

Akrabalık Matrisinin Tersinin Henderson Metodu İle Hesaplanması ve Hesaplamada Python Programlama Dili Kullanımı

Kemal YAZGAN

Harran Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 63050, Şanlıurfa, Türkiye

*Sorumlu yazar: E-mail: kemalyazgan@gmail.com

Geliş Tarihi (Received): 07.11.2016

Kabul Tarihi (Accepted): 05.01.2017

Bu çalışmanın iki amacı olup, bunlardan birincisi Python programlama diliyle yazılmış ve hayvan ıslahında oldukça önemli bir parametre olan akrabalık matrisinin direk tersinin Henderson metoduyla (Akrabalık ilişkisini göz ardı eden) hesaplanmasının anlatımı ve ikincisi ise bu hesaplamada kullanılabilen Python programlama dili ile yazar tarafından yazılmış bir programın tanıtımıdır. Program Python 2.7.11 sürümü ile yazılmış olup Numpy, Pandas, Xlrd ve XlsWriter modülleri kullanılmıştır. Diğer programlardan farklı olarak ilk üç sütununa hayvan, baba ve anaya ait kulak numaraları yazılmış bir MS Excel dosyası programda direk girdi dosyası olarak kullanılabilir. Programın çalışması tamamlandıktan sonra program tarafından akrabalık matrisinin tersinin kaydedildiği "txt" uzantılı ve NRM isimli bir dosya ve aynı zamanda soy kütüğü ile ilgili çeşitli bilgi ve istatistikleri içeren üç farklı "xlsx" uzantılı çıktı dosyaları oluşturulur. Program <http://www.kemalyazgan.com.tr/indexturk.htm> adresinden edinilebilir. Hayvan ıslahında kullanılan birçok program Fortran programlama dili ile yazılmıştır. Bu çalışma Python programlama dilinin Fortran'a iyi bir alternatif olabileceğini göstermesi açısından ve ülkemizde damızlık değer tahmini yapan yerli programların geliştirilmesi için öncülük edebilir.

Anahtar Kelimeler: Python programlama dili, Akrabalık matrisi, Pedigri analizi, Hayvan ıslahı

Calculating of Inverse Relationship Matrix via Henderson Method and Using Python Programing Language for This Calculation

There were two aims of this study. One of them was about with calculating inverse of relationship matrix, which is one of the most important parameter for animal breeding, by means of Henderson method (ignoring inbreed). That presenting a utilize software for this calculation which was written using Python programing language 2.7.11 version by author was the other aim. Numpy, Pandas, Xlrd and XlsWriter modules were used with Python. As different from other programs, an MS Excel file of which first three column were filled with animal, sire and dam tag number respectively, can be used input file directly. When program terminate, it create called NRM.txt file which is an output file included invers of relationship matrix. Apart from this, it creates three xlsx extension output files and they include some information and statistics for pedigree. Program available from <http://www.kemalyazgan.com.tr/indexturk.htm>. Many programs which interested in animal breeding were written via Fortran Programing language. As it shown with this study, Python can be a good alternative to Fortran. Also, it is pioneering for domestic software developing in Turkey which estimates breeding value.

Key Words: Python programing language, Numerator relationship matrix, Pedigree analysis, Animal breeding

Giriş

Geriye doğru 4-6. generasyonda ortak ebeveyni bulunan hayvanlar akraba olup, bunlar arasındaki yetiştirme tarzına akrabalı yetiştirme denir (Tüzemen ve ark., 2003). Akrabalık ve popülasyonda bundan kaynaklı genetik benzerliğin artması ya da azalması verime doğrudan etki eder.

İki bireyde ortak atadan dolayı özdeş genlerin toplam genlere oranına akrabalık derecesi denir. Anne ve babası akraba olan bireylere akrabalı

yetişmiş veya akrabalı yetiştirilmiş denir. Bir diğer deyişle bir A bireyinin bir lokusunda bulunan iki genin, ortak soydan dolayı özdeş olması olasılığına, A bireyinin akrabalı yetiştirme katsayısı adı verilir (Elandt-Johnson, 1971; Toktamış, 1978). Bununla birlikte damızlık değer tahmininin esası bireyler arasındaki genetik benzerlikten yararlanmadır. Bu ise damızlık değer tahmin edilmesi için bireyler arasındaki genetik benzerliği hesaplamayı gerektirir.

Bireyler arasındaki genetik kovaryansın üç bileşeni vardır. Bunlar; eklemeli varyans, dominantlık varyansı ve epistatik varyanstır. BLUP'in doğruluk derecesi ve yansızlığı çok büyük ölçüde söz konusu bu akrabalık ilişkilerinin hesaplanmasına bağlıdır.

Seleksiyonda isabet derecesinin artırılması için başta kalıtım derecesi olmak üzere diğer genetik parametre tahminlerinin sağlıklı olarak tahminlenmesi oldukça önemlidir. Bireylere ait şansa bağlı etkilerin tahmin edilmesinde kullanılan ve Henderson (1973) tarafından uygulamaya sokulan karışık model eşitliklerinde (Eşitlik 1) akrabalık matrisinin tersi kullanılmaktadır (Mrode ve Thompson, 2007).

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} b \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (1)$$

Burada;

- X : Sabit etkiler matrisini
Z : Bireyleri içeren matrisini
y : Gözlem değerlerini
 σ_e^2 : Hata varyansını
 σ_a^2 : Genotipik varyansı
b : Sabit etkileri
a : Şansa bağlı etkileri
 A^{-1} : Akrabalık matrisinin tersini

ifade etmektedir.

Python, Guido Van Rossum adlı Hollandalı bir programcı tarafından yazılmış dünyada popülerliği giderek artmakta olan, açık kaynak kodlu ve ücretsiz bir programlama dilidir. Geliştirilmesine 1990 yılında başlanmıştır. Mühendisliğin birçok alanında etkin bir şekilde kullanılmakta olup www.python.org adresinden edinilebilir.

Bu çalışma da hayvan ıslahında damızlık değer tahmininde çok önemli bir ögeyi oluşturan akrabalık matrisinin tersinin Henderson metodu ile hesaplanmasının nasıl yapıldığının anlatımı ve bu hesaplamada Python programlama dili ile yazılmış bir programın tanıtımı amaçlanmıştır

Materyal ve Yöntem

Akrabalık matrisinin direk tersini veren Henderson metodunun uygulanışı

Akrabalık matrisinin tersi Henderson metodu adı da verilen bir yöntemle hesaplanabilmektedir. Yöntemde bireylerin birbirleri ile olan akrabalık dereceleri hesaba katılmamakta sadece bireylerin

ebeveynlerinin bilinip bilinmemesi değerlendirilmektedir.

Burada Çizelge 1 deki değerler gözetilerek ve aşağıdaki işlem basamakları ve kurallar izlenerek hesaplanmanın nasıl yapıldığı açıklanmıştır¹ (Mrode ve Thompson, 2007). Buna göre işlem basamakları aşağıda verildiği gibidir;

i. hayvanın anası ve babası biliniyor ise matristeki;

- (i, i) koordinatlarında bulunan elemana (köşegen üzerindeki) α_i değeri eklenir.
- (baba², i), (i, baba³), (ana⁴, i) ve (i, ana⁵) koordinatlarındaki elemanlara $-\alpha_i/2$ değeri eklenir
- (baba, baba), (baba, ana), (ana, baba) ve (ana, ana) koordinatlarındaki elemanlara $\alpha_i/4$ değeri eklenir.

i. hayvanın anasından veya babasından sadece birisi biliniyor ise matristeki;

- (i, i) koordinatlarında bulunan elemana (köşegen üzerindeki) α_i değeri eklenir
- Sadece babası biliniyor ise (baba, i), (i, baba) koordinatlarındaki elemanlara veya sadece anası biliniyor ise (ana, i) ve (i, ana) koordinatlarındaki elemanlara $-\alpha_i/2$ değeri eklenir.
- Sadece babası biliniyor ise (baba, baba) koordinatlarındaki elemana veya sadece anası biliniyor ise (ana, ana) koordinatlarındaki elemana $\alpha_i/4$ değeri eklenir.

i. hayvanın anasının ve babasının hiçbiri bilinmiyor ise matristeki;

- (i, i) koordinatlarında bulunan elemana (köşegen üzerindeki) α_i değeri eklenir

¹ Burada verilen akrabalık matrisinin direk hesaplama ilişkin metot Mrode ve Thompson (2007) tarafından yayınlanan "Linear models for the Prediction of Animal Breeding Values, Second Edition" isimli kitabın 29 ve 30. sayfalarındaki örnekten Türkçeye çevrilerek aktarılmıştır.

² i. hayvanın babasının bulunduğu satırı belirtir

³ i. hayvanın babasının bulunduğu sütunu belirtir

⁴ i. hayvanın anasının bulunduğu satırı belirtir

⁵ i. hayvanın anasının bulunduğu sütunu belirtir

Çizelge 1. Altı adet buzağıya ait ana ve baba bilgileri* (Mrode ve Thompson, 2007).
Table 1. Pedigree records for six calves (Mrode and Thompson, 2007).

Buzağı No	Baba	Ana
1	0	0
2	0	0
3	1	2
4	1	0
5	4	3
6	5	2

*: Bilinmeyen hayvanlar 0 değeri ile gösterilmektedir.

Öncelikle, Çizelge 1’de verilen örnekte hayvan sayısı 6 olduğu için 6x6’lık bir matris oluşturulacaktır. Matrisin diyagonal elementleri bireyin her iki ebeveyni de biliniyor ise 2, biri biliniyor ise 4/3 ve hiçbir ebeveyn bilinmiyor ise 1’e eşitlenir (Çizelge 1). Bu değer α olarak nitelendirilir (Diğer bir deyişle α sırasıyla 2, 4/3 ve 1 değerini almaktadır).

1 ve 2 numaralı hayvanların ebeveynleri bilinmemektedir. Bu yüzden $\alpha_1 = \alpha_2 = 1$ olup köşegenlerine 1 yazılır. 3. hayvan için her iki ebeveynde bilinmektedir. Bu yüzden $\alpha_3 = 2$ ’dir ve köşegene yazılır. (3,1)⁶, (1,3), (3,2) ve (2,3)’e -1 değeri, (1,1), (1,2), (2,1) ve (2,2) no’lu elemanlara ise $\frac{1}{2}$ değeri eklenir. 4. hayvan için sadece tek ebeveyn bilindiğinden $\alpha_4 = \frac{4}{3}$ dür. Bundan hareketle (4,4) elemanına $\frac{4}{3}$ değeri, (1,1) elemanına $\frac{1}{3}$ değeri ve (4,1) ve (1,4) elemanına da $-\frac{2}{3}$ değeri eklenir. Sonuç olarak ilk dört hayvandan sonraki Çizelge Çizelge 3 de verilen şekilde gerçekleşir (Mrode ve Thompson, 2007).

Tüm işlemler bittikten ve değerler yerine yazıldıktan sonra matrisin son hali Çizelge 4 de verilmiştir. Bu Çizelge 1’de verilen soy kütüğü ile oluşturulmuş akrabalık matrisinin direk tersidir. Direk tersini alma işlemi neticesinde oluşturulan matrisin tekrar tersini almaya gerek olmayacağından bu yöntem çok yüksek sayıda hayvan içeren soy kütükleri kullanıldığında hesaplamalarda kolaylık sağlar.

Python programlama dilinin akrabalık matrisinin tersinin hesaplanmasında kullanımı⁷

Python programlama dili kendi internet sayfasından (www.python.org) indirilerek 2,1’de anlatılan Henderson metodunun işleyişine uygun bir algoritma yazılmıştır. Sonraki aşamada program “cx_Freeze-4.3.3” yazılımı (<https://sourceforge.net/projects/cx-freeze/>) kullanılarak derlenmiş ve exe uzantılı hale getirilmiştir. Programda ayrıca kullanılan modüller ve işlevleri Çizelge 5’de verilmiştir.

Programın girdi dosyasının hazırlanması

Programda girdi dosyası diğer bir deyişle hayvan baba ve ana kulak numaralarını içeren soy kütüğü dosyası direk MS Excel biçiminde hazırlanabilir ve hazırlanan bu dosya başka hiçbir formata veya dosyaya dönüştürülmeden direk program tarafından kullanılabilir. Bu da kullanıcıya kolaylık sağlayan bir durumdur. Ancak bu dosyanın ismi sadece veri.xlsx olmalıdır (Şekil 1). Bundan farklı isimlerde program çalışmayacaktır. Sayfa ismi standart MS Excel tarafından verilen “Sayfa1” olmalıdır. Dosyada ilk satır başlık olarak kullanılabilir. Eğer başlık olarak kullanılmayacaksa ilk satır boş bırakılmalıdır. Ayrıca hayvan numaraları ana ve baba numaralarından büyük olmalı kulak numarası bilinmeyen hayvanlar “0” olarak kodlanmalıdır. Aksi halde program hata verir (Şekil 1).

⁶ Parantez içindeki değerlerden ilki Çizelgeler de verilen matrislerdeki satır numarasını, ikincisi ise sütun numarasını temsil etmektedir.

⁷ Bu kısımda adı geçen tüm ticari ve ticari olmayan yazılımlardan yazar hiçbir suretle çıkar sağlamamaktadır. Hiçbir suretle reklam amaçlanmamaktadır. Sadece bu çalışmada kullanımı gerektiği ve bu çalışma için alternatiflerinden istenilen verim alınmadığından bu yazılımlar kullanılmıştır.

Çizelge 2. Köşegen elemanları belirlenmiş matris.

Table 2. Example matrix of which diagonal elements were calculated

Buzağı no	1	2	3	4	5	6
1	1					
2		1				
3			2			
4				4/3		
5					2	
6						2

Çizelge 3. İlk dört hayvan için yapılan hesaplamalar

Table 3. Calculations for first-four animals

Buzağı no	1	2	3	4	5	6
1	$1+(1/2)+(1/3)$	1/2	-1	-2/3		
2	1/2	$1+1/2$	-1			
3	-1	-1	2			
4	-2/3			4/3		
5					2	
6						2

Programın edinilmesi ve çalıştırılması

Yazılım

<http://www.kemalyazgan.com.tr/indexturk.htm>

adresinden "Akrabalik.exe" linki tıklanarak indirilebilir. İndirilen "rar" uzantılı dosya bilgisayarın C diskine çıkarılmalıdır. Burada bulunan PROG_AKR klasörü içindeki "Akrabalik.exe" dosyasına tıklamak programın çalıştırılması için yeterlidir. Program aynı klasör

içerisine kaydedilmiş ve nasıl hazırlanacağı yukarıda anlatılan veri.xlsx isimli dosyayı okur ve aşağıda isimleri ve içerikleri açıklanan dört farklı çıktı dosyasını aynı klasör içinde oluşturarak sonlanır. Bu klasör başka yere taşınmamalıdır. Sadece C diskinde bulunmalıdır. Ayrıca klasör içerisindeki exe uzantılı dosya da bu klasör dışında hiçbir yerde çalışmamaktadır. Programın bilgisayarda bulunduğu yer "C:\PROG_AKR\Akrabalik.exe" adresindeki gibi olmalıdır.

Çizelge 4. Akrabalık matrisinin tersi

Table 4. Invers of relationship matrix

Buzağı no	1	2	3	4	5	6
1	1.83	0.5	-1	-0.6666	0	0
2	0.5	2	-1	0	0.5	-1
3	-1	-1	2.5	0.5	-1	0
4	-0.6666	0	0.5	1.83333	-1	0
5	0	0.5	-1	-1	2.5	-1
6	0	-1	0	0	-1	2

Çizelge 5. Python programlama dili için hazırlanmış ve bu çalışmadaki yazılımda da kullanılan modüller
Table 5. Modules prepared for Python programming language and used in this software

Modül/Kütüphane adı	Versiyon	Açıklama	Kaynak
Numpy	1.11.0	Çeşitli sayısal işlemleri gerçekleştirmek için kullanılan kütüphane	http://www.numpy.org
Pandas	0.18.0	Veri düzenleme analiz ve istatistik kütüphanesi	http://pandas.pydata.org
Xlrd	0.9.4	MS Excel sayfalarından veri yazmak için kullanılan modül	http://www.python-excel.org/
XlsxWriter	0.8.6	MS Excel sayfalarına veri yazmak için kullanılan modül	https://github.com/jmcnamara/XlsxWriter

	A	B	C	D
1	hayvan	baba	ana	
2	1145	425	825	
3	1178	425	826	
4	1198	0	0	
5	1223	536	828	
6	1221	536	829	
7	1245	578	0	
8	1246	598	831	
9	1248	601	832	
10	1250	0	833	
11	1254	601	834	
12	1257	601	246	
13	1258	601	836	
14	1259	259	837	
15	1260	601	178	
16	1265	601	839	
17	1280	0	0	
18	1298	536	841	
19	1299	645	842	
20	1301	259	0	
21				

Şekil 1. "veri" isimli girdi dosyası
Figure 1. input file called "veri"

	A	B	C	D
1	1_Yeni_Hayvan_No	2_Yeni_Baba_No	3_Yeni_Ana_No	
2	23	4	10	
3	24	4	11	
4	25	0	0	
5	26	5	13	
6	27	5	12	
7	28	6	0	
8	29	7	14	
9	30	8	15	
10	31	0	16	
11	32	8	17	
12	33	8	2	
13	34	8	18	
14	35	3	19	
15	36	8	1	
16	37	8	20	
17	38	0	0	

Şekil 3. Soydef_yeni dosyası
Figure 3. Soydef_yeni file

	A	B	C	D
1	Eski_no	Yeni_no		
2	0	0		
3	178	1		
4	246	2		
5	259	3		
6	425	4		
7	536	5		
8	578	6		
9	598	7		
10	601	8		
11	645	9		
12	825	10		
13	826	11		
14	828	12		
15	829	13		
16	831	14		

Şekil 2. Soydef_num dosyası
Figure 2. Soydef_num file

	A	B	C
1	Soykutugunde verim kaydi bulunan birey sayisi:	19	
2	Soykutugunde bulunan baba sayisi:	7	
3	Soykutugunde bulunan ana sayisi:	15	
4	Verim kaydi bulunan fakat hem babasi hemde anasi bilinmeyen birey sayisi:	2	
5	Verim kaydi bulunan fakat sadece babasi bilinmeyen birey sayisi:	1	
6	Verim kaydi bulunan fakat sadece anasi bilinmeyen birey sayisi:	2	
7	Verim kaydi bulunan ve hem anasi hemde babasi bilinen birey sayisi:	14	
8	Soykutugunde bulunan toplam birey sayisi:	41	
9			

Şekil 4. Soydef_ist dosyası
Figure 4. Soydef_ist file

Çıktı Dosyaları

Soydef_num.xlsx: Soy defterinde (Soy kütüğünde) bulunan tüm hayvanların orijinal ve programın vermiş olduğu yeni numaraları içerir (Şekil 2).

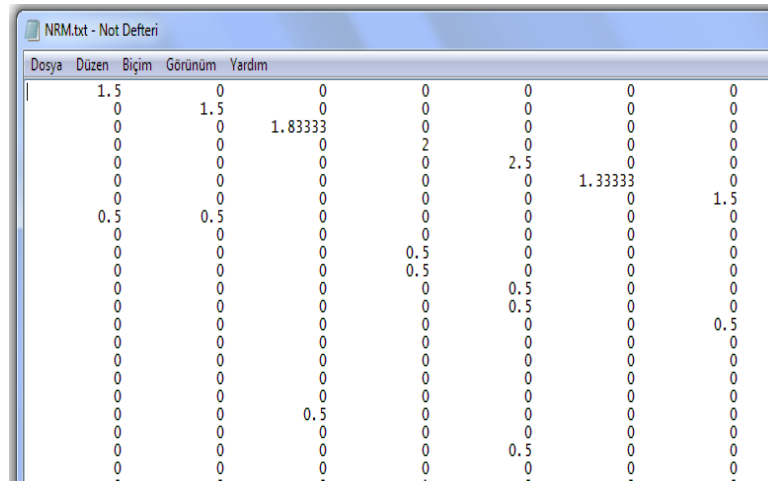
Soydef_yeni.xlsx: Yeniden numaralandırılmış hayvanlarla oluşturulan yeni soy defterini içerir (Şekil 3).

Soydef_ist.xlsx: Soy defterine ilişkin detaylı istatistiksel bilgi içerir (Şekil 4).

NRM.txt: Program tarafından hazırlanan akrabalık matrisinin tersinin yazıldığı metin dosyasıdır (Şekil 5).

Bulgular ve Tartışma

Daha öncede belirtildiği gibi direk ters alma işlemi oluşturulan matrisin tekrar tersini almaya gerek olmayacağından çok yüksek sayıda hayvan içeren soy kütükleri kullanıldığında bu yöntem hesaplamalarda kolaylık sağlar. Bununla birlikte bu çalışmada popülerliği giderek artan Python programlama dili ile hayvancılıkta ıslah çalışmaları için gerekli olan bir takım yazılımların yazılabileceği gösterilmiştir. Bununla birlikte farklı programlama dilleri ile yazılmış ve aynı hesaplamayı yapan programlar olmasına karşın, bu programlar incelendiğinde bu çalışmada tanıtılan programın kullanımının direk xlsx uzantılı dosyaları girdi dosyası olarak kullanıp ve çıktı dosyalarını da xlsx uzantılı olarak verebilmesi kullanımının daha kolay olduğunu göstermektedir..



The screenshot shows a window titled "NRM.txt - Not Defteri" with a menu bar containing "Dosya", "Düzen", "Biçim", "Görünüm", and "Yardım". The main area displays a matrix of numerical values. The values are arranged in a grid, with some cells containing non-zero values like 1.5, 1.83333, 2, 2.5, 1.33333, 0.5, and 1.5, while most cells contain 0.

Dosya	Düzen	Biçim	Görünüm	Yardım
1.5	0	0	0	0
0	1.5	0	0	0
0	0	1.83333	0	0
0	0	0	2	0
0	0	0	0	2.5
0	0	0	0	0
0	0	0	0	1.33333
0	0	0	0	0
0.5	0.5	0	0	1.5
0	0	0	0	0
0	0	0	0.5	0
0	0	0	0.5	0
0	0	0	0	0.5
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0.5	0	0
0	0	0	0	0
0	0	0	0	0.5
0	0	0	0	0
0	0	0	0	0

Şekil 5. NRM.txt dosyasının bir parçası
Figure 5. A detail from NRM.txt file

Sonuç

Python dilinin nesne tabanlı, kullanımı kolay ve ücretsiz bir programlama dili olması, diğer programlama dillerinde olduğu gibi noktalı virgül, süslü parantez gibi ayrıntıların olmaması, Windows, Linux, Mac, Unix gibi tüm işletim sistemi üzerinde çalışabilir olması, çok geniş bir kütüphaneye sahip olması ve ayrı bir derleyici programa ihtiyaç duymaması gibi özellikleri sayesinde bu programlama dili hayvan yetiştirme ve ıslah alanında gerekli programların yazımında da Fortran programlama diline iyi bir alternatif olabilir.

Bu çalışma da tanıtımı yapılan yazılımın kullanım dilinin Türkçe olması girdi ve çıktı dosyalarının Excel formatında olması kullanıcıya kolaylık sağlamaktadır. Ülkemizde bu alanda yazılım geliştirme yok denecek kadar azdır. Bu ve bu gibi kolay kullanılabilir pratik programların hazırlanıp kullanımlarının yaygınlaştırılması bunların sahada da kullanım olanaklarını artıracak ve daha isabetli seleksiyon çalışmaları yapılmasını sağlayacaktır. Bundan dolayı bu çalışma bu alanda öncü bir nitelik taşıyabilir.

Kaynaklar

Tüzemen, N., M. Yanar,, Ö. Akbulut,. 2003. Hayvan ıslahı. Atatürk Üniversitesi Ziraat Fakültesi Ders Yayınları No: 230, Atatürk Üniversitesi Ziraat Fakültesi Ofset Tesisi, Erzurum, s.234-250.

Elandt-Johnson, R.C, 1971. Probability models and statistics. John Wiley, Newyork.

Toktamış, Ö, 1978. Akrabalık ve Akrabalı Yetiştirme Katsayılarını Hesaplama ve Tahmin Yöntemlerinin incelenmesi, Yayınlanmamış Doktora Tezi, Ankara.

Henderson, C.R, 1973. Sire evaluation and genetic trends. Proc. Anim. Breed. Genet. Sym. Honor of Lush, ASA and ADSA, Champaign, Il.

Mrode, R., R. Thompson, 2007. Linear models for the prediction of animal breeding values. Second Edition. CAMBRIDGE, MA 02139 USA.

<http://www.python.org> (Erişim Tarihi: 01.02.2016).

<https://sourceforge.net/projects/cx-freeze/> (Erişim Tarihi: 01.02.2016).

<http://www.numpy.org> (Erişim Tarihi: 01.02.2016).

<http://pandas.pydata.org> (Erişim Tarihi: 21.03.2016).

<http://www.python-excel.org/> (Erişim Tarihi: 21.01.2016).

<https://github.com/jmcnamara/XlsxWriter> (Erişim Tarihi: 01.02.2016).