

Nohut (*Cicer arietinum* L.) Bitkisinde Verime Etki Eden Bazı Karakterlerin Alternatif Regresyon Yöntemleriyle Karşılaştırılması

Ufuk KARADAVUT¹ Aşır GENÇ² Abdurrahman TOZLUCA³ İsmail KINACI²
Şeref AKSOYAK¹ Çetin PALTA¹ Ahmet PEKGÖR²

Geliş Tarihi: 26.05.2004

Öz: Regresyon, bağımlı değişkenler ile bağımsız değişkenler arasındaki ilişkinin matematiksel ifadesidir. Bağımlı değişkenler, bağımsız değişkenlere regresyon fonksiyonu denilen bir fonksiyonla bağlantı içindedirler. Basit regresyon olarak tanımlanan regresyonda biri bağımlı diğeri ise bağımsız olmak üzere iki değişken arasındaki ilişkiler incelenmektedir. Çalışmamızda tane ağırlığına etki eden değişkenler ile tane ağırlığı arasındaki doğrusal ilişkiler incelenmiştir. Denemeler, Bahri Dağdaş Uluslararası Tarımsal Araştırma Enstitüsü deneme alanlarında Damla nohut çeşidi kullanılarak yürütülmüştür. Nohut bitkisinde tane ağırlığına etki eden bitki boyu, dal sayısı, ilk bakla yüksekliği, bakla sayısı, tane sayısı, hasat indeksi ve biyolojik verim gibi karakterler ölçülmüştür. Bu ölçümler $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$ şeklinde bir çoklu lineer regresyon modeli olarak ele alınmıştır.

Bu modelin parametrelerinin tahmin değerleri En Küçük Kareler Yöntemi'nin yanı sıra bu yöntemle alternatif olarak Ridge Regresyon ve M-Regresyon yöntemleri ile hesaplanmış ve sonuçlar karşılaştırılarak en uygun model önerilmiştir.

Anahtar Kelimeler: Basit regresyon, nohut, alternatif yöntemler

Comparing the Alternative Regression Methods on Some Yield Traits of Chickpea (*Cicer arietinum* L.)

Abstract: Regression express that the mathematical relationships between independent variables and dependent variables. Dependent variables related to independent variables that is named regression function. In simple regression, relationships among one dependent and another independent variables are investigated. In this study linear relationships between effective variables on seed weight and seed weight were investigated. The experiments were carried out in Bahri Dağdaş International Agricultural Research Institute's experimental areas on Damla Chickpea cultivar. In Chickpea plants the factors that effects seed yield plant high, number of branch, first bean height, pod number, seed number, harvest index and biological yields were measured. These measurements were inquired as multiple linear regression model ($Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$). Estimated values of this model, least square method and its alternatives Ridge regression and M-Regression methods were calculated and this results compared. This study of results were suggested the most appropriate model.

Key Words: Simple regression, chickpea, alternative methods

Giriş

Tarımsal araştırmalarda konu olan birçok bitkide morfolojik, biyolojik, genetik ve çevresel etkenler gibi pek çok faktörün verime katkısı son derece önemlidir. Populasyonlar da doğrudan verimi almak yerine verime en çok etki eden karakterleri almak daha yararlı olacaktır (Cinsöy ve Yaman 1998).

İncelemeye konu olan verim ve bitki boyu olarak ele alınan iki değişkenden birisi diğeri bir fonksiyonu olarak düşünülebilir. Bir bitkinin boyu arttıkça kur madde birikimi de zamana bağlı olarak artmaktadır. Yani bitki boyu ile kuru madde birikimi arasında pozitif bir ilişki söz konusudur. Burada kuru madde birikimi bitki boyunun bir fonksiyonudur. Kuru madde birikiminin bir sonucu olan verim bağımlı değişken, bitki boyu ise bağımsız değişken olarak ele alınabilir (Kesici ve Kocabaş 1998). Verim ile bitki, boyunun yanı sıra bitkide bakla sayısı, bakla ağırlığı ve bitkide tane ağırlığı gibi çok sayıda faktöre bağlı olarak

sebeup-sonuç ilişkisi ortaya konulabilir. Bu ilişki çoklu regresyon yöntemleri ile istatistiksel olarak irdelenebilir. Dolayısıyla verim, bağımlı değişken olup diğer faktörler bağımsız değişken olarak ele alınabilir. Çoklu regresyon yönteminde bağımlı değişkenin gerçek ölçüleri ile bağımsız değişkenlerden elde edilen kestirim ölçüleri arasındaki uzaklığı en küçük yapan regresyon katsayılarının tahmini En Küçük Kareler kestirimi (EKK) ile hesaplanmaktadır (İpek 2002, Akdeniz ve Öztürk 1996).

Tarımsal açıdan benzer ekolojilerden alınan verimlerin farklı bitkilere ait regresyon denklemlerinin diğer benzerlik gösteren ekolojilerden alınan verimlerin bağımsız olarak ele alınması, yani EKK yönteminin uygulanması acaba katsayı tahminlerinin sapmalı sonuçlar vermesine neden olabilir mi?" sorusuna aranan cevap doğal olarak farklı tahmin yöntemlerinin kullanılması ile elde edilebilecektir (Öğüt ve Üçdoğru 1999).

¹ Bahri Dağdaş Uluslararası Tarımsal Araştırma Enstitüsü, Karatay-Konya

² Selçuk Üniv. Fen Edebiyat Fakültesi İstatistik Bölümü, Kampüs-Konya

³ Selçuk Üniv. Ziraat Fak. Zootekni Bölümü, Kampüs-Konya

Bir regresyon denklemi ilerde açıklanacak klasik regresyon varsayımlarını sağlarsa yapılan iş istenen sonuçları verecektir (Akdeniz ve Öztürk 1996). Aksi takdirde tahmin değeri sapma gösterecektir. Parametre tahmin edicisinin yansız olmamasını gerektirir. Yani yanlış tahmin edici olduğunu gösterir. Bu durumda tahmin edicinin minimum varyanslılık özelliğini etkiler. Aynı zamanda da regresyon modelindeki değişkenlerin bağımsız olması özelliğini bozar. Diğer bir deyişle çoklu bağlantı probleminin varlığını gösterir. Katsayı tahminlerinin sapmalı sonuçlar vermesinden kasıt; tahmin edicinin yansız tahmin edici olmamasıdır.

Bu nedenle nohut verimine etki eden karakterlere ait verilere lineer regresyonda en küçük kareler yöntemine alternatif olan Parametrik Olmayan Regresyon, Bayesci Regresyon, Robust Regresyon ve benzeri yöntemler uygulanabilir. Bu çalışmada en küçük kareler yöntemine alternatif olan ve tarımsal araştırmalarda uygulaması sık görülmeyen Ridge Regresyon ve Robust Regresyon ele alınmıştır. Bu yöntemlere göre tahminlerin hassasiyetleri karşılaştırılarak tarımsal çalışmalara uygun tahmin yöntemi önerilmiştir.

Çoklu lineer regresyon modelinin tanıtımı ve bazı gösterimler: \underline{Y} bağımlı değişken, $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_p$ açıklayıcı (bağımsız) değişkenler, $\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_p)' \in \mathcal{R}^{p+1}$ parametre vektörü, $\varepsilon_i \in \mathcal{R}$ gözlenemeyen hata terimleri ve model fonksiyonu,

$$f(x_t; \underline{\beta}) = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_p x_{pt}, \quad t = 1, 2, \dots, n$$

olmak üzere lineer regresyon modeli,

$$y_t = f(x_t; \underline{\beta}) + \varepsilon_t, \quad t = 1, 2, \dots, n \quad (1)$$

şeklinde tanımlanır. Bu model,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

gösterimi altında vektör gösterimi ile

$$\underline{Y} = X \underline{\beta} + \underline{\varepsilon} \quad (2)$$

şeklinde ifade edilebilir. Burada $\underline{Y}_{n \times 1}$ bağımlı değişken vektörü, X

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix}_{n \times (p+1)}$$

şeklinde tasarım matrisi, $\underline{\beta}_{(p+1) \times 1}$ bilinmeyen parametre vektörü ve $\underline{\varepsilon}_{n \times 1}$ hata vektörüdür.

Regresyon modellerinin değişkenler arasındaki ilişkiyi açıklamak için yeterli olup olmadığı veya ilişkiyi iyi açıklayıp açıklaymadığı konusunda karar vermek için hata terimleri ile ilgili temel varsayımların da geçerli olması gerekmektedir. Varsayımların geçerli olmaması durumunda, tahmin edicilerde bulunmasını istediğimiz bazı özellikler sağlanmayabilir. Lineer regresyon modeli olarak verilen (1) eşitliği üzerindeki temel varsayımlar aşağıdaki gibi sıralanabilir (Akdeniz ve Öztürk 1996, Graybill 1961):

- $\varepsilon_i, i = 1, 2, \dots, n$ hata terimlerinin her biri normal dağılıma sahiptir,
- $\varepsilon_i, i = 1, 2, \dots, n$ hata terimlerinin beklenen değeri sıfırdır. Yani, $E(\varepsilon_i) = 0$ dir,
- $\varepsilon_i, i = 1, 2, \dots, n$ hata terimleri sabit varyanslıdır. Yani, $E[\varepsilon_i - E(\varepsilon_i)]^2 = \sigma^2$ (sabit) dir,
- $\varepsilon_i, i = 1, 2, \dots, n$ hata terimleri bağımsızdır. Yani,

$$E[(\varepsilon_i - E(\varepsilon_i))(\varepsilon_j - E(\varepsilon_j))] = E(\varepsilon_i \varepsilon_j) = 0, \quad i \neq j$$

dir. Hata terimleri arasında ilişki olması otokorelasyon olarak adlandırılır. İstenen, otokorelasyonun olmamasıdır,

- X tasarım matrisinin kolonları arasında lineer bağımlılık yoktur,
- Bağımsız değişkenler rasgele değişken (r.d.) değildirler. Bağımsız değişken \underline{X}_i 'lerin r.d. olması durumunda \underline{X}_i ile ε_i bağımsızdır,
- $(X'X)$ matrisi singüler olmayan bir matristir.
- Model doğru olarak kurulmuştur.

Materyal ve Yöntem

Damla Nohut çeşidinin kullanıldığı bu çalışma Bahri Dağdaş Uluslararası Tarımsal Araştırma Enstitüsünün deneme alanlarında 2002-2003 yetiştirme yıllarında yürütülmüştür. Ekim işlemleri sıra arası 30 cm, sıra üzeri ise 10 cm olacak şekilde düzenlenmiş ve ekimle birlikte banda 15 kg/da DAP (Diamonyum fosfat) taban gübresi verilerek nisan ayında yapılmıştır. Çalışmada bitkiler 50 m² 'lik üç eşit parselden rastgele örnekleme suretiyle bitkide tane ağırlığı, bitki boyu, bitkide dal sayısı, ilk bakla yüksekliği, bitkide bakla sayısı, bitkide tane sayısı, biyolojik verim ve hasat indeks değerleri ölçülmüştür.

Verimin temel belirleyicisi olan tane ağırlığına diğer değişkenlerin (bitki boyu, bitkide dal sayısı, ilk bakla yüksekliği gibi) ne şekilde etki ettiğini doğru bir şekilde

belirleyebilmek için, verim ile diğer değişkenlerin oluşturduğu lineer regresyon modelinin parametreleri EKK, Ridge ve Robust yöntemlerinden M yöntemi uygulanarak tahmin edilmiştir.

Eşitlik (1) ile belirtilen lineer regresyon modelinde çoklu doğrusal bağlantının olduğu durumlarda (çoklu doğrusal bağlantının belirlenmesinde $X'X$ matrisinin en büyük özdeğerinin en küçük özdeğerine oranı koşul sayısını vermektedir. Bu kriter verilerdeki küçük değişimlere karşı regresyonun duyarlılığını ölçmektedir. Hesaplanan koşul sayısı 100'den küçük ise çoklu doğrusal bağlantının önemsiz olduğu, 100-1000 arasında orta derecede, 1000'den büyük ise yüksek derecede olduğuna karar verilir (Montgomery ve Peck 1992) bağımsız değişkenleri değiştirmeden modelde tutarak çoklu doğrusal bağlantı problemini aşmaya yönelik yanlı tahmin sonuçları vermelerine karşılık parametre varyansları düşük Temel Bileşenler Regresyonu ve Ridge Regresyonu gibi parametre tahmin yöntemleri bulunmaktadır.

Temel Bileşenler Regresyonu (TBR) orijinal bağımsız değişkenlere ait doğrusal bileşenlerin bulunması ve daha az değişken kullanarak toplam değişkenin taşıdığı varyasyonun açıklanmasını sağlar. Amaç birbirine dik belli sayıda yeni bileşen elde ederek boyut azaltmaktır. TBR, EKK yöntemine göre örneklemdeki bilginin daha azını kullandığı gerekçesiyle çoklu doğrusal bağlantının etkilerinden kurtulmak için uygulanması doğru olmayabilir. Bu nedenle uygulamalarda daha çok Ridge Regresyon Yöntemi kullanılmaktadır (Sinan 2003).

Çoklu doğrusal bağlantı olduğunda yanlı tahmin yöntemlerinden Ridge Tahmin Yöntemi gerekli olan tüm değişkenlerin modele alınmasını sağlar. Bu yöntem çoklu doğrusal bağlantı olduğunda EKK tahminlerinden daha küçük varyanslı parametre tahminlerinin elde edilmesini ve modelden gereksiz değişkenlerin çıkarılmasını amaçlamaktadır (Marguarat ve Snee 1975).

En küçük kareler (EKK) yöntemi: Eşitlik (1) ile verilen doğrusal regresyon modeli göz önüne alınsın. $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ hata vektörü için,

$$E(\varepsilon_i) = 0, \quad i = 1, 2, \dots, n$$

$$Cov(\varepsilon_i, \varepsilon_j) = \begin{cases} 0 & , \quad i \neq j \\ \sigma^2 & , \quad i = j \end{cases}, \quad i, j = 1, 2, \dots, n$$

veya vektör gösterimi ile,

$$E(\underline{\varepsilon}) = \underline{0}$$

$$Cov(\underline{\varepsilon}) = \sigma^2 I$$

olduğu varsayılacaktır. $\underline{\beta}$ parametre vektörünün EKK tahmini,

$$Q(\underline{\beta}) = \sum_{t=1}^n [Y_t - f(\underline{x}_t, \underline{\beta})]^2$$

hata kareler toplamının $\underline{\beta}$ ya göre minimize edilmesinden elde edilir. Vektör gösterimi ile (3) eşitliği,

$$Q(\underline{\beta}) = [\underline{Y} - \underline{f}(\underline{\beta})]' [\underline{Y} - \underline{f}(\underline{\beta})] \quad (3)$$

yada,

$$Q(\underline{\beta}) = [\underline{Y} - \underline{X}\underline{\beta}]' [\underline{Y} - \underline{X}\underline{\beta}] \quad (4)$$

şeklinde yazılabilir. Burada,

$$\underline{f}(\underline{\beta}) = (f(x_1, \underline{\beta}), f(x_2, \underline{\beta}), \dots, f(x_n, \underline{\beta}))'$$

dir. Yukarıdaki (4) eşitliği minimum yapan $\hat{\underline{\beta}}$ değeri,

$$\left. \frac{\partial Q(\underline{\beta})}{\partial \underline{\beta}} \right|_{\underline{\beta}=\hat{\underline{\beta}}} = 0$$

olup buradan da,

$$\hat{\underline{\beta}} = (X'X)^{-1} X'Y \quad (5)$$

eşitliği elde edilir. Burada, elde edilen $\hat{\underline{\beta}}$ tahmin edicisi (2) lineer regresyon modelinin $\underline{\beta}$ parametresinin EKK tahmin edicisidir.

Normallik varsayımı altında,

$$\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$$

olmak üzere $\hat{\underline{\beta}}$ kestiricisi en çok olabilirlik tahmin edicisine eşit olur (Graybill, 1961)

Ridge regresyon: Ridge regresyon yöntemi genellikle modeldeki iki yada daha fazla bağımsız değişkenler arasında yüksek dereceden ilişki olması durumunda yani varsayım e-f-g'nin bozulması durumunda kullanılır. Bu yöntemde uygulanırken ilk adım olarak bağımsız değişkenler standartlaştırılır. Standartlaştırılmamış orijinal değişkenlerin bulunduğu model,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

şeklinde verilmiştir. Bu modeldeki bağımsız değişkenler standartlaştırılarak,

$$Y_i = \mu + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \dots + \gamma_p z_{ip} + \varepsilon_i \quad (6)$$

modeli elde edilir. Matris notasyonu ile

$$\underline{Y} = \underline{1}\mu + \underline{Z}\underline{\gamma} + \underline{\varepsilon}$$

şeklinde yazılabilir. Burada,

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \underline{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{np} \end{bmatrix}, \quad \underline{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix},$$

$$\underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

şekindedir. Eşitlik (6)'daki $\underline{\mu}$ ve $\underline{\gamma}$ parametrelerinin EKK tahmin edicileri,

$$\hat{\underline{\mu}}_{LS} = \bar{Y}$$

$$\hat{\underline{\gamma}}_{LS} = (Z'Z)^{-1} Z'Y \quad (7)$$

şekindedir. Z'Z matrisi singüler olması durumunda $\underline{\mu}$ ve $\underline{\gamma}$ parametrelerinin Ridge tahmin edicileri ise

$$\hat{\underline{\mu}}_R = \bar{Y}$$

$$\hat{\underline{\gamma}}_R = (Z'Z + kI)^{-1} Z'Y \quad (8)$$

olarak elde edilir. Burada $k = p \frac{\hat{\sigma}_{LS}^2}{\|\hat{\underline{\gamma}}\|^2}$ Ridge parametresi

, $\hat{\sigma}_{LS}^2$, σ^2 nin EKK tahmin edicisi ve I, $p \times p$ boyutlu birim matristir (Birkes ve Dodge 1993).

Robust Regresyon: Lineer en küçük kareler tahmin edicisi ekstrem değerlerden çabuk etkilenmektedir. Özellikle hatalar normal dağılmadığı durumda EKK kötü davranışlar sergileyebilmektedir. Robust regresyon analizi, temel varsayımların verilere başarılı bir şekilde uygulanmadığı zaman en küçük kareler regresyon modeline bir alternatif oluşturur (Huber 1981, Yaffee 2002). Bir araştırmacı istatistiksel regresyon modellerini tahmin ve varsayımlarını test ettiği zaman sık sık varsayımların sağlanmadığını görür. Bu durumda değişkenlerine bir şekilde dönüşümler uygular. Ancak, dönüşüm uç nokta ve etkin nokta gibi etkileri her zaman yok edemediği için öngörü yada parametre tahminleri yanlış olarak elde edilir (Chatterly ve Machler 1995). Bu şartlar altında, etkin noktalara dirençli olan Robust regresyon yöntemi kullanılabilir. En çok bilinen Robust tahmin edicileri M-Regresyon, Ağırlıklı en küçük kareler (WLS),

En küçük medyan kare (EKM), En küçük mutlak sapma (LAD) vb. dir. En yaygın olan Robust tahmin edicisi M regresyon tahmin edicisidir. Bu nedenle, bu çalışmada M-Regresyon yöntemi ele alınmıştır.

M-Regresyon: Lineer regresyon modelinin parametrelerinin $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ tahminleri EKK yönteminde $\sum \hat{e}_i^2$ hata kareler toplamının, en küçük mutlak sapma (LAD) yönteminde ise $\sum |\hat{e}_i|$ ifadesinin minimize edilmesiyle elde edilir. M yönteminde ise bu fikirler genelleştirilir ve $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ tahminleri $\sum \rho(\hat{e}_i)$ nin minimizasyonundan elde edilir. Burada $\rho(e)$, e 'nin bir fonksiyonudur.

Huber, parametrelerin M tahmini için e^2 ve $|e|$ 'nin birlikte yer aldığı bir $\rho(e)$ fonksiyonu kullanmıştır. LAD tahmininin EKK tahminine göre esas avantajı uç değerlere karşı fazla duyarlı olmamasıdır. Eğer uç noktalar yoksa EKK tahminleri daha uygun olmaktadır. Huber bu yüzden bu iki avantajı birleştirerek $\rho(e)$ fonksiyonunu; e sıfıra yakınsa e^2 ve e sıfırdan uzaksa $|e|$ (veya $|e|$ 'nin bir fonksiyonu) şeklinde yani,

$$\rho(e) = \begin{cases} e^2 & , \quad -k \leq e \leq k \\ 2k|e| - k^2 & , \quad e < -k, e > k \end{cases}$$

şeklinde tanımlamıştır. Huber buradaki k değerini $k = 1,5\hat{\sigma}$ olarak önermiştir. Buradaki $\hat{\sigma}$ değeri $\hat{\sigma} = 1,483m$ şeklinde ve m değeri $|\hat{e}_i|$ mutlak sapmalarının medyanı şeklinde tanımlanmaktadır.

$\beta_0, \beta_1, \dots, \beta_p$ parametrelerinin tahmin edilmesi için $\sum \rho(\hat{e}_i)$ ifadesinin minimize edilmesi gerekmektedir. Bunun için bir $\underline{\beta}^0$ başlangıç noktasına göre $e_i^0 = y_i - (\underline{\beta}^0)'x_i$ sapmaları ve bu sapmalardan faydalanarak $\hat{\sigma}^0$ değeri hesaplanır. Burada $y_i = (\underline{\beta}^0)'x_i + e_i^0$ yazabiliriz. Şimdi $y_i^* = (\underline{\beta}^0)'x_i + e_i^*$ modelini tanımlayalım. Burada e_i^* , e_i^0 'ın sınırlandırılmasıyla elde edilen düzeltilmiş sapmalardır ve hiçbir sapma $1,5\hat{\sigma}^0$ 'dan büyük değildir. Yani eğer $e_i^0, -1,5\hat{\sigma}^0$ ve $1,5\hat{\sigma}^0$ arasında ise $e_i^* = e_i^0$ (yani $y_i^* = y_i$), eğer $-1,5\hat{\sigma}^0$ 'dan küçükse

$e_i^* = -1,5\hat{\sigma}^0$ ve eğer $1,5\hat{\sigma}^0$ 'dan büyükse $e_i^* = 1,5\hat{\sigma}^0$ dir. Daha sonra $\underline{\beta}$ parametre vektörünün M tahminleri \underline{y} yerine \underline{y}^* kullanılarak EKK yöntemi ile elde edilir (Birkes ve Dodge 1993).

Bulgular ve Tartışma

EKK yöntemi: Tane ağırlığına etki eden değişkenlere ilişkin regresyon modeli

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

şeklindedir. Bu modeldeki parametrelerin tahmin değerlerine ilişkin sonuçlar aşağıdaki Tablo 1'de verildiği gibi elde edilmiştir.

Tablo 1'deki p değerlerine baktığımızda bu değerlerin hepsinin $\alpha=0.05$ 'den daha büyük olduğunu görülür. Bu modeldeki bütün parametrelerin sifıra eşitliği hipotezinin kabul edilmesi anlamına gelmektedir. Bu da istenmeyen bir durumdur. Bu durumu ortadan kaldırmak için ilk akla gelen işlem modelden p değeri en büyük değişkenleri çıkartmaktır. Buna göre modelden Dal sayısı (x_2) ve Bakla Sayısı (x_4) değişkenleri çıkartılmıştır. Bu durumda yeni model,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

şeklinde oluşacaktır. Oluşan bu yeni modele ilişkin tahmin sonuçları Tablo 2'de verildiği gibi elde edilmiştir.

Tablo 1. Tüm değişkenlerin bulunduğu model için tahmin değerleri

Değişkenler	Katsayılar	St. Hata	T	P
Sabit Terim	59,79	25,69	2,33	0,081
Boy	-0,3951	0,1599	-2,47	0,069
Dal	-0,0635	0,5310	-0,12	0,911
İlk Bakl	0,6326	0,2727	2,32	0,081
Bakla Sa	0,2113	0,7384	0,29	0,789
Tane say	-1,0012	0,4114	-2,43	0,072
Biyo ver	-0,9008	0,3756	-2,40	0,074
Hasat in	0,7503	0,3593	2,09	0,105
S = 0,8688	R ² = 81,6%			

Tablo 2. Değişkenler çıkartıldıktan sonraki tahmin değerleri

Değişkenler	Katsayılar	St.Hata	T	P
Sabit Terim	64,10	17,64	3,63	0,011
Boy	-0,4182	0,1160	-3,61	0,011
İlk Bakla y.	0,6051	0,2073	2,92	0,027
Tane sayısı	-0,9771	0,3332	-2,93	0,026
Biyo verim	-0,9534	0,2681	-3,56	0,012
Hasat ind.	0,8061	0,2416	3,34	0,016
S = 0,7210	R ² = 81,0%			

Tablo 2'deki p değerlerine bakıldığında hiçbir p değerinin $\alpha=0.05$ 'den daha büyük olmadığı görülür. Bu da modelde yer alan bütün bağımsız değişkenlerin bağımlı değişkene olan etkisinin istatistiksel olarak anlamlı olduğu anlamına gelmektedir.

Ridge Regresyon: Tane ağırlığına etki eden değişkenlere ilişkin regresyon modeli

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

şeklindedir. Bu modeldeki parametrelerin Ridge tahmin değerlerini elde edebilmek için önce bağımsız değişkenler standartlaştırılarak

$$Y_i = \beta_0 + \beta_1 z_{i1} + \beta_3 z_{i3} + \beta_5 z_{i5} + \beta_6 z_{i6} + \beta_7 z_{i7} + \varepsilon_i$$

modeli oluşturulur. Koşul sayısı 792 olarak bulunmuştur. Bu da orta derecede çoklu bağlantının varlığını göstermektedir. Daha sonra bu modeldeki parametrelerin tahmin değerleri (7) eşitliğinin kullanılmasıyla elde edilerek Ridge yönteminde kullanılacak olan k değeri $k=0,13839$ olarak hesaplanmıştır. Bu değer (8) eşitliğinde yerine koyularak modeldeki parametrelerin Ridge tahmin değerleri hesaplanmış ve Tablo 3'de verilmiştir.

Tablo 3'deki p değerlerine bakıldığında hiçbir p değerinin $\alpha=0.05$ 'den daha büyük olmadığı görülür. Bu da modelde yer alan bütün bağımsız değişkenlerin bağımlı değişkene olan etkisinin istatistiksel olarak önemli olduğu anlamına gelmektedir.

M-Regresyon: Tane ağırlığına etki eden değişkenlere ilişkin regresyon modeli

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

şeklindedir. Bu modeldeki parametreleri M-yöntemi ile tahmin edebilmek için $\underline{\beta}$ parametre vektörüne bir başlangıç değerinin verilmesi gerekmektedir. Bu başlangıç değeri olarak parametrelerin EKK kareler tahmin değerleri yani $\underline{\beta}^0 = \hat{\underline{\beta}}_{EKK}$ şeklinde alınmıştır. Bu tahmin değerleri kullanılarak k değeri $k=0,48939$ olarak bulunmuştur. Bu k değeri kullanılarak elde edilen yeni bağımlı değişke Y_i^* 'ın kullanıldığı

Tablo 3. Ridge yöntemine ilişkin tahmin sonuçları

Değişkenler	Katsayılar	St. Hata	T	P
Sabit Terim	13,2250	0,2081	63,55	0,000
Boy	-1,4234	0,4480	-3,18	0,001
İlk Bakl	0,7273	0,2778	2,62	0,008
Tane say	-1,0470	0,4281	-2,45	0,014
Biyo ver	-2,0717	0,6441	-3,22	0,001
Hasat in	1,6815	0,5707	2,95	0,003
S = 0,7739	R ² = 78,1%			

$$Y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

regresyon modeli EKK yöntemi ile tahmin edilerek parametrelerin M tahminleri elde edilmiştir ve Tablo 4'de verilmiştir.

Tablo 4'deki p değerlerine bakıldığında hiçbir p değerinin $\alpha=0.05$ 'den daha büyük olmadığı görülür. Bu da modelde yer alan bütün bağımsız değişkenlerin bağımlı değişkene olan etkisinin istatistiksel olarak önemli olduğu anlamına gelmektedir.

Nohut bitkisindeki tane ağırlığına etki eden bitki boyu, dal sayısı, ilk bakla yüksekliği, bakla sayısı, tane sayısı, biyolojik verim ve hasat indeksi gibi değişkenlerin bağımsız değişken, tane ağırlığının da bağımlı değişken olduğu regresyon modeli kurulmuş ve bu modelin parametre tahmin değerleri EKK, Ridge ve M yöntemleri ile elde edilmiştir. Tüm bağımsız değişkenlerin bulunduğu model EKK ile tahmin edildiğinde dal sayısı ve bakla sayısı değişkenlerinin bağımlı değişkene olan etkisinin istatistiksel olarak anlamsız olduğu görülmüş ve analizlerden çıkartılmıştır. Daha sonra kalan diğer bağımsız değişkenlerin bulunduğu regresyon modeli EKK, Ridge ve M yöntemleri ile tahmin edilmiştir. Bu yöntemlere göre tahmin edilen tane ağırlığı ve gerçek tane ağırlığı değerlerinin grafiği Şekil 1'de verildiği gibi elde edilmiştir.

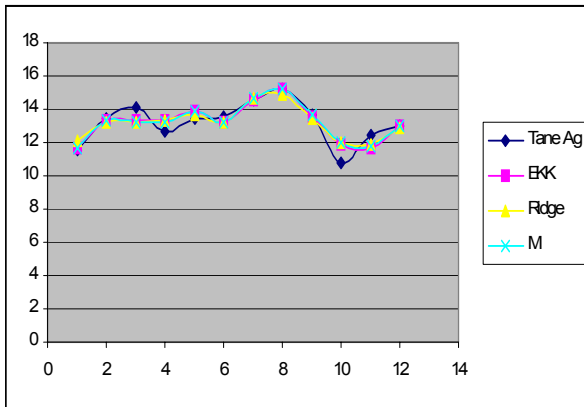
Lineer regresyon modeli için bu üç yöntemle elde edilen modelin standart hataları (S) ve R^2 değerleri Tablo 5'te verilmiştir.

Tablo 4. M yöntemine ilişkin tahmin sonuçları

Değişkenler	Katsayılar	St.Hata	T	P
Sabit Terim	63,65	11,03	5,77	0,001
Boy	-0,4079	0,0725	-5,63	0,001
İlk Bakl	0,5958	0,1297	4,59	0,004
Tane say	-0,9895	0,2084	-4,75	0,003
Biyo ver	-0,9352	0,1677	-5,58	0,001
Hasat in	0,7874	0,1511	5,21	0,002
S = 0,4508		$R^2 = 91,3\%$		

Şekil 1. Gözlenen ve tahmin edilen tane ağırlığının grafiği

Tablo 5. Her bir yöntemle elde edilen S ve R^2 değerleri



Yöntem	S	R^2
EKK	0,7210	%81,0
Ridge	0,7739	%78,1
M	0,4508	%91,3

Sonuç

Çalışmada EKK yöntemine alternatif bazı tahmin ediciler verilmeye çalışılmıştır. Burada M-regresyon ve Ridge regresyon yöntemlerine değinilmiştir. M-regresyon bir robust yöntem olup, bu yöntem etkin noktalara direnç gösterir. Tablo 5'de M-regresyon için $S=0.4508$ ve $R^2= \% 91.3$ bulunmuştur. Ridge yöntemi yanlış bir kestirim yöntemi olup etkin noktalardan kolay etkilenmektedir. Burada Ridge yönteminde $S=0.7739$ ve $R^2= \% 78.1$ bulunmuştur. Oysa EKK yöntemi yansız bir tahmin edicidir. Ancak, verilerdeki her türlü ekstrem noktalardan çok kolay etkilenir. Burada EKK 'de $S=0.7210$ ve $R^2= \% 81.0$ olarak hesaplanmıştır.

Bütün bu açıklamaların ışığında M-regresyon yönteminin uygun bir tahmin edici olarak tercih edilebileceği söylenebilir.

Kaynaklar

- Akdeniz, F. ve F. Öztürk. 1996. Lineer Modeller, A.Ü.F.F. Döner Sermaye İşletmesi Yayınları, No:38, Ankara.
- Birkes, D. and Y. Dodge. 1993. Alternative Methods of Regression, John Wiley&Sons, New York.
- Chatterjee, S. and M. Machler. 1995. Robust Regression: A Weighted Least Squares Approach, Communications in Statistics, Theory and Methods 26: 1381-1394.
- Cinsoy, A. S. ve M. Yaman. 1998. Nohutta bazı özellikler arası ilişkilerin path analiz ile değerlendirilmesi. Anadolu Ege Tarımsal Araştırma Enstitüsü Yayını. 8 (1) :116-126.
- Graybill, F. A. 1961. An Introduction to Linear Statistical Models, McGraw Hill, New York.
- Huber, P. 1981. Robust Statistics. John Wiley & Sons: New York.
- İpek, O. 2002. Yanlı Regresyon Yöntemlerinden Ridge Regresyon. Kara Harp Okulu Dergisi.
- Kesici, T. ve Z. Kocabaş. 1998. Biyoistatistik. Ankara Üniv. Eczacılık Fak. Yayınları No: 79. Ankara
- Marquard, D. W. and R. D. Sneer. 1975."Ridge Regression in Practice" the American Statistician, vol:29 (1).
- Montgomery, D. C. and E. A. Peck. 1992. Introduction to Linear Regression Analysis, 2nd ed., John Wiley and Sons, New York.
- Öğüt, C. ve Ş. Üçdoğruk. 1999. Farklı tahmin yöntemleri ile elde edilen regresyon katsayıları ve ekonomik ölçütlerin kıyaslanması. 4. Ulusal Ekonomtri ve İstatistik Kongresi. 14-16 Mayıs 1999. Antalya.
- Sinan, A. 2003. Lineer Regresyonda Ridge Tahmin Edicileri ve Bir Uygulama. Selçuk Üniv. Fen Bilimleri Enst. Yüksek Lisans Tezi (Basılmamış).
- Yaffee, R. A. 2002. Robust Regression Analysis: Some Popular Statistical Package Options. Statistitscs, Social Science and Mapping Group, Academic Computing Service.

İletişim adresi:

Ufuk KARADAVUT
Bahri Dağdaş Uluslar arası Tarımsal Araştırma Enstitüsü
Karatay/KONYA