Atatürk Üniv. Ziraat Fak. Derg.
33 (4), 417-423, 2002

Teşhis Edilen Anormal Gözlemlerin İterasyon ve Kovaryans Yöntemleri İle Yeniden Tahmini

İrfan ÖZTÜRK Zeki DOĞAN
Harran Üniversitesi Ziraat Fakültesi Zootekni Bölümü, Şanlıurfa

Necati YILDIZ
Atatürk Üniversitesi Ziraat Fakültesi Zootekni Bölümü, Erzurum

Geliş Tarihi :

ÖZET : Bazı testler sonucunda anormal gözlem (Outlier) olduğuna karar verilen gözlem değerlerinin veri setinden çıkarılması gerekmektedir. Bu durumda, gözlem değerleri eksileceği için planlı bir denemede olması gereken dengeli yapı bozulacak ve normal varyans analizi yapılamayacaktır. Dolayısıyla eksilen gözlem değerlerinin, uygun tahmin yöntemleri ile tahmin edilerek çözüme gidilmesi gerekmektedir. Bu çalışmada Anscombe-Tukey testi ile anormal gözlemler teşhis edilmiştir. Söz konusu anormal gözlemler denemeden çıkarılmış ve kayıp değer varsayılarak iterasyon ve kovaryans analiz teknikleri vasıtasıyla ilgili gözlemlerin yerine yeni tahmin değerleri elde edilmiştir. Her iki yöntem birbirine yakın sonuçlar vermiştir. Ancak kayıp değer sayısı iki ve ikiden az ise basit deneme planlarında iterasyon yöntemi ile çözüme gitmek daha kolay olmaktadır. Buna karşılık kayıp değer sayısı ikiden fazla ve deneme planın karışık olması durumunda bilgisayar yardımı ile kovaryans analiz tekniğinin kullanılması daha avantajlıdır.

Anahtar Kelimeler : Anormal gözlem, Kayıp Gözlem, Tekrarlı çözüm, covaryans.

Dedection of Outlers and Estimation of Them As a missing Observation Using Teration and Covariance methods

ABSTRACT : Some values which is deceded as an outlier by using some statistical tests should be omited from the given data set. The balance of ex perimental design might be changed since some observation will be discarded. Therefore the missing observations should be revaluated using appropriate statistical technique in order to replace within a new data set. In this study the outlier was determined using anscombe-Turkey test. Then the determined outlier were considered to be missing absorvation and new data set were estimated instead of outlier by using teration oud covariance technigues. Both missing observation estimation methods, gave similar results. However if the number of missing data is ouly one or two iteration method is recomanded in simple experimental designs. On the other hand if the missing observations is more than two the caviriance analyses method with computer is recomended.

Key words: Outlier missin observation, iteration covariance

GİRİŞ

Deneme planlarının dengeli yapısını kayıp gözlemler olduğu gibi anormal gözlemler de (Outlier) bozmaktadır. Çünkü testler sonucunda anormal olduğuna karar verilen gözlem değerlerinin deneme dışında bırakılması gerekmektedir. Deneme planı tesadüf parselleri olması halinde, muamelelere düşen fert sayılarının farklı olması analizi güçleştirmediği halde; tesadüf blokları, latin kare ve bunlara ilişkin kompleks deneme planlarında kayıp değerlerin varlığı, genel varyans analizini zorlaştırmaktadır. Bu durum dengeli yapıdaki bir deneme planının dengesini bozacağından ilgili gözlem değeri tekrar uygun bir tahmin yöntemi ile tahmin edilmesi gerekmektedir.

Bazı durumlarda iterasyon yöntemi ile kayıp değerleri ve neticesi olan sapmaları tahmin etmek için gerekli işlemler çok yorucu olabilmektedir. Bu nedenle denemelerinde kayıp veri bulunan bazı araştırmacılar, bu değerleri istatistiki yöntemlerle tahmin etmek yerine, istatistiki temele dayalı olmayan bazı basit ortalamalar almak suretiyle sözü edilen kayıp değerlerin yerini doldurdukları bilinmektedir. Araştırmacıların bu gibi hatalı durumlara sürüklenmemeleri için, bütün deneme planlarına tatbik edilebilir daha genel ve pratik bir yöntemin sunulması gerekmektedir. Bu nedenle bu araştırmada kayıp veya anormal gözlemin bulunması halinde iterasyon ve Kovaryans analiz yöntemi ile nasıl tahmin edilebileceği ve uygulamada hangi yöntemin daha pratik olduğu belirlenmeye çalışılacaktır.

Bu konuda Hartley (1956)'in yaptığı bir araştırmada, Yates (1933) tarafından ortaya atılan ve kayıp değerin tahmini için iterasyon metodunun bazı durumlarda kullanışlı olduğunu ancak kayıp değerin tahmininde genel bir metodun kullanılmasının daha pratik olacağını belirtmiş ve kayıp değerin tahminde kullanılabilecek bir non-iteratif metot önermiştir.

Coons (1957) , tesadüf blokları ve Latin kare gibi basit deneme planlarında tek bir kayıp değer için kolaylıkla tahmin formülleri türetilebileceğini, kayıp değerlerin tahmini yapıldıktan sonra da yine bunlara ait sapmalar hesaplanıp, daha sonra verilerin analizinin kolaylıkla yapılabileceğini belirtmiştir. Ancak bazı durumlarda kayıp değerleri ve bunların sonucu olan sapmaları tahmin etmek için gerekli işlemlerin çok yorucu olabileceğini, ayrıca genel bir formülün elde edilemediği durumlarda da işin daha da zorlaşacağını belirtmiştir. Mesela bölünerek tekrarlanmış denemelerde (bölünen-bölünmüş parseller, iç içe sınıflandırılmış faktöryel denemeler, gibi ....) hatanın tahminini her zaman aynı yolla yapılmadığını ve bir kayıp değerin tahmini, hata teriminin fonksiyonu olduğu için, her problem için tekrar yeni bir formül çıkarmak zorunda kalınacağını belirtmiştir. Bu nedenle kayıp gözlemlerin kolaylıkla tahmin edilebilmesinde bazı genel metotların kullanılması gerektiğini ve bu genel metotlardan birisinin kovaryans analiz metodu olduğunu ifade etmiştir.

Christensen ve Ronald (1987), herhangi bir çok faktörlü dengeli bir denemede bir gözlem değerinin kaybolmasıyla dengenin bozulup, analizin oldukça komplike olacağını belirtmiştir. Normal varyans analizini yapabilmek için eksik gözlemin tahmin edilmesi gerektiğini bunun için ise en güvenilir yolun kovaryans analizi olduğunu belirtmiştir. Burada matris notasyonuyla kovaryans modelindeki hata kareler toplamının silinen eksik modeldeki hata kareler toplamına eşit olduğunu göstermek istemiştir.

Kovaryans analizi bağımlı değişkenleri (kovaryetleri) içerdiğinden, modelin

$$\mathrm{Y} = \mathrm{X}\cdot\beta + \mathrm{Z}\cdot\gamma + \mathrm{e}$$

şeklinde olduğunu göstermiştir. Matris notasyonu ile bu modelin,

$$Y = \begin{bmatrix} X_{n-r} & 0 \\ X_r & I_r \end{bmatrix} \bullet \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + e$$

şeklinde yazılabileceğini belirterek, matrislerin yardımıyla kovaryans analizi sonucu elde edilen HKT ve HÇT 'nı kullanarak kayıp değerleri tahmin etmiştir.

Bek (1988), herhangi bir denemedeki kayıp değerin mevcudiyetinin, planlı bir denemede olması gereken dengeli yapıyı bozacağını, dolaysıyla normal varyans analizi yapılamayacağını ifade etmiştir. Bu durum karşısında, ya dengesiz bir planda en küçük kareler yöntemini uygulayarak sonuca varmak gerektiğini veya kayıp verilerin bulunduğu deneme ünitesine ait verinin tahmin edilerek normal varyans analizi yapılması gerektiğini belirtmiştir. Birinci yol uzun hesaplamaları gerektirdiğinden, uygulamada araştırıcıların ikinci yolu tercih ettiklerini ifade etmiştir. Ayrıca iterasyon yöntemi ile kayıp verilerin tahmini ve bu tahmin değerleri yerine konulduktan sonra yapılan varyans analizinde, varyasyon kaynaklarına ait kareler toplamlarındaki sapmalar (bias) üzerinde durmuştur.

MATERYAL ve METOT

Uygulamada kullanılacak veriler; farklı rasyonların çeşitli sütçü sığır ırklarında, süt veriminde meydana getirdikleri verim farkları esas alınarak tesadüf blokları deneme planına göre türetilmiştir.

Anormal gözlemlerin tespitinde Anscombe -Tukey metodu

Anormal gözlemlerin tespit edilmesinde birçok yöntem olmakla birlikte, bu araştırmada Anscombe-Tukey testi kullanılacaktır. Çünkü bu test normal bir gözlemin anormal kabul edilip ret edilmesine karşılık, belirli bir emniyet payı kullanması bakımından elverişli bir testtir.

Anscombe ve Tukey (1963) tarafından geliştirilen metot, $C\sqrt{HKO}$ kritik değerinden mutlak değerce büyük olan hata terimlerini anormal gözlem olarak kabul etmektedir. Burada, HKO: hata kareler ortalaması olup, C ise aşağıdaki şekilde hesaplanır.

1) Önce standardize edilmiş Z cetveli kullanılarak aşağıdaki eşitlikten Z_1 hesaplanır

$$P(Z_0 > Z_1) = \frac{HSD}{N}\cdot P \qquad (1)$$

Elde edilen Z_1 eşitlikte yerine konularak k değeri hesaplanır, $k = 1.40 + 0.85\, Z_1$

2) $$C = k\cdot\left(1-\frac{k^2-2}{4\cdot HSD}\right)\cdot\sqrt{\frac{HSD}{N}} \qquad (2)$$

olarak C katsayısı elde edilir. Burada; **HSD**: hata serbestlik derecesi, **N**: toplam gözlem sayısı, ***P*** : normal gözlemin anormal olarak reddini önlemek için alınan emniyet payı (P=0.025).

Eğer $\left|e_{ij}\right| > C\sqrt{HKO}$ ise ilgili gözlem anormal gözlem olarak kabul edilir ve bunun sebebi araştırılır. Yapılacak herhangi bir durum yok ise, gözlem değeri analiz dışında bırakılır. Bu durumda deneme planının dengeli yapısı bozulacağından ilgili gözlem değerlerinin yerine kayıp gözlem (missing observation) yöntemleri ile yeni gözlem değerlerinin tahmin edilmesi gerekmektedir. Kayıp gözlem değerleri veya analizden çıkarılması gereken anormal gözlemler sırasıyla iterasyon ve kovaryans analiz yöntemleri ile tahmin edilecektir.

Kayıp değerin iterasyon metodu ile tahmini

Tesadüf blokları deneme planında j.ci bloktaki i.ci muamelenin gözlem değeri kayıpsa, Y_{ij} * 'nin tahmin eşitliği aşağıdaki şekilde yazılabilir,

$$Y_{ij}{*} = \hat{\mu} + \hat{t}_i + \hat{b}_j \qquad (3)$$

Burada: $\hat{\mu}$, $\hat{t}_i$ ve $\hat{b}_j$ sırasıyla; genel ortalama, i.ci muamelenin ve j.ci blokun etkisi olup, eşitliğin sağındaki terimlerin en küçük kareler tahminleri aşağıda gösterildiği gibi elde edilir.

$$\hat{\mu} = \frac{G + Y_{ij}{*}}{p\cdot n} \qquad (4)$$

olup. ($\hat{t}_i$) i.ci muamelenin etki payı ve ($\hat{b}_j$) j.ci bloğun etki payı olup sırasıyla,

$$\hat{t}_i = \frac{T_i + Y_{ij}{*}}{n} - \frac{G + Y_{ij}{*}}{p\cdot n} \quad \text{ve}$$

$$\hat{b}_j=\left(\frac{B_j+Y_{ij}*}{p}-\frac{G+Y_{ij}*}{p\cdot n}\right) \tag{5}$$

şeklinde tahmin edilir. Bu tahmin değerleri (3) no'lu eşitlikte yerine yazılacak olursa,

$$Y_{ij}*=\frac{G+Y_{ij}*}{p\cdot n}+\left(\frac{T_i+Y_{ij}*}{n}-\frac{G+Y_{ij}*}{p\cdot n}\right)$$

$$+\left(\frac{B_j+Y_{ij}*}{p}-\frac{G+Y_{ij}*}{p\cdot n}\right) \tag{6}$$

olur. Burada ; G, T_i ve B_j sırasıyla; Genel toplam, i.ci muamele toplamı ve j.ci blok toplamı dır. Bu toplamların hepsi mevcut parseller üzerinden alınmaktadır.

(6) no'lu eşitliğin paydaları eşitlenip gerekli sadeleştirmeler yapılacak olursa,

$$Y_{ij}*=\frac{(n\cdot B_j+p\cdot T_i)\text{-}G}{(n\text{-}1)\cdot(p\text{-}1)} \tag{7}$$

olur. Ancak kayıp değer sayısının birden fazla olması halinde genel ortalama formülüyle,

$$Y_{ij}=\left(\frac{T_i}{n-1}+\frac{B_j}{p-1}\right)\Big/2 \tag{8}$$

kayıp değer sayısının bire indirgenmesi gerekmektedir. Kalan tek kayıp değer (7) no'lu formül ile tahmin edilir ve bulunan değer yerine yazılır. Bundan sonra genel ortalama formülü ile tahmin edilen değer silinerek, yeniden (7) no'lu formüle göre tahmin değeri hesaplanır. Tekrar birinci tahmin silinerek, yeni eksik toplamlara göre (7) nolu formüle ile tahminde bulunulur. Bu şekilde devam ettirilecek tekrarlı çözüm, aynı kayıp değer için arka arkaya yapılacak iki tahmin birbirine eşit oluncaya kadar devam edilir. Tahmin edilen kayıp değerler yerlerine yazıldıktan sonda, normal varyans analizi yapılır. Düz.HKO elde etmek için HSD' sinden kayıp değer sayısınca bir birim çıkartılır. Ayrıca kayıp gözlem içeren bir tablodan hesaplanan muamele kareler ortalaması, asıl değerinden daha yüksek tahmin edilmektedir. Bu nedenle MKO'dan,

$$S=\frac{\left(B_j-(p-1)\cdot Y_{ij}*\right)^2}{p(p-1)^2} \tag{9}$$

formülü ile elde edilecek sapma değeri çıkarılarak Düz.MKO elde edilir Bek (1987), [Yıdız (1994)].

Burada her bir deneme planına ait sapmayı hesaplayabilmek için ayrı bir formülün yeniden türetilmesi gerekmektedir.

Kovaryans analiz yöntemiyle iki yönlü tablolarda kayıp veri tahmini

Bu metodun geliştirilmesindeki amaç bir istatistiki denemede bir veya birden fazla kayıp gözlemin bulunması halinde kullanılabilecek genel bir metodun bütün detaylarıyla gösterilmesidir. Orijinali Bartlet, (1937) tarafından ortaya konulan ve daha sonra Anderson,(1946) tarafından tanımlanan metot mümkün olacak bütün avantajları sağlayabilecek kadar eksiksiz bir metot değildir[Coons(1957)].

Bununla birlikte John ve Prescott (1975), herhangi bir varyans analizinde kayıp değeri tahmin etmek için kovaryans analizinde uydurulmuş değişkenlerin (kovaryetlerin) kullanılması büyük miktarda hesaplamayı gerektirdiği için zaman zaman literatürlerden çıkarıldığını belirtmiştir. Ancak yaptığı araştırmada, aksine olarak kovaryans tekniğinin kullanılmasını, iteratif metotlara nazaran bazen bilgisayar analizlerinde tercih edilebileceğini göstermiştir.

Kovaryans analizinde tesadüf blokları deneme planına ait matematik model,

$$Y_{ij}=\mu+t_i+b_j+b\,(X_{ij}-\overline{X})+e_{ij} \tag{10}$$

şeklindedir. Burada sırasıyla t_i = muamele etkilerini, b_j = blok etkilerini, b =X_{ij} 'nin Y_{ij} 'ye göre regresyon katsayısını ve X_{ij} ise yardımcı değişkeni tanımlamaktadır.

Gözlem değerlerinin genel ortalamadan ayrılışları,

$$(Y_{ij}-\overline{Y}..)=(\overline{Y}_{i}.-\overline{Y}..)+(\overline{Y}._{j}-\overline{Y}..)$$

$$+(Y_{ij}-\overline{Y}_{i}.-Y._{j}+\overline{Y}..) \tag{11}$$

şeklinde yazılabilir. Bu eşitliğin her iki tarafının karesi alınarak i ve j için toplam yapılırsa, kovaryanslarda 0 alınarak,

$$\sum\sum(Y_{ij}-\overline{Y}..)^2=n\sum(\overline{Y}_i.-\overline{Y}..)^2+p\sum$$

$$(\overline{Y}._j-\overline{Y}..)^2+\sum\sum(Y_{ij}-\overline{Y}_i.-\overline{Y}._j+\overline{Y}..)^2 \tag{12}$$

eşitliği elde edilir. Eşitliğin her iki yanında yer alan terimler üzerinde gerekli sadeleştirmeler yapılarak, varyasyon kaynaklarına ait kareler toplamları aşağıdaki gibi bulunur.

$$GKT(Gyy)=\sum\sum Yij^2-\frac{Y..^2}{p\cdot n} \tag{13}$$

$$BKT(Byy)=\frac{\sum Y.j^2}{p}-\frac{(Y..)^2}{p\cdot n} \tag{14}$$

$$MKT(Tyy)=\frac{\sum Yi.^2}{n}-\frac{(Y..)^2}{p\cdot n} \tag{15}$$

$$HKT(Eyy)=Gyy-(Tyy+Byy \tag{16}$$

Benzer şekilde kovaryetler (X_{ij}) için 'de K.T' 'ları tahmin edilecek olursa üste verilen formüllerdeki Y değerlerinin yerine X yazılır.

Burada G, B, T, E sırasıyla, genel, blok, muamele, hata olup; $Y._j$ = j. blokun toplamı; $Y_i.$= i. muamelenin toplamı; $Y..$= Genel toplam; Y_{ij}= i. muamele ve j. bloka ait gözlem değeridir. Bu tanımlamalar Y değişkenleri için yapılmış olup, Y yerine X konulduğunda X değişkenleri içinde aynısı geçerlidir.

Kovaryans analizi için, karakterlerin (X ve Y) birlikte değişimleri (çarpımlar toplamları esasına dayanarak) notasyon olarak gösterilecektir. Tek karakterin değişim ölçüsü olan varyansa paralel olarak kovaryans;

$$Syx = \sum^{n}(X_i - \overline{X})(Y_i - \overline{Y})/(n-2) \qquad (17)$$

dır. Kareler toplamına paralel olarak (17) sayılı formülün payına çarpımlar toplamı denir ve

$$\sum dxdy = \sum XY - \frac{\sum X \cdot \sum Y}{n} \qquad (18)$$

kare çarpımının özel bir halidir. Üsteki formülde (n-2) çarpımlar toplamının serbestlik derecesidir. İki karaktere ait ortalama, parametre olarak kullanılmakta olduğundan SD 'si iki azaltılır. Buna göre varyasyon kaynaklarına ait çarpımlar toplamı aşağıdaki gibidir.

$$GÇT(Gxy) = \sum\sum X_{ij} \cdot Y_{ij} - \frac{X.. \cdot Y..}{p \cdot n} \qquad (19)$$

$$BÇT(Bxy) = \frac{\sum(X._j \cdot Y._j)}{p} - \frac{X.. \cdot Y..}{p \cdot n} \qquad (20)$$

$$MÇT(Txy) = \frac{\sum(X_i. \cdot Y_i.)}{n} - \frac{X.. \cdot Y..}{p \cdot n} \qquad (21)$$

$$HÇT(Exy) = Gxy - (Txy + Bxy) \qquad (22)$$

Kovaryans analizinde elde edilen bu terimlerin Doğru F testini yapabilmek için X' e göre düzeltilmiş Y değerlerinin bulunması gerekir. Bunun için gerekli formüller altta verilmiştir.

$$Düz.MKT = Tyy - \frac{(Txy)^2}{Txx} \qquad (23)$$

$$Düz.HKT = Eyy - \frac{(Exy)^2}{Exx} \qquad (24)$$

$$Düz.(HKT + MKT) = (Eyy + Tyy) - \frac{(Exy + Txy)^2}{(Exx + Txx)} \qquad (25)$$

Kayıp değer ise HÇT 'nın HKT 'na bölümünden elde edilen regresyon katsayısının (-1) ile çarpımından tahmin edilmektedir.

$$\hat{\beta} = (-1) \cdot \frac{Exy}{Exx} \qquad (26)$$

Kayıp gözlem sayısı birden fazla olması durumunda, çoklu kovaryans analizi uygulanmalıdır. Kayıp olan " Y " gözlemlerine 0 değeri verilerek, her kayıp gözlem için bir bağımlı değişken (kovaryet) düzenlenir. Bu kovaryetleri X_m ile gösterecek olursak, X_m 'lerin bütün hücrelerine 0, yalnız kayıp (Y) gözlemlerinin bulunduğu hücreye karşılık gelen X_m 'e bir (1) değerleri verilir. Eğer (f) kadar gözlem değeri kayıpsa, (Y) gözlem değerleri kayıp (f) gözlem sayısınca oluşturulacak X_m kovaryetleri ile çoklu kovaryans analizine tabi tutulur.

Önce kayıp değer sayısınca oluşturulan X kovaryetleriyle, Y gözlem değerlerinin düzeltilmemiş Ç.T ve K.T 'ları ayrı ayrı hesaplanır. Sonra birden fazla kayıp değer bulunduğundan β' ları tahmin etmek için çoklu regresyon denklemlerinden faydalanılır. Bunun için Y karakteri ile X_1 , X_2 ,...,X_k karakterleri arasındaki ilişki en basit olarak,

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \qquad (27)$$

şeklinde ifade edilir ki bu model, bağımsız değişkenlerin bağımlı değişkenle düz (interaksiyonsuz) ilişki halinde olduklarını gösterir. Buradaki β' ları tahmin edebilmek için en küçük kareler metodu uygulanır. Örneğin kayıp değer sayısı iki tane ise, model;

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (28)$$

şeklinde olur. Bilindiği üzere β katsayıları, $\sum$ (Y-$\hat{Y}$)^2 ifadesini minimum yapacak şekilde hesaplanır. Bu ifadedeki $\hat{Y}$ yerine (28) 'deki eşitlik konulduğunda,

$$KT = \sum (Y - \hat{Y})^2 = \sum [Y - (\beta_0 + \beta_1 X_1 + \beta_2 X_2)]^2 \qquad (29)$$

elde edilir. Bu ifadenin β' lar için $(\beta_0, \beta_1, \beta_2)$ ayrı ayrı kısmi türevleri alınıp 0' a eşitlenirse, β_0 için,

$$\sum Y = n \cdot \hat{\beta}_0 + \hat{\beta}_1 \sum X_1 + \hat{\beta}_2 \sum X_2 \qquad (30)$$

$$\beta_1 \text{ için, } \sum X_1 Y = \hat{\beta}_0 \sum X_1 + \hat{\beta}_1 \sum X_1^2 + \hat{\beta}_2 \sum X_1 X_2 \qquad (31)$$

$$\beta_2 \text{ için, } \sum X_2 Y = \hat{\beta}_0 \sum X_2 + \hat{\beta}_1 \sum X_2 X_1 + \hat{\beta}_2 \sum X_2^2 \qquad (32)$$

eşitlikleri elde edilir. Elde edilen bu üç eşitliğin yardımıyla β 'ların hesaplama tekniği aşağıda verilmiştir. Evvela (30) sayılı eşitlikten β_0 'ın eşiti bulunur.

$$\beta_0 = \frac{\sum Y}{n} - \beta_1 \frac{\sum X_1}{n} - \beta_2 \frac{\sum X_2}{n} \qquad (33)$$

(31) sayılı eşitlikteki β_0 yerine bulunan (33) 'deki eşiti konulduğunda,

$$\sum X_1\left(\frac{\sum Y}{n} - \hat{\beta}_1\frac{\sum X_1}{n} - \hat{\beta}_2\frac{\sum X_2}{n}\right) +$$

$$\hat{\beta}_1\sum X_1^{\;2} + \hat{\beta}_2\sum X_1X_2 = \sum X_1Y \qquad (34)$$

$$\hat{\beta}_1\left(\sum X_1^{\;2} - \frac{(\sum X_1)^2}{n}\right) + \hat{\beta}_2\left(\sum X_1X_2 - \frac{\sum X_1\sum X_2}{n}\right) =$$

$$\sum X_1Y - \frac{\sum X_1\sum Y}{n} \qquad (35)$$

elde edilir ve özet olarak model,

$$\beta_1\sum d^2X_1 + \beta_2\sum dX_1dX_2 = \sum dX_1dY \qquad (36)$$

şeklinde yazılabilir. Aynı işlem (32) sayılı eşitlik içinde yapıldığında,

$$\beta_1\sum dX_1dX_2 + \beta_2\sum d^2X_2 = \sum dX_2dY \qquad (37)$$

elde edilir. (36) ve (37) sayılı iki eşitlikten iki bilinmeyenli (β_1 ve β_2) denklem vasıtasıyla kayıp değerler kolayca hesaplanır [Öztürk (1998), Federer (1977)].

TARTIŞMA VE BULGULAR

Tablo 1.'de verilen gözlem değerlerinde kayıp gözlem bulunmaktadır. Ancak verilerin içinde kayıp gözlem sayılabilecek ve tekrar tahmin edilmesi gereken anormal gözlem değerlerinin olup olmadığını belirlemek için Anscombe-Tukey testi yapılacaktır. Bunun için önce orijinal verilere ait varyans analizini yapılmış ve Anscombe-Tukey testi için gerekli HSD= 25 ve HKO= 31428 olarak elde edilmiştir. Buradan da ilgili eşitliklerin yardımıyla kritik değer aşağıdaki gibi elde edilir.

$$P(Z_0 > Z_1) = \frac{25}{36}\cdot 0.025 = 0{,}01735 \text{ olup,}$$

buradan Z_1 =2,11 dir.

Ikinci formülde elde edilen Z_1 değeri yerine yazılarak,

k = 1.40+(0.85) (2.11) = 3.19

$$C = 3.19\cdot\left(1 - \frac{3.19^2 - 2}{4\cdot 25}\right)\cdot\sqrt{\frac{25}{36}} = 2.441$$

şeklinde hesaplanır. Bu değer $C\sqrt{HKO}$ kritik değerini elde etmek için yerine yazıldığında,

$$\text{Kritik Değer} = 2.441\cdot\sqrt{31428} = 432.7$$

olur. Bu kritik değer bütün gözlem değerleri için $e_{ij} = Y_{ij} - \hat{\mu} - t_i - b_j$ eşitliği ile hesaplanan tablo 1.'deki hata terimleri ile karşılaştırıldığında, (e_{25} = 733.1)> 432.7 olduğundan. Tablo 1. 'de verilen Y_{25}=3802 kg değeri anormal gözlem olarak kabul edilir. Bu nedenle Y_{25} değeri veri setinden çıkartılır. Eksilen Y_{25} gözlem değeri kayıp değer varsayılarak iterasyon veya kovaryans analiz metotları ile tahmin edilmesi gerekmektedir.

Kayıp verilerin tahmini için aynı veriler kovaryans analizinde de kullanılacağı için tablo 2. kovaryans analizine uygun olarak hazırlanmıştır. Yani her bir kayıp gözlem değerine karşılık bir X kovaryeti uydurulmuştur.

Tablo 1. Irklara ilişkin süt verimleri (Y_{ij}) ve hata terimleri (e_{ij})

Irk / Yem	Karacabey esmeri		Ayrshire		Jersey		Holstein		Guernsey		Brown Swiss	
	Y_{i1}	***e_{i1}***	**Y_{i2}**	***e_{i2}***	**Y_{i3}**	***e_{i3}***	**Y_{i4}**	***e_{i4}***	**Y_{i5}**	***e_{i5}***	**Y_{i6}**	***e_{i6}***
A	3757	*16.6*	3651	*44.4*	3590	*24.6*	3655	*23.7*	3580	*-113.9*	3705	*4.7*
B	2958	*-157.4*	2840	*-141.6*	2818	*-122.4*	2858	*-148.3*	**3802**	***733.1***	2912	*-163.3*
C	3288	*4.2*	3180	*30.1*	3165	*56.2*	3195	*20.4*	3090	*-147.3*	3280	*36.4*
D	3955	*67.2*	3785	*31.1*	3715	*2.2*	3800	*21.4*	3652	*189.3*	3915	*67.4*
E	3650	*32.6*	3495	*11.4*	3450	*7.6*	3555	*46.7*	3445	*-125.9*	3605	*27.7*
F	3335	*36.9*	3189	*24.7*	3155	*31.9*	3225	*36.1*	3095	*-156.6*	3285	*27.1*

Tablo2. Süt verimine ilişkin verilerin, kovaryans analiz metoduna uygun hazırlanması (kg)

Irk / Yem	Karacabey esmeri		Ayrshire		Jersey		Holstein		Guernsey		Brown Swiss		Σ	
	Y	**X**	**Y**	**X**	**Y**	**X**	**Y**	**X**	**Y**	**X**	**Y**	**X**	**Y**	**X**
A	3757	0	3651	0	3590	0	3655	0	3580	0	3705	0	**21938**	**0**
B	2958	0	2840	0	2818	0	2858	0	**0**	**1**	2912	0	**14386**	**1**
C	3288	0	3180	0	3165	0	3195	0	3090	0	3280	0	**19198**	**0**
D	3955	0	3785	0	3715	0	3800	0	3652	0	3915	0	**22822**	**0**
E	3650	0	3495	0	3450	0	3555	0	3445	0	3605	0	**21200**	**0**
F	3335	0	3189	0	3155	0	3225	0	3095	0	3285	0	**19284**	**0**
Σ	**20943**	**0**	**20140**	**0**	**19893**	**0**	**20288**	**0**	**16862**	**1**	**20702**	**0**	**118828**	**1**

Y_{25} gözlem değerini İterasyon metoduyla tahmin etmek için (7) no'lu eşitlikten yararlanarak kayıp değer,

$$Y_{ij}* = \frac{nB_j + pT_i - G}{(p-1)(n-1)} \quad =>$$

$$Y_{25} = \frac{6 \cdot 16862 + 6 \cdot 14386 - 118828}{(6-1)\cdot(6-1)} = 2746.4$$

olarak hesaplanmıştır.

İterasyon yöntemi ile tahmin ettiğimiz kayıp değer yerine konulmuş ve varyans analizi yapılarak sonuçlar tablo 3.'deki gibi elde edilmiştir. Düzeltilmiş serbestlik dereceleri hesaplanırken genel ve hata serbestlik derecesinden kayıp değer sayısınca bir birim çıkartılmıştır.

Muamele kareler ortalamasından sapmayı gidermek için (9) nolu eşitlikten hesaplanan sapma,

$$\text{Sapma} = \frac{(B_j - (p-1)Y_{ij}*)^2}{p(p-1)^2} =$$

$$\frac{(16862 - (6-1)\cdot 2746.4)^2}{6(6-1)^2} = 65312.7$$

değeri muamele kareler ortalamasından çıkarılıp düzeltilmiş MKO elde edilmiştir.

Düz.MKO = 737513 - 65312.7 => 672200.3

Düzeltilmiş kareler ortalamasına göre F testi uygulanacak olursa,

$$F_h = \frac{Düz.MKO}{Düz.HKO} = \frac{672200.3}{495} = 1358\ **$$

elde edilir. Bu değeri elde edeceğimiz F cetvel değeriyle karşılaştırdığımızda,

F_{Cetvel} = F (MSD, Düz.HSD) => 0.01 (5, 24) => 3.9

$F_C < F_h$ değerinden küçük olduğundan çeşitler arasındaki farkın çok önemli olduğu söylenebilir.

Tesadüf Blokları Deneme Planında Kayıp Değerin Kovaryans Yöntemiyle Tahmini

Kayıp değeri kovaryans metoduyla tahmin edebilmek için Tablo 2.'de *Y* gözlemlerine karşılık *X* kovaryetleri uydurulmuştur. *Y* sütunundaki kayıp değerin yerine (0) onun karşısındaki *X* (kovaryet) ' in yerine (1), diğer bütün yerlere de (0) değerleri verilmiştir.

$\hat{\beta}$ değerini diğer bir ifade ile kayıp değeri tahmin edebilmek için *X* ve *Y* 'lerin kareler toplamları ve çarpımlar toplamları yukarıda verilen ilgili eşitliklerden hesaplanmış ve sonuçları tablo 4' de verilmiştir. Kayıp değer ($\hat{\beta}$) ise (26) sayılı eşitlik vasıtasıyla

$$\hat{\beta} = (-v) \cdot \frac{Exy}{Exx} = (-1)\ \frac{-1907.22}{0.6946} \cong 2745.78$$

olarak tahmin edilmiştir. Elde edilen KT ve ÇT düzeltilmemiş olduğundan, muameleler arasındaki önem farkını test edebilmek için *X* 'e göre düzeltilmiş *Y* değerlerinin hesaplanması gerekir. Bunun için (23), (24), (25) sayılı eşitliklerden yararlanılarak ilgili düzeltmeler yapılmış ve sonuçları tablo 4.'de verilmiştir.

Düzeltilmiş kareler ortalamalarından yararlanarak Düz.F aşağıdaki gibi hesaplanır.

$$\text{Düz. F} = \frac{\text{Düz. MKO}}{\text{Düz. HKO}} = \frac{344887.9}{544.7} \cong 633.2$$

Elde edilen bu değeri F cetvel değeri ile karşılaştırılacak olursak,

F cetvel = $F_{0.01(\text{Düz.MSD, Düz.HSD})}$ => $F_{0.01(5, 24)}$ => 3.90

Düz. $F_h > F_C$ olduğundan, çeşitler arasındaki fark çok önemlidir denir.

Tablo 3. Kayıp Değer Tahmin Edildikten Sonra, Elde Edilen Varyans Analiz Tablosu

Var. Kaynakları	SD	K.T.	K.O.	Düz. SD	Düz. KO	F
Blok (Irk)	5	206046	41209	5		
Muamele (Yem)	5	3687566	737513	5	672200.3	1358**
Hata	25	11888	476	25-1=24	495	
Genel	35	3905500		35-1=34		

** ; P< 0.01 düzeyinde çok önemli

Tablo 4. Tesadüf bloklarına göre yürütülen süt verimlerine ait kovaryans analiz sonuçları

	X ve Y'nin kareler ve çarpımlar top.				X 'e göre Düzeltilmiş Y değerleri			
Var.Kay	**SD**	$\sum X^2$	$\sum X_1Y$	$\sum Y^2$	**SD**	**KT**	**KO**	**F**
Blok(Irk)	5	0.1388	-490.44	1852359.9				
Mua.(Yem)	5	0.1388	-903.11	7600575.6				
Hata	25	0.6946	-1907.2	5249882.7	24	13072.9	544.7	
Düz.Mua	5				5	1724439.6	344887.9	633.2**

Her iki yönteme göre tahmin edilen kayıp gözlem değerleri birbirine çok yakın sonuçlar vermiştir. Basit bir deneme planında bir veya iki tane kayıp değerin tahmininde her iki yöntemin uygulanabilirliğinin pratik olduğu söylenebilir. Ancak deneme planlarının karmaşık bir hal alması durumunda, genel bir yöntem olan kovaryans analizi ile kayıp değerin tahmin edilmesi daha avantajlıdır. Kayıp değer sayısının ikiden fazla olması durumunda her iki yöntemde de zorluklar ortaya çıkmaktadır. Bunlar; iterasyon yöntemi iki kayıp değere kadar verileri tahmin etmeyi öngörürken. Kovaryans analizinde bu sınırlama olmamakla birlikte kayıp değer sayısı artıkça, ikiden fazla bilinmeyenli denklemleri çözebilmek ağır matematiksel işlemleri beraberinde getirmektedir. Ancak bilgisayar desteği ile hazır matematik paket programlarından yararlanılarak, elde edilen denklemler çözülecek olursa; kovaryans analiz yöntemi ile çok sayıdaki kayıp değerleri tahmin edebilmek kaçınılmaz bir yöntemdir.

KAYNAKLAR

Anderson, R. L., 1946. Missing Plot Technigues , Biometrics, 2, 41 - 47.

Anscombe, F.J. and Tukey, J.W., 1963. The examiniation of residuals. Technometrics; 5, 141-160, 247-248.

Bartlett, M. S.,1937. Some examples of statistical methods of research in agriculture and applied botany. J. Roy. Statist. Soc.. B4: 137-170.

Bek, Y., 1987. Denemelerde eksik veriler tahmini ve bununla ilgili sapmanın (bias) hesaplanması. Çukurova Üni. Ziraat Fak. Dergisi, 2: 64-78.

Christensen, R., 1987. Plane Answers to Complex Questions. 1-MTS, Wiley Ss. 266, New York - Berlin.

Coons, I., 1957. The analysis of designed experiments with missing observations. Appl. Statist. 27: 38-46.

Federer, W. T., 1977. Experimental Design. Macmillan, Ss. 545, New York.

Hartley H.D.,1956. Programming Analysis of Variance for General Purpose Computers, Biometrics, 12, 110-112.

John, J. A. ve Prescott, P., 1975. Estimating in Experiments , Appl. Statist., 24, 190-192.

Öztürk, İ. Doğan Z.,Yıldız N., 1998. Deneme planlarında eksik verilerin kovaryans analiz metodu ile tahmini. Harran Üni. Ziraat Fak. Dergisi 2(1): 127-136.

Yates, F., 1933. The analysis of replicated experiments when the field results are incomplete. Emp. J. Exp. Agric. 1: 129-142.

Yıldız N., 1994. Araştırma ve Deneme Metotları, Atatürk Üni. Yayınları 697, Ziraat Fak. No: 305, Ders Kitabı No: 57, Erzurum.