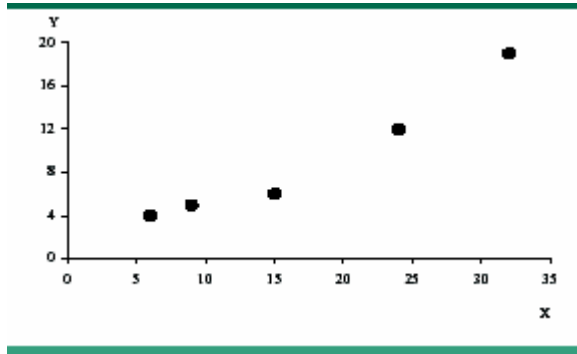
Bir regresyon modeli oluşturulurken genelde en-küçük kareler ve en büyük olasılık (maximum likelihood) teknikleri olarak bilinen iki yaklaşımdan birisi kullanılır. Eğer hata teriminin normal dağılım göstermesi şeklinde bir varsayım varsa en büyük olasılık, hata

teriminin dağılışı ile ilgili herhangi bir varsayım söz konusu değilse en-küçük kareler tekniği kullanılarak parametreler tahmin edilir.¹

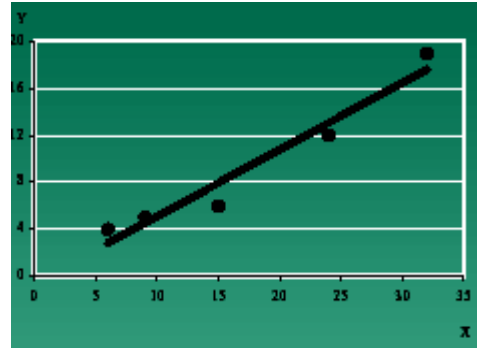
En-küçük kareler tekniği kullanılarak parametrelerin nasıl tahmin edildiğini örnek bir veri grubu üzerinde kısaca özetleyelim.

<u>Boy(cm) (X)</u>	<u>Çevre(cm) (Y)</u>
9	5
15	6
6	4
24	12
32	19

- Tabloda verilen X ve Y değişkenlerine ait beş gözlem çifti, koordinat eksenlerine yerleştirildiğinde elde edilen serpmeye diyagramının Şekil (a)'daki grafik elde edilir.



(a)



(b)

Şekil (a)'da verilen noktaları temsil eden regresyon doğrusu oluşturulursa Şekil (b) elde edilir. Uydurulan regresyon doğrusu ile gözlem noktaları arasındaki fark hata (ϵ) olarak isimlendirilir. Regresyon doğrusuna ait parametreler öyle tahmin edilmelidir ki; doğru ile gözlem noktaları arasındaki fark (hata) en az olsun. Bunu sağlayacak teknik ise **en-küçük kareler tekniği**dir.

Regresyon Analizi

İki değişken arasındaki korelasyonun matematiksel ifadesini tespit etmek için yapılan analiz.

Bu matematiksel ifadeye de regresyon denklemi diyoruz.

¹ http://www.mku.edu.tr/genel/fakulte/ziraat/zootekni/personel_veri/suat/Bolum_9.pdf

Regresyon denkleminin genel ifadesi: $Y' = a + bX$

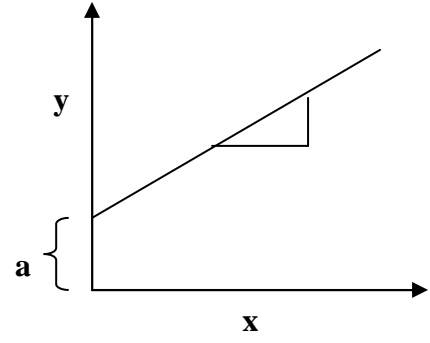
X: seçilen bağımsız değişkenin değeri

Y': seçilmiş X değerine için tahmin edilen Y değeri

a: doğrunun Y eksenini kestiği noktanın değeri

b: doğrunun eğimi

a ve b: regresyon katsayıları



Regresyon analizi

Bir kriter değişkeni ile bir veya daha fazla sayıda tahmin değişkenleri arasındaki ilgiyi sayısal hale dönüştürmede kullanılan istatistiksel analizdir. Regresyon analizi esas olarak değişkenler arasında ilişkinin niteliğini saptamayı amaçlar. Tahmin değişkeni olarak bir değişken kullanılırsa basit regresyon, tahmin değişkenleri olarak iki veya daha fazla değişken kullanılırsa çoklu regresyon analizinde söz etmek mümkündür. Amaç her tahmin değişkeninin kriter değişkenindeki toplam değişmeye olan katkısının saptanması ve dolayısıyla tahmin değişkenlerinin doğrusal kombinasyonunun değerinden hareketle kriter değerinin tahmin edilmesidir.²

Örnek: Tüketicilerin gelir düzeyleri ile A malının satışları arasındaki ilişkinin doğrusal olduğu varsayılarak iki değişken arasındaki ilişki matematiksel olarak gösterilebilir.

İki ya da daha çok değişken arasında ilişki olup olmadığını, ilişki varsa yönünü ve gücünü inceleyen “**korelasyon analizi**” ile değişkenlerden birisi belirli bir birim değiştiğinde diğerinin nasıl bir değişim gösterdiğini inceleyen “**regresyon analizi**” sağlık bilimlerinde çok kullanılan istatistiksel yöntemlerdir.³

Regresyon analizi, bilinen bulgulardan, bilinmeyen gelecekteki olaylarla ilgili tahminler yapılmasına izin verir. Regresyon, bağımlı ve bağımsız değişken(ler) arasındaki ilişkiyi ve doğrusal eğri kavramını kullanarak, bir tahmin eşitliği geliştirir. Değişkenler arasındaki ilişki belirlendikten sonra, bağımsız değişken(ler)in skoru bilindiğinde bağımlı değişkenin skoru tahmin edilebilir.

Bağımlı Değişken (y)

Bağımlı değişken, regresyon modelinde açıklanan ya da tahmin edilen değişkendir. Bu değişkenin bağımsız değişken ile ilişkili olduğu varsayılır.

Bağımsız Değişken (x)

Bağımsız değişken, regresyon modelinde açıklayıcı değişken olup; bağımlı değişkenin değerini tahmin etmek için kullanılır.

² http://www.frekans.com.tr/tr_analizler.html

³ http://www.toraks.org.tr/mse-ppt-pdf/Kenan_KOSE3.pdf

• Değişkenler arasında doğrusal ilişki olabileceği gibi, doğrusal olmayan bir ilişki de olabilir. Bu nedenle, saçılım grafiği yapılmadan (ilişki yok/doğrusal ilişki var/doğrusal olmayan ilişki var) ve değişkenler arasında korelasyon varlığına rastlanmadan regresyon analizine karar verilmemesi gerekir.

• Bu bilgiler doğrultusunda, tek/çok değişkenli doğrusal regresyon analizlerinin yanı sıra, tek/çok değişkenli doğrusal olmayan regresyon analizleri de mevcuttur.



(a) (+) yönlü Doğrusal ilişki



(b) (-) yönlü Doğrusal ilişki



(c) Doğrusal Olmayan ilişki



(d) ilişki Yok

Bağımlı değişken ile bağımsız değişken arasındaki doğrusal ilişkiyi açıklayan tek değişkenli regresyon modeli aşağıdaki gibidir:

$$Y = a + bx$$

Burada;

y = Bağımlı değişkenin değeri

a = Regresyon doğrusunun kesişim değeri (Sabit değer)

b = Regresyon doğrusunun eğimi

x = Bağımsız değişkenin değerini göstermektedir.

Örnek : Kardiyoloji kliniğine başvuran erkek hastalar üzerinde yapılan bir araştırmada, yas(x) ve kolesterol(y) değişkeni arasındaki korelasyondan yola çıkılarak kurulan regresyon modeli aşağıdaki gibi elde edilmiştir:

$$Y=3,42+0,326x$$

Bu modele göre, yastaki bir birimlik artışı, kolesterol değerinde 0.326 birimlik bir artışa neden olacağı, yeni doğan bir erkeğin (X=0) kolesterol değerinin ise 3.42 olacağı söylenebilir. Kurulan bu modele göre, 50 yaşında bir erkeğin kolesterol değerinin ne kadar olacağını tahmin edebiliriz.

X=50 için :

$$Y=3,42+0,326*(50)=19,52$$

50 yaşında bir erkeğin kolesterol değerinin **19.52** olacağı söylenebilir.

Tek Değişkenli Ve Çok Değişkenli Regresyon Analizi

Regresyon analizi bağımlı değişken ile bir veya daha çok bağımsız değişken arasındaki ilişkiyi incelemek amacıyla kullanılan bir analiz yöntemidir. Bir tek bağımsız değişkenin kullanıldığı regresyon **tek değişkenli regresyon analizi**, birden fazla bağımsız değişkenin kullanıldığı regresyon analizi de **çok değişkenli regresyon analizi** olarak adlandırılır.

Regresyon analizi ile ; bağımlı ve bağımsız değişkenler arasında bir ilişki var mıdır? Eğer bir ilişki varsa bu ilişkinin gücü nedir? Değişkenler arasında ne tür bir ilişki vardır? Bağımlı değişkene ait ileriye dönük değerleri tahmin etmek mümkün müdür ve nasıl tahmin edilmelidir? Belirli koşulların kontrol edilmesi durumunda özel bir değişken veya değişkenler grubunun diğer değişken veya değişkenler üzerindeki etkisi nedir ve nasıl değişir? gibi sorulara cevap aranmaya çalışılır.⁴

Tek Değişkenli Regresyon Analizi 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.

Çok Değişkenli Regresyon Analizi İçinde bir adet bağımlı değişken ve birden fazla bağımsız değişkenin bulunduğu regresyon modelleri çok değişkenli regresyon analizi olarak bilinir.

Regresyon analizi, birçok alanda veri analizi için başvurulan önemli bir istatistiksel teknik olup değişkenler arasındaki ilişkiyi açıklamak için kullanılır. Kısaca regresyon analizi

⁴ http://istatistikanaliz.com/regresyon_analizi.asp

,bağımlı bir değişken ile bağımlı değişken üzerinde etkisi olduğu varsayılan bağımsız değişken(ler) arasındaki ilişkinin matematiksel bir model ile açıklanmasıdır.

Basit doğrusal regresyon analizinde bir bağımlı ve bir bağımsız değişken söz konusu iken,**çoklu doğrusal regresyon analizinde** ise bir bağımlı değişken varken iki ya da daha fazla bağımsız değişken vardır ve her iki analizde de değişkenler arasında doğrusal bir ilişki vardır.Ayrıca bağımlı ve bağımsız iki değişken arasında eğrisel bir ilişki var ise değişkenler arasındaki ilişki **eğrisel regresyon modeli** ile açıklanır.

Korelasyon analizinde,değişkenlerin bağımlı ve bağımsız değişken olarak belirlenmesi hesaplamaların sonucu açısından önemli değilken,regresyon analizinde ise değişkenlerin hangisinin bağımlı hangisinin bağımsız değişken olduğunu tespit etmek çok önemlidir.

Örneğin ,ithalat ve ihracat yapan bir işletmenin karlılık düzeyini işletmenin bulunduğu ülkedeki döviz kurları etkileyebilir.O halde,işletmenin karlılık düzeyi bağımlı(y),ülkedeki döviz kurları ise bağımsız(x) değişken olarak ele alınmalıdır.⁵

Tek Değişkenli Regresyon Analizi

Tek değişkenli regresyon analizi bir bağımlı değişken ve bir bağımsız değişken arasındaki ilişkiyi inceleyen analiz tekniğidir.Bu analizle bağımlı ve bağımsız değişkenler arasındaki doğrusal (lineer) ilişkiyi temsil eden bir doğru denklemi formüle edilmektedir.Korelasyon analizinde olduğu gibi ,regresyon analizinde üzerinde durulan ilişki,değişkenler arasındaki doğrusal ilişkidir.Bu doğrunun hesaplanması ise en küçük kareler metodu yardımıyla yapılmaktadır.

Regresyon analizi sonuçlarının yorumlanmasında birçok araştırmacı ve öğrenci tarafından ciddi hatalar yapılmaktadır.En yaygın hata,regresyon analizi sonuçlarının yorumlanmasında,x bağımsız değişkeninin y bağımlı değişkenine sebep olduğu şeklindeki yorumdur.Bağımsız değişkenlerin bağımlı değişkendeki değişimi açıklıyor olması sebepselliği gerekli kılmaz.Başka bir ifade ile,bağımlı ve bağımsız değişkenler arasında (pozitif ve negatif) bir ilişkinin olması her zaman bağımsız değişken(lerin) bağımlı değişkenin sebebi olduğu sonucunu doğurmayacaktır.

İki değişken arasında bir ilişkinin olabilmesi için sebepsellik şart değildir.İlişkinin sebebi belki de iki değişkenin üçüncü bir değişkenle olan ilişkilerinden kaynaklanıyor olabileceği gibi,söz konusu ilişki tamamen tesadüfi olarak da ortaya çıkmış

⁵ Bilimsel Araştırma Süreci ve SPSS ile Veri Analizi , **SPSS 12.0 for Windows** , Ayhan URAL,2005

olabilir. Sebepsellik ile ilişkiselliğin aynı şeyler olmadığı unutulmamalıdır. Regresyon analizi değişkenler arasındaki ilişkinin yapısı ve derecesi ile ilgilenmektedir.

Çok Değişkenli Regresyon Analizi

Bir bağımlı değişken ve birden fazla bağımsız değişkenin yer aldığı regresyon modellerine **çok değişkenli regresyon analizi** denir. Çok değişkenli regresyon analizinde bağımsız değişkenler eş zamanlı olarak (aynı anda) bağımlı değişkendeki değişimi açıklamaya çalışmaktadır. Hesaplama ve yorum bakımından tek değişkenli regresyon analizine benzemektedir. Çok değişkenli regresyon analizinin yorumu tek değişkenli regresyon analizine benzemektedir. Ancak bazı farklılıklar vardır. Örneğin, tek değişkenli regresyon analizindeki karşılığı çoklu regresyon katsayısı R (multiple R) olarak ifade edilmektedir. Çoklu regresyon katsayısı R, bir bağımlı değişkendeki değişim ile eşzamanlı (aynı anda) ele alınan birden fazla bağımsız değişkendeki değişim arasındaki ilişkinin derecesini göstermektedir. Daha basit bir ifade ile, bağımlı değişken ile birlikte ele alınan bir grup bağımsız değişkendeki değişimin ilişkisinin (korelasyonunun) bir göstergesidir.

Çok değişkenli regresyon analizi sosyal bilimlerin birçok dalında kullanım alanı bulmaktadır. Pazarlama, sosyoloji ve psikoloji gibi bilim dallarında davranışsal hareketlerin belirlenmesinde, ekonomide zaman serisi türü ekonomik değişkenleri etkileyen faktörlerin tespiti ve geleceğe yönelik projeksiyonlarında (tahmininde) kullanım alanı bulmaktadır.

Çoklu Regresyon Analizi Modelleri

Basit regresyon analizinde bağımlı ve bağımsız değişken arasındaki ilişkiler analiz ediliyordu. Bazı durumlarda ise; incelenen bir bağımlı değişkeni birçok bağımsız değişken etkileyebilir. Bir bağımlı değişkeni birçok bağımsız değişkenin etkileme durumlarına kısa örnekler verelim. Okuldaki bir öğrencinin başarısını, zekası, çalışma üresi, çalışma ortamı, aile ortamı vs. gibi birçok faktörler etkiler. Bir hastanın iyileşme süresini aldığı ilaçların dozu, yaşı, cinsiyeti, hastalığın ağırlık derecesi gibi birçok faktörler etkiler.

Çoklu regresyon analizinden olumlu sonuç alınabilmesi için, bağımlı ve bağımsız değişkenlerin sayısal olarak ve aynı ölçüm birimiyle ölçülmesi iyi sonuçlar verebilir. Sınıflandırma ve aralıklı ölçümlerle, ve hele ölçüm birimleri de değişik olursa, iyi yorumlanabilir sonuçlar alınamaz. Çünkü niteliksel özellikler sayısal olarak gösterilse bile, mesela zengine 1 yoksula 0 kodu verildiğinde, 1 ile 0 arasındaki farkın gerçekte zengin ile yoksul arasındaki farkı gösterip göstermediği tartışılabilir. Zenginle yoksul arasındaki fark 0

ile 1 arasındaki fark gibi değil, 0 ile 20,30,50 arasındaki fark gibidir.Öte yandan meslek gruplarını, mezun olunan okulları, cinsiyeti kodlamanın hiçbir anlamı yoktur.

Bağımsız değişkeni seçerken de ,kendi aralarında yüksek korelasyona sahip bağımsız değişkenlerden sadece birisinin alınması yoluna gidilmelidir.Mesela,kişinin yaşı,meslekteki kıdemi,aldığı ücret ayrı değişkenlermiş gibi görünmelerine rağmen,aslında her üçü de birbiriyle sıkı sıkıya bağlantılıdır.Çoklu regresyon analizinde böyle değişkenlerden sadece biri seçilmelidir.Bunun için de analize başlamadan bütün değişkenlerin korelasyon matrisine bakıp,aralarında yüksek korelasyon alanlardan biri seçilmelidir.

Çoklu regresyonda bazen hangi bağımsız değişkenin daha önemli olduğunu ve bağımlı değişkeni daha çok etkilediğini bilmek gerekir.Bunun için önce korelasyonlara bakarız.Yüksek korelasyon,daha güçlü doğrusal ilişkiyi gösterir.İkinci yol,bağımsız değişkenin bağımlı değişkeni,bir kez ikili değişken formülünde bir kez de ikili ,bir kez de çoklu regresyon içinde nasıl etkilediğine bakmaktır.

Çoklu Regresyon Metodları

- **Enter metodu** :Bağımsız değişkenleri bir blok olarak tek adımda girilip değerlendirildiği metod.
- **İleri Doğru Seçim Metodu (Forward selection)** :Bağımlı değişken ile en yüksek pozitif veya negatif korelasyonu olan bağımsız değişken ilkönce seçilir.Daha sonra girilen değişkenin katsayısının 0 olduğu hipotezi F testi ile yoklanır.Burada elde edilen F değeri,SPSS'in öngörülen F değerleri ile karşılaştırılır.SPSS'in iki F ölçütü vardır.

1. F değeri sizin belirleyeceğiniz minimum bir F değeri ile karşılaştırılır.(**F-to-enter,FIN**).Normal ayar **3,84**'tür.(Bunu ayar yeri "Linear regression" penceresinden "options" anahtarı ile geçilen pencerede "Use F Value" kısmındadır.

2. F istatistiği ile bağlantılı ihtimalin ayarlanması (Probability of F-to-enter,PIN).Bunun normal ayarı da **0,05**'tir.

Eğer elde edilen F değeri bu ölçütlere eşit veya küçükse,o bağımsız değişken regresyon değerlendirmesine alınır ve seçim ileri doğru devam eder;yoksa işlem orada durdurulur.

Bir değişken seçilip işleme alındığında,geride kalan bağımsız değişkenlerle bağımlı değişken arasındaki korelasyonlara bakılır ve en yüksek korelasyona sahip bağımsız değişken bir sonraki aday olur.Bu ,aynı zamanda en geniş F değerine sahip değişkenin de seçimi olur.

- **Geriye Doğru Eleme (Backward Elimination) Metodu** :İleri doğru seçimin tersine,burada önce bütün bağımsız değişkenler seçilir;sonra sıra ile belli ölçütlere göre eleme yapılır.SPSS ; eleme için iki ölçüt koymaktadır.

1. Değişkenin formülde kalabilmesi için en küçük kareler F değeri (F-to-remove,FOUT).Normal ayar **2,71**'dir.
2. En büyük F ihtimali (Probability of F-to-remove,POUT).Bunun normal ayarı da **0,10**'dur.

İlk önce en küçük kısmi korelasyon katsayısına sahip değişken incelenir.Öngörülen değerlerden,büyük değere sahip değişken elenir.

- **Adım Adım Seçme (Stepwise Selection) Metodu** :İlk olarak bağımsız değişken seçilir;eğer bu –ileri doğru seçmedeki- FIN veya PIN gereklerini yerine getirirse ikinci değişken seçilir;yoksa işlem orada biter.İkinci değişken olarak en yüksek kısmi korelasyona sahip değişken alınır.Seçimler yüksek korelasyondan düşüğe doğru yapılır.Bağımsız değişkenler ölçütlere uyarsa regresyon analizine başlanır.İlk değişken seçildikten sonra Adım Adım seçme,ileri doğru seçmeden farklılaşır.İlk değişkenin,geriye doğru elemadaki gibi FOUT veya POUT ölçütlerine uyup uymadığı kontrol edilir.Yani adım adım seçmede,hem ileri doğru seçme hem e geriye doğru eleme işlemleri yapılır.⁶

Çoklu Doğrusal Bağlantı

Basit regresyon analizindeki tüm varsayımlar çoklu regresyon analizinde de geçerlidir.Bu varsayımları hatırlarsak bağımlı değişken x_1 tesadüfi bir değişkendir ve normal dağılım göstermektedir.Tahmin hataları tesadüfidir ve normal dağılım gösterirler.Hatalar birbirinden bağımsızdır.**(otokorelasyon yoktur.)** .Her bağımsız değişkenin değerlerine ait olan bağımlı değişken değerlerinin alt setleri varyansları birbirine eşittir.**(eşit varyanslılık-homoscedasticity)**.Çoklu regresyon analizinde bu varsayımlara bir dördüncüsü eklenmektedir.Bu varsayım bağımsız değişkenler arasında basit doğrusal ilişkilerin olmaması şeklinde ifade edilebilir.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.⁷

⁶ Bilimsel Araştırmalarda Bilgisayarlarla İstatistik Uygulamaları,SPSS for Windows,Prof.Dr.Mustafa ERGÜN

⁷ Uygulamalı Regresyon Ve Korelasyon Analizi,**Doç.Dr.Neyran ORHUNBİLGE**, İ.Ü İşletme Fakültesi