

1. Giriş

Değişkenler arasındaki ilişkiyi incelemek bilimin uğraşlarından birisi olagelmıştır. Bunu doğal karşılamak gerekir. Çünkü gerek günlük hayatımızda gerekse bilimsel araştırmalarda karşılaştığımız sorunların çoğunluğu iki (veya daha çok) değişken arasında bir ilişki olup olmadığının saptanması ile ilgilidir.

İki değişken arasında bir ilişki bulunup bulunmadığı, eğer varsa bu ilişkinin derecesinin saptanması da istatistiksel çözümlemelerde sık sık karşılaşılan bir sorundur. Değişkenler arasındaki ilişkinin incelenmesinde regresyon ilk akla gelen tekniktir.

İstatistiksel anlamda iki değişken arasındaki ilişki, bunların değerlerinin karşılıklı değişimleri arasında bir bağıllık şeklinde anlaşılır. Gerçekten X değişkeninin değerleri değişirken buna bağlı olarak Y değişkeninin değerleri de değişiyorsa, bu ikisi arasında bir ilişki bulunduğu söylenebilir.

Regresyonda değişkenlerin bağımlı değişken ve bağımsız değişken(ler) olarak iki gruba ayrılması bir zorunluluktur. Bağımlı değişken, bağımsız değişken(ler) tarafından açıklanmaya çalışılan değişkendir. Regresyonda bağımlı değişken Y ve bağımsız değişken(ler) de X ile gösterilir.

Regresyonda, amaçlardan biri bağımlı değişkenle bağımsız değişken(ler) arasındaki ilişkilerin ortaya çıkarılmasıdır. Örneğin Y ile X arasında $Y_i = \alpha + \beta X_i + \epsilon_i$ ($i=1,2,3,...$) gibi doğrusal bir ilişki öngörülüyorsa ilk adım modelin bilinmeyen α ve β parametrelerinin tahmin edilmesidir. Modelin bilinmeyen parametreleri tahmin edildiğinde bağımsız değişken(ler)'in farklı değerleri için bağımlı değişkenin alacağı değeri tahmin etmek regresyonda bir diğer amaçtır. Bağımsız değişken(ler)'in her farklı değer(ler)'i için bağımlı değişkenin değeri sabit ise ortada araştırılacak bir problem yoktur.

Hem regresyonda hem de varyans çözümlemesinde bağımlı değişken en az eşit aralıklı düzeyde ölçülür. Regresyonda, bazı özel durumlar dışında, bağımsız değişken(ler) de en az eşit aralıklı düzeyde ölçülmelidir. Varyans çözümlemesinde ise bağımlı değişken yine en az eşit aralıklı düzeyde ölçülürken bağımsız değişken(ler) sınıflama ya da sıralama ölçme düzeyinde ölçülür. Bu nedenle regresyon ve varyans çözümlemesinde amaç aynı olmakla beraber hangisinin kullanılacağı bağımsız değişken(ler)in ölçme düzeyine bağlıdır.

Regresyon analizi, bağımsız değişken sayısına göre;

1. Basit regresyon analizi (Tek bağımsız değişken).
2. Çoklu regresyon analizi (Birden çok bağımsız değişken).

Fonksiyon tipine göre;

1. Doğrusal regresyon analizi.
2. Doğrusal olmayan regresyon analizi (Eğrisel).

Verilerin kaynağına göre;

1. Ana kütle verileriyle regresyon analizi.
2. Örnek verileri ile regresyon analizi.
3. Zaman serilerinde regresyon analizi (Eşleştirilmiş zaman serileri). Şeklinde gruplandırılır.

2. Basit Doğrusal Regresyon Analizi

Birçok istatistiksel çalışmada olduğu gibi regresyon analizinde de ana kütle verilerinin tümü yerine bu ana kütlede seçilen örnek verileri ile analiz yapılır. Daha sonra elde edilen sonuçlar ana kütledeki ilişkinin tahmininde kullanılır. Bilindiği gibi, ana kütle birimi sayısı çok fazla olduğundan , zamandan ve araştırma masraflarından tasarruf amacıyla tüm ana kütle birimleri yerine, bu ana kütlelerden tesadüfi olarak belirli sayıda birim (n) seçilerek istatistik analizler yapılır. Ana kütle ve örnek verileriyle yapılan istatistik araştırmalarda tekniklerinin uygulanmasında farklılık yoktur. Ancak teknikler uygulandıktan sonra örnekleme teorisinden yararlanılarak ana kütle parametrelerinin testleri ve tahminleri yapılır.

Regresyon analizinde de uygulama aynı şekilde olmaktadır. Büyük harfler ana kütle, küçük harfler ise örneğe ait verileri ve istatistik ölçüleri göstermekte kullanılmaktadır.

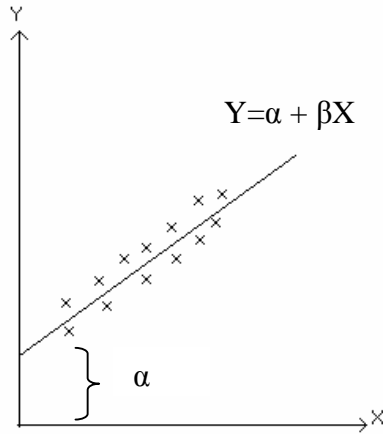
Basit doğrusal regresyon analizi, Y bağımlı değişkeninin tek bir bağımsız (açıklayıcı) değişken X ile arasındaki ilişkinin doğrusal fonksiyonla ifade edilmesine dayanmaktadır. Basit doğrusal regresyon modeli, tek bir serbest değişken içeren;

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (2.1)$$

modelidir. Bu modelin α ve β parametrelerini bulmak için X serbest değişkeni, Y bağılı değişkeni ve ε hata terimi ile ilgili gözlemlere gerek duyulur.

Ana kütle içinde birer α ve β değeri varken, bu ana küttelen çekilen her bir örneklem için ayrı birer $\hat{\alpha}$ ve $\hat{\beta}$ elde edilmektedir. İşte bu $\hat{\alpha}$ ve $\hat{\beta}$ normal bölünmeye sahip olup beklenen değerleri sırasıyla α ve β dir. Uygulamada tek bir örneklem alınmakta ve bu örneklem yardımıyla ana kütle parametreleri tahmin edilmektedir.

α doğrusal fonksiyonun sabitidir. $X=0$ olduğunda regresyon doğrusunun dikey eksen olan Y ile kesiştiği noktayı göstermektedir. β (β_{yx} ile de gösterilebilir) ise doğrusal fonksiyonun eğimidir. Regresyon analizinde bağımsız değişken X deki bir birimlik değişiminin bağımlı değişken Y 'de (Y cinsinden) ne kadarlık bir değişime yarattığını gösteren regresyon katsayısıdır. Fonksiyon tipinin belirlenmesi için regresyon analizine serpilme diyagramı çizilerek başlanır. Aşağıdaki serpilme diyagramında gözlem noktalarının dağılımının doğrusal bir eğilimde olduğu açıkça görülmektedir.



Şekil 1: Basit Doğrusal Regresyon Doğrusu

α ve β parametrelerinin gösterdiği grafikte regresyon doğrusunun eğiminin pozitif olduğu anlaşılmaktadır. β 'nin işareti iki değişken arasındaki ilişkinin yönünü göstermektedir. Her iki değişken birlikte artıyor veya azalıyorsa β 'nin işareti pozitif (+), değişkenlerden biri artarken diğeri azalıyorsa β nin işareti negatif (-) olacaktır. β nin sıfır (0) olması ise iki değişkenin arasında bir ilişki olmadığını göstermektedir. Sıfır

(0) dan farklılık ise iki değişken arasında belirli bir ilişkinin varlığını ifade etmektedir. Regresyon katsayısının alt sınırı (0) vardır, ancak belirli bir üst sınırı yoktur. Bu nedenle regresyon doğrusuna bakarak ilişkinin gücü hakkında kesin bir şey söylemek mümkün değildir.

Regresyon modeline açıkça dahil edilemeyen diğer değişkenleri temsil etmek üzere $Y_i = \alpha + \beta X_i + \epsilon_i$ modelinde yer verilen ϵ hata terimini gözlemek hiçbir zaman mümkün olmaz. Dolayısıyla ϵ hata terimi hakkında aşağıda değineceğimiz bazı varsayımları yapmak zorunlu hale gelir.

“Y ve X arasındaki gerçek ilişki” ; $Y = \alpha + \beta X_i + \epsilon$ iken “gerçek regresyon doğrusu” : $E(Y_i) = \alpha + \beta X_i$ eşitliğidir. Öte yandan, “Tahmin edilen ilişki”:

$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i + e_i$ şeklinde gösterilmektedir. Tahmin edilen regresyon doğrusu ise şudur: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$

Yukarıdaki eşitliklerde :

Y_i Y değişkeninin gözlenen değerini,

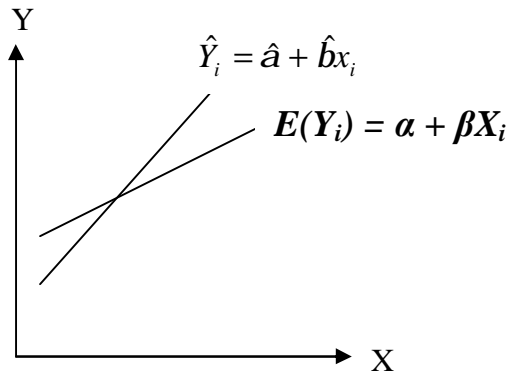
\hat{Y}_i X değişkeninin belli bir değeri veri iken Y değişkeninin tahmin edilen değerini,

$\hat{\alpha}$ a gerçek kesim noktasının tahminini,

$\hat{\beta}$ b gerçek parametresinin tahminini,

E e hata teriminin gerçek değerinin tahminini ifade eder.

Gerçek ve tahmin edilen regresyon doğruları aşağıdaki grafikte gösterilmiştir:



Şekil 2: Gerçek ve Tahmin Edilen Regresyon Doğruları

Doğrusal regresyon modeli bazı varsayımlara dayanmaktadır. Söz konusu varsayımlar “hata teriminin bölünmesi”, “hata terimi ile serbest değişkenler arasındaki” ve “serbest değişkenler arasındaki ilişki” ile ilgilidir. Bu varsayımları “stokastik⁽¹⁾ varsayımlar” ve “diğer varsayımlar” olmak üzere iki grupta toplayabiliriz.

2.1. Varsayımlar

VARSAYIM 1: “*Hata terimi normal dağılıma sahiptir*”. Diğer bir deyişle her X_i değeri için hata teriminin değerleri kendi ortalamaları etrafında çan eğrisi biçiminde simetrik bir bölünme gösterir.

$$\varepsilon \sim N(0, \sigma^2)$$

VARSAYIM 2: “*Hata terimlerinin ardışık değerleri birbirlerinden bağımsızdır*”. Diğer bir deyişle, birbirini izleyen hata terimleri arasında otokorelasyon yoktur.

Bu varsayıma göre $i \neq j$ olmak üzere e_i ve e_j ’nin kovaryansı sıfıra eşittir:
 $\text{Cov}(e_i, e_j) = 0$. u’nun altındaki i ve j indisleri yerine t ve $t-1$ indisleri de koyarak
 $\text{Cov}(e_t, e_{t-1}) = 0$ da yazabiliriz.

VARSAYIM 3: “*Hata teriminin varyansı X değerlerine göre değişmez yani sabittir*”. Diğer bir deyişle, bütün X değerleri için ε hata terimleri kendi ortalamaları etrafında aynı değişkenliğe sahiptir.

Hata teriminin varyansı ayrıca bağlı değişkenin varyansına da eşittir:
 $\text{Var}(Y) = \text{Var}(\varepsilon) = \sigma^2$. “Sabit varyans varsayımı” adı da verilen bu varsayıma göre, hata teriminin ortalaması etrafındaki varyansı bütün X değerleri için sabittir.

α ve β ’nin tahminleri olan $\hat{\alpha}$ ve $\hat{\beta}$ ’nin hesaplanmasında “En Küçük Kareler Tekniği”nden yararlanılabilir.

2.2. Doğrusal Regresyon Denkleminin Yazılışı

Regresyon analizi, herhangi bir değişkenin değişkenin, (bağımlı değişken) bir veya birden fazla birden fazla değişkenle (bağımsız-açıklayıcı değişken) arasındaki ilişkinin matematiksel bir fonksiyon şeklinde yazılmasıdır. Bu fonksiyona regresyon denklemi adı verilmektedir. Regresyon denklemi yardımıyla açıklayıcı değişkenlerin çeşitli değerlerine karşılık bağımlı değişkenin ulaşacağı değer tahmin edilir. Bağımlı değişkeni etkileyen açıklayıcı değişkenlerin saptanmış olması da bağımlı değişken üzerinde geliştirilecek politikalarda hangi değişkenlerin önem kazandığını ortaya çıkarmaktadır.

Çeşitli X değerleri karşısındaki Y değerlerinin dağılımını gösteren (Şekil -1) serpilme diyagramları incelendiğinde doğrusal bir eğilim görünüyorsa X'in Y'ye göre matematik fonksiyonunun doğrusal olduğuna (kesin olmasa da) karar verilebilir. Ancak gözlem noktaları arasından çok sayıda doğrusal fonksiyon geçirilebilir. Bu fonksiyonlardan en uygunu (tüm doğrusal fonksiyonlar arasından) Y gözlem değerine en yakın tahmini (teorik) Y' değerini (minimum hata) veren doğrusal fonksiyon olacaktır. Yani hatası $e = Y - Y' = Y_i - \hat{a} - \hat{b}x = \min$ olan fonksiyon seçilmelidir. Tüm gözlem değerleri için bu durumun geçerli olması gerektiğine göre;

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left(Y_i - \hat{Y}_i \right)^2 = \sum_{i=1}^n \left(Y_i - \hat{a} - \hat{b}X_i \right)^2 = \min \quad (2.2.1)$$

yapılması gerekir. Bu yönteme “En Küçük Kareler Yöntemi” adı verilmektedir. Elde edilen basit doğrusal regresyon denklemi de “En Küçük Kareler Yöntemi ile Basit Doğrusal Regresyon Denklemi” olacaktır. Bu fonksiyonun minimum olabilmesi için \hat{a} ve \hat{b} parametrelerine göre kısmi türevlerinin alınması ve bunların 0'a eşitlenmesi gerekir.

$$\sum_{i=1}^n \left(Y_i - \hat{a} - \hat{b}X_i \right)^2 = S \quad \text{dersek;}$$

$$\frac{\partial S}{\partial \hat{a}} = 2 \sum (Y_i - \hat{a} - \hat{b}X_i)(-1) = 0$$

$$2 \sum (-Y_i + \hat{a} + \hat{b}X_i) = 0$$

$$\sum (-Y_i + \hat{a} + \hat{b}X_i) = 0$$

$$-\sum Y_i + n\hat{a} + \hat{b} \sum X_i = 0$$

$$\boxed{\sum Y_i = n\hat{a} + \hat{b} \sum X_i}$$

$$\frac{\partial S}{\partial \hat{b}} = 2 \sum (Y_i - \hat{a} - \hat{b}X_i)(-X_i) = 0$$

$$2 \sum (-X_iY_i + \hat{a}X_i + \hat{b}X_i^2) = 0$$

$$\sum (-X_iY_i + \hat{a}X_i + \hat{b}X_i^2) = 0$$

$$-\sum X_iY_i + \hat{a} \sum X_i + \hat{b} \sum X_i^2 = 0$$

$$\boxed{\sum X_iY_i = \hat{a} \sum X_i + \hat{b} \sum X_i^2}$$

Minimum değ er i in yeterli  art fonksiyonun ikinci kısmi t revlerinin pozitif olması oldu undan S fonksiyonunun \hat{a} ve \hat{b} ’ya g re ikinci kısmi t revlerinin hesaplanması gerekir.

$$\frac{\partial^2 S}{\partial \hat{a}^2} = 2n > 0$$

$$\frac{\partial^2 S}{\partial \hat{b}^2} = 2 \sum X_i^2 > 0$$

Basit do rusal regresyon modelindeki α ve β ’nın tahminleri \hat{a} ve \hat{b} ’nın bulunmasında yararlanılacak normal denklemler  unlardır:

$$\sum Y_i = n\hat{a} + \hat{b} \sum X_i \quad (2.2.2)$$

$$\sum X_i Y_i = \hat{a} \sum X_i + \hat{b} \sum X_i^2 \quad (2.2.3)$$

Bu denklemdeki X_i ve Y_i değerleri 0 orijinine göre ifade edilmişlerdir. X_i ve Y_i değerleri yerine bunların kendi aritmetik ortalamalarından sapmalarının, kısacası “ortalamalar orijinine göre x_i ve y_i değerleri” nin konulmasıyla daha kısa zamanda sonuca götüren şu denklemlere ulaşılır:

$$\sum y_i = n\hat{a} + \hat{b} \sum x_i \quad (2.2.4)$$

$$\sum x_i y_i = \hat{a} \sum x_i + \hat{b} \sum x_i^2 \quad (2.2.5)$$

Normal denklemleri \hat{a} ve \hat{b} için çözümlerse gözlem değerleri yardımıyla en küçük kareler tahminleri bulunur :

$$\hat{a} = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (2.2.6)$$

$$\hat{b} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (2.2.7)$$

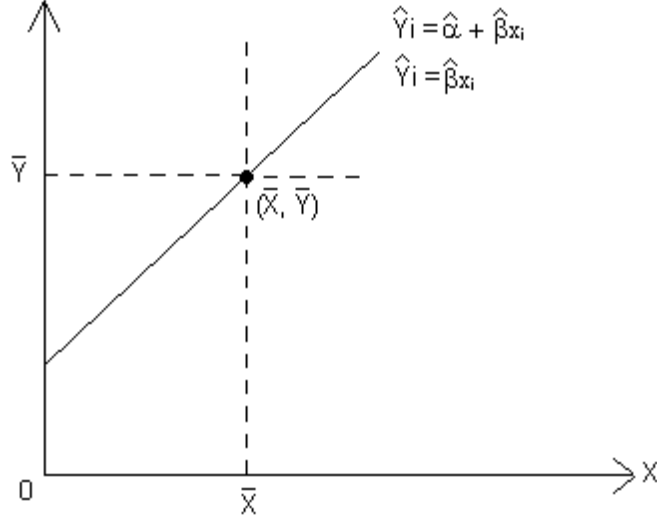
Terimlerin aritmetik ortalamadan farklarının cebirsel toplamının sıfır olduğu varsayımı altında, değişkenlerin toplamını ifade eden terimler sıfır olur:

$$\sum (y - \bar{y}) = 0 \quad \text{ve} \quad \sum (x - \bar{x}) = 0 \quad \text{eşitliklerinden ;}$$

$$\hat{a} = 0$$

$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{y}_i = \hat{b} x_i \quad \text{elde edilir.}$$



Şekil 3: Sıfır ve Ortalamalar Orijinlerine Göre Regresyon Doğruları

x ve y değişkeninin birlikte değişmesinden x değişkeninin kendi içindeki değişkenlik arındırıldığı için, “regresyon katsayısı “ adı verilen ve regresyon doğrusunun eğimini belirten b parametresi x’deki bir birimlik değişmenin y’de ne kadarlık bir değişme yarattığını göstermektedir.

Bu formüller ile bulunan \hat{a} ve \hat{b} değerleri, ana kütle regresyon doğrusundaki a ve b parametrelerinin kestirim değerleridir. En küçük kareler yöntemi ile bulunan \hat{a} ve \hat{b} kestiricileri sistematik hata içermeyen (tarafsız) kestiricilerdir. Diğer bir deyişle ;

$$E(\hat{a}) = a \quad \text{ve} \quad E(\hat{b}) = b \quad \text{olur.}$$

İspat 1: $E(\hat{b}) = b$

$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i y_i \quad (\text{Burada } k_i = \frac{x_i}{\sum x_i^2} \text{ 'dir.})$$

$$\sum k_i = 0, \quad \sum k_i x_i = 1 \quad \text{ve} \quad \sum k_i^2 = \frac{1}{\sum x_i^2} \quad \text{'dir.}$$

$$\hat{b} = \sum k_i (a + b x_i + e_i)$$

$$\hat{b} = b + \sum k_i e_i$$

Her iki tarafın beklenen değerleri alındığında;

$$E(\hat{b}) = E(b) \Rightarrow E(\hat{b}) = b \quad \text{elde edilir. Çünkü tüm } i \text{ 'ler için } E(u_i) = 0 \text{ 'dır.}$$

$E\{\hat{b} - E(\hat{b})\}E(\hat{b}) - b$ ifadesini deęerlendirmede toplama iřlemlerindeki kurallar geerlidir. Sabit olan herhangi bir faktör sola kaydırılarak beklenen deęer iřaretinin önüne arpan olarak yazılabilir. Bir sabitin beklenen deęeri, bu sabitin kendisidir. Dolayısıyla;

$$\begin{aligned} E\{\hat{b} - E(\hat{b})\}E(\hat{b}) - b &= [E(\hat{b}) - b]E\{\hat{b} - E(\hat{b})\} \\ &= [E(\hat{b}) - b]\{E(\hat{b}) - E(\hat{b})\} \\ &= 0 \end{aligned}$$

İspat 2: $E(\hat{a}) = a$

$$\hat{a} = Y - \hat{b}\bar{x}$$

$$\hat{a} = a + b\bar{x} + \bar{e} - \hat{b}\bar{x}$$

$$\hat{a} = a - (\hat{b} - b)\bar{x} + \bar{e} \quad (E(\hat{b}) = b \text{ ve } E(\bar{e}) = 0 \text{ olduęu için})$$

Her iki tarafın beklenen deęerleri alındığında;

$$E(\hat{a}) = E(a) - E((\hat{b} - b)\bar{x}) + E(\bar{e})$$

$$E(\hat{a}) = E(a) \Rightarrow E(\hat{a}) = a \text{ elde edilir.}$$

2.3. Regresyon Denklemiyle Yapılacak Tahminlerin Standart Hatası:

Ana kütle regresyon doğrusundaki a ve b parametrelerinin kestirimini hata kareleri toplamını minimum yapacak \hat{a} ve \hat{b} deęerlerini bularak yaptığımızı göre, bu kestirim sürecinde e deęişkeninin standart hatası olacaktır. Bu deęişkenin standart hatası;

$$S_e = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-2}} \quad (2.3.1)$$

olur. Kullanılan her bir istatistik için (\hat{a} ve \hat{b}) bir serbestlik derecesi kaybedeceęinden serbestlik derecemiz $(n-2)$ olacaktır. Dięer taraftan gözlenen ve hesaplanan y deęerleri arasındaki farkların toplamı sıfır olacaęından (2.3.1) numaralı formül;

$$S_e = \sqrt{\frac{\sum e_i^2}{n-2}} \quad (2.3.2)$$

$$= \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{a} - \hat{b}x_i)^2}{n-2}}$$

$$S_e = \sqrt{\frac{\sum y^2 - \hat{a} \sum y - \hat{b} \sum xy}{n-2}} \quad (2.3.3)$$

şekillerinde de yazılabilir.

2.4. Regresyon Katsayısının Testi:

Regresyon denklemlerindeki parametrelerin testi, örnek verileriyle elde edilen değişkenler arasındaki ilişkinin gerçekten ana kütle için geçerli mi yoksa sadece örnekler için mi (tesadüfen) geçerli olduğunu ortaya çıkarmak için yapılır. Regresyon katsayıları testlerinde $n \geq 30$ olduğunda Z (Normal Dağılım), $n < 30$ olduğunda (n-2) serbestlik dereceli t (Student) testleri uygulanır.

2.4.1. Hipotezlerin Yazılması:

$H_0 : b = 0$ (Ana kütlede X'deki bir birimlik değişme Y'de hiçbir değişme yaratmamaktadır.) (İki değişken arasında ilişki yoktur.)

$H_1 : b \neq 0$ (Ana kütlede X'deki bir birimlik değişme Y'de önemli(anlamlı) bir değişme yaratmaktadır.) (İki değişken arasında ilişki vardır.)

2.4.2. α Anlamlılık Seviyesinin Belirlenmesi:

α , H_0 doğru olduğu (İki değişken arasında ilişki yoktur.) halde H_0 reddedildiğinde yapılan hatayı göstermektedir. Regresyon analizi testlerinde α mümkün olduğunca küçük tutulur.

2.4.3. Örnek Regresyon Katsayılarının Standart Değişken Şekline Dönüştürülmesi:

$$n < 30 \text{ ise } t_h = \frac{\hat{b}}{\hat{S}_b}, \quad t_h = \frac{\hat{a}}{\hat{S}_a} \text{ formülü,}$$

$$n \geq 30 \text{ ise } z_h = \frac{\hat{b}}{\hat{S}_b}, \quad z_h = \frac{\hat{a}}{\hat{S}_a} \text{ formülü kullanılır.}$$

H_0 hipotezinde $b = 0$ olduğu için;

$$t \text{ (veya } z) = \frac{\hat{b}}{\hat{S}_b / \sqrt{\sum (x - \bar{x})^2}}$$

$$t \text{ (veya } z) = \frac{\hat{a}}{\hat{S}_a / \sqrt{\sum (x - \bar{x})^2}} \text{ şeklinde yazılabilir.}$$

$$\hat{S}_a = S_e \sqrt{\frac{\sum x_i^2}{n \sum x_i'^2}} \text{ ve } \hat{S}_b = S_e \sqrt{\frac{1}{\sum x_i'^2}}$$

2.4.4. Karar:

$z_{a/2}$ veya $t_{a/2; n-2}$ tablo değeri z_h veya t_h değerinden büyükse H_0 kabul edilir.

Yani iki değişken arasında ilişki yoktur (örnek regresyon doğrusu tahminlerinde kullanılmaz).

Tablo değerleri z_h veya t_h 'dan küçükse, H_0 ret edilir. Yani iki değişken arasında ilişki vardır (örnek regresyon denklemi tahminlerde kullanılabilir).

2.5. Ana Kütle Regresyon Doğrusunun Parametreleri için Aralık Kestirimi:

\hat{b} İçin Güven Aralığı:

$$P(\hat{b} - t\hat{S}_b \leq b \leq \hat{b} + t\hat{S}_b) = 1 - \alpha$$

\hat{a} İçin Güven Aralığı:

$$P(\hat{a} - t\hat{s}_{\hat{a}} \leq a \leq \hat{a} + t\hat{s}_{\hat{a}}) = 1 - \alpha$$

2.6. Verilen Bir X Değeri İçin m_y 'nin Güven Aralığı:

Örneklemden kestirilen $\hat{y}_i = \hat{a} + \hat{b}x_i$ regresyon doğrusundan, verilen bir $X=x_0$ değeri için hesaplanan \hat{y}_i değeri, m_y için bir nokta kestirimidir. Ancak aynı x_0 değeri için çeşitli örnekler temel alınarak hesaplanacak \hat{y}_i değerleri değişkenlik gösterecektir. Bu değişkenliğin bir ölçüsüne gereksinim vardır. Bu değişkenlik ölçüsü verilen bir X değeri için m_y 'nin kestiriminin standart hatasıdır. Bu standart hata, aynı zamanda X'in değerini sabit tuttuğumuzda Y'nin koşullu dağılımından hesaplanan ortalama y değerinin standart hatasıdır. Normallik varsayımı altında kestirimi yapılan regresyon modeli için Y değişkeninin koşullu ortalamasının standart hatası:

$$\hat{s}_{(m_y / X=x_0)} = S_e \sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}} \right)} \quad (2.6.1)$$

formülü ile hesaplanır. Böylece, verilen bir X değeri için Y değişkeninin ortalamasının $(1-\alpha)$ 'lık güven aralığı:

$$\hat{y} - t_{\alpha/2; n-2} \hat{s}_{(m_y / X=x_0)} \leq m_y \leq \hat{y} + t_{\alpha/2; n-2} \hat{s}_{(m_y / X=x_0)} \quad (2.6.2)$$

2.7. Verilen Bir X Değeri İçin Y Değerinin Güven Aralığı:

$$\hat{s}_{(Y / X=x_0)} = S_e \sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}} \right)} \quad (2.7.1)$$

şeklindedir. Böylece verilen bir X değeri için Y değişkeninin alacağı değer (1- α) olasılıkla:

$$\hat{y} - t_{\alpha/2; n-2} \hat{S}_{(Y/X=x_0)} \leq Y \leq \hat{y} + t_{\alpha/2; n-2} \hat{S}_{(Y/X=x_0)} \quad \text{olacaktır.} \quad (2.7.2)$$

2.8. Belirlilik Katsayısı (r^2):

Parametreleri tahmin ettikten ve en küçük kareler doğrusunu belirledikten sonra bu doğrunun x ve y'nin gözlemlerine ne kadar iyi uyduğunu bilmemiz gerekir. Yani gözlemlerin regresyon doğrusu etrafındaki dağılmalarını ölçmemiz gerekir. Gözlemler, doğruya ne kadar yakınsa, yani y'deki değişimin bağımsız değişkendeki değişmelerle açıklanması o kadar iyidir.

Bu uyum iyiliğinin bir ölçüsü, bağımlı değişkendeki toplam değişimin yüzde kaçının bağımsız değişken x tarafından açıklanabildiğini gösteren korelasyon katsayısının karesi olan belirlilik katsayısı (r^2) dir. Aşağıdaki formüller yardımıyla belirlilik katsayısı hesaplanabilir:

$$\text{i) } r^2 = \frac{(\sum x'_i y'_i)^2}{\sum x_i'^2 \sum y_i'^2} \quad (2.8.1)$$

$$\text{ii) } r^2 = \frac{\hat{b} \sum x'_i y'_i}{\sum y_i'^2} \quad (\text{eğim katsayısı yardımıyla}) \quad (2.8.2)$$

$$\text{iii) } r^2 = \hat{b}^2 \frac{\sum x_i'^2}{\sum y_i'^2} \quad (\text{eğim katsayısı yardımıyla}) \quad (2.8.3)$$

$$\text{vi) } r^2 = 1 - \frac{\sum e_i^2}{\sum y_i'^2} \quad (\text{artıklar yardımıyla}) \quad (2.8.4)$$

BASİT DOĞRUSAL REGRESYONA İLİŞKİN ÇÖZÜMLÜ ÖRNEKLER

Örnek 1: Tablo 1’de 1936 yılında İngiltere’nin 48 eyaletindeki buğday ve patates ürün miktarı gösterilmiştir. Aşağıdaki verilere göre regresyon doğrusu denklemi elde ediniz.

Eyaletler	Buğday (x) (0,4 hektarda ton)	Patates (y) (her 0,4 hektarda ton)
Bedfordshire	16,0	5,3
Huntingdonshire	16,0	6,6
Combridgeshire	16,4	6,1
Ely	20,5	5,5
Suffolk, West	18,2	6,9
Suffolk, East	16,3	6,1
Essex	17,7	6,4
Hertfordshire	15,3	6,3
Middlesex	16,5	7,8
Norfolk	16,9	8,3
Lincs (Holland)	21,8	5,7
Lincs (Kesteven)	15,5	6,2
Lincs (Lindsey)	15,8	6,0
Yorkshire (East Riding)	16,1	6,1
Kent	18,5	6,6
Surrey	12,7	4,8
Sussex, East	15,7	4,9
Sussex, West	14,3	5,1
Berkshire	13,8	5,5
Hampshire	12,8	6,7
Isle of Wight	12,0	6,5
Nottinghamshire	15,6	5,2
Lencestershire	15,8	5,2
Rutland	16,6	7,1
Northamptonshire	14,3	4,9
Peterborough	14,4	5,6
Buckinghamshire	15,2	6,4
Oxfordshire	14,1	6,9
Warwickshire	15,4	5,6
Shoropshire	16,5	6,1
Worcestershire	14,2	5,7
Gloucestershire	13,2	5,0
Wiltshire	13,8	6,5
Herefordshire	14,4	6,2
Somersetshire	13,4	5,2
Dorsetshire	11,2	6,6
Devonshire	14,4	5,8
Cornwall	15,4	6,3
Northumberland	18,5	6,3
Durham	16,4	5,8
Yorkshire (Nort Riding)	17,0	5,9
Yorkshire (West Riding)	16,9	6,5
Cumberland	17,5	5,8
Westmorland	15,8	5,7
Lancashire	19,2	7,2
Cheshire	17,7	6,5
Derbyshire	15,2	5,4
Stoffordshire	17,1	6,3

KAYNAK: The Advanced Theory Of Statistics (Mourice G. KEndall, Alan Stuart, 1967).

Çözüm 1:

$$\Sigma x_i = 758,0 \quad , \quad \mu_x = 15,792$$

$$\Sigma y_i = 291,1 \quad , \quad \mu_y = 6,095$$

$$\Sigma x_i^2 = 12170,48$$

$$\Sigma y_i^2 = 1791,03$$

$$\Sigma x_i y_i = 4612,64$$

$$\hat{a} = \frac{\Sigma x_i^2 - \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{(12170,48)(291,1) - (758,0)(4612,64)}{48(12170,48) - (758,0)^2}$$

$$\hat{a} = 4,8285$$

$$\hat{b} = \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{48(4612,64) - (758,0)(291,1)}{48(12170,48) - (758,0)^2}$$

$$\hat{b} = 0,078$$

Doğrusal Regresyon Denklemi : $\hat{y}_i = \hat{a} + \hat{b}x_i$

$$\hat{y}_i = 4,8285 + 0,078 x_i$$

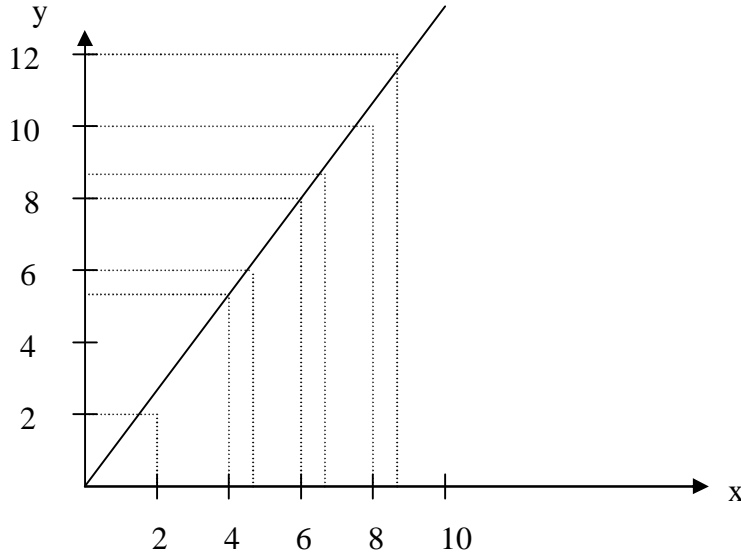
Örnek 2: Bir üretim merkezinde üretilen bir maddenin sertlik ve dayanıklılık bakımından ilişkisi araştırılmaktadır. Bu amaç için üretilen maddeler arasından 10 parça seçilmiştir. Sertlik ve dayanıklılık testi yapılmıştır. Veriler aşağıdaki gibidir. x ve y arasındaki ilişkiye ait regresyon denklemini bulunuz.

Parça No	x (Sertlik)	y (Dayanıklılık)	x^2	xy	y^2	\hat{y}	$y - \hat{y}$
1	7	10	49	70	100	9	1
2	9	12	81	108	144	11,5	0,5
3	5	6	25	30	36	6,5	-0,5
4	8	9	64	72	81	10,25	-1,25
5	6	8	36	48	64	7,25	0,25
6	9	11	81	99	121	11,5	-0,5
7	7	10	49	70	100	9	1
8	4	5	16	20	25	5,25	-0,25
9	8	10	64	80	100	10,25	-0,25
10	7	9	49	63	81	9	0
TOPLAM:	70	90	514	660	852	90	0

İlişkinin yapısı henüz bilinmediğinden (x,y) gözlem çiftleri için serpilme diyagramı yapılır.

Çözüm 2:

$$\hat{y} = 0,25 + 1,25x$$



Serpilme diyagramından x ve y arasındaki ilişkinin lineer olacağı açıktır.

Regresyon katsayıları $\hat{\alpha}$ ve $\hat{\beta}$:

$$\hat{\beta} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{10.660 - 70.90}{10.517 - 70^2} = \frac{6600 - 6300}{5140 - 4900} = 1,25$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = \frac{90}{10} - 1,25 \frac{70}{10} = 0,25$$

ve regresyon denklemi: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i = 0,25 + 1,25x_i$

(Kaynak: Olasılık ve İstatistik: Prof. Dr. Fikri Akdeniz)

Örnek 3: Bir akaryakıt şirketi ısınmak amacıyla sattığı gaz miktarı ve günlük sıcaklık arasındaki ilişkiyi belirlemek istemiştir. x günlük ortalama sıcaklığı, y (bin litre olarak) gaz satış miktarını göstermek üzere satış ve sıcaklık ile ilgili aşağıdaki tablo verilmiştir.

Günlük Ortalama Sıcaklık (x)	Satışlar (bin litre) (y)
-17	44
-14	35
-11	33
-10	26
-7	26
-6	19
-4	17
-1	10
3	8
7	4

x üzerinde y'nin regresyon doğrusunun denklemini bulunuz.

Çözüm 3:

Verilerden; $\Sigma x_i = -60$, $\Sigma y_i = 222$, $\Sigma x_i y_i = -2183$, $\Sigma x_i^2 = 866$

$$\hat{\beta} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{10(-2183) - (60).(222)}{10.866 - (60)^2} = \frac{-21830 + 13320}{8660 - 3600}$$

$$\hat{\beta} = \frac{-8510}{5060} \cong -1,68$$

$$\hat{\alpha} = \frac{\Sigma y - \hat{\beta} \Sigma x}{n} = \frac{222 - (-1,68).(-60)}{10} = \frac{222 - 100,8}{10} = 12,12$$

Tahmin edilmiş regresyon doğrusu:

$$\hat{y}_i = 12,12 - 1,68x_i$$

Örnek 4: Bir manav muzun kilogram fiyatı ile haftalık istek arasında doğrusal bir bağlantı olduğunu hissettirmektedir. Böyle bir bağlantı olup olmadığını incelemek amacıyla 8 hafta için aşağıdaki değerleri kaydetmiştir.

a) x üzerinde y'nin regresyon denklemini bulunuz.

b) x = 20 için \hat{y} 'yi bulunuz.

Fiyat (100 milyon) (x)	İstek (kg) y	x^2	xy	y^2
10	230	100	2300	52900
12	230	144	2760	52900
13	210	169	2730	44100
13	220	169	2860	48400
14	220	196	3080	48400
15	220	225	3300	48400
13	230	169	2990	52900
17	190	289	3230	36100
Toplam: 107	1750	1461	23250	384100

Çözüm 4:

$$a) \quad \hat{b} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} = \frac{8.23250 - (107)(1750)}{8.1461 - (107)^2} = \frac{186000 - 187250}{11688 - 11449}$$

$$\hat{\beta} = -\frac{1250}{239} = -5,23$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{1750}{8} - (-5,23)\frac{107}{8} = 218,75 + 69,95 = 288,7$$

$$\text{Regresyon denklemi: } \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i = 288,7 - 5,23x_i$$

$$b) \quad x = 20 \text{ için } \hat{y}_{20} = \hat{\alpha} + \hat{\beta} (x=20) = 288,7 - 5,23 (20) = 184,1$$

Örnek 5: Büyük bir mağazanın satış müdürü her ay reklam için harcadığı parayı kullanarak aylık satışlarını kestirmek istemektedir. Bu kestirimi yapmak için geçmiş 12 aya ait satış bedellerini ve reklam giderlerini kaydediyor. Bu veriler aşağıda verilmiştir, x üzerinde y'nin regresyon doğrusunun denklemini bulunuz. Her bir x değerine karşılık gelen aylık satışları kestiriniz.

Reklam (x)	Satış (y)	\hat{y}
1200	2,480,000	2499833,75
1300	2,450,000	2534132,75
1000	2,450,000	2431235,75
1200	2,530,000	2499833,75
2400	2,900,000	2911421,75
1700	2,590,000	2671328,75
1900	2,800,000	2739926,75
2300	2,880,000	2877122,75
1700	2,730,000	2671328,75
1500	2,650,000	2602730,75
1600	2,570,000	2637029,75
1300	2,580,000	2534132,75

Çözüm 5:

$$\sum x_i = 19100$$

$$\sum y_i = 31610000$$

$$\sum x_i y_i = 51036000000$$

$$\sum x_i^2 = 32510000$$

$$\hat{b} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} = \frac{12.(51036000000) - (19100)(31610000)}{12.(32510000) - (364810000)}$$

$$\hat{\beta} = \frac{8681000000}{25310000} = 342,99$$

$$\hat{\alpha} = \hat{y} - \hat{\beta} \bar{x} = 26334166,667 - (342,99) \cdot (1591,667) = 2088245,75$$

Tahmin edilen regresyon doğrusu: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i = 2088245,75 + 342,99x_i$

Örnek 6: Nem oranı buharlaşmayı etkilediğinden, suda çözülebilir boyaların püskürtme süresince çözücü dengesi nemlilikten etkilenecektir. Nemlilik (x) ve çözücü buharlaşma (y) arasındaki ilişkiyi incelemek için kontrollü bir çalışma yapılmıştır. Bu ilişkinin bilinmesi boyacının boya tabancasını nemliliğe göre ayarlaması için yararlı olacaktır. Relatif nemlilik (x) ve çözücü buharlaşma (y) ile ilgili yapılan n = 25 ölçüm sonucunda,

$$\Sigma x_i = 1314,9 ; \Sigma y_i = 235,7 ; \Sigma x_i y_i = 11824,44 ; \Sigma x_i^2 = 76308 ; \Sigma y_i^2 = 2286,7$$

elde edilmiştir. Uygun basit doğrusal regresyon denklemini bulunuz.

Çözüm 6:

$$\hat{b} = \frac{n \Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{25 \cdot (11824,44) - (1314,9)(235,7)}{25 \cdot (76308) - (1728962,01)}$$

$$\hat{\beta} = \frac{-14310,93}{178737,99} = -0,08$$

$$\hat{\alpha} = \hat{y} - \hat{\beta} \bar{x} = \frac{235,7}{25} - (-0,08) \cdot \frac{1314,9}{25} = 9,428 + 4,208$$

$$\hat{\alpha} = 13,636$$

Tahmin edilen basit doğrusal regresyon denklemi:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \Rightarrow \hat{y}_i = 13,636 - 0,08x_i$$

Örnek 7: Bir araştırma grubu, akıl hastalarına sürekli olarak uygulanan, ancak uygulaması zor, pahalı ve zaman alıcı olan standart bir test yerine yeni bir test geliştiriyor ve elde edilen yeni test sonuçlarından, çok yaygın olarak kullanılan standart test sonuçlarını bir denklem yardımıyla elde etmek istiyor. Bu amaçla, bir hasta grubuna ilişkin test ölçümleri her iki test ile elde ettikten sonra, yeni test (x_i) sonuçları 50, 55, 60, ..., 100 olan hastalar arasından birer hasta rasgele seçiliyor. İlgili sonuçlar aşağıdaki tabloda verilmiştir.

Yeni Test Sonuçları (x ₁)	Standart Test Sonuçları (y)
50	61
55	61
60	59
65	71
70	80
75	76
80	90
85	106
90	98
95	100
100	114

- a) Regresyon katsayılarını belirleyerek örneklem regresyon doğrusunu elde ediniz.
- b) Yeni test sonuçları, standart test sonuçlarını yeterli derecede açıklayabilmekte midir? (Artıklarla olan ilişkisinden yararlanarak açıklayınız)
- c) Eğim katsayısının anlamlı olup olmadığına %5 A.B.'de karar veriniz.

Çözüm 7:

a) $\Sigma x_i = 825$ $\Sigma y_i = 916$
 $\Sigma x_i^2 = 64625$ $\Sigma y_i^2 = 80076$
 $\bar{x}_1 = 75$ $\bar{y} = 83,27$

$\Sigma x_i y_i = 71790$

$$\hat{a} = \frac{n\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{(64625)(916) - (825)(71790)}{11(64625) - (825)^2} = -1,0$$

$$\hat{b} = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{11(71790) - (825)(916)}{11(64625) - (825)^2} = 1,124$$

Örneklem regresyon doğrusu: $\hat{y}_i = -1,0 + 1,124 x_i$

- b) x_i ve y değerleri yardımıyla artıklar hesaplanır. ($e_i = y_i - \hat{a} - \hat{b} x_i$)

e_i	e_i^2	$y_i'^2$	$x_i'^2$
5,818	33,85	495,95	625
0,2	0,04	495,95	400
-7,418	55,03	589,03	225
-1,04	1,082	150,55	100
2,345	5,501	10,69	25
-7,27	52,89	52,85	0
1,11	1,232	45,29	25
11,49	132,04	516,65	100

-2,13	4,54	216,97	225
-5,75	33,06	279,89	400
2,64	6,97	944,33	625
$\Sigma e_i^2 = 326,235$	$\Sigma y_i'^2 = 3798,15$	$\Sigma x_i'^2 = 2750$	

$$r^2 = 1 - \frac{e_i^2}{y_i'^2} = 1 - \frac{326,235}{3798,15} = 0,914$$

Yeni test sonuçları standart test sonuçlarını 0,914 oranında yeterli derecede açıklamaktadır.

c) $H_0 : \beta = 0$ (Katsayı anlamlı değildir)

$H_1 : \beta \neq 0$ (Katsayı anlamlıdır)

$$Se^2 = \frac{\Sigma e_i^2}{n - k} = \frac{326,235}{11 - 2} = 36,25$$

$$\hat{s}_b^2 = se^2 \cdot \frac{1}{\Sigma x_i'^2} = 36,25 \cdot \frac{1}{2750} = 0,013 \Rightarrow \hat{s}_b = 0,11$$

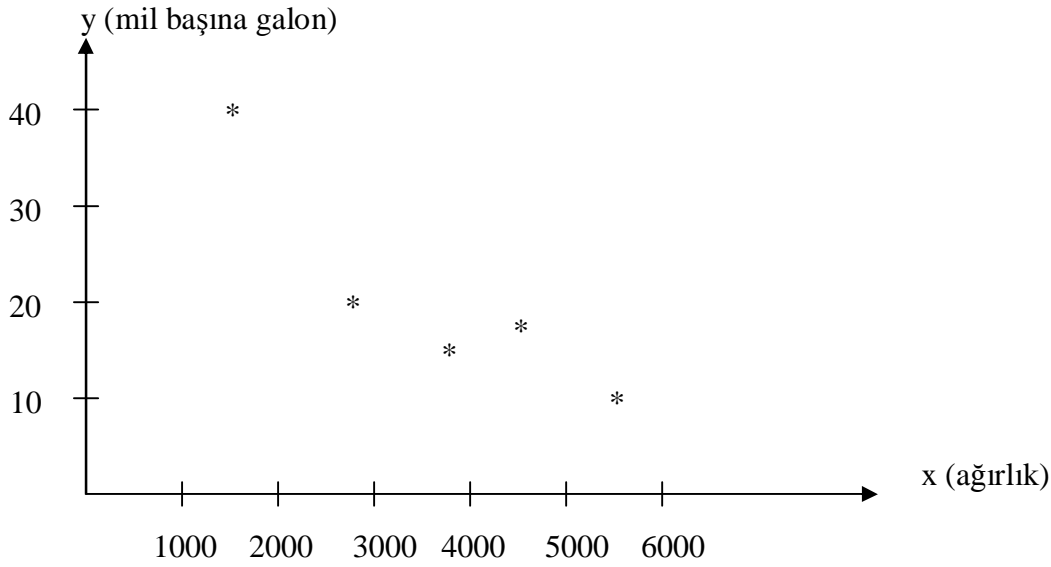
$$t_h = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} = \frac{1,124}{0,11} = 9,79$$

$$t_{0,05/2; 11-2} = t_{0,025; 9} = 2,26$$

$t_h > t_t$ olduğu için H_0 ret edilir. Yani %5 A.D.'de eğim katsayısının anlamlı olduğuna karar verilir.

Örnek 8: Donanmaya kira karşılığı verilecek 5 araba tesadüfi olarak seçilmiştir. Her bir araba için ağırlık ve 100 millik bir sürüşte mil başına harcadığı yakıt tutarı galon cinsinden sonuçlar verilmiştir. Serpilme diyagramını çizerek ağırlık ve 1 milde harcadığı yakıt arasındaki regresyon doğrusunu belirleyiniz.

Araba	Ağırlık (x)	Mil başına galon (y)	xy	x^2	y^2
1	2743	21,4	58700,2	7524049	457,96
2	3518	15,2	53473,6	12376324	231,06
3	1855	38,9	72159,5	3441025	1513,21
4	5214	12,7	66217,8	27185796	161,29
5	4341	17,8	77269,8	18844281	316,84
TOPLAM	17671	106	327820,9	69371475	2680,34

Çözüm 8:

$$\hat{b} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} = \frac{5.(327820,9) - (17671)(106)}{5.(69371475) - 312264241}$$

$$\hat{\beta} = \frac{-234021,5}{34593134} = -0,007$$

$$\hat{\alpha} = \hat{y} - \hat{\beta} \bar{x} = \frac{106}{5} - (-0,007) \cdot \frac{17671}{5} = 21,2 + 24,74 = 45,94$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} \bar{x}_i \Rightarrow \hat{y} = 45,94 - 0,007x_i$$

(Kaynak: Understanding Business statistics “John E.Hanke, Arthur G.Reitsch”)

Örnek 9: Doğumsal (congenital) kalp hastalığı olan 16 çocuğa ilişkin kalpten akciğerlere giden (pulmonary) kan akışı (PBF) ile kalpten akciğerlere giden kan hacmi (PBV) bulguları aşağıda verilmiştir. Regresyon denklemini bulunuz ve anlamlılığını test ediniz.

y	x			
PBV (m/-sqm)	PBF (L/min/sqm)	$x_i'^2$	$y_i'^2$	e_i^2
168	4,31	24,8	32627,2	10098,31
280	3,40	34,69	4710,08	684,54
391	6,20	9,55	1795,22	8477,66
420	17,30	64,16	5093,68	3326,27
303	12,30	9,06	2082,1	8865,48
429	13,99	22,09	6459,34	21,42

350	8,73	0,31	1,88	106,76
422	8,90	0,15	5383,16	4335,36
224	5,87	11,7	15532,64	4845,78
291	5,00	18,4	3321,22	129,92
233	3,51	33,41	13370,3	511,1
370	4,24	25,5	456,68	10534,34
531	19,41	102,4	33258,82	374,34
516	16,61	53,58	28012,72	2444,07
211	7,21	4,33	18942,02	10859,61
439	11,60	5,34	8166,74	2821,28

Çözüm 9:

$$\Sigma x_i = 148,58 \quad \Sigma y_i = 5578 \quad \Sigma x_i y_i = 58553,88$$

$$\Sigma x_i^2 = 1799,24 \quad \Sigma y_i^2 = 2123844 \quad \Sigma x_i'^2 = 419,47$$

$$\bar{x} = 9,29 \quad \bar{y} = 348,63 \quad \Sigma y_i'^2 = 179213,8$$

$$\Sigma e_i^2 = 70432,24$$

$$\hat{a} = \frac{n \Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{(1799,24)(5578) - (148,58)(58553,88)}{16(1799,24) - (148,58)^2} = 199,08$$

$$\hat{b} = \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{16(58553,88) - (148,58)(5578)}{16(1799,24) - (148,58)^2} = 16,10$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \Rightarrow \hat{y}_i = 199,08 + 16,10 x_i$$

$$H_0 : \alpha = 0 \text{ (Katsayı anlamlı değildir)}$$

$$H_1 : \alpha \neq 0 \text{ (Katsayı anlamlıdır)}$$

$$s_e^2 = \frac{\Sigma e_i^2}{n - k} = \frac{70432,24}{16 - 2} = 5030,87$$

$$\hat{\sigma}_{\hat{\alpha}}^2 = s_e^2 \cdot \frac{\Sigma x_i^2}{\Sigma x_i'^2} = 5030,87 \cdot \frac{1799,24}{419,47} = 21579$$

$$\hat{\sigma}_{\hat{\alpha}}^2 = 146,9$$

$$t_n = \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}} = \frac{199,08}{146,9} = 1,355$$

$$t_{0,05/2; (16-2)} = t_{0,025; 14} = 2,145$$

$t_n = 1,355 < t_t = 2,145$ olduğu için %5 A.D'de H_0 kabul edilir. Yani katsayı anlamlı değildir.

$$H_0 : \beta = 0 \text{ (Eğim katsayısı anlamlı değildir)}$$

$$H_1 : \beta \neq 0 \text{ (Eğim katsayısı anlamlıdır)}$$

$$\hat{\sigma}_{\hat{\beta}}^2 = Se^2 \cdot \frac{1}{\sum x_i^2} = 5030,87 \cdot \frac{1}{419,47} = 11,99 \Rightarrow \hat{\sigma}_{\hat{\beta}}^2 = 3,46$$

$$t_h = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} = \frac{16,10}{3,46} = 4,65$$

$t_{0,025; 14} = 2,145 < t_h = 4,65$ olduğundan H_0 ret edilir. %5 A.D.'de eğim katsayısının anlamlı olduğuna karar verilir.

3. ÇOKLU DOĞRUSAL REGRESYON ANALİZİ

Özellikle ekonomi ve işletmecilik alanlarında herhangi bir ekonomik değişkeni tek bir bağımsız değişkenle açıklamak mümkün değildir. Birçok ekonomik değişken bir araya gelerek bir değişkeni etkileyebildikleri gibi, kendi aralarında da birbirlerini etkileyebilmektedir. Birden fazla bağımsız değişkenli analize “Çoklu Regresyon Analizi “(Multiple Regression Analysis) adı verilmektedir.

Basit regresyon analizinde bağımlı değişken Y ile gösterilirken, bağımsız değişken X ile gösterilmekteydi. Çoklu regresyon analizinde de bağımlı değişken Y ile fakat bağımsız değişkenler X_1, X_2, \dots, X_k ile gösterilecektir.

$$Y_i = b_0 + b_1x_{1i} + \dots + b_kx_{ki} + e_i, i=1,2,\dots,n$$

3.1. Doğrusal Çoklu Regresyon Analizinde Varsayımlar:

VARSAYIM 1: Tahmin hataları ($e = Y - Y'$) tesadüfidir ve normal dağılım gösterirler.

$$\varepsilon \sim N(0, \sigma^2)$$

VARSAYIM 2: Tahmin hataları birbirinden bağımsızdır. Yani hata terimleri arasında otokorelasyon yoktur.

$$\text{Cov}(e_t, e_{t-1})=0$$

VARSAYIM 3: Her bağımsız değişkenin değerlerine ait olan bağımlı değişken değerlerinin alt setleri varyansları birbirine eşittir (eşit varyanslılık=homoscedasticity)

VARSAYIM 4: Bağımsız değişkenler arasında basit doğrusal ilişkiler yoktur. 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. Bu nedenle açıklayıcı değişkenler seçilirken, bunların bağımlı değişkenlerle basit doğrusal korelasyon katsayılarının yüksek (1'e yakın) ancak birbirleri arasındaki basit doğrusal korelasyon katsayılarının düşük (0 veya 0'a yakın) olmasına dikkat edilmelidir.

3.2. İki Bağımsız Değişkenli Doğrusal Regresyon Analizi:

Bir bağımsız değişkenli, bir bağımlı değişkenli doğrusal regresyon çok sayıda bağımsız değişkenler için genişletilebilir. Örneğin y değişkeninin x_1 ve x_2 bağımsız değişkenlerinin fonksiyonu olduğunu kabul edelim. O halde aşağıdaki denklemi düşünebiliriz.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad (3.2.1)$$

(3.2.1) denklemi üç boyutlu uzayda bir düzlem gösterir.

En uygun düzlemde bulunan y 'nin tahmin edilmiş değeri \hat{y} olsun. Bu durumda;

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 \quad (3.2.2)$$

denklemine x_1 ve x_2 üzerinde y 'nin regresyon denklemi denir ve

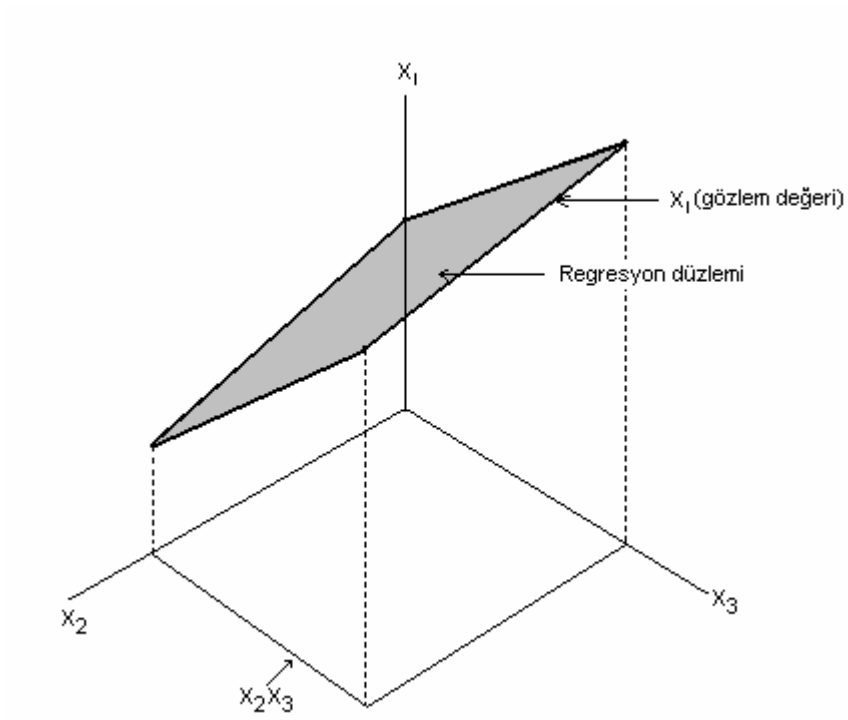
$$m_{y/x_1, x_2} = b_0 + b_1 x_1 + b_2 x_2 \quad (3.2.3)$$

kitle regresyon denkleminin bir tahminidir.

En Küçük Kareler Yöntemi (3.4) denklemindeki \hat{b}_0 , \hat{b}_1 ve \hat{b}_2 'nin değerini bulmak için kullanılır. Y gözlem değerlerine en yakın tahmini değerleri (Y' 'leri) verecek olan çoklu regresyon denklemi hataların kareleri toplamını

$\sum e^2 = \sum (Y - Y')^2 \Rightarrow \min$ yapan fonksiyondur. Bu fonksiyona ulaşabilmek için,

\hat{b}_0 ve iki kısmi regresyon katsayısına (\hat{b}_1 ve \hat{b}_2) göre birinci dereceden türevlerinin alınıp sıfıra eşitlenmesi gerekir. Böylece elde edilen fonksiyona “En Küçük Kareler Yöntemiyle Çoklu Regresyon Denklemi” adı verilir. Denklemin elde edilmesi aşağıda gösterilmektedir.



Şekil 4: İki Bağımsız Değişkenli Regresyon Analizi

$$\sum_{i=1}^n e^2 = \sum_{i=1}^n (y - y')^2 = \sum_{i=1}^n (y - \hat{b}_0 - \hat{b}_1 x_1 - \hat{b}_2 x_2)^2$$

$$\frac{\partial e}{\partial \hat{b}_0} = 2 \sum (-1) (y - \hat{b}_0 - \hat{b}_1 x_1 - \hat{b}_2 x_2) = 0$$

$$= -\sum y + n\hat{b}_0 + \hat{b}_1 \sum x_1 + \hat{b}_2 \sum x_2 = 0$$

$$\frac{\partial e}{\partial \hat{b}_1} = 2 \sum (-x_1) (y - \hat{b}_0 - \hat{b}_1 x_1 - \hat{b}_2 x_2) = 0$$

$$= -\sum yx_1 + \hat{b}_0 \sum x_1 + \hat{b}_1 \sum x_1^2 + \hat{b}_2 \sum x_1 x_2 = 0$$

$$\frac{\partial e}{\partial \hat{b}_2} = 2 \sum (-x_2) (y - \hat{b}_0 - \hat{b}_1 x_1 - \hat{b}_2 x_2) = 0$$

$$= -\sum yx_2 + \hat{b}_0 \sum x_2 + \hat{b}_1 \sum x_1 x_2 + \hat{b}_2 \sum x_2^2 = 0$$

Elde edilen üç denklemde ilk terimlerin(-) işaretlerinden kurtulmak için bu terimler eşitliğin sağına geçirildiğinde aşağıdaki üç “Normal Denklem” elde edilmektedir:

$$\sum y = n\hat{b}_0 + \hat{b}_1 \sum x_1 + \hat{b}_2 \sum x_2^2 \quad (3.2.4)$$

$$\sum yx_1 = \hat{b}_0 \sum x_1 + \hat{b}_1 \sum x_1^2 + \hat{b}_2 \sum x_1x_2 \quad (3.2.5)$$

$$\sum yx_2 = \hat{b}_0 \sum x_2 + \hat{b}_1 \sum x_1x_2 + \hat{b}_2 \sum x_2^2 \quad (3.2.6)$$

Bu denklemlerdeki y , x_1 , x_2 değerleri yerine bunların kendi aritmetik ortalamalarından sapmalarının, kısacası ” ortalamalar orijinine göre y_i , x_{1i} ve x_{2i} değerleri”nin yazılmasıyla daha kısa zamanda sonuca götüren gözlem değerlerine göre şu denklemlere ulaşılır.

$$\sum y_i = n\hat{b}_0 + \hat{b}_1 \sum x_{1i} + \hat{b}_2 \sum x_{2i} \quad (3.2.7)$$

$$\sum y_i x_{1i} = \hat{b}_0 \sum x_{1i} + \hat{b}_1 \sum x_{1i}^2 + \hat{b}_2 \sum x_{1i}x_{2i} \quad (3.2.8)$$

$$\sum y_i x_{2i} = \hat{b}_0 \sum x_{2i} + \hat{b}_1 \sum x_{1i}x_{2i} + \hat{b}_2 \sum x_{2i}^2 \quad (3.2.9)$$

Bu üç denklemdeki x_i ve y_i değerleri yerine her x_i ve y_i değerinin aritmetik ortalamasından farkları ($x'_i = x_i - \bar{x}_i$ ve $y'_i = y_i - \bar{y}_i$) yerleştirilir. Terimlerin aritmetik ortalamadan farklarının cebirsel toplamının 0 olduğu varsayımı altında aşağıdaki denklemlere ulaşmak mümkündür:

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}_1 - \hat{b}_2 \bar{x}_2 \quad (3.2.10)$$

$$\hat{b}_1 = \frac{\left(\sum x_{1i}' y_i'\right) \left(\sum x_{2i}'^2\right) - \left(\sum x_{2i}' y_i'\right) \left(\sum x_{1i}' x_{2i}'\right)}{\left(\sum x_{1i}'^2\right) \left(\sum x_{2i}'^2\right) - \left(\sum x_{1i}' x_{2i}'\right)^2} \quad (3.2.11)$$

$$\hat{b}_2 = \frac{\left(\sum x_{2i}' y_i'\right) \left(\sum x_{1i}'^2\right) - \left(\sum x_{1i}' y_i'\right) \left(\sum x_{1i}' x_{2i}'\right)}{\left(\sum x_{1i}'^2\right) \left(\sum x_{2i}'^2\right) - \left(\sum x_{1i}' x_{2i}'\right)^2} \quad (3.2.12)$$

Ayrıca gözlem değerlerine göre matrislerden yararlanarak da \hat{b}_0 , \hat{b}_1 ve \hat{b}_2 değerlerine ulaşmak mümkündür:

$$|\Delta| = \begin{vmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^2 \end{vmatrix}$$

$$|\Delta_0| = \begin{vmatrix} \sum y_i & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} y_i & \sum x_{1i}^2 & \sum x_{2i} x_{1i} \\ \sum x_{2i} y_i & \sum x_{1i} x_{2i} & \sum x_{2i}^2 \end{vmatrix} \quad \hat{b}_0 = \frac{|\Delta_0|}{|\Delta|}$$

$$|\Delta_1| = \begin{vmatrix} n & \sum y_i & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i} y_i & \sum x_{1i} x_{2i} \\ \sum x_{2i} & \sum x_{2i} y_i & \sum x_{2i}^2 \end{vmatrix} \quad \hat{b}_1 = \frac{|\Delta_1|}{|\Delta|}$$

$$|\Delta_2| = \begin{vmatrix} n & \sum x_{1i} & \sum y_i \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} y_i \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i} y_i \end{vmatrix} \quad \hat{b}_2 = \frac{|\Delta_2|}{|\Delta|}$$

3.3. Katsayılar İçin Anlamlılık Sınaması:

$$\hat{S}_{\hat{b}_0}^2 = S_e^2 \left(\frac{1}{n} + \frac{\bar{x}_1^2 \sum x_{2i}'^2 + \bar{x}_2^2 \sum x_{1i}'^2 - 2\bar{x}_1 \bar{x}_2 \sum x_{1i}' x_{2i}'}{\sum x_{1i}'^2 \sum x_{2i}'^2 - (\sum x_{1i}' x_{2i}')^2} \right) \quad (3.3.1)$$

$$\hat{S}_{\hat{b}_1}^2 = \frac{S_e^2 \sum x_{2i}'^2}{\sum x_{1i}'^2 \sum x_{2i}'^2 - (\sum x_{1i}' x_{2i}')^2} \quad (3.3.2)$$

$$\hat{S}_{\hat{b}_2}^2 = \frac{S_e^2 \sum x_{1i}'^2}{\sum x_{1i}'^2 \sum x_{2i}'^2 - (\sum x_{1i}' x_{2i}')^2} \quad (3.3.3)$$

$$H_0 : b_0 = 0$$

$$H_0 : b_1 = 0$$

$$H_0 : b_2 = 0$$

$$H_1 : b_0 \neq 0$$

$$H_1 : b_1 \neq 0$$

$$H_1 : b_2 \neq 0$$

$$t_h = \frac{\hat{b}_0}{\hat{S}_{\hat{b}_0}}$$

$$t_h = \frac{\hat{b}_1}{\hat{S}_{\hat{b}_1}}$$

$$t_h = \frac{\hat{b}_2}{\hat{S}_{\hat{b}_2}}$$

Hesaplanan t istatistik değerleri $t_{\alpha/2; n-k}$ tablo değeri ile karşılaştırılarak karar verilir.

3.4. Çoklu Belirlilik Katsayısı:

Çoklu belirlilik katsayısı, bağımsız değişkenlerin bağımlı değişkeni açıklama oranıdır ve aşağıdaki gibi hesaplanır.

$$R^2 = \frac{\sum \hat{y}_i'^2}{\sum y_i'^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad (3.4.1)$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum y_i'^2} \quad (3.4.2)$$

$$R^2 = \frac{\hat{b}_1 \sum y_i' x_{1i}' + \hat{b}_2 \sum y_i' x_{2i}'}{\sum y_i'^2} \quad (3.4.3)$$

3.5. Düzeltilmiş Belirlilik Katsayısı:

R^2 nin (3.4.1), (3.4.2) ve (3.4.3) formülleri, fonksiyona eklenen yeni bağımsız değişkenlerden doğan serbestlik derecesini hesaba katmaz. Bu sakıncayı gidermek için fonksiyona yeni değişkenler katıldığında azaldığı belli olan serbestlik derecesini hesaba katmak üzere düzeltilmiş çoklu belirlilik katsayısı hesaplanır:

$$R_D^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

$$R_D^2 = 1 - \left(\frac{\sum e_i^2 / (n-k)}{\sum y_i'^2 / (n-1)} \right)$$

ÇOKLU DOĞRUSAL REGRESYON ANALİZİ İÇİN ÇÖZÜMLÜ ÖRNEKLER

Örnek 1: Bir pazarlama firması gazete ve radyo reklamları ile uygun bir karışım hazırlayarak brüt satışları etkilemeye karar vermiştir. Yönetim iki benzer pazar alanındaki üretimin iki medya reklamları için harcanan para miktarları arasındaki farkları test etmiştir. Harcamalar ve satış sonuçları cetvel halinde aşağıda verilmiştir. Çok değişkenli lineer regresyon denklemini bulunuz.

Pazar Alanı	Radyo (x_1) \$ (10^{-2})	Gazete (x_2) \$ (10^{-2})	Brüt Satışlar (y) \$ (10^{-6})
A	42	14	24
B	45	16	29
C	44	15	27
D	43	15	27
E	45	13	25
F	43	13	26
G	46	14	28
H	44	16	30
I	45	16	28
J	44	15	28
TOPLAM	441	147	272

Elde edilecek regresyon denklemi : $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

\hat{y} = Brüt satış tahmini

x_1 = Radyo reklamı için harcanan miktar

x_2 = Gazete reklamı için harcanan miktar

Normal denklemler yardımıyla çözüldüğünde $\hat{\beta}_0$, $\hat{\beta}_1$ ve $\hat{\beta}_2$ aşağıdaki eşitliklerden bulunacaktır.

(Kaynak: Probability and Statistics Exam File.)

Çözüm 1:

$$\Sigma y = n \cdot \hat{\beta}_0 + \hat{\beta}_1 \Sigma x_1 + \hat{\beta}_2 \Sigma x_2$$

$$\Sigma x_1 y = \hat{\beta}_0 \Sigma x_1 + \hat{\beta}_1 \Sigma x_1^2 + \hat{\beta}_2 \Sigma x_1 x_2$$

$$\Sigma x_2 y = \hat{\beta}_0 \Sigma x_2 + \hat{\beta}_1 \Sigma x_1 x_2 + \hat{\beta}_2 \Sigma x_2^2$$

x_1^2	x_1x_2	x_1y	x_2y	x_2^2
1764	588	1008	336	196
2025	720	1305	464	256
1936	660	1188	405	225
1849	645	1161	405	225
2025	585	1125	325	169
1849	559	1118	338	169
2116	644	1288	392	196
1936	704	1320	480	256
2025	720	1260	448	256
1936	660	1232	420	225
19416	6485	12005	4013	2173

$$272 = 10.\hat{\beta}_0 + 441\hat{\beta}_1 + 147\hat{\beta}_2$$

$$12005 = 441\hat{\beta}_0 + 19461\hat{\beta}_1 + 6485\hat{\beta}_2$$

$$4013 = 147\hat{\beta}_0 + 6485\hat{\beta}_1 + 2173\hat{\beta}_2$$

Denklemler çözüldüğünde;

$$\hat{\beta}_0 = 13,828 \quad \hat{\beta}_1 = 0,564 \quad \hat{\beta}_2 = 1,099$$

Regresyon denklemi: $\hat{y} = 13,828 + 0,564x_1 + 1,099x_2$

Örnek 2: Bir hastane yöneticisi, A hastalığı nedeni ile ameliyat olan hastaların ameliyat sonrası hastanede kalış gün sayısını (y), hastaların tıbbi sorun sayısı (x_1) ve ameliyat öncesi hastanede kalış gün sayısı (x_2) ile ilişkili olup olmadığını, eğer bağımlı ve bağımsız değişkenler arasında bir ilişki varsa ameliyat sonrası hastanede kalış gün sayısını bu iki bağımsız değişkenden hangisinin daha fazla etkilendiğini öğrenmek ve de bu iki bağımsız değişkeni kullanacak hastaların ameliyat sonrası hastanede kalış gün sayısını önceden kestirmek istiyor. Rasgele seçilen 20 hastaya ilişkin veriler aşağıdaki tabloda verilmiştir.

- Regresyon denklemini matrisler yardımıyla belirleyiniz.
- Modelin genel anlamlılığını varyans analizi ile %5 A.D’de belirleyiniz.
- Regresyon katsayılarının anlamlılığını matrisler yardımıyla belirleyiniz.
- Regresyon katsayılarının güven aralıklarını ve çoklu belirlilik katsayısını hesaplayınız.

No	y	x ₁	x ₂	No	y	x ₁	x ₂
1	6	1	1	11	13	1	4
2	6	2	1	12	9	1	2
3	11	2	2	13	17	3	3
4	9	1	3	14	17	2	4
5	16	3	3	15	12	4	1
6	16	1	5	16	6	1	1
7	4	1	1	17	5	1	1
8	8	3	1	18	12	3	2
9	11	2	2	19	8	1	2
10	13	3	2	20	9	2	2

Çözüm 2:

Hastane verisine ilişkin bazı istatistikler aşağıdadır.

$$\begin{aligned}
\Sigma y_i &= 208 & \Sigma x_{1i} &= 38 & \Sigma x_{2i} &= 43 \\
\Sigma y_i^2 &= 2478 & \Sigma x_{1i}^2 &= 90 & \Sigma x_{2i}^2 &= 119 \\
\Sigma x_{1i}y_i &= 430 & \Sigma x_{2i}y_i &= 519 & \Sigma x_{1i}x_{2i} &= 79 \\
\bar{y} &= 10,4 & \bar{x}_1 &= 1,9 & \bar{x}_2 &= 2,15
\end{aligned}$$

$$y = \begin{bmatrix} 6 \\ 6 \\ . \\ . \\ 9 \end{bmatrix} \quad x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ . & . & . \\ 1 & 2 & 2 \end{bmatrix} \quad x^1 = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ 1 & 1 & \dots & 2 \end{bmatrix}$$

$$x'x = \begin{bmatrix} n & \Sigma x_{1i} & \Sigma x_{2i} \\ \Sigma x_{1i} & \Sigma x_{1i}^2 & \Sigma x_{1i}x_{2i} \\ \Sigma x_{2i} & \Sigma x_{1i}x_{2i} & \Sigma x_{2i}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ 1 & 1 & \dots & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ . & . & . \\ 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 20 & 38 & 43 \\ 38 & 90 & 79 \\ 43 & 79 & 119 \end{bmatrix}$$

$$[x'x]^{-1} = \begin{bmatrix} 0,480228 & -0,12089 & -0,09327 \\ -0,12089 & 0,05706 & 0,005803 \\ -0,09327 & 0,005803 & 0,038255 \end{bmatrix}$$

$$x'y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ 1 & 1 & \dots & 2 \end{bmatrix} \cdot \begin{bmatrix} 6 \\ 6 \\ . \\ . \\ 9 \end{bmatrix} = \begin{bmatrix} 208 \\ 430 \\ 519 \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_{1i}y_i \\ \Sigma x_{2i}y_i \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} 0,480228 & -0,12089 & -0,09327 \\ -0,12089 & 0,05706 & 0,005803 \\ -0,09327 & 0,005803 & 0,038255 \end{bmatrix} \begin{bmatrix} 208 \\ 430 \\ 519 \end{bmatrix} = \begin{bmatrix} -0,50398 \\ 2,402321 \\ 2,948635 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Buna göre regresyon kestirim denklemi;

$$\hat{y}_i = -0,50398 + 2,402321 x_{1i} + 2,948635 x_{2i}$$

Buna göre, örneğin, ameliyat öncesi hastanede kalış gün sayısı sabit tutulduğunda, hastanın tıbbi sorun sayısındaki 1 birimlik artış, ameliyat sonrası hastanede kalış gün sayısında 2,402321 (2,4) günlük bir artışa neden olmaktadır.

b) Toplam Değişim Kareler Toplamı = $[y_1 \ y_2 \ \dots \ y_{20}] \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_{20} \end{bmatrix} = n \cdot \bar{y}^2$

$$\text{TDKT} = [6 \ 6 \ \dots \ 9] \begin{bmatrix} 6 \\ 6 \\ \cdot \\ \cdot \\ 9 \end{bmatrix} - 20(10,4)^2 = 2478 - 2163,2 = 314,8$$

$$\text{Regresyon Kareler Toplamı} = \text{RKT} = [-0,50498 \ 2,402 \ 2,948] \begin{bmatrix} 208 \\ 430 \\ 519 \end{bmatrix} - 20(10,4)^2$$

$$\text{RKT} = 2458,511755 - 2163,2 = 295,31279$$

$$\text{Açıklanamayan Değişim} = \text{Artık K.T} = \text{TDKT} - \text{RKT}$$

$$\text{AKT} = 314,8 - 295,31279 = 19,48721$$

Varyans Analizi Tablosu

Değişkenlik Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması
Regresyon	$k - 1 = 2$	295,31279	147,65639
Artıklar	$n - k = 17$	19,48721	1,14631
Toplam	$n - 1 = 19$	314,8	-

$$H_0 : \beta_0 = \beta_1 = \beta_2 = 0$$

$$H_1 : \text{En az bir } \beta_i \neq 0$$

$$F_h = \frac{147,65639}{1,14631} = 128,81$$

$F_h = 128,81 > F_{0,05; 2,17} = 3,59$ olduğu için doğrusal ilişkinin anlamlı olduğu yada y değişkenindeki değişimin x_1 ve x_2 değişkenlerince açıklanabildiği söylenir.

c) Regresyon denkleminin standart hatası;

$$\hat{\sigma}_{(\beta_i)} = 1,07066 \sqrt{\begin{bmatrix} 0,480228 & & \\ & 0,05706 & \\ & & 0,038255 \end{bmatrix}} = \begin{bmatrix} 0,74194 \\ 0,25575 \\ 0,209408 \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_{\hat{\beta}_0} \\ \hat{\sigma}_{\hat{\beta}_1} \\ \hat{\sigma}_{\hat{\beta}_2} \end{bmatrix}$$

olarak elde edilir ve regresyon katsayılarının anlamlılığı aşağıdaki gibi elde edilir.

$$H_0 : \beta_0 = 0 \quad (\text{Katsayı anlamlı değildir})$$

$$H_1 : \beta_0 \neq 0 \quad (\text{Katsayı anlamlıdır})$$

$$t_h = \frac{\hat{\beta}_0}{\hat{\sigma}_{\hat{\beta}}} = \frac{-0,503976}{0,741949} = 0,679$$

$\alpha = 0,05$ için $t_{0,025; 17} = 2,11 > t_h = 0,679$ olduğundan %5 A.D.'de H_0 kabul edilir. Yani katsayı anlamlı değildir.

$$H_0 : \beta_1 = 0 \quad (\text{Katsayı anlamlı değildir})$$

$$H_1 : \beta_1 \neq 0 \quad (\text{Katsayı anlamlıdır})$$

$$t_h = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}}} = \frac{2,402321}{0,255750} = 9,393 > t_t = 2,11 \text{ olduğundan } \%5 \text{ A.D.'de } H_0 \text{ reddedilir.}$$

Yani katsayı anlamlıdır.

$$H_0 : \beta_2 = 0 \quad (\text{Katsayı anlamlı değildir})$$

$$H_1 : \beta_2 \neq 0 \quad (\text{Katsayı anlamlıdır})$$

$$t_h = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{2,948635}{0,209408} = 14,081 > t_t = 2,11 \text{ olduğundan } \%5 \text{ A.D.'de } H_0 \text{ reddedilir.}$$

Yani katsayı anlamlıdır.

Sonuç olarak, her iki bağımsız değişkenin ameliyat sonrası hastanede kalış gün sayısı üzerine etkisi anlamlıdır.

d) $\alpha = 0,05$ için regresyon katsayılarına ilişkin güven aralıkları aşağıdaki gibidir.

β_0 için Güven Aralığı:

$$\hat{\beta}_0 - t_t \cdot \hat{\sigma}_{\hat{\beta}_0} \leq \beta_0 \leq \hat{\beta}_0 + t_t \cdot \hat{\sigma}_{\hat{\beta}_0}$$

$$-0,503976 - (2,11) (0,741949) \leq \beta_0 \leq -0,503976 + (2,11) (0,741949)$$

$$-2,069 \leq \beta_0 \leq 1,061$$

β_1 için Güven Aralığı:

$$\hat{\beta}_1 - t_t \cdot \hat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_t \cdot \hat{\sigma}_{\hat{\beta}_1}$$

$$2,402321 - (2,11) (0,25575) \leq \beta_1 \leq 2,402321 + (2,11) (0,25575)$$

$$1,863 \leq \beta_1 \leq 2,942$$

β_2 için Güven Aralığı

$$\hat{\beta}_2 - t_t \cdot \hat{\sigma}_{\hat{\beta}_2} \leq \beta_2 \leq \hat{\beta}_2 + t_t \cdot \hat{\sigma}_{\hat{\beta}_2}$$

$$2,948635 - (2,11) (0,209408) \leq \beta_2 \leq 2,948635 + (2,11) (0,209408)$$

$$2,506 \leq \beta_2 \leq 3,390$$

Buna göre, örneğin, ameliyat öncesi hastanede kalış gün sayısı için evren regresyon katsayısı %95 güvenilrlikte 2,5 ile 3,4 arasında değişir.

Çoklu Belirlilik Katsayısı:

$$R^2 = \frac{RKT}{TDKT} = \frac{295,31279}{314,8} = 0,938 \text{ olarak bulunur. Buna göre } x_1 \text{ ve } x_2$$

değişkenleri y'deki değişimi %93,8 oranında açıklamaktadır.

Düzeltilmiş Belirlilik Katsayısı;

$$R^2_D = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (0,062) \frac{(20-1)}{(20-3)} = 0,93 \text{ olur.}$$