Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi,
Tarım Bilimleri Dergisi (J. Agric. Sci.),
2001, 11(1):11-17

Geliş Tarihi: 15.09.2000

Hayvancılıkta Hata Varyanslarının Heterojen Olduğu Durumlarda Log-Lineer Modeller Kullanılarak Varyans-Kovaryans Unsurlarının Tahminlenmesi

Serhat ARSLAN [1] M. Kazım KARA[1]
Hayrettin OKUT[1] Levent TÜRKMUT[2]

Özet: Hata varyanslarının heterojen olduğu durumlar için karışık doğrusal modelde kullanılan bir metot tanıtılmıştır. Metot hata varyansları için Log-lineer Modeller esasına uygun olarak Marjinal En Yüksek Olabilirlik Fonksiyonunu kullanarak işlem yapmaktadır. Hata varyanslarının heterojenliğinin testinde Bartlett testi kullanılmıştır. Hesaplama tekniği sahadan alınmış bir örnek üzerinde uygulanarak, elde edilen sonuçlar tartışılmıştır.

Anahtar kelimeler: Varyansların heterojenliği, log-lineer model, marjinal maksimum olabilirlik

Using Log-Linear Models for Estimating Variance-Covariance Components in the Case of Heterogeneity of Residual Variance in Animal Study

Abstract: A statistical method is presented for heterogeneity of residual variance in mixed linear models. The method is based on Log-Linear Model for the residual variances from which parameters can be estimated and hypotheses tested using the Marginal Maximum Likelihood Function. It can be tested for heterogeneity variance that using a Bartlett's test. Computing procedures are presented and illustration is applying with experimental or field data. These analysis results are discussed under the methods.

Key words: Heterogeneity of variances, log-linear model, marginal maximum likelihood

Giriş

Biyolojik çalışmalarda, araştırmaya konu olan faktörlere ait hata varyanslarının homojenliği, parametrik analiz yöntemlerinin esas aldığı temel varsayımlardan en önemlisidir. Özellikle hayvancılıkta yürütülen çalışmalarda varyansların homojen olmadığı durumlarla sıklıkla karşılaşılmaktadır (Mc Cullagh ve Nelder, 1983; Hill,1984; Henderson, 1984; Gianola, 1986; Foulley ve ark.,1989). Wilkeman ve Schaeffer (1988), hatalar arasındaki heterojenliğin önemsiz derecede küçük olduğu durumlarda kullanılan metodun bunu bertaraf edebileceğini bildirmişlerdir. Buna örnek olarak De Veer ve Van Vleck (1987); ve Winkelman ve Schaeffer (1988), bazı verimlere ait yıl etkisine ait kovaryansları REML (Restricted Maximum Likelihood) algoritması kullanarak tahminlemişlerdir. Bu çalışmada, önemsiz sayılabilecek heterojenliğin metot tarafından giderildiği bildirilmiştir. Gianola ve ark. (1986), yaptıkları benzer bir çalışmada aynı bildirişte bulunmaktadırlar. Ancak bu tip çalışmalarda asıl konu faktör seviyeleri nedeniyle ortaya çıkan alt gruplar arasında hesaplanan varyansların heterojenliğidir. Bu durum, klasik istatistik metotlarla parametre tahminlerinin arzulanan şekilde elde edilmesini güçleştirmektedir. Zira, faktörler arası farkın genetik nedenlerden ya da çevresel varyanslardan doğan sapmalar dışında özdeş olduğu varsayımı geçerli olmamaktadır (Henderson, 1984; Garrick ve Van Vleck, 1987; Foulley ve ark., 1987).

Bu çalışmada varyansların heterojen olduğu durumlarda kullanılabilecek doğrusal karışık bir modelin teorik esasları verilmiştir. Önerilen modelin bir veri setine uygulanışı gösterilerek, elde edilen sonuçlar tartışılmıştır. Burada verilen model, değişik varyasyon kaynaklarına göre istenilen şekilde değiştirilebilmektedir.

Materyal ve Yöntem

Çalışmada, Ceylanpınar Devlet Üretme Çiftliği'nde yetiştirilmekte olan Siyah-Alaca ırkı süt sığırlarına ait 1995-1996 yılları arasında tutulmuş 2 yıllık kayıtlar; yıl, baba, doğum mevsimi etkilerini içerecek şekilde düzenlenmiş ve 305 günlük laktasyon süt verimlerine ait toplam 1208 adet gözlemden oluşan bir veri seti kullanılmıştır.

[1]Yüzüncü Yıl Üniversitesi Ziraat Fakültesi Zootekni Bölümü, 65080 - VAN
[2]Ege Üniversitesi Ziraat Fakültesi Zootekni Bölümü, Bornova, İZMİR

Model

Tek değişkenli karışık bir modelin matris yazılımı klasik olarak;

$Y=X\beta + Zu + e$ [1]

şeklindedir. Burada Y, β, u, e sırasıyla gözlem değerleri, sabit (yıl ve doğum mevsimi) ve şansa bağlı etkilere (baba) ve bireysel olarak gözlenen deneysel hata değerlerine ait vektör olmaktadır. X ve Z ise sabit ve şansa bağlı etkiler için oluşturulan desen matrisi olmaktadır. Bu model için iki varsayım söz konusudur;

$Y \mid \beta, u, \Sigma_e \sim \mathbf{N}(X\beta+Zu, \Sigma_e)$ [2]

$u \mid \Sigma u \sim \mathbf{N}(0, \Sigma_u)$ [3]

Burada, $\sum_e = \sigma_e^2$ ve $\sum_u = \sigma_u^2$ olmaktadır. e ve u'nun bağımsız olduğu, başka bir söyleyişle Cov(e,u)=0 olarak kabul edilmektedir.

Faktörlerin tüm olası seviyelerini i gibi bir indise atarsak (i=1,2, ...,I); örneğin, BabaxYıl kombinasyonları herhangi bir baba için bulunduğu yıldaki döl verim kayıtlarını içerecektir. Bu durumu n_ix1 boyutlu bir vektörde Y_i. Olarak özetleyebiliriz. Burada 2 nolu eşitlikteki varsayımı göz önünde bulundurarak, u'lar bilindiğinde i durumlarına ait gözlemlerin dağılışları:

$Y_i \mid \beta, u, \sigma_e^2 \sim N(X_i\beta + Z_{iu}, \sigma_e^2); i = 1,2,....I$ [4]

olmaktadır. Burada X_i ve Z_i sırasıyla, daha önceden verilen şekilde ve $n_i$xp ve $n_i$xq boyutlu β ve u seviyeleri için oluşturulmuş katsayı matrisleridir. Kısaca;

$X_i\beta + Z_iu = T_i\theta$ yazılabilir. Bu durum için u'nun bilindiği tüm dağılışlar bağımsızdır. Böylece;

$Y' = (Y'_1, Y'_2, \; ...., Y'_i \; , \; ...., Y'_I,).$ [5a]

$X' = (X'_1, X'_2, \; ...., X'_i \; , \; ...., X'_I,).$ [5b]

$Z' = (Z'_1, Z'_2, \; ...., Z'_i \; , \; ...., Z'_I,)$ [5c]

şeklinde oluşturulur ve $\sum_e = \bigoplus_{i=1}^{I} R_i\sigma_e^2$ [6]

yazılabilir. Burada R_i; n_ixn_i boyutlu pozitif tanımlı bir matristir ve genellikle birim matris olmaktadır. σ_e^2 şansa bağlı varyans vektörünü içeren varyans-kovaryans matrisi Σ_e olarak tanımlanmaktadır.

Hata varyansları esas alınarak, I farklı durum için (genellikle büyük veri setlerinde) heterojenliği tanımlayacak en az sayıda parametre içeren bir model kullanmak amaçlanmaktadır. Böyle bir model Box ve Meyer (1986), Davidan ve Carroll (1987) ve Nair ve Pregidon (1988) tarafından hata varyansları:

$\log \sigma_e^2 = W_i'\gamma$; i= 1,2,......I [7]

şeklinde tanımlanmıştır: Burada γ, Kx1 boyutlu, hata varyanslarını etkileyen (K ≤ I) dağılış parametreleri vektörü ve W_i' , 1xK boyutlu söz konusu parametrelerin bulunma olasılıklarına ait W_i vektörünün transpozu olmaktadır. 7 nolu eşitlikteki model Mc Cullah ve Nelder (1983) tarafından logaritmik link fonksiyonu kullanılarak yazılmış ve dağılış parametreleri ile genişletilmiş olan doğrusal karışık bir modeldir. Varyanslar tanımlanırken daima, pozitif tanımlı varyansların üssel fonksiyonları kullanılmıştır. Bu durum γ'nın içerdiği parametrelerin işaretinin olumsuz (negatif) yönde etkilenmesini engellemektedir.

$\sigma_e^2 = \left\{\sigma_{e_i}^2\right\}$ olduğuna göre, $\log \sigma_e^2 = \log \left\{\sigma_{e_i}^2\right\}$

yazılabilir. $W'_{1xK} = (W_1, W_2, ..., W_I)'$ şeklindeki tanımlamayla 7 nolu eşitlik

$\log \sigma_e^2 = W\gamma$ [8]

olarak yazılabilir. Burada 1≤rank(W)≤I ve tam sütün rankı olmaktadır. Varyanslar homojense W matrisinin rankı 1'e eşit olmaktadır. Varyansların heterojen olduğu durumlarda rank I'ya eşit olmaktadır. Böylece i = 1,2..........I tanımlamasında tüm i'ler için;

$\log(\sigma_{e_i}^2) = \mu$

veya

$\log(\sigma_{e_i}^2) = \mu_i \qquad \mu_i \neq \mu_{i'} \quad (i \neq i'$ için)

şeklinde olacaktır. Yazılan bu model **Heterojen Model** olarak adlandırılmaktadır ve faktörler arası hata varyanslarının heterojenliğini tanımlamada dağılış için ihtiyaç duyulan olası tüm parametreleri içermektedir.

$\theta = (\beta', u')'$ ile yer ölçülerini içeren parametre vektörünü ve u'lara ait dağılış unsurlarını, $\tau = (\sigma_u^2, \gamma)'$ olarak tanımlarsak; Gianola ve ark (1986) ve Foulley ve ark. (1990) tarafından θ ve τ için gösterilen nokta tahminlerini tahminlerini dikkate alarak:

$\tau^* = \text{Arg}\ \tau\ \text{Max} \log P(\tau \mid Y)$ [9a]

$\theta^* = \text{Arg}\ \theta\ \text{Max} \log P(\theta \mid Y, \tau=\tau^*)$ [9b]

yazabiliriz. Arg Max ortalamalarına ait değerler fonksiyonu maksimum yapacak olan, P(τ| Y)'nin maksimize edilmesiyle bulunur. 9a'daki eşitlikte, fonksiyonun maksimize edilmesiyle, τ'nin marjinal sonraki (posterior) yoğunluk fonksiyonu elde edilir. Bu, Bayes Analizi'nde, τ ve β'lar için önceki (prior) olasılıklar bulunduğunda, τ'ya ait marjinal olasılık fonksiyonun bir parçası olarak gözlenmektedir. Benzer şekilde 9b'de θ için çözüm yapılırsa, θ'ya ait eklemeli yoğunluk fonksiyonu elde edilmektedir ($\tau=\tau^*$). Buradan, Bayes Analizi'nde τ ve β için önceki yoğunlukları τ^* istatistiği için τ'nın Marjinal Maksimum Olasılık (MML) tahmini olmaktadır. Aynı şekilde θ^*, θ'nın Maksimum Posterior (MAP) Tahmini olmaktadır. MAP(θ) ve MML(τ) için hesaplama yöntemleri bir çok literatürde tanımlanmıştır. Louis (1982) buna ek olarak bir çalışmada P(τ| Y)'nin bir ve ikinci mertebeden türevlerinden faydalanarak bir algoritma geliştirmişlerdir. Logaritmik Marjinal Olasılık Fonksiyonu;

$$L(\sigma_u, \gamma; Y) = \log(Y \mid \sigma_u, \gamma) \qquad [10]$$

olarak yazılabilir. Bu amaç için 10'daki eşitliğin kullanılması bir avantajdır. Çünkü ilişkiler göz önünde tutulduğunda hesaplama kolaydır. 10'daki eşitliğin γ ye göre kısmi türevi alırsak :

$$\frac{\partial}{\partial_\gamma} L(\sigma_u, \gamma_i, y = E_c\left[\frac{\partial}{\partial_\gamma}\log \quad p(y \mid \beta, u, \gamma)\right] \qquad [11a]$$

$$\frac{\partial^2}{\partial_\gamma \partial'_\gamma} L(\sigma_u, \gamma, y) = E_c\left[\frac{\partial^2}{\partial_\gamma \partial'_\gamma}\log \quad p(y \mid \beta, u, \gamma)\right]$$

$$+Var_c\left[\frac{\partial}{\partial_\gamma}\log \quad p(y \mid \beta, u, \gamma)\right] \qquad [11b]$$

E_c(.) ve Var_c(.) sırasıyla, (θ| y, σ_u, γ)'nin sonraki dağılışları verildiğinde, varyans-kovaryanslara ait beklenen değerlerdir. 11a ve 11b eşitlikleri marjinal olasılık dağılışına bağlı değildir. Bu durum 2 nolu eşitlikte verilen normal dağılıştaki yoğunluk fonksiyonlarının logaritmasıyla çalışma olanağı vermesi açısından avantaj sağlamaktadır. σ_u bilindiği durumlarda sadece, γ'nin tahminlenmesi için kullanılabilir. 11a ve 11b'nin kullanımı, 'σ_u'nun bilindiği' varsayımının geçerli olduğu durumlar için geçerli olmaktadır. Bu varsayım geçerli olduğunda hesaplamada hiçbir sıkıntı olmamaktadır. Newton-Raphson algoritmasından yararlanarak 9a'da tanımlanan eşitliğin maksimizasyonuyla 9b'den üç ayrı algoritma yazabiliriz.

Veri analizlerinin en önemli aşaması, hetorejenliğin anlamlı kaynaklarına bölünerek test edilmesini içeren hipotez testleri olmaktadır. Örneğin, eklemeli veya interaksiyonu içeren faktörlerin varyanstaki payının araştırıldığı 8 nolu eşitlikte verilen modeldeki , γ doğrusal parametreleri bu amaçla kullanılabilmektedir. Bu durumda, H_o hipotezi: H_o: $\mu'\gamma$ = m [12]

olarak yazılmaktadır. μ' γ, r kadar bağımsız doğrusal bir şekilde tanımlanmış tahminlenebilir γ fonksiyonu ve m, rx1 boyutlu bir vektördür. 'Varyanslar heterojendir' şeklinde kurulan H_o hipotezini, tüm i'ler için, H_o: log $\sigma^2_{e_i}$ =μ_i = μ olarak yazılabilir. Bu hipotezin test edilmesinde oran testleri kullanılabilmektedir. Bununla beraber, Marjinal fonksiyonların kullanımı tüm olasılık fonksiyonunun kullanımından daha avantajlı olduğu için testin marjinal fonksiyonun kullanılarak yapılması önerilmektedir. Zira, marjinal fonksiyonlar, sabit etkileri içeren β'nın tahmininde yapılacak olan hataları göz önünde bulundurmaktadır. τ'nin Marjinal Yoğunluk Fonksiyonuna göre tanımlanmış olmalıdır. Bu yoğunluk fonksiyonu şu şekilde yazılabilir:

$$P(y|\delta_u,\gamma) = (2\pi)^{-(N-P)/2} \ |\Sigma_u|^{-1/2} \ |\oplus R_i \exp(W_i^1 \ \gamma)|^{-1/2}$$

x

$$\sum_i (T_i' R_i^{-1}{}_i - 1 T_i \exp(-W_i'\gamma)) + \sum^{-}|^{-1/2}$$

$$\sum_i (T_i' R_i^{-1}{}_i - 1 T_i \exp(-W_i'\gamma)) + \sum^{-}|^{-1/2}$$

$$\sum_i (T_i' R_i^{-1}{}_i - 1 T_i \exp(-W_i'\gamma)) + \sum^{-}|^{-1/2} \qquad [13]$$

Burada, $\hat{\theta}$ ve T_i önceden tanımlandığı gibi olmaktadır ve $\sum^{-}$ ise :

$\begin{vmatrix} 0 & 0 \\ 0 & \sum_u^{-1} \end{vmatrix}$ şeklinde bir hesaplamayla matrisin genelleştirilmiş tersi olmaktadır. H_o hipotezinin H_1 karşıt hipotezine göre testi;

$$\lambda = 2 \ [\text{Max } \Gamma_1 L(\sigma_u, \gamma; Y) - \text{Max}\Gamma_0 L(\sigma_u, \gamma; Y)] \qquad [14]$$

ve H_0 altında r serbestlik dereceli Khi-Kare Dağılışı göstermektedir. α önem seviyesinde, λ değerinin, λ_c kritik değerini geçtiği durumlarda H_0 hipotezi red edilmektedir. Bu Pr($\chi^2 \geq \lambda_c$) = α olduğu durumlarda geçerli olmaktadır. 14 nolu eşitlikte Γ_0 ve Γ_1 , H_0 ve H_1 altında parametreleri ifade etmektedir. Genişlikleri K-r ve K olmaktadır. Max Γ L(σ_u , γ;Y) aynı hipotez altında maksimize edilmiş olasılık değerleridir. λ'nın gözlenen değeri, Genelleştirilmiş Doğrusal Model'de genellikle **sapma** (deviance) olarak adlandırılmaktadır.

Bulgular ve Tartışma

Çalışmada kullanılan veri seti yöntem kısmında belirtilen algoritmaya göre SAS (1998) istatistik paket programının PROC GENMOD prosedürü kullanılarak analiz ve tahminleme işlemleri tamamlanmıştır. En iyi model seçimi için sapmalardan faydalanılmıştır. Buna göre elde edilen sonuçlar Çizelge 2'de özetlendiği gibidir. Yıllar (Y)ve Doğum Mevsimi (M) sabit, babalar (B) ise şansa bağlı olarak kabul edilmiştir. Bu şekilde çapraz olarak sınıflandırılmış olan doğrusal modelde interaksiyon etkisinin olmadığı varsayılmıştır. Ayrıca, varyansların normal dağılıştan geldiği varsayılmıştır. Her bir muamele kombinasyonu $Y_i x M_j x B_k$ gibi olacağından bu kombinasyona ait hata varyansı $\sigma^2_{e_{ijk}}$ olacaktır. Bu şekilde tüm kombinasyonlar için, hata varyansı hesaplanmıştır. Seçilen bir faktör kombinasyonunun varyansı her bir faktörün varyansları çarpımı olmaktadır. Genel olarak, tanımlanan temel varyanslar kombinasyon grupları için :

$(T_1 x H_1 x S_1)$ σ^2_{323} = Tem Var x (Var T_i/Var T_1)x (VarH_j /H_1)x (Var S_k/VarS_1) olarak Logaritmik Link Fonksiyonu:

$\log \sigma^2_{ijk}$ = μ $+\alpha_i$ $+\beta_j$ + σ_k tanımlanmasıyla bulunur. Burada μ; ortalamayı, α_i; yıl β_i ; doğum mevsimi ve σ_k ise şansa bağlı baba etkisine karşılık gelmektedir. Böylece hata varyansındaki heterojenlik doğum mevsimi, yıl ve baba etkilerinin çarpımıyla bulunur. 8 nolu eşitlikteki γ parametre vektörü:

$$\gamma' = (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_1, \sigma_2)$$

elemanlarından oluşmaktadır. Burada, $\alpha_1 = \beta_1 = \sigma_1 = 0$ kısıtlamasına gidildiğinde, γ' yeniden düzenlenirse:

$$\gamma' = (\mu, \alpha_2 - \alpha_1, \beta_2 - \beta_1, \beta_3 - \beta_1, \beta_4 - \beta_1, \sigma_2 - \sigma_1)$$

şeklinde bir parametre vektörü yazmak mümkün olmaktadır.

Toplam 1208 gözlemi içeren veri seti çizelge 1'de özetlenmiştir. Burada gözlemler ve gözlem sayıları çapraz sınıflama grupları için eklemeli olarak yazılmıştır. Hata varyansları ve tanıtıcı istatistikler her bir grup için verilmiştir. Bu örneğin gerçekten varyans heterojenitesini içermesini sağlamak bakımından veriler buna göre seçilmiştir. Örneğin 3. kombinasyonla, 6. kombinasyon arasında hata varyansları bakımından 26 katlık bir fark mevcuttur. Ancak burada heterojenliğe neyin sebep olduğu tam olarak belirlenememiştir.

Algoritmanın hata sınırları

Y +B + DM etkilerini içeren modelde hem yer ölçülerine ait parametrelerin hem de hata varyanslarının hesaplanmasında her bir tahminlemede bir önceki tahmin değeri ile son tahmin arasında:

$$\left(\sum_i \Delta\hat{\gamma}_i^2 / \sum_i \hat{\gamma}_i^2\right)^{1/2} \leq 10^{-8}$$

sağlanınca iterasyona son verilmiştir. Yaklaşım (convergence) sağlanması için ihtiyaç duyulan iterasyon sayıları Çizelge 2'de verilmiştir. Tüm parametre tahminlerinde Newton-Raphson Algoritması kullanılmasına rağmen sırasıyla 2 ve 3'den 2 yada 3 kat daha hızlı yaklaşım elde edilmiştir. Yaklaşım sağlanması muameleler içi varyansların esas alındığı γ başlangıç değeri ile sıkı bir ilişki içerisindedir.

Çizelge 1. Çalışmada kullanılan veri setinin yapısı

Yıl (Y)	Baba(B)	Doğ. Mevsimi (DM)	N	Σy	Minimum	Maksimum	σ^2_e
1	1	1	62	274465	2651	6079	594403.99
		2	60	282385	3297	6335	415501.42
		3	43	205487	3712	6261	370627.71
		4	265	1257395	1135	7131	804931.95
	2	1	81	384336	1646	9270	1036812.6
		2	68	303844	2367	6737	656704.39
		3	61	262622	2666	6500	47816.1
		4	344	1578306	1290	8124	76192.56
2	1	1	10	46220	4350	4670	149382.25
		2	6	28315	4398	5237	302737.64
		3	4	17907	3897	4670	75717.98
		4	63	303814	3975	7329	154835.95
	2	1	16	73348	3861	4860	19284.87
		2	10	44699	3450	4670	464374.10
		3	8	35960	4370	4670	85834.35
		4	107	522746	3269	7976	493753.56

Veri seti için "alt gruplar arası hesaplanan varyanslar homojendir" şeklinde kurulan H_0 hipotezi red edilmiştir. Tüm 6 durumda da yaklaşım aynı noktada sağlanmıştır. Bu durumda, yöntemin algoritma ve başlangıç değeri seçiminde doğru olduğu kabul edilebilir.

Çizelge 2. Algoritmalara göre Yaklaşım elde etmek için iterasyon sayıları (Yıl x Baba x Doğum Mevsimi'nin dikkate alındığı model için)

Algoritma[1]	Grup İçi Varyanslar	Genel
1	3	7
2	14	14
3	21	23

[1] Algoritmalar sırasıyla Newton-Raphson, Genelleştirilmiş EM, İteratif Ağırlıklı En küçük Kareler

En iyi model seçimi

Hata varyanslarının dikkate alındığı 7 model yazılabilmektedir. En basit model sadece bir parametre içeren, yani sadece γ'yı içeren modeldir. Burada varyansların homojen olduğu hipotezi öne sürülmüştür. Daha çok bilgi veren bir başka model 3 faktörü (Y,B ve M)'de içeren bir model veya Y+B, Y+M, B+M etkilerinin gözlendiği bir model olabilmektedir. Buna üç faktörlü model demekte mümkündür. Her durumda modelde tüm veri seti Karışık Doğrusal Model Eşitliği kullanılarak analiz edilmiştir. Uyum yetmezliği sapmalardan yola çıkarak belirlenmiş ve en iyi model seçimi yapılmıştır. Oluşturulan modellerin tamamında uyum yetmezliği testi yapılarak en iyi modellerin Y +B ve Y +B + M etkilerini içeren modeller olduğu belirlenmiştir. Bu iki model dağılışı tatminkar bir biçimde açıklayabilmiştir. Varyansların homojen olduğu hipotezi ise P<0.01 seviyesinde red edilmiştir. Uyum yetmezliğine ait test sonuçları Çizelge 3'te verilmiştir. Bu sonuçlara benzer şekilde tek bir faktörün yer aldığı modelde varyansların heterojenliği yetersiz bilgi nedeniyle önemsiz bulunmuştur. Her bir faktörün parçalar halinde modele sokulması ile yapılan hata varyanslarının heterojenliğine ait Log-Olabilirlik Link Fonksiyonu ve Genelleştirilmiş Varyans Analizi kullanılarak yapılmış olan analizlerden elde edilen sonuçlar Çizelge 4'te özetlenmiştir.

Çizelge 3. Farklı etkilerin modele alınmasıyla oluşturulan modellere ait uyum yetmezliği test sonuçları

Model	Sd.	Sapma	P Değeri	Önem Seviyesi
Heterojenlik[1]	16-16=0	0		
Y+B+M[2]	16-6=10	8.005	0.3805	Önemsiz
Y+B	16-4=12	11.002	0.1265	Önemsiz
Y+M	16-5=11	12.004	0.0366	*
B+M	16-6=10	25.456	0.0012	**
Y	16-2=14	26.365	0.0000	**
B	16-2=12	36.365	0.0000	**
M	16-4=12	56.654	0.0000	**
Ortalama	16-1=15	79.345	0.0000	**

[1] Tam Heterojenlik
[2] Y=Yıl; B= Baba; M=Doğum Mevsimi
[*] P<0.05
[**] P<0.01

Parametre Tahminleri

Y +B + M ve Y +B modelleri için γ parametre tahminleri Maksimum Olabilirlik Fonksiyonu (ML) Marjinal Yoğunluk fonksiyonuna göre düzenlenmesinden elde edilen Marjinal Maksimum Olabilirlik (MML) eşitlikleri kullanılarak 2 model için (Y + B ve Y + B + M) tahminlenmiştir. Sonuçlar Çizelge 5'te Özetlenmiştir. Tahminlerin standart hataları, parametrelerin MML tahminleyicilerinin 2. derecede türevlerinden yola çıkılarak elde edilmiştir. μ =log 1208 = 3.082 varyans tabanı kullanılarak hata varyansları için pozitif tahminler elde edilmiştir. 2, 3 ve 4'üncü doğum mevsimleri birinci mevsime göre düzeltilmiş olduğundan, 2. yıla ait varyanslar daha yüksek bulunmuştur.

Bu çalışmada çok yaygın kullanım alanı bulması nedeniyle Logaritmik Link Fonksiyonunun kullanılmasıyla, muamele grupları arasındaki hata varyanslarının homojen olmadığı durumlarda analiz ve tahminleme aşamaları özetlenmiştir. Log-Lineer Model'in dışında kullanılabilecek değişik modeller bulunmaktadır. Örneğin GEE (Generelazition Estimation Equation) bunlardan biridir (SAS,1998).

Log-Lineer Model'de parametre tahminleri ve hipotez testleri kullanılan parametrelerin entegrasyonu yoluyla sabit ve şansa bağlı etkilerin modelde kesin olmaması nedeniyle, Marjinal Olabilirlik Fonksiyonu baz alınarak analizler tamamlanabilmiştir. Bu sayede model seçimi ve doğru model oluşturulması gibi önemli aşamalarda bir esneklik sağlanmaktadır. Ayrıca, model seçimi ve doğru modelin oluşturulmasında tüm olasılık fonksiyonunun kullanılması güçlüğü ortadan kalkmış olmaktadır. Burada REML (Restricted Maximum Likelihood) yönteminin de temel aldığı hata varyanslarının tek olduğu varsayımına (ortak varyans), K=I ya da K=1 değerlerinde ulaşılmıştır. Bu, parametre yazılımı aşamasında değişmezlikten ileri gelen ML (Maksimum Olabilirlik)' ye ait bir durumdur.

Çizelge 4. Faktörlerin hata varyansları için, varyansların heterojenliği için kullanılan modellere ait karşılaştırma sonuçları

VK	Karşılaştırma	Sd	Test İstatistiği[1]	P Değeri	Önem Seviyesi
Y	B+M ve Y+B+M	2	14.56	0.000	**
B	Y+M ve Y+B+M	1	65.32	0.023	*
M	Y+B ve Y+B+M	2	1.45	0.485	Önemsiz

[1] 14 nolu eşitliğe göre
[*] $P<0.05$
[**] $P<0.01$

Sabit etkilere ait β ve ayrıca γ parametrelerinin tahmini için Bayes Analizi benzer sonuçlar verebilmektedir. REML ve bir çok algoritmada quadratik formlarda yapılacak ufak maniplasyonlarla sonuca ulaşılabilir. Ancak bu metot kendi parametre uzayı içinde tahminleme yaptığı için sapmasız bir tahminleyici olarak kabul edilmemektedir. Bu durumda çoğu zaman hesaplama tekniğinden doğacak güçlüklerle karşılaşılacaktır. Bu nedenle en hızlı ve kullanışlı algoritma olan Newton-Raphson tercih edilmelidir. Zira, basit notasyonlarda çok daha etkin bir kullanıma sahiptir.

Çizelge 5. Kullanılan iki model için parametre tahminleri ve standart hataları

Parametre[1]	Gerçek Değer	Model	
		Y + B + M	Y + B
μ	7.64	7.6256± 0.2565	7.6052±0.2415
$\alpha_2 + \alpha_1$	-2.48	-2.9578±0.4245	-2.8564±0.6523
$\beta_2 + \beta_1$	2.65	2.3895±0.8743	2.0056±0.9845
$\beta_3 + \beta_1$	-0.56	1.0098±0.2636	0.9866±0.3241
$\beta_4 + \beta_1$	-0.00	0.0045±0.0456	-0.0963±0.0025
$\sigma_2 + \sigma_1$	0.65	0.6126±0.3798	0.6245±0.2312

[1]α: yıl; β: Doğum Mevsimi ve σ: babalar olmaktadır.

Benzer kaynaklara ait varyansların heterojenliği söz konusu olduğunda, hata varyansı tahmini ve H_o hipotezinin test edilmesine ilişkin kullanışlı bir metot olan Log-Likelihood tanıtılmıştır. Bu metot sayesinde doğrudan veri setinin kendisinin kullanımı söz konusu olduğu için, transforme edilmiş veri setlerinin yorum ve işlem karmaşasından kaçınılmış olmaktadır. Hayvancılık alanında bir çok çalışma sonucunda elde edilen veri setlerinde benzer sorunla sıklıkla karşılaşılmaktadır. (Gianola ve ark., 1986; Garrick ve Van Vleck, 1987; Gianola ve ark., 1990). Çalışmada kullanılan veri seti de bunlardan birisidir. Transformasyonla 'varyansların homojenliği' varsayımına kısmen ulaşılmaktadır. Ancak, parametre tahminini esas alan ıslah amaçlı çalışmalarda transformasyondan doğacak olan metot hatalarından kaçınmak çoğu zaman mümkün olmamaktadır (Winkelman ve Schaeffer, 1988). Anılan sakıncası nedeniyle bu metot özellikle ıslah amaçlı olarak yürütülmüş olan çalışmalarda en önemli aşama olarak kabul edilen analiz ve varyans unsur tahmini aşamasında büyük kolaylıklar sağlamaktadır. Bunlardan en önemlisi negatif varyans unsur tahmini yapılmamaktadır. Yöntem quasi-likelihood temeline dayalı olduğu için şansa bağlı unsurlar için yapılan varyans unsur tahminlerine yaklaşım sağlanıncaya kadar devam etmekte, ancak yaklaşım sağlanamazsa 0 atayarak işlem durdurulmaktadır. Bu nedenle negatif parametre tahminine izin verilmemektedir. Ancak, kendi parametre uzayı içersinde tahminleme yapan bu metot sapmasız bir tahminleyici olarak kabul edilmemektedir.

Sonuç olarak, çalışmada önerilen metot özellikle hayvan ıslahı çalışmalarında elde edilen veri setlerinde alt gruplar arasında varyansların bir örnek olmadığı durumlarda kullanışlı bir metot olmaktadır. Varyans-kovaryans unsurlarının tahminlemesi aşamasında Newton-Raphson yönteminin kullanılması hesaplamada kolaylık sağlamaktadır.

Kaynaklar

Box, G. E. P. and R. D. Meyer, 1986. Dispersion effects from factorial design. ***Technometrics.,*** 28: 19.

Davidan, M and R. J. Carrol, 1987. Variance function estimation. ***J. Am. Stat. Assoc.,*** .82: 1079.

De Veer, J.C. and L. D. Van Vleck, 1987. Genetic parameters for first lactation milk yields at three levels of herd production. ***J. Dairy SCI.,*** 70:1434.

Foulley, J. L., D. Gianola and S. Im, 1989. A Simple algorithm for computing marginal maximum likelihood estimates of variance components and it's relation to

EM. Proc. 47th Sess. Int. Stat. Inst., Paris, Fr., Aug. 29-Sept 6.

Foulley, J.L., D. Gianola, M. San Cristobal and S. Im, 1990. A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. ***J. Dairy Sci.,*** 73: 1612-1624.

Foulley, J.L., S. Im, D. Gianola and I. Hoschele, 1987. Emprical bayes estimation of parameters for n polygenic binary traits. ***Genet. Sel. Evol.,*** 19: 197.

Foulley, J. L. and R. L. Quaas, 1995, Heterogeneous variances in gaussian linear mixed models. ***Genet. Sel. Evol.***, 27, 211-228.

Garrick, D. J. and L. D. Van Vleck, 1987. Aspects of selection for performance in several environments with heterogeneous variances. ***J. Anim. Sci.,*** 65: 409.

Gianola, D., 1986. On selection criteria and estimation of parameters when the variance is heterogeneous. ***Theor. Appl. Genet.,*** 72: 671.

Gianola, D., J. L. Foulley and R. L. Fernando, 1986. Prediction of breeding values when variances are not known. ***Genet. Sel. Evol.,*** 18: 485.

Henderson, C. R., 1984. ***Application of Linear Models in Animal Breeding.,*** Univ. Guelph, Guelph. Can.

Hill, W. G., 1984. On selection among groups with heterogeneous variance. ***Anim. Prod.,*** 39: 473.

Louis, T. A., 1982. Finding the observed information matrix when using the EM algoritm. ***J.R. Stat. Soc. B.*** 44: 226.

Mc Cullagh, P. and J. A. Nelder, 1983. ***Generalized Linear Models. Chapman and Hall,*** London,UL.

Winkelman, A. and L. R. Schaeffer, 1988. Effect of heterogeneity of variance on dairy sire evaluation. ***J. Dairy Sci.,*** 71: 3033.

SAS, 1998, ***User's Guide:Statistics, Version 12.0 Edition.*** SAS Inst., Inc., Cary, Nc.