5. BASİT DOĞRUSAL REGRESYON MODELİ

Regresyon analizi, belli amaç ve varsayımlara bağlı olarak, bağımsız değişken ya da değişkenlerin bağımlı değişkene nasıl bağlanacağı ile ilgili bir araştırma veya inceleme tekniğidir. Regresyon terimini ilk defa Francis galton kullanmıştır (Galton, 1886). Regresyon analizinin amaçları arasında bağımlı değişken (Y)'deki değişimin açıklanması ve noktalara en iyi eğriyi uydurma vardır. Regresyon analizi, büyük ölçüde popülasyon ortalaması veya bağımlı değişkenin ortalama değerinin tahminini, açıklayıcı değişkenlerin bilinen veya sabit değerlerini esas alarak yapmakla ilgilidir. Ekonometrik analizlerde en yaygın olarak kullanılan model türü olarak regresyon modelleri yer almaktadır.

Regresyon modellerinde değişken sayısı önemlidir. Bu modellerde kullanılan değişken sayısına bağlı olarak modelin basit veya çoklu regresyon modelleri olduğuna karar verilir. Bu bölümde, basit doğrusal regresyon modeli ile yani, biri bağımlı, biri açıklayıcı iki değişken arasında doğrusal bir fonksiyon ile bağlanmış bir model ele alınacaktır (Akkaya, 1990).

5.1. Basit Doğrusal Regresyon Modeli Kavramı

Basit doğrusal regresyon modelinde iki değişken mevcut olup, biri bağımlı diğeri bağımsız değişkendir. Bağımsız değişken; etkileyen ya da açıklayan değişken iken bağımlı değişken etkilenen ya da açıklanan değişkendir. Basit regresyon, değişmelerin bir değişken tarafından açıklanması durumudur (Dikmen, 2012). Modele doğrusal denilmesi ise, modelin fonksiyonel şeklinin doğrusal olmasından ibarettir. Basit regresyon modellerinde, basit ifadesi ise, modelin diğer ekonometrik modellere göre daha kolay analiz ve tahmin edilmesinden kaynaklanmaktadır. Örneğin, buğday verimi ile gübre miktarı arasındaki ilişkiyi açıklamak için oluşturulan doğrusal model, basit doğrusal regresyon modelidir. Diğer bir örnek, bireylerin tüketim harcamalarında görülen toplam değişimin yüzde ne kadarı bireylerin gelirleri ile açıklanabilir?. Bu ilişkileri, matematiksel olarak, y=f(x) bağıntısı ile göstermek mümkündür.

Bağımlı (Y) ve bağımsız (X) değişken olarak iki değişken arasındaki doğrusal ilişkiyi matematiksel bir denklem olarak,

$Y_i = \beta_0 + \beta_1 X_i$ şeklinde ifade etmek mümkündür. Hata teriminin bulunmadığı bir model, iki değişken arasındaki ilişkinin doğrusallığına bir işarettir. Yani, X ve Y değişkenleri arasında deterministik ya da tam bir ilişki söz konusudur (Güriş ve Çağlayan, 2010) (Dikmen, 2012). Ancak, iktisadi olaylarda bu tür ilişkilere pek rastlanmamaktadır. İktisadi değişkenler arasında doğrusallık olmamakla birlikte, ilişkilerin doğrusala yakın olabileceği varsayımdan hareketle, doğrusal modeller kurmak mümkündür.

İki değişken arasında doğrusal bir ilişkinin varlığından hareketle, basit doğrusal anakütle regresyon denklemini,

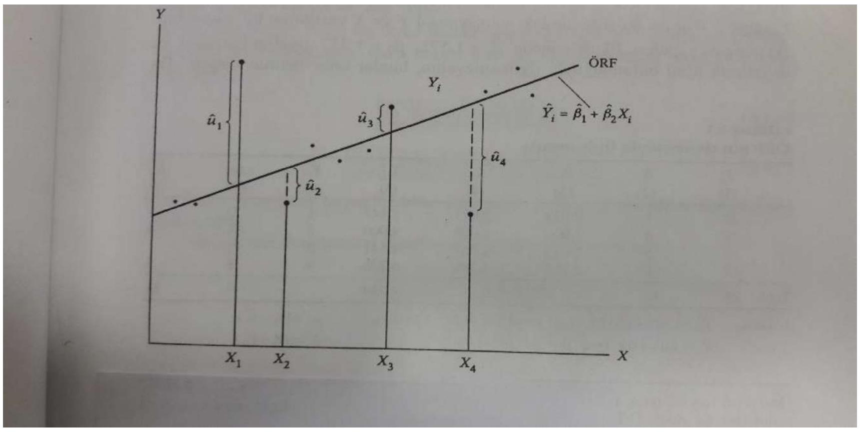
$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ gibi yazabiliriz. Bu denkleme, olasılıklı ya da stokastik model de denir (Güriş ve Çağlayan, 2010) (Dikmen, 2012). Bu regresyon modelinde, eşitliğin sağ tarafında yer alan ve X ile ifade edilen değişkene bağımsız değişken ya da açıklayıcı değişken denilmektedir. Eşitliğin sol tarafında yer alan ve Y ile gösterilen değişken ise, bağımlı veya açıklanan değişkendir.

Modelde yer alan β_0 ve β_1 bilinmeyen regresyon katsayılarıdır. Modelde β_0 sabit terim veya kesim noktası adını alır. β_0 katsayısı regresyon doğrusunun y eksenini kestiği noktayı göstermektedir (Şekil 4). Ayrıca, β_0 değeri ekonometrik açıdan X değişkeni sıfır olduğunda Y bağımlı değişkeninin alacağı değeri gösterir. β_1 katsayısı veya parametresi ise, bağlanım katsayısı adını alır ve doğrunun eğimini gösterir. Bu β_1 katsayısı, istatistiksel olarak, X değişkenindeki bir birimlik değişimin yani artışın veya azalışın bağımlı değişken Y'de yapacağı ortalama değişiklik miktarını vermektedir.

ε_i terimi, hata (yanılgı) terimi veya artık değerdir. i'inci gözlemin artık terimi olarak isimlendirilmekte ve gözlenen değer ile kestirim değeri arasındaki fark $y_i - \hat{y}_i$ olarak bilinmektedir. Bu sapmalar, tesadüfi değişken olan hata terimi tarafından açıklanır. Hata terimi olarak u ya da e harfi de kullanılmaktadır.

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ basit doğrusal anakütle regresyon denkleminin parametreleri bilinmediğinden anakütle ya da popülasyondan çekilen n birimli bir örnekten oluşturulan regresyon denklemi ile tahmin yapılmaktadır. Tahmin sonrası modeli $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$ şeklinde olacaktır. Burada $\hat{\beta}_0$ ve $\hat{\beta}_1$, β_0 ve β_1'in tahmincileridir. e_i ise, hata teriminin örnekten belirlenecek değerlerini ifade etmekte olup artık değer şeklinde de isimlendirilmektedir. Anakütle regresyon denkleminde yer alan ε_i terimi ile e_i terimi birbirinden farklıdır. Çünkü, e_i terimi, ε_i'nin örnekten tahmin edilen değeridir.

X ve Y'ye ait gözlemlerin bir diyagramda gösterilmesi X ve Y arasındaki gerçek ilişkiyi belirlemektedir. Modelde yer alan X ve Y değişkenlerinin aldıkları değerlerin dağılımı Şekil 4'de verilmektedir. Bu dağılımın doğrusal veya doğrusala yakın bir dağılım olduğu varsayımından hareketle noktalar arasından en uygun şekilde regresyon doğrusu geçirilmiştir.



Şekil 4. Modelde yer alan X ve Y değişkenlerinin aldıkları değerler

Regresyon doğrusu X, Y noktaları arasından geçeceğinden dolayı, hata terimleri pozitif veya negatif işaretli olabilir (Güriş ve Çağlayan, 2010) (Dikmen, 2012). Hata terimlerinin beklenen değeri sıfırdır yani

$E(\varepsilon_i) = 0$ olarak ifade edilir. Bu dağılım ortalaması şu şekilde gösteribilir:

$E(\varepsilon_i) = E(y_i - \mu_i) = \mu_i - \mu_i = 0$ 'dır.

Anakütle regresyon modelinin beklenen değeri ise,

$E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 + \beta_1 X_i$ şeklinde olacaktır.

$V(Y_i) = V(\beta_0 + \beta_1 X_i + \varepsilon_i) = V(\varepsilon_i) = \sigma_i^2$ veya

$V(Y_i) = E[Y_i - E(Y_i)]^2 = E[(\beta_0 + \beta_1 X_i + \varepsilon_i) - (\beta_0 + \beta_1 X_i + \varepsilon_i)]^2 = E(\varepsilon_i^2) = \sigma^2$

biçiminde yazılabilir.

5.2. Basit Doğrusal Regresyon Denkleminin Temel Varsayımları

n sayıda gözlemden X bağımsız değişkeni ve Y bağımlı değişkeni arasındaki basit doğrusal regresyon denklemi aşağıdaki gibi,

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ i=1, 2, …, n biçiminde yazılır. Burada β_0, β_1 ve ε_i denklemin bilinmeyenleridir.

Basit doğrusal regresyon denkleminin geçerliliği ve iyi bir kestiriminin elde edilmesinde aşağıdaki varsayımlar geçerlidir:

a) Normallik varsayımı: $\varepsilon_i \sim N(0,\sigma^2), i = 1,2,...,n$ için. Hata terimleri normal dağılım özelliği gösterir.

b) Sıfır ortalama: $E(\varepsilon_i) = 0$, i=1,2,…,n. Hata terimlerinin beklenen değeri sıfırdır.

c) Sabit varyans: $V(\varepsilon_i) = \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2 = \sigma^2$'dir. Regresyon hata terimlerinin varyansı sabittir. Eğer veriler bu varsayımı sağlamıyorsa ortaya değişen varyanslılık sorunu çıkar.

d) Otokorelasyon olmaması: $Cov(\varepsilon_i,\varepsilon_j) = E[(\varepsilon_i - E(\varepsilon_i)][(\varepsilon_j - E(\varepsilon_j)] = 0, \quad i \neq j$. Hata terimleri arasında ilişki olması durumu otokorelasyon (ardışık bağımlılık, özilişki) adını alır. Ancak bu varsayıma göre, bir hata terimi kendinden önce ya da sonra yer alan hata terimlerini etkilememesidir. ε_i'lerin kestirimi olan e_i'lerin birbirleri ile ilişkili olmamaları istenir.

e) Bağımsız değişkenin tesadüfî değişken olmaması: $E(x_i\varepsilon_i) = 0,\ i = 1,2,.....,n$. X'ler raslantı değişkeni olduğunda X ve ε değişkenlerinin bağımsız olmaları gereklidir.

5.3. Basit Doğrusal Regresyon Denkleminin Kestirimi ve Özellikleri

En Küçük Kareler (EKK) yöntemi, Alman Matematikçi Carl Friedrich Gauss (1777-1855) tarafından ortaya atılmış popüler metotlardan biridir (Kutlar, 2007).

İki değişken arasında gerçekte bir ilişki var ise, regresyon denklemi β_0 ve β_1'lerin $y_i - \hat{y}_i$ artıklarını küçük yapabilen kestiricilerin elde edilmesi ile bulunmaktadır. Bu toplamı en küçükleyen ya da minimize eden $\hat{\beta}_0$ ve $\hat{\beta}_1$ kestiricilerinin elde edilmesine ilişkin yöntem En Küçük Kareler (EKK) kestirimi ya da yöntemi adını alır.

En küçük $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ = En küçük $\sum_{i=1}^{n} e_i^2$

β_0 ve β_1'lerin EKK kestirimleri, en küçükleme yöntemi ile;

Artıkların Kareler Toplamı (AKT) = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0$ şeklinde yazılır. Bu terimin, β_0 ve β_1'e göre türevleri alındıktan sonra kısmi türevler sıfıra eşitlenir ve işlemlere aşağıdaki gibi devam edilir:

$$\frac{\partial AKT}{\partial \hat{\beta}_0} = -2\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial AKT}{\partial \hat{\beta}_1} = -2\sum x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Bu iki eşitlik düzenlenirse, aşağıdaki "**normal denklemler"** elde edilir (Johnston, 1991).

$$\sum y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum x_i$$

$$\sum x_i y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2$$

Bu denklemlere normal denklemler adı verilmektedir (Güriş ve Çağlayan, 2010) (Dikmen, 2012) (Johnston, 1991). Normal denklemlerin ortak çözümü sayesinde $\hat{\beta}_0$ ve $\hat{\beta}_1$ tahmin edilmektedir.

5.4. Gerçek Değerler İle Parametre Tahmini

Birinci denklemden $\hat{\beta}_0$ çekildiğinde, β_0 'ın kestiricisi aşağıdaki gibi,

$$\hat{\beta}_0 n + \hat{\beta}_1 \sum x_i = \sum y_i$$

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n}$$

$$\hat{\beta}_0 = \frac{\sum y_i}{n} - \hat{\beta}_1 \frac{\sum x_i}{n}$$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ olarak bulunur.

Elde edilen bu ifade ikinci denklemde yerine konulduğunda ise,

$$\hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$$

$$(\bar{y} - \hat{\beta}_1 \bar{x}) \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$$

$$\bar{y} \sum x_i - \hat{\beta}_1 \bar{x} \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$$

$$\hat{\beta}_1 (\sum x_i^2 - \bar{x} \sum x_i) = \sum x_i y_i - \bar{y} \sum x_i$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i}{\sum x_i^2 - \bar{x} \sum x_i}$$

Pay ve payda $\frac{n}{n}$ ile çarpıldığında, β_1 'in kestiricisi ise

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

(Güriş ve Çağlayan, 2010) (Dikmen, 2012).

Örnek 5.1. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarları verilmiştir. Verimin, kullanılan gübrenin bir fonksiyonu olduğu varsayımı ile oluşturulan basit doğrusal denkleminin parametrelerini En küçük kareler (EKK) yöntemiyle tahmin edelim:

Çizelge 3. Türkiye'de Şekerpancarı Verimi ve Gübre Kullanım Miktarları

Yıl	**Verim (10 Kg/da)**	**Gübre (10 Kg/da)**
2001	354	1.0
2002	444	1.1
2003	401	1.2
2004	429	1.0
2005	452	1.3
2006	446	1.4
2007	415	1.1
2008	482	1.2
2009	533	1.3
2010	545	1.2
2011	548	1.5
2012	532	1.4
2013	566	1.5

Y: Verim (10 Kg/da)
X: Gübre (10 Kg/da)
Kaynak: Türkiye İstatistik Kurumu, 2014.

$$\sum y_i = 6147$$

$$\sum x_i = 16{,}20$$

$$\sum x_i y_i = 7762{,}20$$

$$\sum x_i^2 = 20{,}54$$

a)Normal denklemler ile parametre tahmini,

$$\sum y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum x_i$$

$$\sum x_i y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2$$

6147=13 $\hat{\beta}_0$ +16,20 $\hat{\beta}_1$

7762,20=16.20 $\hat{\beta}_0$ +20,54 $\hat{\beta}_1$

$\hat{\beta}_0$=111,7336

$\hat{\beta}_1$=289,7817

olarak tahmin edilir. Buna göre model,

$\hat{Y}_i = 111,73 + 289,78 X_i$ olur.

b)Gerçek değerler ile parametre tahmini,

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum x_i^2 - n\overline{x}^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Burada,

$$\overline{Y} = \frac{\sum y_i}{n} = \frac{6147}{13} = 473$$

$$\overline{X} = \frac{\sum x_i}{n} = \frac{16,20}{13} = 1,25$$

$$\hat{\beta}_1 = \frac{7762,20 - 13(1,25)(473)}{20,54 - 13(1,25)^2} = 289,78$$

$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$ =473-(289,78)(1,25)=111,73

olarak tahmin edilir. Buna göre model,

$\hat{Y}_i = 111,73 + 289,78 X_i$ olur.

Bu iki yöntemin sonuçlarının aynı çıktığı görülmektedir. Basit doğrusal regresyon modelinde EKK yöntemi ile $\hat{\beta}_0$ ve $\hat{\beta}_1$ olmak üzere iki bilinmeyen katsayı hesaplanır. $\hat{\beta}_0$ değeri sabit terim ya da kesim noktası ve $\hat{\beta}_1$ değeri bağımsız değişken X'in katsayısı ve regresyon doğrusunun eğim derecesi olarak bilinir.

$\hat{\beta}_0$ değeri, denklemin sağ tarafında yer alan X değişkeninin sıfır değeri alması durumunda bağımlı değişken Y'nin alacağı değeri ifade eder. $\hat{\beta}_1$ değeri ise, basit doğrusal regresyon denkleminde bağımlı değişken Y'nin bağımsız değişken X'e göre birinci türevi $\hat{\beta}_1$ katsayısına eşit olur. Bu katsayının pozitif veya negatif değerli olmasına göre yorum yapmak gerekir. Eğer $\hat{\beta}_1$ katsayısı pozif değere sahipse, X'de meydana gelen değişmelerden Y doğru yönde etkilenir aksi halde ters yönde etkilenir.

Şekerpancarı verim ve gübre kullanım miktarları için tahmin edilen model parametrelerini yorumlayalım. Regresyon modeli $\hat{Y}_i = 111{,}73 + 289{,}78 X_i$ olarak tahmin edilmiştir. Modelde $\hat{\beta}_0$=111,73 değeri sabit terimdir ve gübrenin sıfır olması durumunda yani bağımsız değişken X'in sıfır olması halinde şekerpancarı verimi 111.73 kg/da olur.

Regresyon modelinde $\hat{\beta}_1$=289,78'dir. Bu katsayı pozitif bir değere sahiptir. Gübre kullanım miktarı ve şekerpancarı verimi arasında pozitif yönlü ve doğrusal bir ilişkiden bahsedilir. Yani, gübre kullanım miktarları artarsa şeker pancarı verimi de artar. Diğer bir ifadeyle, gübre kullanım miktarlarında %1 bir artış meydana geldiğinde şeker pancarı veriminde %290 oranında bir artış meydana gelmektedir.

5.5. Artık Değerler ve Teorik Değerlerin Hesaplanması

Gözlem değerleri (y_i) ve bu değerlerden hesaplanmış olan tahmin değerleri ($\hat{y}_i$) arasındaki fark $u_i = y_i - \hat{y}_i$'dır. Bu formül yardımı ile artık değerler (hata terimleri) hesaplanmaktadır. $u_i = y_i - \hat{y}_i$ yardımı ile hesaplanmış olunan artık değerlerin bazıları pozitif veya negatif değerleri olmaktadır (Dikmen, 2012).

Örnek 5.2. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarları verilmiştir. Bu tahmin edilmiş modelin teorik değerlerini, regresyon hata terimlerini ve artık değerlerin kareleri toplamını hesaplayalım.

Çizelge 4. Türkiye'de Şekerpancarı Verimi, Gübre Kullanım Miktarları, Hesaplanmış Değerler

Yıl	**Verim(10 kg/da)**	**Gübre (10 Kg/da)**	$\hat{Y}_i$	$u_i = y_i - \hat{y}_i$	u_i^2
2001	354	1.0	401,52	-47,52	2257,70
2002	444	1.1	430,49	13,51	182,43
2003	401	1.2	459,47	-58,47	3418,93
2004	429	1.0	401,52	27,48	755,41
2005	452	1.3	488,45	-36,45	1328,59
2006	446	1.4	517,43	-71,43	5101,96
2007	415	1.1	430,49	-15,49	240,05
2008	482	1.2	459,47	22,53	507,53
2009	533	1.3	488,45	44,55	1984,72
2010	545	1.2	459,47	85,53	7315,10
2011	548	1.5	546,41	1,59	2,54
2012	532	1.4	517,43	14,57	212,34
2013	566	1.5	546,41	19,59	383,92
					$\sum u_i^2 = 23691{,}21$

Kaynak: Türkiye İstatistik Kurumu, 2014.

5.6. Parametrelerin Kovaryansı

Kovaryans iki raslantı değişkeninin birbirine olan yakınlığını ölçen bir istatistiktir. Eğer $S_{x,y}$=0 eşit ise iki değişken arasında bir ilişki yoktur. $S_{x,y}$ negatif işaretli ise değişkenler arasında ters yönlü bir ilişkiden bahsedilirken, $S_{x,y}$ pozitif işaretli ise değişkenler arasında aynı yönlü bir ilişkiden bahsedilmektedir.

X ve Y gibi iki raslantı değişkeni arasında ortak varyans veya kovaryansı, kitle ve örneklem için aşağıdaki gibi yazabiliriz:

Her iki değişkenin ortalamaya göre düzeltilmiş kareler toplamı aşağıdaki gibi açık bir şekilde,

$$ÇT_{XY} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = \sum x_i y_i - n\overline{x}\overline{y}$$

yazılır.

Kitle için kovaryans, $\sigma_{x,y} = \dfrac{ÇT_{XY}}{N}$'e bölümü ile elde edilirken,

Örneklem kovaryansı ise $S_{x,y} = \dfrac{ÇT_{XY}}{n-1}$'e bölümü ile elde edilir.

Örnek 5.3. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarları verilmiştir. Bu tahmin edilen modelin kovaryansını hesaplayalım.

$$S_{x,y} = \frac{ÇT_{XY}}{n-1}$$

$$S_{x,y} = \frac{102,1}{13-1} = 8,51$$

5.7. Parametre Tahmincilerinin Varyansları ve Standart Hataları

$\hat{\beta}_0$ ve $\hat{\beta}_1$ kestiricileri en küçük varyanslılık özelliğini taşımaktadır. $\hat{\beta}_0$ 'nın varyansı,

$$Var(\hat{\beta}_0) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{KT_x}) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2})$$

$\hat{\beta}_1$ 'in varyansı ise,

$V(\hat{\beta}_1) = \dfrac{\sigma^2}{KT_X} = \dfrac{\sigma^2}{\sum (x_i - \bar{x})^2}$ elde edilir (Güriş ve Çağlayan, 2010).

Yukarıdaki eşitliklerde, σ^2 hata terimleri veya artıkların varyansını ifade eder. Artıkların varyansı ise,

Artık Kareler Toplamı= $AKT = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$

Artık Kareler Ortalaması= $AKO = \dfrac{AKT}{n-k} = \dfrac{\sum e_i^2}{n-k} = \sigma^2$ elde edilir. AKO, artıkların varyansı yani σ^2 'nin bir kestirimidir. Formülde n-k değeri, serbestlik derecesi olarak bilinir. Gözlem sayısı n ile gösterilirken, k ise sabit terim dahil diğer regresyon katsayıları toplamını verir. Basit doğrusal regresyon modelinde artıkların hesabında β_0 ve β_1 kullanıldığı için serbestlik derecesi n-k = n-2 olarak bulunur.

Tahmin edilen değerlerin standart hatası $S_{\hat{\beta}_0}$ ve $S_{\hat{\beta}_1}$ olarak hesaplanmaktadır. Burada $S_{\hat{\beta}_0} = \sqrt{V(\hat{\beta}_0)}$ ve $S_{\hat{\beta}_1} = \sqrt{V(\hat{\beta}_1)}$ olarak hesaplanır.

Örnek 5.4. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarlarını kullanarak bu örnek için hata terimleri varyansını ve standart hatasını, parametre tahmincilerinin varyansını ve standart hatalarını hesaplayalım.

Hata terimi varyansı tahmini;

$$AKO = \frac{AKT}{n-k} = \frac{\sum e_i^2}{n-k} = \sigma^2$$

$$AKO = \sigma^2 = \frac{\sum e_i^2}{n-k} = \frac{23691,21}{13-2} = 2153,75$$

Parametre tahmincilerinin varyansları ve standart hataları;

Sabit terim $\hat{\beta}_0$ için,

$$Var(\hat{\beta}_0) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{KT_x}) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2})$$

$$Var(\hat{\beta}_0) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}) = (2153,75)(\frac{1}{13} + \frac{(1,246)^2}{0,3523}) = 9656,82$$

Değişken katsayısı $\hat{\beta}_1$ için,

$$V(\hat{\beta}_1) = \frac{\sigma^2}{KT_X} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$V(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \frac{2153,75}{0,3523} = 6113,39$$

5.8. Değişken Katsayılarının Varyans-Kovaryans Matrisi

$\hat{\beta}$ 'nın varyans-kovaryans matrisi yardımı ile,

$$V(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

$= \sigma^2 (X'X)^{-1}$ olarak yazılır. (k+1)*(k+1) boyutlu bir matrisin köşegen öğeleri katsayıların varyansını vermektedir. Bu da,

$V(b_j) = \sigma^2 (X'X)^{-1}_{jj}$ şeklinde gösterilir. Varyans-kovaryans matrisi ise şöyle olur:

$$V(\hat{\beta}) = \begin{bmatrix} V(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & . & . & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & V(\hat{\beta}_2) & . & . & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ . & . & . & . & . \\ . & . & . & . & . \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & . & . & V(\hat{\beta}_k) \end{bmatrix}$$

Örnek 5.5. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarlarını kullanarak bu örnek için parametre tahmincilerinin varyans değerini ve standart hatasını varyans-kovaryans matrisi yardımıyla hesaplayalım.

$$V(b_j) = \sigma^2 (X'X)^{-1}_{jj}$$

$$AKO = \frac{AKT}{n-k} = \frac{\sum e_i^2}{n-k} = \sigma^2$$

X matrisi ve X matrisinin transpozunun (X') çarpımı ise aşağıdaki gibi,

$$X'X = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} = \begin{bmatrix} 13 & 16{,}20 \\ 16{,}20 & 20{,}54 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 4{,}47772 & -3{,}5316 \\ -3{,}5316 & 2{,}834 \end{bmatrix} \text{ elde edilir.}$$

Varyans değerleri,

$$V(b_j) = \sigma^2 (X'X)^{-1}_{jj}$$

$$V(b_j) = 2153{,}75 \begin{bmatrix} 4{,}47772 & -3{,}5316 \\ -3{,}5316 & 2{,}834 \end{bmatrix} = \begin{bmatrix} 9643{,}89 & -7606{,}18 \\ -7606{,}18 & 6103{,}73 \end{bmatrix}$$

Standart hata değerleri,

$$S(\hat{\beta}_0) = \sqrt{V(\hat{\beta}_0)} = \sqrt{9643{,}89} = 98{,}20$$

$$S(\hat{\beta}_1) = \sqrt{V(\hat{\beta}_1)} = \sqrt{6103{,}73} = 78{,}13 \text{ elde edilir.}$$

5.9. Parametre Tahmincilerinin Güven Aralıkları

İstatistiksel analizler sayesinde elde edilen kestirim değerlerinin gerçek değerlere genelleştirme aralık tahminleri vasıtasıyla yapılır. Regresyon çözümlemesi ile elde edilen $\hat{\beta}_0$ ve $\hat{\beta}_1$ kestirimlerinden kitle ya da popülasyon için güven aralıkları veya aralık tahminlerini yapmak mümkündür. Regresyonda bu aralık tahminleri ile kestirim, normallilik varsayımı sağlanması koşulu altında yapılmaktadır.

Örnek sayısı n değerine bağlı olarak örnekleme dağılımı belirlenir. Yani, n<30 olması durumunda n-2 serbestlik dereceli t dağılımı kullanılmaktadır. $\hat{\beta}_0$'ın aralık kestiricisi,

$$P(\hat{\beta}_0 - t_{\alpha/2;n-2} S_{\hat{\beta}_0} \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2;n-2} S_{\hat{\beta}_0}) = 1-\alpha$$

ve $\hat{\beta}_1$'in aralık kestiricisi,

$$P(\hat{\beta}_1 - t_{\alpha/2;n-2} S_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2;n-2} S_{\hat{\beta}_1}) = 1-\alpha$$ olarak elde edilir.

Örnek 5.6. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarlarını kullanarak tahmin edilen $\hat{\beta}_0$ ve $\hat{\beta}_1$ parametreleri için güven aralıklarını bulunuz.

Standart hata değerleri,

$$S(\hat{\beta}_0) = \sqrt{V(\hat{\beta}_0)} = \sqrt{9643{,}89} = 98{,}20$$

$$S(\hat{\beta}_1) = \sqrt{V(\hat{\beta}_1)} = \sqrt{6103{,}73} = 78{,}13$$ olarak hesaplanır.

α=0,05 için,

sd=n-k=13-2=11

$$t_{\alpha/2;n-2} = t_{0{,}025;11} = 2{,}201$$

$\hat{\beta}_0$'ın aralık kestiricisi,

$$P(\hat{\beta}_0 - t_{\alpha/2;n-2} S_{\hat{\beta}_0} \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2;n-2} S_{\hat{\beta}_0}) = 1-\alpha$$

$$P(111{,}7336 - (2{,}201)(98{,}203)) \leq \beta_0 \leq (111{,}7336 + (2{,}201)(98{,}203)) = 0{,}95$$

$$P(111{,}63279 \leq \beta_0 \leq 327{,}9326) = 0{,}95$$

Bu aralığa göre, %5 hata payı olması durumunda β_0 katsayısı 111,63 ile 327,93 arasında değişir.

$\hat{\beta}_1$'in aralık kestiricisi,

$$P(\hat{\beta}_1 - t_{\alpha/2;n-2} S_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2;n-2} S_{\hat{\beta}_1}) = 1-\alpha$$

$$P(289{,}7817 - 2{,}201(78{,}126)) \leq \beta_1 \leq (289{,}7817 + 2{,}201(78{,}126)) = 0{,}95$$

$$P(117{,}6437 \leq \beta_1 \leq 461{,}9197) = 0{,}95$$

Bu aralığa göre, %5 hata payı olması durumunda β_1 katsayısı 117,64 ile 461,92 arasında değişir.

Ana kütle parametresi, serbestlik derecesi içerisinde (sd=n-k), hesaplanmış olan güven aralığı içinde yer alma olasığını ifade eder.

5.10. Belirtme (Belirlilik) katsayısı

Regresyon denkleminin verilere uyumunun bir göstergesidir. Belirtme katsayısı, açıklanabilen değişimin toplam değişime oranlanmasıyla elde edilir. Aynı zamanda, bağımlı değişkendeki değişimin yüzde ne kadarının bağımsız değişken veya değişkenlerce açıklanabileceğini ifade etmektedir.

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

$$R^2 = \frac{BKT}{KT_y} = \frac{b.ÇT_{XY}}{\sum(y_i - \bar{y})^2}$$ olarak bulunur.

Belirtilememe yüzdesi ise, açıklanamayan kısmın toplam değişime oranlanması ile edilir. Elde edilen bu yüzde, bağımlı değişkendeki değişimin %(1-R^2) kadarı denklem dışı diğer etkenler tarafından açıklanabilen kısmıdır.

Bu ise,

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{AKT}{KT_y} = 1 - \frac{\sum u_i^2}{\sum(y_i - \bar{y})^2}$$ olarak bulunur.

Örnek 5.7. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarları verilmiştir. Belirtme (Determinasyon) katsayısını hesaplayarak yorumlayınız.

$$R^2 = \frac{BKT}{KT_y} = \frac{29584{,}48}{53275{,}69} = 0{,}555309 \cong 0{,}56$$ olarak

veya,

$$R^2 = 1 - \frac{AKT}{KT_y} = 1 - \frac{\sum u_i^2}{\sum (y_i - \bar{y})^2}$$ olarak bulunur.

$$R^2 = 1 - \frac{AKT}{KT_y} = 1 - \frac{\sum u_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{23691,21}{53275,69} = 0,555309 \cong 0,56$$

olarak bulunur.

Şekerpancarı verimindeki değişmelerin %56'sı kullanılan gübre miktarındaki değişmeler ile açıklanabilirken %44'ü diğer etmenlere bağlıdır.

5.11. Korelasyon (İlişki) Katsayısı

Korekasyon katsayısı, belirtme katsayısının kareköküne eşit olan bir ölçüdür. Bunu,

$r = \mp\sqrt{R^2}$ ile gösterilir. Korekasyon katsayısı, Y ve X değişkenleri arasınaki ilişkinin derecesini göstermekte olup negatif değer alabilir. R^2'nin kök dışına nasıl çıkacağına (pozitif ya da negatif değerli olarak) eğim katsayısının işaretine bakarak karar verilebilir. Eğim katsayısı pozitif ise korelasyon katsayısı pozitif, negatif ise korelasyon katsayısı negatif işaretli olur (Akkaya, 1990).

Korelasyon çözümlemesinin amacı, değişkenler arasındaki ilişkinin hem yönü hem de derecesi hakkında bilgi vermektir. Korelasyon, iki ya da daha çok sayıda değişken arasındaki ilişkiyi göstermekte olup, ilgi miktarı ise korelasyon katsayısı ile belirlenmektedir. Korelasyon katsayısı, 0 ile ±1 arasında değer alır. Eğer iki değişken ters yönde ilişkili ise eksi, aynı yönde ilişkili ise artı, ilişki yok ise sıfır ya da sıfıra yakındır (Akkaya, 1990).

Örneklem ilişki katsayısı,

$$r = \frac{ÇT_{xy}}{\sqrt{KT_X KT_Y}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum x_i - \bar{x})^2 (\sum y_i - \bar{y})^2}}$$

şeklinde bulunur (Johnston, 1991).

Örnek 5.8. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarları kullanılarak korelasyon katsayısını hesaplayarak, yorumlayınız.

$$r = \frac{ÇT_{xy}}{\sqrt{KT_X KT_Y}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{102,09}{\sqrt{(0,3523)(53275,69)}} = 0,75$$

Şekerpancarı verimi ve gübre kullanım miktarı arasında aynı yönde bir ilişki mevcuttur.

5.12. Anlamlılık Testleri

Regresyon çözümlemesi bir örnekleme çalışması olduğu için, örneklemden elde edilen model denklemlerine ilişkin katsayıların gerçek katsayılara veya sıfıra eşitliğinin incelenmesi gerekmektedir. Verilerin bu örneklem modeli sayesinde temsil edilip edilemeyeceğini göstermektedir.

$\hat{\beta}_1$ doğrusallığı gösteren bir terim olduğu için, regresyonda ağırlık olarak $\hat{\beta}_1$'nın testi üzerinde durulmaktadır. $\hat{\beta}_0$ doğrusallığı gösteren bir terim değil, modelin sadece bir katsayısıdır ve bunun testini yapmak çoğu zaman gereksiz gözükmektedir.

Burada verilecek bütün testlerin geçerliliği için, en küçük kareler varsayımları geçerlidir. Bunun için, gözlemlerin birbirinden bağımsız olduğu, hata ya da artık terimlerin sıfır ortalamaya ve σ^2 varyansı ile normal dağılıma sahip olması gerekmektedir.

5.12.1. t testi

Amacımız regresyon doğrusunun anlamlılık testi olduğunda,

Temel hipotez eşitliği $H_0 : \beta_1 = 0$ ve alternatif hipotez $H_1 : \beta_1 \neq 0$ şeklinde oluşturulur. β_1'nın standart hatası $S_{\hat{\beta}_1}$ şeklinde daha önce verilmişti. H_0 reddedilir ise $\beta_0 \neq 0$ ve regresyon doğrusu anlamlıdır. Elde ettiğimiz doğrusal regresyon modelini amacımıza uygun olarak kullanabiliriz. Bu bağlamda, veri noktalarından yaklaşık olarak bir doğru çizmek mümkün olabilir. H_0'ın kabul edilmesi durumunda ise, örneklem regresyon denkleminin amaca yani doğrusallığa uygun olarak kullanılamayacağı ortaya çıkar. Bu durumu, iki değişken arasında doğrusal bir ilişkinin yokluğu şeklinde yorumlamak mümkündür.

Buna göre H_0 önsavı,

$t = \dfrac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$ ile test edilir. $t > t_{\alpha;n-2}$ ise H_0 reddedilir.

Örnek 5.9. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarlarını kullanarak $\hat{\beta}_0\, ve\, \hat{\beta}_1$ katsayılarının anlamlılıklarını test edelim ve yorumlayalım.

Temel hipotez eşitliği $H_0 : \beta_1 = 0$ ve alternatif hipotez $H_1 : \beta_1 \neq 0$ önsavlarını test edelim.

$$t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{289,7817}{78,126} = 3,71$$

$$t_{\alpha;n-2} = t_{0,05;13-2} = 1,80$$

$t > t_{\alpha;n-2}$ olduğundan H_0 reddedilir. %95 güvenilirlik düzeyinde $\hat{\beta}_1$ istatistiksel bakımdan anlamlıdır.

5.12.2. Varyans Çözümlemesi ve F Testi

$H_0 = \hat{\beta}_1 = 0$ şeklinde verilen önsavın testini varyans çözümlemesi ile de yapmak mümkündür. F testinde tüm bağımsız değişkenlerin katsayıları birlikte test edilmektedir. Ancak, sabit katsayının testi yapılmamaktadır. Basit doğrusal regresyon için F ve t testi aynı sonucu vermektedir. Çünkü basit doğrusal regresyonda tek bağımsız değişken mevcuttur.

Temel hipotez eşitliği $H_0 = \beta_1 = 0$ ve alternatif hipotez $H_1 = \beta_1 \neq 0$ şeklinde oluşturulur. F testinde tek taraflı test söz konusu değildir (Güriş ve Çağlayan, 2010) (Dikmen, 2012).

Açıklanamayan ve açıklanabilen değişimlerin varyans göstergeleri olarak artık kareler ortalaması (AKO) ve regresyon doğrusuna uyumun kareler ortalaması (BKO) olarak göstermek mümkündür. Buradan F test istatistiği,

$$F = \frac{BKT/1}{AKT/(n-2)} = \frac{BKO}{AKO} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2/1}{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/n-2}$$ şeklinde

hesaplanır.

R²'nin formülü ile F test istatistiği karşılaştırıldığında aralarında bir ilişki görülecektir. Bu ilişki,

$$F = \frac{R^2}{1-R^2}\frac{n-2}{1} = \frac{(n-2)R^2}{1-R^2}$$ şeklinde olacaktır (Güriş ve Çağlayan, 2010) (Dikmen, 2012).

Bu hesaplanan F test istatistikleri ile Çizelge değerleri karşılaştırılırsa,

$F > F_{\alpha;1,n-2}$ ise H_o hipotezi red

$F < F_{\alpha;1,n-2}$ ise H_o hipotezi kabul edilir.

Ayrıca, bağımlı değişkendeki değişimin iki kaynağı yani AKO ve BKO alınarak varyans çözümleme Çizelgesi aşağıdaki gibi kurulmaktadır. Bu çizelge sayesinde,

$H_0 = \beta_1 = 0$ (Doğrusal regresyon uygun değildir)

$H_1 = \beta_1 \neq 0$ ($Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ denklemine uyum anlamlıdır) önsavları test edilmektedir.

Çizelge 5. Varyans Çözümleme Çizelgesi

Değişim Kaynağı (D.K.)	Serbestlik Derecesi (S.D.)	Kareler Toplamı (KT)	Kareler Ortalaması (KO)	F
Toplam Değişme (KT_y)	n-1	$\sum_{i=1}^{n}(Y_i - \overline{Y})^2$		
Açıklanan Değişme (BKT)	1	$\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$	$\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 / 1$	$\frac{BKO}{AKO} = \frac{BKT/1}{AKT/n-2}$
Açıklanamayan Değişme (AKT)	n-2	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ /n-2	

Buradan hesaplanan F hesap değeri ile $F_{\alpha;1,n-2}$ Çizelge değeri karşılaştırıldığında,

$F > F_{\alpha;1,n-2}$ ise H_o hipotezi red

$F < F_{\alpha;1,n-2}$ ise H_o hipotezi kabul edilir.

Çizelgeden görüleceği üzere, F değeri ile t^2 aynıdır. Bu ise,

$$F = \frac{BKO}{AKO} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 / 1}{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / n-2} = \frac{\hat{\beta}_1.ÇT_{XY}}{\hat{\sigma}^2} = t^2$$ şeklinde gösterilir.

Örnek 5.10. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarlarını kullanarak varyans çözümleme çizelgesini oluşturarak modelin anlamlılığını sınayınız.

Çizelge 6. Varyans Çözümleme Değerleri

D.K.	S.D.	KT	KO	F
KT_y	12	53275,69		
BKT	1	29584,48	29584,48	13,74
AKT	11	23691,21	2153,75	

$F = 13,73628 > F_{0,05;1,13-2} = 9,65$ olduğundan H_o hipotezi reddedilir.

5.13. Esneklik

Esneklik, bağımsız değişkende meydana gelen %1'lik artışa bağlı olarak bağımlı değişkende meydana gelen artış veya azalışın yüzde değeridir. Bağımlı değişken (Y)'nin bağımsız değişken (X)'e göre esnekliği,

$$E_{XY} = \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta Y}{\Delta X}.\frac{X}{Y}$$

olarak bulunur. Tek bir nokta için esneklik ise,

$$E_{XY} = \frac{dY / Y}{dX / X} = \frac{dY}{dX}.\frac{X}{Y} \text{ olur.}$$

Tahmin edilen basit doğrusal regresyon modelimiz $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ şeklinde olsun. Bu modelde $\hat{\beta}_1$ katsayısı eğimi verirken, $\hat{\beta}_0$ kesim noktasını vermektedir. Burada bağımlı değişken (Y)'nin bağımsız değişken (X)'e göre esnekliği,

$$E_{XY} = \frac{dY}{dX}.\frac{X}{Y} = \hat{\beta}_1.\frac{X}{Y}$$ olarak bulunur. Ortalama esneklik ise, bağımlı ve bağımsız değişkenlerin birlikte kullanımı sonucunda aşağıdaki gibi,

$$E_{XY} = \frac{dY}{dX}.\frac{X}{Y} = \hat{\beta}_1.\frac{\overline{X}}{\overline{Y}}$$ olacaktır (Güriş ve Çağlayan, 2010).

Örnek 5.11. 2001-2013 yılları arasındaki Şekerpancarı verimi ve buna etki eden gübre kullanım miktarlarını kullanarak ortalama esneklik değerini hesaplayalım.

$$E_{XY} = \frac{dY}{dX}.\frac{X}{Y} = \hat{\beta}_1.\frac{\overline{X}}{\overline{Y}}$$

$$E_{XY} = \hat{\beta}_1.\frac{\overline{X}}{\overline{Y}} = 289{,}78 * \frac{1{,}25}{473} = 0{,}76 \text{ olur.}$$

Gübrede %1 artış olması halinde ortalama olarak şekerpancarı verimi %0.76 artmaktadır.

5.14. Basit Doğrusal Regresyon Modelinde Matris Gösterimleri

Tek bağımsız değişken ve n sayıda gözlemin olduğu normal denklemlerin çözümü doğrudan yapılmaktadır. Ancak, değişken sayısının artmasına paralel olarak hesaplama ve aynı zamanda gösterim güçlükleri ile karşılaşılır. Bunun için, çok değişkenli regresyon denklemine geçilmeden önce iki değişken üzerinden regresyonun matrisler ile gösterimi aşağıdaki şekilde ifade edilecektir.

Tek bağımsız değişkenin olduğu model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ şeklindedir. Bu modelin matrisler ile gösterimi n gözlem için aşağıdaki gibi yazılabilir:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ 1 & X_n \end{bmatrix} \quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \qquad \varepsilon_i = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Anakütle regresyon denklemi matris türünden, $Y = X\beta + \varepsilon$ yazılır. Burada, X, nx2 boyutlu girdi ya da gözlem değerlerini; Y, nx1 boyutlu yanıt vektörünü; $\hat{\beta}$, 2x1 boyutlu bilinmeyen katsayılar vektörünü; ε, nx1 boyutlu hata terimi ya da yanılgı vektörünü tanımlamaktadır.

En küçük kareler kestirim yöntemine göre normal denklemler,

$\hat{\beta}_0 n + \hat{\beta}_1 \sum x_i = \sum y_i$

$\hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$ şeklinde elde edilmişti. Elde edilen normal denklemleri, matris türünden aşağıdaki gibi yazabiliriz:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$ yazılır.

Burada,

$$X'X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}, \; X'Y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Yukarıdaki matris ise, kapalı bir biçimde,

$(X'X)\hat{\beta} = X'Y$ olacaktır. Bu eşitliği, soldan $(X'X)^{-1}$ ile çarparsak,

$\hat{\beta} = (X'X)^{-1} X'Y$ olarak bulunur. $\hat{\beta}$ matrisinden, $\hat{\beta}_0$ ve $\hat{\beta}_1$ katsayı kestirimleri elde edilir (İşyar ve Kip, 1981).

Örnek 5.12. 2001-2013 yılları arasındaki Şekerpancarı verimine etki eden gübre kullanım miktarları verilmiştir. Bu değerleri kullanarak basit doğrusal denkleminin parametrelerini matris yöntemi kullanarak hesaplayalım.

Hesaplama aşamaları şunlardır:

Çizelge 7. Türkiye'de Şekerpancarı Verimi ve Gübre Kullanım Miktarları, 2001-2013

Yıl	**Verim (10 Kg/da)**	**Gübre (10 Kg/da)**
2001	3.542	10
2002	4.444	11
2003	4.014	12
2004	4.290	10
2005	4.524	13
2006	4.464	14
2007	4.154	11
2008	4.829	12
2009	5.332	13
2010	5.459	12
2011	5.488	15
2012	5.325	14
2013	5.666	15

Kaynak: Türkiye İstatistik Kurumu, 2014.

a. Gözlem değerlerinden yararlanarak X ve Y matrisi oluşturulur.

b. $X'X\, ve\, X'Y$ matrisi hesaplanır.

c. $X'X$ matrisinin tersi bulunur.

d. Formül gereği $X'X$ matrisinin tersi ile $X'Y$ matrisi çarpılır.

e. Buradan elde sonuçlardan yararlanarak tahmin edilen regresyon denklemi yazılır.

$$X'X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}, \; X'Y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} = \begin{bmatrix} 13 & 16,20 \\ 16,20 & 20,54 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 4,48 & -3,53 \\ -3,53 & 2,83 \end{bmatrix} \text{ elde edilir.}$$

$$X'Y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} 6147 \\ 7762,20 \end{bmatrix}$$

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

$$\hat{\beta} = \begin{bmatrix} 4{,}48 & -3{,}53 \\ -3{,}53 & 2{,}83 \end{bmatrix} \begin{bmatrix} 6147 \\ 7762{,}20 \end{bmatrix} = \begin{bmatrix} 111{,}73 \\ 289{,}78 \end{bmatrix}$$

$Y_i = 111{,}73 + 289{,}78 X_i$ olur.

Şekerpancarı verim ve gübre kullanım miktarları için tahmin edilen model parametrelerini yorumlayalım. Regresyon modeli $\hat{Y}_i = 111{,}73 + 289{,}78 X_i$ olarak tahmin edilmiştir. Modelde $\hat{\beta}_0$=111,73 değeri sabit terimdir ve gübrenin sıfır olması durumunda yani bağımsız değişken X'in sıfır olması halinde şekerpancarı verimi 111,73 kg/da olur.

Regresyon modelinde $\hat{\beta}_1$=289,78'dir. Bu katsayı pozitif bir değere sahiptir. Gübre kullanım miktarı ve şekerpancarı verimi arasında pozitif yönlü ve doğrusal bir ilişkiden bahsedilir. Yani, gübre kullanım miktarları artarsa şeker pancarı verimi de artar. Diğer bir ifadeyle, gübre kullanım miktarlarında %1 bir artış meydana geldiğinde şeker pancarı veriminde %290 oranında bir artış meydana gelmektedir.

Normal denklemler, gerçek değerler ve matris yöntemi ile elde edilen sonuçlar tamamıyla aynı sonuçları vermektedir.