



Eksik Veri Analizinde Çoklu Atama Yönteminin Değerlendirilmesi: Hayvancılıkta Tekrarlı Ölçüm Verisi Üzerine Bir Uygulama

Gazel Ser*, Cafer Tayyar Bati

Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootehni Bölümü, Biyometri ve Genetik Ana Bilim Dalı, 65080 Van, Türkiye

MAKALE BİLGİSİ

Geliş 09 Ağustos 2015
Kabul 23 Ekim 2015
Çevrimiçi baskı, ISSN: 2148-127X

Anahtar Kelimeler:
Eksik gözlem yapıları
Çoklu atama
Tekrarlı veri
Karışık model
Norduz koyunu

* Sorumlu Yazar:
E-mail: gazelser@gmail.com

Ö Z E T

Bu çalışmada, eksik gözlem yapısının şansa bağlı ve tamamen şansa bağlı olması durumunda çoklu değer atama yöntem performansının, genel doğrusal karışık model yaklaşımında değerlendirilmesi amaçlanmıştır. Uygulama verisini, 7 aylık yaşta toplam 77 baş Norduz erkek toklusu oluşturmuştur. Kesimden sonra, beş farklı zaman noktasında ölçülen pH değerleri, bağımlı değişken olarak belirlenmiştir. Bununla beraber modellere bağımsız değişken olarak, sıcak karkas ağırlığı, kas glikojen düzeyi ve açlık süreleri dahil edilmiştir. Eksik gözlem içermeyen bağımlı değişkende, belirli oranlarda (%10 ve %25) gözlemlerin silinmesiyle tamamen şansa bağlı (MCAR-missing completely at random) ve şansa bağlı (MAR-missing at random) olmak üzere iki eksik gözlem yapısı oluşturulmuştur. Sonrasında eksik gözlem yapılarına sahip veri setlerinde, çoklu atama yöntemi (MI-multiple imputation) uygulanarak tam veri setleri elde edilmiştir. MI yöntemi kullanılarak tamamlanan veri setlerine, genel doğrusal karışık model uygulanarak elde edilen sonuçlar, tam veriye ilişkin sonuçlarla karşılaştırılmıştır. Tam veri ve MI veri setlerine uygulanan karışık modelde varyans-kovaryans yapıları aynı, parametre tahmin sonuçları ve standart hatalar ise tam veriye oldukça yakın sonuçların elde edilmesini sağlamıştır. Sonuç olarak, bu çalışmada her iki eksik gözlem yapısı ve oranlarında, eksik gözlem atama yöntemi olarak MI'nin tercih edilmesi durumunda karışık modelde güvenilir bilgilerin elde edilmesi sağlanmıştır.

Turkish Journal Of Agriculture - Food Science And Technology, 3(12): 926-932, 2015

Evaluation of Multiple Imputation in Missing Data Analysis: An Application on Repeated Measurement Data in Animal Science

ARTICLE INFO

Article history:
Received 09 August 2015
Accepted 23 October 2015
Available online, ISSN: 2148-127X

Keywords:
Missing observation structures
Multiple imputation
Repeated data
Mixed model
Norduz sheep

* Corresponding Author:
E-mail: gazelser@gmail.com

ABSTRACT

The purpose of this study was to evaluate the performance of multiple imputation method in case that missing observation structure is at random and completely at random from the approach of general linear mixed model. The application data of study was consisted of a total 77 heads of Norduz ram lambs at 7 months of age. After slaughtering, pH values measured at five different time points were determined as dependent variable. In addition, hot carcass weight, muscle glycogen level and fasting durations were included as independent variables in the model. In the dependent variable without missing observation, two missing observation structures including Missing Completely at Random (MCAR) and Missing at Random (MAR) were created by deleting the observations at certain rations (10% and 25%). After that, in data sets that have missing observation structure, complete data sets were obtained using MI (multiple imputation). The results obtained by applying general linear mixed model to the data sets that were completed using MI method were compared to the results regarding complete data. In the mixed model which was applied to the complete data and MI data sets, results whose covariance structures were the same and parameter estimations and standard estimations were rather close to the complete data are obtained. As a result, in this study, it was ensured that reliable information was obtained in mixed model in case of choosing MI as imputation method in missing observation structure and rates of both cases.

Giriş

Tekrarlı ölçümlü denemeler, belirli bir zaman veya koşul altında aynı deneme ünitesinden elde edilen ardışık ölçümler için kullanılan bir deneme desenidir. Tekrarlı ölçüme sahip denemelerde ardışık olarak elde edilen gözlemler arasında bir ilişki söz konusudur. Bu ilişkinin varlığı, varyans analizinin ön şartı olan, gözlemlerin bağımsızlığı varsayımının sağlanamamasına neden olmaktadır. Örneğin, laktasyonun üçüncü ayında yüksek verime sahip bir ineğin, muamele etkisini dikkate almaksızın laktasyonun dördüncü ayında da yüksek verim verebileceği düşünülmektedir. Dolayısıyla, aynı hayvandan alınan ardışık ölçümler birbirinden bağımsız değildir. Bu nedenle tekrarlı ölçüme sahip verilerin analizinde aynı bireyden elde edilen gözlemler arasındaki ilişkiyi ve heterojen varyansların, kullanılacak model(ler)de tanımlanması gerekmektedir. Bu tanımlama, ardışık gözlemler arasındaki ilişkileri hesaba katan uygun varyans-kovaryans yapılarının kullanılmasıyla mümkün olmaktadır (Kaps ve Lamberson, 2004; Wang ve Goonewardene, 2004).

Bu tip denemelerin analizinde kullanılan yöntemler, modern ve geleneksel olmak üzere iki başlık altında incelenmektedir. Geleneksel yöntemler, tekrarlı ölçümlü varyans analizi (Repeated ANOVA) ve tekrarlı ölçümlü çok değişkenli varyans analizi (Repeated MANOVA) şeklindedir. Modern yöntemler ise karışık model (Mixed Model) ve genelleştirilmiş tahmin denklemleri (GEE-Generalized Estimating Equations) yaklaşımlarıdır (Wolfinger, 1992; Ser, 2011). Geleneksel yöntemler, varyans-kovaryans yapılarına karşı olan kısıtlayıcılıkları, dengeli veri yapısı gerektirmesi gibi özelliklerinden dolayı analizlerde zorluklara neden olmaktadır. Karışık model veya genelleştirilmiş tahmin denklemleri yaklaşımları ise eksik gözlem olması durumunda uygulanabilir olması, tekrarlı ölçümler arasındaki ilişki yapısını dikkate alan uygun varyans-kovaryans yapılarının tanımlanabilmesi gibi avantajlar sağlamaktadır (Wang ve Goonewardene, 2004; Bahçecitapar, 2006; Hedeker ve Gibbson, 2006; Davidian, 2007).

Tekrarlı ölçüme sahip denemelerde eksik gözlem durumuyla sıklıkla karşılaşmaktadır. Çalışmadaki eksik gözlemlerin sayısı, çalışma sonucunu etkileyebilecek durumda iki yol tercih edilebilir. İlki, eksik gözlem durumunu dikkate alan modellerin tercih edilmesi, ikinci olarak da eksik gözlem problemi için uygun değer atama veya çözüm yöntemlerinin kullanılarak eksik gözlem sorununun giderilmesidir. Her iki yolun tercihinden önce, veri içerisindeki eksik gözlemlerin şansa bağlılığının sorgulanması gerekmektedir.

Veri içerisinde bulunan eksik gözlemlerin şansa bağlılığı üç farklı yapı ile tanımlanmıştır. İlki, eksik gözlemlerin, gözlenen ve gözlenemeyen verilerden bağımsız olduğu tamamen şansa bağlı yapı (missing completely at random, MCAR), ikinci olarak eksik gözlemlerin sadece gözlenen verilere bağlı olduğu şansa bağlı yapı (missing at random, MAR) ve eksik gözlemlerin şansa bağlılık göstermediği şansa bağlı olmayan (missing not at random, MNAR) yapı söz konusudur (Little ve Rubin, 1987). Özellikle MAR yapısı, genellikle denemede desenlerindeki eksikliklerden meydana gelmesinden dolayı, sıklıkla karşılaşılan bir eksik

gözlem yapısıdır (Twisk, 2004). Değer atama yöntemlerinden günümüzde en çok tercih edileni çoklu atama (MI) yöntemidir. Yöntemde, belirlenen atama sayısı kadar eksik değerler yerine, birden fazla değer atanarak tamamlanmış veri setleri üzerinden analizler yapılmaktadır (Baygül, 2007; Erdoğan, 2012). Uygulanacak değer atama ve istatistik modellerin seçiminde, eksik gözlem yapıları belirleyici olmaktadır. Örneğin veri setinde bulunan eksik gözlemlerin yapısı MCAR ise standart GEE yönteminin tercih edilmesi daha geçerli tahminlerin elde edilmesini sağlamaktadır. Bununla beraber, hem karışık model hem de MI yöntemi eksik gözlem yapısının MAR ve MCAR olduğu durumda geçerli sonuçların elde edilmesine olanak sağlamaktadır (Twisk, 2004; Molenberghs ve Kenward, 2007; Kenward ve Carpenter, 2009; Twisk ve ark., 2013).

Bu çalışma, son yıllarda diğer yöntemlere göre araştırmacılara sağladığı avantajlı yönleriyle ön plana çıkan çoklu atama yönteminin, iki farklı eksik gözlem mekanizmasında (MAR ve MCAR) uygulama prensiplerinin irdelenmesini ve hem eksik gözlem durumunda hem de tam gözlem durumunda kullanılabilen genel doğrusal karışık modelde performansını değerlendirilmeyi amaçlamıştır.

Materyal ve Metot

Çalışma Materyali

Çalışmanın uygulama materyalini, kesim öncesi 0, 12, 24 ve 48 saat süre ile su kısıtlaması olmaksızın aç bırakma uygulamasına tabi tutulan 7 aylık yaşta 77 baş Norduz erkek toklusu karkaslarından elde edilen pH, sıcak karkas ağırlığı ve kas glikojen düzeyi parametreleri oluşturmuştur. Etin pH'sı 45.dakika (pH_{45dk}), 4. saat (pH_{4sa}), 8. saat (pH_{8sa}), 24. saat (pH_{24sa}) ve 72. saat (pH_{72sa}) olmak üzere beş farklı zamanda belirlenmiştir. pH ölçümleri, sol yarımda *m. longissimus thoracis* (LT) üzerinden 12-13. kaburgalar arasından karkas üzerinde gerçekleştirilmiştir. Kas glikojen düzeyinin belirlenmesine yönelik olarak ise kesim sonrası ilk 30 dakika içerisinde karkasın sağ tarafında LT kasından (12-13. kaburgalar arası) ve karaciğer merkez lobunun orta noktasından örnekleme yapılmıştır. Bu çalışmada kullanılan verilerin elde edilişi ile ilgili detaylı bilgiler, Karaca ve ark. (2015) tarafından tanımlanmıştır.

Çalışmada Kullanılan Veri Setlerinin Oluşturulması

Çalışmada, beş farklı zaman noktasında elde edilen pH ölçümleri bağımlı değişken, açlık süreleri (kategorik yapıya sahip), sıcak karkas ağırlığı ve kas glikojeni ise bağımsız değişkenler olarak belirlenmiştir. Eksik gözlemlerin oluşturulmasında uygulanan silme işlemi, bağımlı değişkende (pH ölçümleri) yapılmış olup, bağımsız değişkenlerde silme işlemi yapılmamıştır. Çalışmada kullanılan veri setleri aşağıdaki şekilde oluşturulmuştur.

- Çalışma içerisinde "Tam veri seti" olarak ifade edilen veri setinde eksik gözlem bulunmamaktadır.
- Çalışmada kullanılan tam veri setinde, iki farklı eksik gözlem yapısının oluşturulması amacıyla gözlemler silinmiştir. Bu işlem iki şekilde

yapılmıştır:

1. Bağımlı değişkende yapılan silme işleminde, veri setinde bulunan değişkenlerin hiçbiri dikkate alınmadan, tamamen şansa bağlı (MCAR) olarak yapılmıştır (Twisk ve De Vente, 2002; Alpar, 2003; Eydur, 2009). Örneğin, bir hayvandan elde edilen pH ölçümlerinin 4.'cü saatteki ölçümü silinirken, diğer hayvanın 4.'cü, 8.'ci ve 72.'ci saatlerine ait pH ölçümlerine ilişkin üç gözlem silinmiştir. Buna göre çalışmada 45.'ci dakika 4, 8, 24 ve 72.'ci saatlerdeki pH ölçümleri tekrarlı olarak alt alta yerleştirildikten sonra, %10 ve %25 oranlarında silme işlemi gerçekleştirilmiştir. Böylece MCAR yapısına sahip iki veri seti oluşturulmuştur. Bu veri setleri çalışma içerisinde %10 MCAR ve %25 MCAR olarak ifade edilmiştir.
 2. Tam veri seti üzerinden şansa bağlı (MAR) eksik gözlem yapısını oluşturmak amacıyla, pH ölçümlerinin alındığı zaman noktaları dikkate alınarak silme işlemi gerçekleştirilmiştir. Örneğin, "t" ölçümlerinin alındığı zaman noktalarını ifade ettiği varsayalım. Buna göre, bir hayvandan 5 zaman noktasından (t=45.'ci dakika, 4, 8, 24 ve 72.'ci saatler) elde edilen pH ölçümlerinde, t=45.'ci dakikaya ilişkin ölçümlerde silme işlemi yapılmazken t=8.'ci saatte elde edilen pH ölçümlerinde en yüksek değere sahip olan bireylerin (erkek toklu) sonraki zaman noktasındaki (t = t+1) ölçüm değerleri silinerek, MAR yapısı oluşturulmuştur (Twisk ve De Vente, 2002). Silme işlemi %10 ve %25 oranlarında gerçekleştirilerek, iki veri seti oluşturulmuştur. Bu veri setleri, çalışma içerisinde %10 MAR ve %25 MAR olarak ifade edilmiştir.
- Çalışmada, yukarıda oluşturulma yöntemlerinin açıklandığı %10-%25 MAR veri setleriyle, %10-%25 MCAR veri setlerindeki eksik gözlemler, sonrasında çoklu atama (MI) yöntemiyle tahmin edilerek tam veri setleri elde edilmiştir. Bu veri setleri ise çalışma içerisinde %10 MCAR-MI, %25 MCAR-MI, %10 MAR-MI ve %25 MAR-MI, şeklinde ifade edilmiştir. Bu aşamada 4 veri seti elde edilmiştir.
 - Özetle, tam veri, eksik gözlemlerin oluşturulduğu 4 veri seti (%10-%25 MAR ile %10-%25 MCAR) ve eksik gözlemlerin tahmin edilerek oluşturulan 4 tam veri seti (%10- %25 MAR-MI ve %10-%25 MCAR-MI) olmak üzere toplamda 9 veri seti üzerinden çalışma yürütülmüştür.

Yukarıda bahsedilen veri setlerinin değerlendirilmesinde, genel doğrusal karışık model kullanılmıştır. Bununla beraber, genel doğrusal karışık modellerde tekrarlı ölçümler arası varyans-kovaryans yapısının belirlenmesinde kullanılan, bileşik simetri (CS-compound symmetry), yapısal olmayan (UN-unstructured), birinci dereceden otoregresif (AR(1)-first order autoregresif), varyans bileşenleri (VC-variance component) ve Toeplitz (TOEP-banded) kovaryans yapıları uygulanarak, en uygun varyans-kovaryans yapısı belirlenmiştir. Veri setine uygun varyans-kovaryans

yapısının seçiminde, ölçümlerin alındığı zaman aralıkları da belirleyicidir. Örneğin, CS, UN, VC eşit ve eşit olmayan zaman aralıklarında tercih edilirken, AR(1) yapısı eşit olmayan zaman aralıklarında çok fazla tercih edilmemekle beraber, AR(1) yapısının özel bir durumu olan Toeplitz yapısı eşit olmayan zaman aralıklarında da anlamlı sonuçlar vermektedir (Kincaid, 2005; Bahçecitapar, 2006; Davidian, 2007; Liang, 2015). Bu çalışmada, zaman aralıkları eşit olmasına rağmen AR(1) yapısı da dikkate alınmıştır. Çünkü AR(1) yapısı, birbirine yakın zaman aralıklarında elde edilen ölçümler arasında daha güçlü ilişkinin olduğunu varsayarken, birbirine uzak zamanlarda bu ilişkinin tüssel olarak azaldığını varsaymaktadır (Aktaş, 2005; Kincaid, 2005).

Veri setine en uygun varyans-kovaryans yapısının belirlenmesinde AIC (Akaike Information Criteria) ve BIC (Bayesian Information Criteria) uyum ölçütleri kullanılmıştır. En küçük AIC ve BIC uyum ölçütünü sağlayan varyans-kovaryans yapısı, veri setine en uygun yapı olarak belirlenmiştir. Eksik gözlem yapılarının oluşturulması ve çoklu atama yöntemine ilişkin analizler, SPSS paket programının "Missing Value Analysis" modülünde (SPSS, 2013) yapılırken, genel doğrusal karışık model analizleri ise SAS paket programının PROC MIXED prosedürü kullanılarak yapılmıştır (SAS, 2014).

Bulgular ve Tartışma

Eksik Gözlem Yapılarının Oluşturulmasına İlişkin Sonuçlar

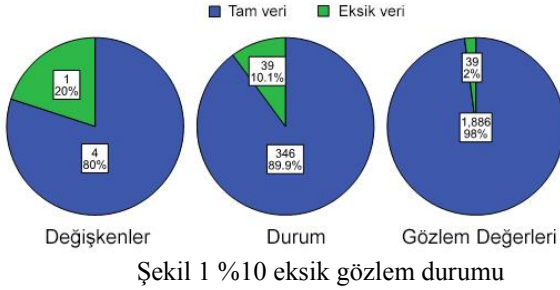
SPSS yazılım programında "Missing Value Modülü" kullanılarak öncelikli olarak pH bağımlı değişkeninin tam veri ve eksik gözlem yapılarındaki gözlem sayıları (N), ortalama ve standart sapmaları Çizelge 1'de verilmiştir.

Çizelge 1'de, bağımlı değişkenin tekrarlı ölçüm özelliğine sahip olmasından dolayı toplam $5 \times 77 = 385$ ölçüm değeri elde edilmiştir. Bağımlı değişkende, MAR ve MCAR yapılarının oluşması için yaklaşık %10 (39 gözlem) ve %25 (96 gözlem) oranında gözlemler silinmiştir. Buna göre bağımlı değişkendeki eksik gözlem durumunu görsel olarak özetleyen pasta grafikler Şekil 1 ve 2'de verilmiştir.

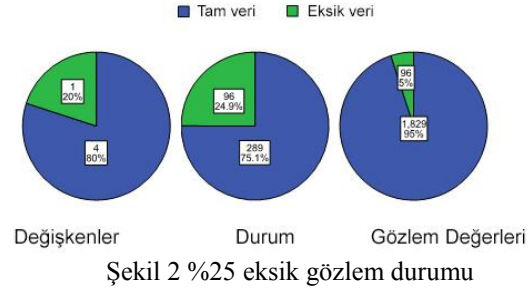
Şekil 1 ve Şekil 2'de değişken durumunu ifade eden pasta grafiğinde, veri setindeki 5 değişkene ait eksik gözlem durumu gösterilmiştir. Buna göre 5 değişkenin 1'inde yani %20'lik kısmında, eksik gözlem bulunmaktadır, diğer 4 değişkende ise eksik gözlem bulunmaktadır. Bu değişkenler, %80'lik dilimi oluşturmuştur. Şekil 1'de, sadece eksik gözlemin bulunduğu değişkenin durumunu gösteren grafik de ise pH özelliğine ait gözlem sayısının %10,1'lik kısmında eksik gözlem bulunurken, %89,9'lük kısmında eksik gözlem bulunmamaktadır. Benzer şekilde Şekil 2'deki durum pasta grafiğinde, pH özelliğinde 96 eksik gözlem bulunmaktadır. Veri setinde bulunan beş değişkene ait gözlem değerlerini kullanarak oluşturulan gözlem değerleri grafiğinde ise Şekil 1 için 1925 gözlemden, 39 tane gözlem değeri eksik ($385 \times 5 = 1925 - 39 = 1886$), Şekil 2'de ise 1925 gözlemden, 96 ölçüm değeri eksiktir ($385 \times 5 = 1925 - 96 = 1829$). Eksik gözlemlerin tamamen şansa bağlı (MCAR) yapıya sahip olup-olmadığının test edilmesinde uygulanan Little'in MCAR testi sonuçları Çizelge 2'de verilmiştir.

Çizelge 1 pH bağımlı değişkenine ilişkin tanıttıcı istatistikler

Bağımlı Değişken	Eksik Gözlem Yapıları	Eksik		N	Ortalama	Standart Sapma
		N	%			
pH	Tam Veri	-	-	385	6,0328	0,32745
	MAR	39	10,1%	346	6,0322	0,32424
	MAR	96	24,9%	289	6,0239	0,32024
	MCAR	39	10,1%	346	6,0420	0,32870
	MCAR	96	24,9%	289	6,0417	0,33042



Şekil 1 %10 eksik gözlem durumu



Şekil 2 %25 eksik gözlem durumu

Çizelge 2 Little'in MCAR testine ilişkin sonuçlar

Eksik gözlem oranları	Eksik gözlem yapıları	Little's MCAR testi	
		Ki-kare(X^2)	P
%10 oranında eksik	MCAR	0,233	P>0,05
%25 oranında eksik	MCAR	0,356	P>0,05

Çizelge 2'de Little'in MCAR testine ilişkin ki-kare sonuçları incelendiğinde, $p>0.05$ olması durumunda verideki eksik gözlemlerin tamamen şansa bağlı (MCAR) olduğu kabul edilmiştir (SPSS, 2013). Eksik gözlem yapıları belirlendikten sonra bu veri setlerindeki eksik gözlemler çoklu atama yöntemiyle tamamlanarak, tam veri setleri elde edilmiştir. Çoklu atama yönteminde, atama modeli olarak doğrusal regresyon modeli kullanılmış ve modele veri setinde bulunan değişkenlerin (pH, açlık süresi, zaman, kas glikojen düzeyi ve sıcak karkas ağırlığı) tamamı dahil edilmiştir. Çoklu atama yönteminde, atama sayısı oldukça önemlidir. Bu çalışmada, atama sayısı ($m=5$) olarak alınmıştır.

Genel Doğrusal Karışık Modellerle İlişkin Sonuçlar

Genel doğrusal karışık modeller tam veri, %10-%25 eksik gözlem oranlarında MAR ve MCAR yapılarındaki veri setlerine MI yöntemi uygulanarak, eksik gözlemler tahmin edilmiştir. MI yöntemiyle tamamlanan veri setleri, %10-%25 MAR-MI ve %10-%25 MCAR-MI şeklinde ifade edilmiştir. Tam veri ve MI veri setlerinde, tekrarlı ölçüm arası varyans-kovaryans yapısını belirlemek amacıyla AIC ve BIC uyum ölçütlerine ilişkin sonuçlar Çizelge 3'de verilmiştir.

Çizelge 3 incelendiğinde, tam veri setinde tekrarlı ölçümler arası varyans-kovaryans yapısı için beş ayrı kovaryans (CS), (UN), (AR(1)), (VC), (TOEP) yapısı karışık modelde tanımlanmıştır. Buna göre tam veride, en küçük AIC ve BIC uyum ölçütü Toeplitz yapısından elde edilmiştir. Bununla birlikte eksik gözlemlerin tahmin edildiği MAR-MI ve MCAR-MI veri setlerinde de en uygun yapı Toeplitz olarak belirlenmiştir. Bu yapının tam veri setiyle aynı olması, dikkat çekicidir. MI, eksik veriler için uygun atamalar yaparak belirlenen atama sayısı kadar tam veri setleri oluşturur. Bu işlem yapılırken eksik

gözleme sahip veri setinin dağılımı ve veri setindeki değişkenler arasındaki ilişki yapısını koruyarak atama işlemini yapmaktadır (Schafer ve Graham, 2002; Kristman ve ark., 2005).

Buna göre Çizelge 4'de Toeplitz yapısından elde edilen kovaryans parametre tahminleri ve önemlilik durumları verilmiştir.

Toeplitz kovaryans yapısında parametre sayısı, verideki tekrarlı ölçüm sayısı kadar olduğundan Çizelge 4'de beş parametre bulunmaktadır. TOEP(2), TOEP(3), TOEP(4) ve TOEP(5), aynı denek (erkek toklu) üzerinde birinci, ikinci, üçüncü ve dördüncü zamanlarda, ölçümler arasındaki kovaryansları göstermektedir. Hayvanlar içi ölçümlerin (hata) değişkenliği yani varyans, tam veride ve MI (MAR-MI ve MCAR-MI) veri setlerinde önemli bulunmuştur ($p<0.0001$). Bununla beraber ölçümler arası kovaryanslar da TOEP(2), TOEP(3), TOEP(4) ve TOEP(5), tam veri ve MAR-MI ve MCAR-MI tahminlerinde önemli bulunmuştur. Tam veri setinde beş ölçümlü Toeplitz yapısının matris formu, aşağıdaki şekilde oluşturulmuştur. Kovaryans matrisindeki sütun numaraları, ölçüm zamanlarını ifade etmektedir.

$$TOEP = \begin{bmatrix} 0.1052 & -0.05204 & 0.01167 & 0.05491 & -0.08245 \\ -0.05204 & 0.1052 & -0.05204 & 0.01167 & 0.05491 \\ 0.01167 & -0.05204 & 0.1052 & -0.05204 & 0.01167 \\ 0.05491 & 0.01167 & -0.05204 & 0.1052 & -0.05204 \\ -0.08245 & 0.05491 & 0.01167 & -0.05204 & 0.1052 \end{bmatrix}$$

Çizelge 5'de tam, MAR-MI ve MCAR-MI veri setlerinde Toeplitz kovaryans yapısından elde edilen sabit etkilerin önemliliklerine ilişkin sonuçlar verilmiştir.

Çizelge 3 AIC ve BIC uyum ölçütlerine ilişkin sonuçlar

Veri Setleri	AIC	BIC
Tam veri		
CS	221,9	226,6
UN	198,5	163,4
AR(1)	170,2	174,9
VC	237,1	239,4
TOEP	60,6	72,3
%10MAR-MI		
CS	205,6	210,3
UN	186,2	151,1
AR(1)	155,4	160,1
VC	224,1	226,4
TOEP	50,4	62,1
%25MAR-MI		
CS	211,2	215,9
UN	155,0	119,9
AR(1)	156,0	160,7
VC	231,9	234,2
TOEP	59,0	70,7
%10MCAR-MI		
CS	227,2	231,9
UN	174,6	139,4
AR(1)	175,9	180,6
VC	242,9	245,2
TOEP	70,6	82,3
%25MCAR-MI		
CS	206,5	211,1
UN	141,6	106,4
AR(1)	153,4	158,1
VC	227,1	229,4
TOEP	70,6	82,4

Çizelge 4 Toeplitz yapısına ait varyans unsurları tahminlerine ilişkin kovaryans parametre tahminleri

Veri Setleri	Kovaryans Parametresi	Birey	Tahmin	Standart Hata	Z	Pr>Z
Tam veri	TOEP(2)	toklu(açlık)	-0,05204	0,007656	-6,80	<0,0001
	TOEP(3)	toklu(açlık)	0,01167	0,003994	2,92	0,0035
	TOEP(4)	toklu(açlık)	0,05491	0,007613	7,21	<0,0001
	TOEP(5)	toklu(açlık)	-0,08245	0,01048	-7,87	<0,0001
	Hata		0,1052	0,01029	10,23	<0,0001
	%10MAR-MI	TOEP(2)	toklu(açlık)	-0,05080	0,007362	-6,90
TOEP(3)		toklu(açlık)	0,009101	0,003872	2,35	0,0187
TOEP(4)		toklu(açlık)	0,05350	0,007536	7,10	<0,0001
TOEP(5)		toklu(açlık)	-0,08055	0,01021	-7,89	<0,0001
Hata			0,1016	0,009966	10,20	<0,0001
%25MAR-MI		TOEP(2)	toklu(açlık)	-0,05384	0,007642	-7,05
	TOEP(3)	toklu(açlık)	0,01138	0,004167	2,73	0,0063
	TOEP(4)	toklu(açlık)	0,05085	0,007533	6,75	<0,0001
	TOEP(5)	toklu(açlık)	-0,07953	0,01048	-7,59	<0,0001
	Hata		0,1032	0,01017	10,15	<0,0001
	%10MCAR-MI	TOEP(2)	toklu(açlık)	-0,05328	0,007814	-6,82
TOEP(3)		toklu(açlık)	0,01259	0,004153	3,03	0,0024
TOEP(4)		toklu(açlık)	0,05369	0,007682	6,99	<0,0001
TOEP(5)		toklu(açlık)	-0,08331	0,01065	-7,82	<0,0001
Hata			0,1067	0,01042	10,24	<0,0001
%25MCAR-MI		TOEP(2)	toklu(açlık)	-0,05162	0,007492	-6,89
	TOEP(3)	toklu(açlık)	0,01109	0,004145	2,68	0,0075
	TOEP(4)	toklu(açlık)	0,04924	0,007506	6,56	<0,0001
	TOEP(5)	toklu(açlık)	-0,07897	0,01031	-7,66	<0,0001
	Hata		0,1025	0,01000	10,25	<0,0001

Çizelge 5’de açlık süresi, veri setlerinin tamamında önemli bulunmuştur ($P<0,05$; $P<0,001$). Sıcak karkas ağırlığı %10 MCAR-MI veri setinin dışında, tüm veri setlerinde önemli bulunmuştur ($p<0,05$). Zaman faktörü %25 MCAR-MI veri setinde önemli ($p<0,05$) bulunurken, kas glikojen düzeyi tüm veri setlerinde önemsiz bulunmuştur. Bununla birlikte, Çizelge 5’de verilen F değerleri arasında bazı farklılıklar söz konusu olmuştur.

Çizelge 6’da karışık modelde tanımlanan, Toeplitz yapısından elde edilen regresyon katsayılarına ilişkin tahminler ve standart hataların önemlilik durumlarına ilişkin sonuçlar verilmiştir.

Çizelge 6’da regresyon katsayılarına ilişkin tahminler, tam veriye göre MAR-MI ve MCAR-MI veri setlerinde oldukça değişken olduğu belirlenmiştir. Ancak, standart hatalar dikkate alındığında, standart hata tahminleri parametre tahminlerine göre daha kararlı sonuçlara sahiptir. MAR-MI ve MCAR-MI veri setlerinde

regresyon katsayılarının standart hataları, tam verideki standart hatalara oldukça yakın bulunmuştur. Ancak her iki yapı ve eksik gözlem oranları arasında bir karşılaştırma yapılmak istenmesi durumunda, özellikle MCAR-MI(%10) yapısından elde edilen standart hatalar, tam verideki standart hatalara daha yakın elde edilmiştir.

Twisk ve ark. (2013) tarafından uzun süreli veri yapısında karışık model uygulanmadan önce MI yönteminin gerekli olup-olmadığını ortaya koymak amacıyla yaptıkları çalışmada, hem bağımlı değişkende hem de kovaryet etkilerde belirli oranlarda silme işlemi yapılarak MCAR, MAR ve MNAR veri setleri elde edilmiştir. MCAR yapısına sahip veri setlerinde, MI yönteminin tam veriye göre performansının, MAR ve MNAR yapısındaki veri setlerine göre daha iyi bulunduğu belirlenmiştir. Bununla beraber, uzun süreli veri yapısına sahip çalışmalarda, karışık model uygulamasından önce MI yöntemine gerek olmadığı bildirilmiştir.

Çizelge 5 Sabit etkilerin önemliliklerine ilişkin sonuçlar (F ve olasılık değerleri)

Veri setleri	Sabit etkiler	F	Pr>F
Tam veri	Zaman	2,20	0,1393
	Açlık Süreleri	3,97	0,0113
	Kas Glikojeni	0,17	0,6839
	Sıcak Karkas Ağırlığı	4,37	0,0403
%10MAR-MI	Zaman	2,59	0,1086
	Açlık Süreleri	6,52	0,0006
	Kas Glikojeni	0,74	0,3924
	Sıcak Karkas Ağırlığı	4,46	0,0381
%25MAR-MI	Zaman	0,65	0,4224
	Açlık Süreleri	5,34	0,0022
	Kas Glikojeni	0,01	0,9239
	Sıcak Karkas Ağırlığı	3,86	0,0533
%10MCAR-MI	Zaman	0,80	0,3727
	Açlık Süreleri	3,74	0,0149
	Kas Glikojeni	0,26	0,6094
	Sıcak Karkas Ağırlığı	3,35	0,0713
%25MCAR-MI	Zaman	4,22	0,0408
	Açlık Süreleri	6,31	0,0007
	Kas Glikojeni	0,49	0,4884
	Sıcak Karkas Ağırlığı	4,88	0,0304

Çizelge 6 Tam veri, MAR-MI ve MCAR-MI verilerinden elde edilen regresyon katsayıları ve standart hata sonuçları

Parametreler	Tam Veri	%10MAR-MI	%25MAR-MI	%10MCAR-MI	%25MCAR-MI	
	Tahmin (Std. Hata)	Tahmin (Std. Hata)	Tahmin (Std. Hata)	Tahmin (Std. Hata)	Tahmin (Std. Hata)	
Sabit	6,2031 (0,08461)****	6,1868 (0,07980)****	6,1735 (0,07869)****	6,1764 (0,08420)****	6,2198 (0,07987)****	
Zaman	-0,00050 (0,000339) ^{ön}	-0,00055 (0,000339) ^{ön}	-0,00027 (0,000337) ^{ön}	-0,00031 (0,000347) ^{ön}	-0,00073 (0,000357)*	
Kas Glikojeni	0,003501 (0,008342) ^{ön}	0,006850 (0,007860) ^{ön}	-0,00066 (0,007752) ^{ön}	0,004337 (0,008297) ^{ön}	0,005565 (0,007860) ^{ön}	
Sıcak Karkas Ağırlığı	-0,00854 (0,004085)*	-0,00813 (0,003849)*	-0,00746 (0,003796)*	-0,00744 (0,004063) ^{ön}	-0,00850 (0,003850)*	
Açlık Süreleri	0 saat	-0,03378 (0,03037) ^{ön}	-0,04625 (0,02862) ^{ön}	0,000556 (0,02823) ^{ön}	-0,03606 (0,03021) ^{ön}	-0,04981 (0,02862) ^{ön}
	12 saat	-0,05558 (0,03088) ^{ön}	-0,04911 (0,02910) ^{ön}	-0,02223 (0,02870) ^{ön}	-0,06609 (0,03072)*	-0,06609 (0,02910)*
	24 saat	-0,1008 (0,03080)**	-0,1261 (0,02902)****	-0,09443 (0,02863)**	-0,09457 (0,03064)**	-0,1234 (0,02903)****
	48 ¹ saat	0,000	0,000	0,000	0,000	0,000

¹: referens parametresi * : p<0.05; ** : p<0.01; *** : p<0.001; **** : p<0.0001; ön: önemsiz

Yapılan bazı çalışmalarda ise eksik gözlemlerin MI yöntemiyle tahmin edilmeden, uygulanan karışık model analizlerinde standart hataların, beklenenden daha küçük olacağı ifade edilmiştir (Mazumdar ve ark., 2007; Kenward ve Carpenter, 2009).

Zhu (2014) tarafından yapılan uzun süreli veri analizinde, MCAR ve MAR yapılarına sahip veri setlerinde eksik gözlemlerin tamamlanması amacıyla liste bazında veri silme (CC-complete case), ortalama atama (MS-mean substitution), en son gözlemi ileriye taşıma (LOCF-last observation carried forward) ve çoklu atama (MI-multiple imputation) yöntemleri kullanılmıştır. Sonuç olarak, MI yönteminin diğer yöntemlere daha güvenilir sonuçlar verdiği saptanmıştır.

Sonuç

Tam veriye uygulanan karışık model sonuçları ve MI veri setlerindeki sonuçlar birbirine oldukça yakın performans göstermiştir. Ayrıca çalışmada, her iki eksik gözlem yapısı ve oranlarında, eksik gözlem atama yöntemi olarak MI'nı tercih edilmesi, karışık modellerde güvenilir sonuçların elde edilmesini sağlamıştır. Bununla beraber, büyük örnek gruplarında eksik gözlem oranının çok fazla olması (örneğin, %50 gibi) durumunda, bu yöntemlerin performanslarının incelenmesi konuyla ilgili bilgi birikimini ve bu çalışmada elde edilen sonuçların daha iyi bir şekilde değerlendirilmesine olanak sağlayacaktır.

Teşekkür

Bu araştırmayı 2014-ZF-B111 numaralı proje ile maddi olarak destekleyen, Yüzüncü Yıl Üniversitesi Bilimsel Araştırma Projeleri Başkanlığı'na teşekkür ederiz. Bununla beraber yazarlar, bu çalışma verilerinin kullanımına izin veren, 2012-ZF-B015'nolu projede yer alan tüm araştırmacılara teşekkürü bir borç bilirler.

Kaynaklar

Aktaş A. 2005. Genelleştirilmiş eşitlik kestirimi ("GEE"). Yüksek lisans tezi. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Programı, Ankara, 50s.

Alpar R. 2003. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1. 2. Baskı, Nobel Yayınları. Ankara, 411s., ISBN: 975-591-431-5.

Bahçecitapar M. 2006. Uzun süreli verilerin analizinde kullanılan istatistiksel modeller. Yüksek lisans tezi. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 145s.

Baygöl A. 2007. Kayıp veri analizinde sıklıkla kullanılan etkin yöntemlerin değerlendirilmesi. Yüksek lisans tezi. İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü, İstanbul, 35s.

Davidian M. 2007. ST732 Applied Longitudinal Data Analysis. <http://www.stat.ncsu.edu/people/davidian/courses/st732/>. (Erişim Tarihi, 12.02.2015).

Erdoğan DB. 2012. Çoklu atama yöntemlerinin Rasch modelleri için performansının benzetim çalışması ile incelenmesi. Doktora tezi. Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara, 115s.

Eyduran E. 2009. Tekrarlanan ölçümlü denemelerin analizinde kullanılan farklı yaklaşımların karşılaştırılması. Doktora tezi. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 131s.

Hedeker D, Gibbson R. 2006. Longitudinal Data Analysis. John Wiley&Sons, Inc.,Hoboken, New Jersey, USA, 337 pp, ISBN: 13 978-0-47 1-42027-9.

Kaps M, Lamberson WR. 2004. Biostatistics for Animal Sciences. CABI Publishing. Cambridge, USA, 445 pp, ISBN: 0 85199 820 8.

Karaca, S., Erdoğan, S., Kor, D., Kor, A. 2015. Farklı rasyonlarla beslenen toklularda kesim öncesi aç bırakma süresinin bazı kan ve rumen parametreleri ile et kalitesi üzerine etkisi. Proje Sonuç Raporu. Yüzüncü Yıl Üniversitesi Bilimsel Araştırma Projeleri Başkanlığı, Proje No: 2012-ZF-B015, Van.

Kenward MG, Carpenter JR. 2009. Chap.21. Multiple imputatin. Longitudinal Data Analysis. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Taylor & Francis Group, CRC Press. New York, pp: 477-499, ISBN: 978-1-58488-658-7.

Kincaid C. 2005. Guidelines for selecting the covariance structure in mixed model analysis. Statistics and Data Analysis, 30: 1-8.

Kristman VK, Manno M, Cote P. 2005. Methods to account for attrition in longitudinal data: Do they work? A simulation study. European Journal of Epidemiology, 20: 657-662.

Liang Y. 2015. Contributions to estimation and testing block covariance structures in multivariate normal models. PhD thesis. Department of Statistics, Stockholm University, Sweden, 54s.

Little JR, Rubin D. 1987. Statistical Analysis with Missing Data. John Wiley & Sons, Inc., New York, 278 pp.

Mazumdar S, Tang G, Houck PR, Dew MA, Begley AE, Scott J, Mulsant BH, Reynolds CF. 2007. Statistical Analysis of longitudinal psychiatric data with dropouts. J. Psychiatr Res., 41: 1032-1041.

Molenberghs G, Kenward MG. 2007. Missing Data in Clinical Studies. John Wiley&Sons, Inc., New York, 504 pp, ISBN:13 978-0-470-84981-1.

SAS 2014. SAS/STAT. Statistical analysis system for Windows. Release 9.4. SAS Institute Inc.

Schafer JL, Graham JW. 2002. Missing data: Our view of the state of the art. Psychol. Methods, 7: 147-177.

Ser G. 2011. Eksik gözlemler uzun süreli (longitudinal) verilerde marjinal ve marjinal olmayan çok seviyeli genelleştirilmiş doğrusal karışık modellerde optimizasyon tekniklerinin karşılaştırılması ve model seçimi. Doktora tezi. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van, 111s.

SPSS 2013. Missing Value Analysis for Windows:Version 22.00, Chicago.

Twisk J, De Boer M, De Vente W, Heymans M. 2013. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J. Clin. Epidemiol., 66: 1022-1028.

Twisk J, De Vente W. 2002. Attrition in longitudinal studies: how to deal with missing data. Journal of Clinical Epidemiology, 55: 329-337.

Twisk JWR. 2004. Longitudinal data analysis. A comparison between generalized estimating equations and random coefficient analysis. European Journal of Epidemiology, 19: 769-776.

Wang Z, Goonewardene LA. 2004. The use of MIXED models in the analysis of animal experiments with repeated measures data. Canadian Journal of Animal Science, 84: 1-11.

Wolfinger RD. 1992. SAS® Technical Support Document 260, A Tutorial on Mixed Models. SAS Institute, Inc. Cary, NC.

Zhu X. 2014. Comparison of four methods for handling missing data in longitudinal data analysis through a simulation study. Open Journal of Statistics, 4: 933-944.