

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

**Mustafa Ersen BEBEK**

**ÇATALBURUN KÖPEKLERİNDE LOJİSTİK REGRESYON  
ANALİZİ İLE SAFLIK TAYİNİ**

**ZOOTEKNİ ANABİLİM DALI**

**ADANA-2021**

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**ÇATALBURUN KÖPEKLERİNDE LOJİSTİK REGRESYON ANALİZİ  
İLE SAFLIK TAYİNİ**

**Mustafa Ersen BEBEK**

**YÜKSEK LİSANS TEZİ**

**ZOOTEKNİ ANABİLİM DALI**

Bu Tez 16/12/2021 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/  
Oyçokluğu ile Kabul Edilmiştir.

.....  
Prof. Dr. Zeynel CEBECİ  
DANIŞMAN

Prof. Dr. G. Tamer KAYAALP  
ÜYE

Prof. Dr. Ercan EFE  
ÜYE

Bu Tez Enstitümüz Zootekni Anabilim Dalında hazırlanmıştır.

**Kod No:**

**Prof. Dr. Sadık DİNÇER**  
**Enstitü Müdürü**

**Bu Çalışma Ç.Ü. Bilimsel Araştırma Projeleri Birimi Tarafından  
Desteklenmiştir.**

**Proje No: FLY-2018-10284**

**Not:** Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ

YÜKSEK LİSANS TEZİ

ÇATALBURUN KÖPEKLERİNDE LOJİSTİK REGRESYON ANALİZİ  
İLE SAFLIK TAYİNİ

Mustafa Ersen BEBEK

ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
ZOOTEKNİ ANABİLİM DALI

Danışman : Prof. Dr. Zeynel CEBECİ  
Yıl: 2021, Sayfa: 72  
Jüri : Prof. Dr. G. Tamer KAYAALP  
: Prof. Dr. Ercan EFE

Bu tez çalışmasında yaygın olarak Mersin İli Tarsus ilçesinde yetiştirilen yerli bir av köpeği ırkı olan Çatalburun köpeklerinin çeşitli kantitatif özelliklerinden lojistik regresyon ile saflık tayini yapılması amaçlanmıştır. Çalışmada Mersin Büyükşehir Belediyesi'ne bağlı Çatalburun Üretim ve Rehabilitasyon Merkezi'nde koruma altına alınan 70 adet av köpeğine ait 18 farklı morfolojik özellik ile uzman veteriner hekim tarafından yapılan saflık sınıfı verisi kullanılmıştır. Tek ve çok değişkenli modellerle yapılan analiz sonuçlarına göre bağımsız değişkenlerin tümü alınarak oluşturulan tam modelin AUC ölçütünün 0.98 ile mükemmel olduğu ve ayrıca ROC eğrisinin de ideale çok yakın bir davranış gösterdiği saptanmıştır. Tam modelin sınıflama doğruluğu yaklaşık %92.9 düzeyinde saptanmıştır. Tam model 0.972 AUC değeri ve %88.6 doğruluk değeriyle burun çapı, burun bir kanat uzunluğu, iki şakak arası genişlik, iki kulak arası uzunluk, göğüs çevresi, arka pati bilek kalınlığı ve kulak genişliği değişkenlerinden oluşan indirgenmiş modelin izlediği görülmüştür. AUC ve doğruluk dereceleri itibariyle tam model en iyi başarıyı göstermesine karşık indirgenmiş modelin daha düşük AIC ve BIC değerleri vermesi nedeniyle Çatalburun köpek ırkının saflık tayininde kullanılabileceği sonucuna varılmıştır.

**Anahtar Kelimeler:** Lojistik regresyon, Saflık tayini, Köpek ırkları, Çatalburun, Gen kaynaklarının korunması



## ABSTRACT

### MSc THESIS

# PURITY CLASSIFICATION OF THE ÇATALBURUN DOG BREED WTH LOGISTIC REGRESSION ANALYSIS

**Mustafa Ersen BEBEK**

**ÇUKUROVA UNIVERSITY  
INSTITUTE OF NATURAL AND APPLIED SCIENCES  
DEPARTMENT OF ANIMAL SCIENCE**

Supervisor : Prof. Dr. Zeynel CEBECİ  
Year: 2021, Pages: 72

Jury : Prof. Dr. G. Tamer KAYAALP  
: Prof. Dr. Ercan EFE

This thesis aimed to determine the purity of Çatalburun dogs, a local hunting dog breed reared in Tarsus district of Mersin province, by using logistic regression models formed with various quantitative traits. In the study, 18 different morphological traits of 70 hunting dogs kept in Çatalburun Production and Rehabilitation Center affiliated to Mersin Metropolitan Municipality, and purity classification data scored by a specialist veterinarian were used. According to the results obtained with the examined univariate and multivariate models, it was determined that the performance of full model including all the independent variables was perfect with a very high AUC value of 0.98, and the ROC curve this model also showed a behavior very close to the ideal situation. The classification accuracy of the full model was also found to be approximately 92.9%. The full model was followed by the reduced model consisting of the variables of nose diameter, nose one wing length, bi-temporal width, bi-auricular length, chest circumference, hind paw wrist thickness, and ear width with an AUC value of 0.972 and an accuracy of 88.6%. Although the full model showed the best performance in terms of AUC and accuracy, it was concluded that the reduced model is also recommended for the purity determination of the Çatalburun dog breed, since it also gave lower AIC and BIC values in this study.

**Keywords:** Logistic regression, Purity classification, Dog breeds, Çatalburun, Conservation of biodiversity.



## GENİŞLETİLMİŞ ÖZET

Bu tez çalışmasında Çatalburun ırkı köpeklerin 18 kantitatif özelliği kullanılarak lojistik regresyon analizi ile saflık tayini yapılması amaçlanmıştır. Araştırma materyali olarak Mersin Büyükşehir Belediyesi'nin Tarsus Çatalburun Üretim ve Rehabilitasyon Merkezi'nde koruma altına alınan ve yaşları 2-5 arasında değişen 29 dişi ve 41 erkek olmak üzere toplam 70 Çatalburun av köpeği kullanılmıştır.

Veri analizleri R istatistiksel hesaplama ve grafik analiz ortamında (R Core Team, 2021) yapılmıştır. Lojistik regresyon analizleri R'nin stats paketindeki glm fonksiyonu ile çalışılmıştır. Analizlerin yapılmasında R'nin temel paketlerindeki çeşitli fonksiyonlar hesaplama ve görselleştirme fonksiyonları kullanılmıştır. Lojistik regresyon analizleri stats paketindeki glm fonksiyonu ile yapılmıştır. Modellere girecek bağımsız değişkenlerin seçiminde bilgi değerlerini hesaplamak ve değişken önem değerlerini hesaplamak üzere caret paketindeki (Kuhn, 2020) varImp fonksiyonu kullanılmıştır. Modellerin tahmin başarımlarını AUC ölçütüne göre değerlendirmede ve ROC eğrilerinin çiziminde ROCR paketi (Sing ve ark., 2005), diğer görseller için corplot paketi (Wei ve Simko, 2017) ile çalışılmıştır.

Model başarımlarını ölçütleri olarak AUC, AIC, BIC, ACC, McFadden  $R^2$ , Cox-Snell  $R^2$  ve Nagelkerke  $R^2$  gibi ölçütler kullanılmıştır. Model başarımlarını karşılaştırmak için karışıklık matrisleri oluşturularak doğruluk değerleri hesaplanmıştır. Sahte  $R^2$  değerlerini hesaplamak için DescTools paketinden (Signorell ve ark., 2021) PseudoR2 fonksiyonu, pscl paketinden (Jackman, 2020) pR2 fonksiyonu, modEVA paketinden (Barbosa ve ark., 2015) RsqGLM fonksiyonu, ModelMetrics paketinden auc fonksiyonu (Hunt, 2020) yararlanılmıştır.

Çalışmada her bir bağımsız değişken tek tek kullanılarak tek değişkenli lojistik regresyon modelleri yanında değişken seçimine dayalı olarak seçilen değişkenlerle oluşturulan çok değişkenli modellerin başarımları tespit edilmiştir.

Tek ve çok deęişkenli modellerle yapılan analiz sonuçlarına göre bağımsız deęişkenlerin tümü alınarak oluşturulan tam modelin AUC ölçütünün 0.98 ile mükemmel olduęu ve ayrıca ROC eğrisinin de ideale çok yakın bir davranış gösterdiği saptanmıştır. Tam modelin sınıflama doğruluęu yaklaşık %92.9 düzeyinde saptanmıştır. Tam model 0.972 AUC deęeri ve %88.6 doğruluk deęeriyle burun çapı, burun bir kanat uzunluęu, iki şakak arası genişlik, iki kulak arası uzunluk, göęüs çevresi, arka pati bilek kalınlığı ve kulak genişliği deęişkenlerinden oluşan indirgenmiş modelin izledięi görülmüştür.

AUC ve doğruluk dereceleri itibariyle tam model en iyi başarıyı göstermesine karşık indirgenmiş modelin daha düşük AIC ve BIC deęerleri vermesi nedeniyle Çatalburun köpek ırkının saflık tayininde kullanılabileceęi sonucuna varılmıştır.

## TEŐEKKÜR

Yüksek lisans tezi olarak sunduđum bu alıřmada, lisansüstü eđitimim ve tez alıřmam boyunca sınırsız desteđini sunan, tezimin bu ařamaya gelmesinde büyük emeđi bulunan deđerli danıřmanım Prof. Dr. Zeynel CEBECİ'ye, deđerli önerileriyle katkıda bulunan Prof. Dr. G. Tamer KAYAALP'a, ve Prof. Dr. Ercan EFE hocalarıma en derin saygılarımı ve řükranlarımı sunuyorum.

Tez alıřmam boyunca enerjisi, ileri görüşlülüđü, akılcı fikirleri ile bana manevi destek veren hayat arkadařım, biricik can yoldařım eřime, hayata geldiđi andan itibaren benim huzur durađım olan ođluma teőekkürü bir bor bilirim. Mesleđini bir nesil daha ileri tařıma řerefine nail olduđum babama, manevi desteđini esirgemeyen anneme ve sevgili meslektařım Do. Dr. Sefa Küçükler'e ve teőekkür etmenin mutluluđu ve onurunu yařamaktayım.









**TABLULAR DİZİNİ****SAYFA**

Tablo 1. Köpeklerin yaş ve cinsiyete göre dağılımı.....	10
Tablo 2. Bağımlı ve Bağımsız değişkenlerin çizelgesi .....	12
Tablo 3. Karışıklık matrisi yapısı.....	23
Tablo 4. AUC değerlerinin yorumlanması.....	26
Tablo 5. Tek değişkenli modeller için regresyon katsayıları .....	34
Tablo 6. Tek değişkenli modellerin analiz sonuçları .....	36
Tablo 7. Tam modele ait analiz sonuçları .....	37
Tablo 8. Tam modele ait başarımlar ölçütleri .....	38
Tablo 9. Bağımsız değişkenlerin önem değerleri ve sıraları.....	40
Tablo 10. Model S1 analizi sonuçları .....	43
Tablo 11. Model S2 analizi sonuçları .....	44
Tablo 12. Model S3 analizi sonuçları .....	44
Tablo 13. Model S4 analizi sonuçları .....	45
Tablo 14. Model S4a analizi sonuçları.....	45
Tablo 15. Model başarımlarının karşılaştırılması .....	45



## ŞEKİLLER DİZİNİ

## SAYFA

Şekil 1: Çatalburun Üretim ve Rehabilitasyon Merkezi (Mersin) .....	9
Şekil 2: Vücut kitle indeksi ölçme şeridi (sol) ve dijital kumpas (sağ).....	11
Şekil 3: Köpeklerin kayıt altına alınması ve saflık değerlendirmesi.....	11
Şekil 4: Ölçme işlemlerinden bazı görüntüler.....	12
Şekil 5: Lojistik regresyon eğrisi .....	18
Şekil 6: ROC eğrisi örneği.....	25
Şekil 7: Bağımsız değişkenler arasındaki korelasyonlar.....	32
Şekil 8: Tam modele ait ROC eğrisi .....	38
Şekil 9: Değişkenlerin önemleri çubuk grafiği .....	41
Şekil 10: Karşılaştırılan modellere ait ROC eğrileri grafiği .....	47



## KISALTMALAR

<b>ACC</b>	: Accuracy-Doğruluk
<b>AIC</b>	: Akaike Information Criterion-Akaike Ölçütü
<b>AUC</b>	: Area Under Curve-ROC Eğrisi Altında Kalan Alan
<b>BD</b>	: Bağımsız Değişken
<b>BIC</b>	: Bayesian Information Criterion-Bayesçi Bilgi Ölçütü
<b>BK</b>	: Basıklık Katsayısı
<b>CE</b>	: Classification error-Sınıflama hatası
<b>CSR2</b>	: Cox ve Snell sahte $R^2$
<b>CUR2</b>	: Cragg ve Uhler / Nagelkerke sahte $R^2$
<b>ÇK</b>	: Çarpıklık Katsayısı
<b>FN</b>	: False Positive-Yanlış Pozitif
<b>FNR</b>	: False Negative Rate-Yanlış Negatif Oranı
<b>FP</b>	: False Negative-Yanlış Negatif
<b>FPR</b>	: False Positive Rate-Yanlış Pozitif Oranı
<b>GLM</b>	: Generalized Linear Model-Genelleştirilmiş Doğrusal Modeller
<b>MAKS</b>	: Maksimum Değer
<b>MFR2</b>	: McFadden sahte $R^2$
<b>MİN</b>	: Minimum Değer
<b>ORT</b>	: Ortalama
<b>ROC</b>	: Receiver Operating Characteristic-Alıcı İletişim Karakteristiği
<b>SH</b>	: Standart Hata
<b>SS</b>	: Standart Sapma
<b>TN</b>	: True Positive-Doğru Negatif
<b>TNR</b>	: True Negative Rate-Doğru Negatif Oranı
<b>TP</b>	: True Positive-Doğru Pozitif
<b>TPR</b>	: True Positive Rate-Doğru Pozitif Oranı

**VIF** : Variance Inflation Factor-Varyans Şişme Faktörü

**VIMP** : Variable Importance-Önem Derecesi



## 1.GİRİŞ

Çatalburun, Mersin ili Tarsus ilçesi kökenli yerli av köpek ırklarından biridir. Zeki, avına odaklanan, avını takip etme ve avının nereye düşeceğini önceden kestirme yeteneğine sahip olmalarıyla karakterize olan bir köpek ırkıdır. Adını burun yapısından alan bu köpek ırkını diğer av köpeklerinden farklı kılan özelliği burun iç hacminin geniş olmasından kaynaklanmaktadır. Koku alanı olarak da bilinen 'olfactory area' koku alma organının bulunduğu nazal mukoza alanının daha geniş olması anlamına gelmektedir (Dinçer, 2006). Mukusun koku moleküllerinin mukusla daha fazla birleşmesiyle daha iyi koku almayı sağlayarak köpeklerin avını daha kolay bulmalarında yardımcı olmaktadır. Avcıların bu özelliğini bilmeleri Çatalburun av köpeklerinin aranan ırklar arasında yer almasını sağlamıştır. Mersin Büyükşehir Belediyesi bu ırkın sahip olduğu özellikleri basın yayın birimi ve sosyal medya aracılığıyla kamu bu ırkı tanımaya başlamıştır. Güvenlik birimlerinden Jandarma Genel Komutanlığı'na bağlı At-Köpek Eğitimi ve Meslek Edindirme Kurs Merkezi Komutanlığı (JAKEM) tarafından da Çatalburun köpeklerine iz takip, devriye, mayın ve bomba arama, yangın tespit, arama ve kurtarma, narkotik ve psikotrop (bağımlılık yapıcı kimyasal) madde arama, kağıt para arama, keşif ve özel operasyon konularında eğitim verilmektedir.<sup>1</sup> JAKEM in dikkatini çeken Çatalburun av köpeklerinde uzman ekipler, yavru alarak merkezlerinde eğitilmeleri için çalışmışlardır. Koku duyularından yararlanmak üzere patlayıcı madde ve uyuşturucu eğitimi gibi 11 farklı dalda eğitim verilen Çatalburun köpekleri yeterli faydayı sağlamaktadır.<sup>2</sup> Çatalburun köpeği geçmiş yıllarda İngiliz Setter ve German Shorthaired Pointer'i gibi farklı ırklarla melezlenmiştir. Melezlemeler nedeniyle saf

<sup>1</sup> [ensonhaber.com/galeri/maharetli-kopekler-jakemde-yetisiyor](http://ensonhaber.com/galeri/maharetli-kopekler-jakemde-yetisiyor)

<sup>2</sup> [Aa.com.tr/tr/turkiye/yerli-irk-kopekler-nevsehirdeki-jakemde-egitiliyor-/1671530](http://Aa.com.tr/tr/turkiye/yerli-irk-kopekler-nevsehirdeki-jakemde-egitiliyor-/1671530)

köpeklerin sayısı hızla azalmıştır. Ayrıca saflık tayini zor bir ırk haline gelmiştir. Morfolojik olarak farklılık gösteren melez hayvanlar nedeniyle saflıktan gittikçe uzaklaşma olmuştur (Dinçer, 2006).

Günümüzde, kayıtları tutulmamış ve standartları belirlenmemiş bu köpeklerin, sayısının 200 civarında olduğu tahmin edilmektedir. Safılaştırma çalışmaları için bu köpekler Tarsus ilçesi ve civar il ve ilçelerden toplanarak Mersin Büyükşehir Belediyesi'ne bağlı Çatalburun Üretim ve Rehabilitasyon Merkezi'nde yetiştirilmeye başlanmış ve pedigrisi (soy kütüğü) kayıtları tutulmaya başlanmıştır. Av köpeklerinde saflık tayininin pedigrisi kayıtlarından yapılması gerekli olsa da pratikte kaydı tutulmamış hayvanlar için başka yöntemlere ihtiyaç duyulmaktadır. Ayrıca sahadan toplanan köpeklerin saf olup olmadığına karar vermek için bazı yardımcı yöntemler ve tekniklere de ihtiyaç duyulduğu anlaşılmaktadır.

Alternatif bir yöntem olarak başta renk olmak üzere çeşitli kalitatif karakterler itibariyle saflık tayini yapıldığı görülmektedir. Ancak saflık tayininin sadece kalitatif karakterler itibariyle yapılması saflık tayini için yeterli olamamaktadır. Son yıllarda hayvanların genotipik yapısının gelişen moleküler tekniklerle ortaya konması mümkün olabilmektedir (Gürses ve Bayraktar, 2014). Ancak moleküler yöntemlerle hayvanlardan alınan örneklerin yüzlerce hayvan için analiz edilmesi ve genetik olarak izlenebilirliği hala pahalı bir seçenek durumundadır.

Hayvan ırkına ait kalitatif ve kantitatif karakterlerden ırkı karakterize etmek için önemli sayılabilecek olanları saptayarak bu karakterler üzerinde istatistiksel sınıflama yoluyla saflık tayini maliyet açısından uygulanabilirliği yüksek bir seçenek gibi gözükmektedir. Böylece herhangi bir uzman kontrolüne en az düzeyde ihtiyaç duyularak saflık tayini yapılabilmesi sağlanabilecektir.

Günümüzde istatistiksel sınıflayıcı olarak kullanılan çok sayıda yöntem geliştirilmiştir. Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees, CART) yöntemlerinden ID3, C4.5 VE C5.0 başarılı sınıflandırıcılar arasında

yer almaktadır. Son yıllarda Rastlantısal Ormanlar (Random Forests, RF) da gittikçe popüler olmaya başlamış ve Derin Öğrenme (Deep Learning) uygulamaları ile başarımları daha da yükselmiş bulunmaktadır. Bunlardan basitliği ve uygulanabilirliği nedeniyle lojistik regresyon modellemesi en uygun analizlerden biri durumundadır. Bu nedenle bu tez çalışmasında basitliği nedeniyle uygulanabilirliği ve anlaşılabilirliği yüksek sınıflayıcılardan biri olan lojistik regresyon modellemesi ve analizi çalışılmıştır. Böylece ileride diğer gelişmiş yöntemlerle yapılacak sınıflama çalışmaları için bilgi ve deneyim altlığı oluşturacaktır



## 2. ÖNCEKİ ÇALIŞMALAR

Dinse ve Haseman, (1986). Hayvancılıkta lojistik regresyon analizlerinin çoğunlukla tümör oluşumunun tahminlenmesi gibi hayvan hastalıkları için risk faktörlerinin belirlenmesinde analitik bir yöntem olarak kullanıldığı çalışmalar oluşturmaktadır. Süt sığırcılığında döl verimi ile ilgili olan morfometrik karakterlerin etkilerinin araştırıldığı görülmektedir. Lojistik regresyon analizi uygulamalarının sıkça kullanıldığı diğer bir alan yaban hayvanlarının ekolojisi ve habitat seçimi üzerinde yoğunlaşmıştır.

North ve Reynolds (1996) radyo telemetri lokasyonlarını kullanarak mikro habitat analizi yapmışlar ve çok kategorili lojistik regresyon analizini uygulamışlardır.

Burmaoğlu ve ark. (2009), çok değişkenli istatistiksel sınıflandırma yöntemlerinden diskriminant analizi ve lojistik regresyon analizinin başarı sonuçlarını karşılaştırmıştır. Diskriminant analizinde %92,5'lik ve Lojistik Regresyon Analizinde %100'lük sınıflandırma başarısı elde edilmiştir.

Drobnjak ve ark. (2012), Mersin'in Tarsus ilçesinde Çatalburunların görüldüğünü ve sayılarının 200 civarında olduğunu kaydetmiştir. Bu av köpeklerinin eşsiz olmaları sebebiyle hakkındaki yayınların artırmaya yönelik araştırmacıların ilgisini çekmeyi amaçlamıştır.

Diñel ve Petek (2012), Türkiye'ye özgü Çatalburun köpekleri Tarsus ve çevresinde av köpeği olarak yıllardır yetiştirilmekte ve halk arasında Türk Pointer'i olarak tanımakta olduğunu belirtmiştir. Morfolojik ve fizyolojik özelliklerinin belirlenerek ırk hakkında daha fazla bilgiye sahip olunması gerektiğini saptamıştır.

Nemes ve Hartel (2010) çevreyle ilgili çalışmalarda çoğunlukla tepki değişkenlerinin ikili değerlerle (Örneğin; incelenen türlerin varlığı veya yokluğu) ile sonuçlandığı, bu gibi durumlarda lojistik regresyon, çeşitli çevresel faktörlere karşı

organizmaların tepkisini modellemek için kullanılabileceğini belirterek ROC eğrisinin yorumlanması üzerine bir çalışma gerçekleştirmişlerdir.

Lepper ve ark. (2002) Kaliforniya Hayvan Barınağı'ndaki köpekler ve kediler arasında ötenaziye karşı adaptasyon tahminlemede lojistik regresyon analizi uygulamışlardır. Kedilerde yaş, cinsiyet, tüy rengi ve vazgeçme nedeni; köpeklerde yaş, cinsiyet, tüy rengi, bırakma nedeni, cins, saflık ve yaralanma durumu adaptasyon için başlıca belirleyicileri olarak kullanılmıştır.

Eyduran ve ark. (2005), hayvancılıkta lojistik kullanımı üzerine yaptıkları çalışmalarında Hamdani ırkı koyunlarda cinsiyet, doğum tipi, doğum ağırlığı değişkenlerinden 30. 45. 60. ve 75. günlerde vücut ağırlığını tahmin için lojistik regresyon analizinin etkinliğini test etmişlerdir. Çalışmalarında yukarıdaki periyotlarda vücut ağırlığını ortalamanın altında ve üstünde olmak üzere iki kategoriye bölerek bağımlı değişken olarak kullanmışlardır. Model uyum iyiliği için Ki-kare, G-test ve konkordans istatistiklerini kullanmışlardır.

Farhangfar ve ark. (2010) İran yerli Holştayn ineklerinde kurudaki gün sayısını etkileyen bazı çevresel faktörleri lojistik regresyon ile analiz etmişlerdir.

Aygün ve Çelikyürek (2017), yaptıkları çalışmada büyükbaş hayvanlara ait bilgilerin kaydedilmesi için bir bilgisayar paket programı hazırlanmışlardır. Bu programda, küçük ve orta ölçekli işletmelerdeki büyükbaş hayvanlarda tutulan kayıtlar bilgisayar ortamına aktarılarak veri tabanlarında uzun süre depolanmakta ve ihtiyaç anında istenen verilere kolayca ulaşılabilmektedir.

Keshavarzi ve ark. (2017), İran'da süt sığırı sürülerinde yavru atmaya etkili risk faktörlerini değerlendirmek için lojistik regresyon kullanmışlardır.

Pereira ve ark. (2011), etlik piliçlerde refahını çevresel ve davranışsal değişkenlerle ilişkili olarak tahmin etmek için lojistik regresyon kullanmışlardır. Bu çalışmada, üç farklı amonyak uygulamasında yetiştirilen hayvanların çeşitli davranışları kaydedilerek çoklu lojistik regresyon analizi ile değerlendirilmiş ve hayvan refahını sınıflandırmada başarılı şekilde kullanılabileceği gösterilmiştir.

Compton ve ark. (2002) da orman kaplumbağası (*Clemmys insculpta*) habitat seçimini lojistik regresyon ile modellemişlerdir.

Ertuğrul ve Yılmaz (2012), Türkiye’de bulunan yerli köpeklerin bazı ırklara ve tiplerine ait bilgilere yer vermiştir. Mersin’de sayılarının oldukça azaldığını ve koklama yeteneğinin olduğunu kaydetmiştir. Avını araması, bulması ve uçurması gibi özellikleri bakımından yetenekli olduğunu belirtmiştir.

Topal ve ark. (2017), kuzularda mortalite üzerinde etkili genetik olmayan faktörlerin etkilerini değerlendirmede lojistik regresyon analizi uygulamışlardır.

Bircan (2004), ikili sonuç değişkeni ile sürekli ve kesikli değişkenlerden oluşan bağımsız değişkenler kümesi arasındaki ilişkiyi lojistik regresyon analizi ile incelemiştir. Oluşturulan modelin biyolojik olarak kabul edilebilir olması ve doğru sınıflama oranının yeterince iyi olması sebebiyle bu modelin risk faktörlerini belirlemede kullanılabileceği sonucuna varılmıştır.

Çokluk (2010), iki kategorili bağımlı değişken ve bağımsız değişkenler kombinasyonunu belirlemeye yönelik ikili lojistik regresyon analizinin temel kavram ve süreçlerini açıklamıştır.

Girginer ve Cankuş (2008), Eskişehir’de toplu taşıma araçlarından biri olan tramvaya yönelik yolcu memnuniyeti, Eskişehir tramvay sistemi (Estram) örneğinde, Binomial Lojistik Regresyon Analizi ile incelenmiştir. Uygulanan Binomial Lojistik Regresyon Analizi sonucunda; tüm bağımsız değişkenlerin negatif etkileri olduğu belirlenmiştir.

Oğuzlar (2005), Bursa Emniyet Müdürlüğü ile ortak yapılan bir proje kapsamında, ahlak, yankesicilik ve narkotik bürolarına ilişkin verilerden yararlanılarak, suçluların profilinin belirlenmesinde lojistik regresyon analizinden yararlanılmıştır.

Akın ve Şentürk (2012), mutluluk düzeyleri Ordinal Lojistik Regresyon Analizi aracılığıyla sosyo-demografik özelliklere göre incelenmiş ve diğer çalışmaların sonuçları karşılaştırılmıştır.

Aktaş (2009), bağımlı değişkenin iki düzeyli olması durumunda demografik, davranış ve risk faktörüyle ilgili tahmin çalışmalarında lojistik regresyon analizi incelenmiştir. Eskişehir Osmangazi Üniversitesi (ESOGÜ) öğrencileri arasında sigara içme alışkanlığını etkileyen faktörleri belirlemek için lojistik regresyon ve diskriminant denklemi belirlenmiştir.

Arı ve Önder (2013), değişkenler arasındaki ilişkinin fonksiyonel şekli regresyon modellerinden; doğrusal regresyon, lojistik regresyon, negatif binom regresyon, poisson regresyon, temel bileşenler regresyonu, probit regresyon, ridge regresyon, Cox regresyon modellerinin hangi durumlarda kullanılabileceğini incelemiştir.

Kneib ve ark. (2011), kahverengi ayılar ve kuşların habitat seçiminde lojistik regresyon analizinin başarılı şekilde kullanılabileceğini göstermişlerdir.

Tomaszyk ve ark. (2014), kuzu hayatta kalma analizinde sinerji faktörleri başlıklı çalışmalarında hayatta kalmanın, hayvancılıkta genetik iyileştirme programlarında en önemli özelliklerinden biri olduğunu belirterek Polonya Merinosu ve Wielkopolska koyun ırkından kuzularda cinsiyet, akrabalı yetiştirme ve doğum tipi değişkenlerini kullanarak kuzuların hayatta kalma sonuçlarını tahmin etmişlerdir.

Köksal (2011), temelde doğrusal olmayan ancak bazı dönüşümlerle doğrusallaştırılabilen lojistik regresyon modelleri için ROC eğrisi kestiriminin kullanılabilirliği incelenmiştir. Farklı model seçme ölçütlerine göre yapılan değerlendirmede ROC eğrisi kestirimi, kestirilen modeller arasından en iyi ayırım gücüne sahip olan modelin belirlenmesinde en uygun seçim ölçütü olarak saptanmıştır.

### 3. MATERYAL VE METOT

#### 3.1. Materyal

Araştırma verileri Mersin Büyükşehir Belediye Başkanlığı'na bağlı Çatalburun Üretim ve Rehabilitasyon Merkezi'nde (Şekil 1) koruma altına alınan 60 köpek ile Tarsus ilçesindeki saha çalışmalarında 10 köpek olmak üzere toplam 70 köpekten toplanılmıştır.



Şekil 1: Çatalburun Üretim ve Rehabilitasyon Merkezi (Mersin)

Çalışma materyalini oluşturan köpeklerin yaş ve cinsiyete göre dağılımı Tablo1'de verilmiştir. Çalışma verisi toplam yaş ortalaması 3.27 olan 29 dişi ve yaş ortalaması 3.39 yıl olan 41 erkek hayvandan ölçülerek kaydedilmiştir.

Tablo 1. Köpeklerin yaş ve cinsiyete göre dağılımı

<i>Cinsiyet</i>	<i>Yaş</i>				<i>Toplam</i>
	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
<i>Dişi</i>	8	11	4	6	29
<i>Erkek</i>	10	15	6	10	41
<i>Toplam</i>	18	26	10	16	70

Çalışmada Tablo 1'de listesi görülen 18 farklı kantitatif karaktere ait ölçümler yapılmıştır. Ölçümler sırasında uzman tarafından saflık tayini de yapılmıştır. Ölçümlerde Şekil 2'de görülen dijital kumpas ve vücut-kütle indeksi ölçme şeridi kullanılmıştır. Ölçümler ve saflık tayini çalışmalarına ait bazı görüntüler Şekil 3 ve 4'te verilmiştir.

Burun çapı, burun bir kanat uzunluğu, burun iki kanat uzunluğu milimetre (mm) cinsinden ölçülmüştür. Burun ucu alın uzunluğu, iki şakak arası genişlik, iki kulak arası uzunluk, boyun kalınlığı, iki şakak ortası ile iki scapula (kürek kemiği) arası uzunluk, iki scapula ile kuyruk arası uzunluk, göğüs çevresi, cidago yüksekliği (ön bacak ile yer arası uzaklık), kuyruk kalınlığı (1. boğum), ön pati bilek kalınlığı, arka pati bilek kalınlığı, ön bacak uzunluğu, arka bacak uzunluğu, kulak uzunluğu, kulak genişliği santimetre (cm) cinsinden ölçülmüştür.



Şekil 2: Vücut kitle indeksi ölçme şeridi (sol) ve dijital kumpas (sağ)



Şekil 3: Köpeklerin kayıt altına alınması ve saflık değerlendirmesi



Şekil 4: Ölçme işlemlerinden bazı görüntüler

Tablo 2. Bağımlı ve Bağımsız değişkenlerin çizelgesi

<b>Bağımlı değişken</b>	<b>Kod</b>
Saflık (1: Saf, 0: Saf değil)	y
<b>Bağımsız değişkenler</b>	
Burun çapı (mm)	x1
Burun bir kanat uzunluğu (cm)	x2
Burun iki kanat uzunluğu (cm)	x3
Burun ucu alın uzunluğu (cm)	x4
İki şakak arası genişlik (cm)	x5
İki kulak arası uzunluk (cm)	x6
Boyun kalınlığı (cm)	x7
İki şakak ortası ile iki scapula arası uzunluk (cm)	x8
İki scapula ile kuyruk arası uzunluk (cm)	x9
Göğüs çevresi (cm)	x10
Cidago yüksekliği (ön bacak üstü) (cm)	x11
Kuyruk kalınlığı (1. Boğum) (cm)	x12
Ön pati bilek kalınlığı (cm)	x13
Arka pati bilek kalınlığı (cm)	x14
Ön bacak uzunluğu (cm)	x15
Arka bacak uzunluğu (cm)	x16
Kulak uzunluğu (cm)	x17
Kulak genişliği (cm)	x18

## **3.2. Metot**

### **3.2.1. Saflık Tayininde Uzman Tarafından İncelenen Özellikler**

Ölçümler yapılırken verteriner uzman hekim tarafından köpeklerin saflık değerlendirmesi de yapılmıştır. Saf olanlar 1 ve saf olmayanlar 0 ile kodlanarak veri setine eklenmiştir. Saflık tayininde aşağıdaki özellikler dikkate alınmıştır.

#### **3.2.1.1. Burun yapısı ve Özellikleri**

Çatalburun, adını aldığı burun deliklerinin ayrıık olması ve burun iç hacminin fazla olması ile karakterize olan bir köpek ırkıdır. Bu iç hacim yüksekliđi daha fazla koku alma alanı (olfactory area) yol açarak koku alma hassasiyetini arttırmaktadır.

#### **3.2.1.2. Gözlerin Yapısı ve Özellikleri**

Gözler badem şekline benzemektedir. Oldukça sık görülen göz rengi kahverengi iken nadiren ela ve çakır göz renklerine de rastlanmaktadır. Farklı ırk köpekler de olduğu gibi gözün retina tabakasındaki özelleşmiş hücrelerin fazla olması karanlıkta insanlara göre daha iyi görmelerini sağlamaktadır.

#### **3.2.1.3. Kulak Yapısı ve Özellikleri**

Kulak uçları yuvarlak ve geniş olup U harfi biçiminde burun hizasına kadar uzanır. Gözlerin hafif üstünden aşağı doğru sarkar. Kulak ince ve narin yapıdadır. Genellikle vücut ile aynı kısa kıl ve renk yapısına sahiptir.

#### **3.2.1.4. Ağız-Dudak-Çene Yapıları ve Özellikleri**

Yandan bakıldığında çenenin üzerini örten bir yanak, önden bakıldığında ise alt çene ve üst çene aynı hizada görünür. Dudak yapısı, ne V harfi kadar aşağı doğru sivri ne de U harfi kadar ovaldir. Alt ve üst dişler birbirini karşılar ve görünmezler.

### 3.2.1.5. Deri-Kıl ve Renk Özellikleri

Deri uzaktan bakıldığında boyun altı hariç kırıksıklık yapmaz. Fakat elastiki özellikleri ile karakterizedir. Kıllar vücudu tümüyle sarar, düz, kısa, ince sıktır. Kılların dip kısımlarında pamuksu kıl tipi görülmemektedir. Renkleri ise kahve, şampanya, kahve-beyaz oldukça sık görülürken diğer renklerde gözlenir.

### 3.2.2. Lojistik Regresyon

Lojistik regresyon bağımlı değişkenlerin kategorik olması durumunda kullanılmak üzere 1958'de Cox tarafından (Cox, 1958) iki kategorili (veya sınıflı) bağımlı değişkenler için önerilmiştir. Günümüzde iki sınıflı problemler için bu yöntemle oluşturulan modellere ikil lojistik regresyon modeli (binary logistic regression) denilmektedir. Bir başka deyişle ikil lojistik regresyon modellerinde bağımlı değişkenin değerleri doğru/yanlış; evet/hayır; var/yok; yaptı/yapmadı; saf/saf değil gibi yalnız iki kategori/sınıftan biri olur. Bununla birlikte bağımsız değişken ikiden fazla kategori de içerebilir. Bu tür lojistik modellere çok terimli lojistik regresyon (Multinomial logistic regression) denilmektedir. Çok terimli lojistik regresyonda bağımlı değişken sıralama ölçeğinde ise sıralama ölçekli lojistik regresyon (ordinal logistic regression) söz konusudur. Lojistik regresyonda bağımsız değişkenlerin ölçeği için herhangi bir kısıtlama olmadığından modele her ölçekten bağımsız değişken eklenebilmektedir. Lojistik regresyon modelinde yalnız bir adet bağımsız değişken varsa bu tür modellere basit lojistik regresyon modeli, bağımsız değişken sayısı birden fazla ise çok değişkenli lojistik regresyon modeli denilmektedir (Cebeci, 2019).

Lojistik regresyon aslında doğrusal regresyonun uygulanmasına uygun olmayan regresyon problemlerine alternatif bir çözüm olarak önerilmiştir (Nelder ve Wedderburn, 1972). Konuyu anlamak için Eşitlik (1)'deki gibi basit regresyon modeli olduğunu varsayalım.

$$Y = \beta_0 + \beta_1 X \quad (1)$$

Eşitlik 1'deki doğrusal regresyon modelinde bağımlı değişken  $Y$  ve bağımsız değişken  $X$ 'in sürekli değişken oldukları varsayılır. Ancak  $Y$  iki sınıflı bağımlı bir değişken ve  $X$  sürekli bir bağımsız değişken ise doğrusal regresyon analizi ile  $Y$ 'nin değerleri sadece 0 ve 1 şeklinde kategorik olarak tahmin edilemez. Bu nedenle bağımlı değişkenlerin kategorik olması halinde uygulanamaz. Doğrusal regresyon modellerine ait varsayımlar, örneğin hataların normal dağıldığı varsayımı da sağlanamaz. Bu durumda alternatif bir regresyon modeline gereksinim doğar ki lojistik regresyon böyle bir analiz yöntemidir.

Lojistik regresyon Genelleştirilmiş Doğrusal Modeller (GLM: Generalized Linear Models) ailesinden bir analiz yöntemi olarak bu soruna çözüm bulur. Bağımlı değişkenin log oranları bağımsız değişkenlerinin doğrusal bir birleşimi ile modellenir. Bağımlı değişken  $Y$  ve bağımsız değişkenler  $X_1, X_2, \dots, X_k$  ile gösterildiğinde temel olarak Eşitlik (2)'deki gibi ifade edilir.

$$g(E(Y)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (2)$$

Burada  $g$  bir bağıntı fonksiyonudur.  $E(Y)$  bağımlı değişkenin beklenen değeri olup  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$  doğrusal bir kestirimcisine eşittir. Amaç  $\beta_0, \beta_1, \dots, \beta_k$  katsayılarını tahmin ederek bağımsız değişkenler ( $X$ 'ler) verildiğinde  $Y$ 'yi kestirmek için model eşitliğini oluşturmaktır. Bağıntı fonksiyonunun rolü,  $Y$ 'nin beklenen değerini doğrusal tahminciye 'bağlamak'tır.

Lojistik regresyonda, yalnızca bağımlı değişkenin olasılığı (başarı veya başarısızlık) ile ilgilenilir. Yukarıda Eşitlik 2'deki  $E(Y)$  bağıntı fonksiyonu, başarı olasılığı ( $p$ ) ve başarısızlık olasılığı ( $1 - p$ ) kullanılmaktadır. Burada  $p$  için  $0 \leq p \leq 1$  koşulu söz konusudur. Çünkü bu bir olasılık değeri olarak daima negatif olamaz ve daima 1'den küçük bir değerdir. Bu iki koşulu sağlayarak lojistik regresyon modelini

oluşturmak için bağıntı fonksiyonu  $E(Y)$  başlangıçta  $p$  ile gösterilerek bir logit fonksiyonu türetilir. Olasılık daima pozitif olacağına göre doğrusal eşitlik Eşitlik 3'teki gibi üssel forma dönüştürülebilir. Böylece herhangi bir eğim ve bağımsız değişken için üs hiçbir zaman negatif olmayacaktır.

$$p = e^{\beta_0 + \beta_1 X} \quad (3)$$

Olasılığı 1'den küçük yapmak için  $p$ , Eşitlik 4'te gösterildiği  $p$ 'den büyük bir sayıya bölünmelidir. Bunun için paydaya +1 eklenir.

$$p = \frac{e^{\beta_0 + \beta_1 X}}{e^{\beta_0 + \beta_1 X} + 1} \quad (4)$$

Eşitlik 3 ve 4 kullanılarak başarı olasılığı Eşitlik 5'teki gibi yeniden tanımlanabilir.

$$p = \frac{e^Y}{1 + e^Y} \quad (5)$$

Eşitlik 5 bir logit fonksiyonu olup  $p$  başarı olasılığıdır. Bu durumda  $q = 1 - p$ , başarısızlık olasılığı olup Eşitlik 6'daki gibi yazılabilir:

$$q = 1 - p = 1 - \frac{e^Y}{1 + e^Y} \quad (6)$$

Eşitlik 5 Eşitlik 6'ya bölünerek Eşitlik 7 yazılabilir.

$$\frac{p}{1-p} = e^Y \quad (7)$$

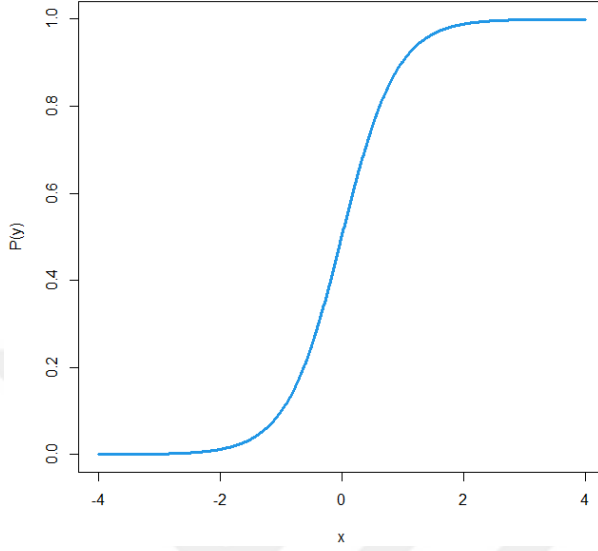
Eşitlik 7'de her iki tarafın logaritması alındığında Eşitlik 8'deki formül elde edilir.

$$Y = \ln\left(\frac{p}{1-p}\right) \quad (8)$$

Eşitlik 8'de  $\ln\left(\frac{p}{1-p}\right)$  bağıntı fonksiyonu olup bağımlı değişkenin logaritmik dönüşümüyle doğrusal olmayan bir ilişkinin doğrusal şekilde modellenmesi sağlanır. Y'nin yerine bu dönüştürme ile elde edilen değeri koyduğumuzda lojistik regresyon modeli Eşitlik 9'daki yazılabilir.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (9)$$

Eşitlik 9'daki lojistik regresyon eşitliğinde  $(p / 1 - p)$  bahis oranı (odds ratio) olarak, logaritması log odd veya logit olarak adlandırılır. Logaritması pozitif olduğunda başarı olasılığı daima %50'den büyüktür. Şekil 5'te tipik bir lojistik model eğrisi görülmektedir. Şekilde olasılığın 0'ın altında ve 1'in üstünde olmadığı açıkça görülmektedir.



Şekil 5: Lojistik regresyon eğrisi

Sonuç olarak, lojistik regresyon bağımsız değişkenlerin bağımlı değişken üzerindeki regresyonunu olasılık olarak hesaplayan bir analiz yöntemidir. Lojistik regresyonda bağımlı değişkenler ile bağımsız değişken arasındaki bağıntı doğrusal olmadığından logit dönüşüm yapılarak aradaki ilişki doğrusal hale getirilmektedir. Böylece doğrusal regresyon analizinin hesaplama tekniklerinden yararlanılabilmektedir.

Lojistik regresyonda doğrusal regresyonda olduğu gibi doğrusallık, normallik, varyans homojenliği ve hataların bağımsız dağılması varsayımlar yoktur. Ancak aşağıdakiler geçerlidir (Schreiber-Gregory ve Bader, 2018):

- Bağımlı değişken nominal veya ordinal ölçekten bir değişken olmalıdır.
- Gözlemler birbirinden bağımsız olmalıdır. Bir başka deyişle tekrarlanan ölçümlerden veya eşli gruplardan gelmemelidir.

- Bağımsız değişkenler arasında çoklu bağımlılık olmamalı ya da çok az olmalıdır. Yani bağımsız değişkenler arasında korelasyon olmamalı veya çok düşük olmalıdır.
- Bağımsız değişkenlerin bağımlı değişkenin log oranları ile ilişkisi doğrusal olmalıdır.
- Örneklem büyüklüğü büyük olmalıdır. Örneğin  $k$  bağımsız değişken var ve en az sıklığa sahip sınıfın beklenen olasılığı  $p$  ise minimum örneklem büyüklüğünün  $n=10 k/p$  olması gerekir.

### 3.2.2.1. Model Uyum İyiliği ve Başarımını Değerlendirme

Lojistik regresyon modellerinin başarımlarını hesaplamak ve yorumlamak için çeşitli görsel yöntemler, ölçütler ve testler kullanılmaktadır. Model değerlendirme ve karşılaştırma işlemlerinde başvurulan ölçütleri aşağıdaki gibi gruplandırmak mümkündür:

- Karışıklık matrisine dayalı ölçütler
- Bilgi teorisine dayalı ölçütler
- Uyum iyiliği ölçütleri

#### Wald Testi

Modeldeki bağımsız değişkenlerin regresyon katsayılarının istatistiksel önemini yani sıfırdan farklı olup olmadığını belirlemek için Wald testi yapılır. Wald istatistiği regresyon katsayısının karesinin katsayının standart hatasının karesine oranı olarak hesaplanır. Wald testi sonucunda sıfır hipotezi reddedilemiyorsa test edilen bağımsız değişkeni modelden çıkarmak model başarımını ciddi düzeyde etkilemeyecektir. Regresyon katsayısının önemliliğini ( $H_0:\beta=0$ ) test etmek için Wald Eşitlik 10'daki  $\chi^2$  istatistiğini kullanmayı önermiştir.

$$Wald\chi^2 = \left(\frac{\hat{\beta}}{s_{\hat{\beta}}}\right)^2 \quad (10)$$

Wald istatistiğinin örnekleme dağılışı yaklaşık  $\chi^2$  dağılışı gösterir ve böylece  $H_0$  hipotezi altında test edilebilir. Ancak bu test R da dâhil bazı yazılımlarda Eşitlik (11)'deki dönüşümle bir  $z$  testi olarak gerçekleştirilmektedir.

$$z = \sqrt{Wald\chi^2} \quad (11)$$

Formüldeki standart hata hesaplaması karmaşık olup yinelemeli bir süreçle maksimum olabilirlik tahmini hesaplanmaktadır. Wald testi çok yaygın kullanılan bir test olmasına rağmen, küçük örneklerde büyük katsayılar için çok tutucu olma eğilimindedir. Lojistik regresyonda katsayılar için güven aralıkları  $CI_\beta = \hat{\beta} \mp 1.96S_{\hat{\beta}}$  olarak hesaplanır. Burada 1.96, iki kuyruklu test ve  $\alpha = 0.05$  olduğunda normal dağılışı için  $z$  kritik değeridir. Güven aralığı sıfır içeriyorsa katsayı önemsizdir. Güven aralıkları odds oranları sunulursa 1 içeren bir aralıklar katsayının önemsiz olduğunu gösterir.

### 3.2.2.1.(1). Sahte (Pseudo) $R^2$ Ölçütleri

Lojistik regresyonda uyum iyiliğini değerlendirmek için doğrusal regresyon analizinde kullanılan  $R^2$  istatistiğine eşdeğer bir istatistik yoktur. Lojistik regresyon analizindeki model tahminleri yinelemeli bir süreç yoluyla ulaşılan maksimum olabilirlik tahminleridir. Varyansı en aza indirecek şekilde hesaplanmadıklarından  $R^2$  uyum iyiliği ölçütü olarak kullanılamaz. Bununla birlikte, lojistik modellerin uyum iyiliğini değerlendirmek için Aldrich-Nelson, Efron, McFadden, Cox-Snell, Nagelkerke/Cragg-Uhler, McKelvey-Zavonia, Veall-Zimmermann ve Tjur gibi birçok sahte  $R^2$  ölçütü geliştirilmiştir (Smith ve McKenna, 2013). Bu ölçütler  $R^2$  gibi  $[0,1]$  aralığında değerler olmakla birlikte bazıları asla 0 veya 1'e ulaşmazlar. Daha yüksek sahte  $R^2$  değerleri daha iyi bir modeli gösterirler. Bu çalışmada McFadden

sahte  $R^2$ , Cragg ve Ulmer sahte  $R^2$  ile Cox ve Snell sahte  $R^2$  değerleri model uyum iyiliğini değerlendirme ölçütleri olarak kullanılmıştır.

Sahte  $R^2$  ölçütlerinin çoğu modellerin boş modele göre sapmasına, bir başka deyişle Eşitlik (12)'de gösterilen log olabilirlik oranına dayanırlar.

$$LLR = G^2 = 2\ln(L(model_i) - L(model_0)) = -2\ln\left(\frac{L(model_0)}{L(model_i)}\right) \quad (12)$$

Eşitlik (12)'de:

$L(model_i)$ : Uydurulan herhangi bir  $i$ . modelin log olabilirlik değeri

$L(model_0)$ : Boş modelin (yalnız kesme yüksekliği içeren) log olabilirlik değeri

Olabilirlik oranı, bir modelin boş model üzerindeki üstünlüğünü gösterir. Oran ne kadar küçükse üstünlük o kadar fazla olur. Model uyumunu değerlendirmek için log olabilirlik fonksiyonu gözlem değerlerini açıklama olabilirliğini maksimuma çıkaran katsayı değerlerini seçer. Fonksiyon değeri 1'e yaklaşırken uyum iyiliğinin yükseldiği, ancak sifıra yaklaşırken modelin verilere uyumunun zayıfladığı anlaşılır. Böylece karşılaştırılan modellerde önemli olmayan bağımsız değişkenler çıkartılarak yeni ve daha sade bir model elde edilir.

McFadden sahte  $R^2$  (MFR2) istatistiği (McFadden, 1979), bir modelin log olabilirliğinin boş modelin log olabilirliğine bölünmesine dayanan bir oran ölçütü olup Eşitlik 13'teki gibi hesaplanır.

$$MFR2 = 1 - \frac{\ln(L(model_i))}{\ln(L(model_0))} \quad (13)$$

Olabilirlikler  $[0,1]$  aralığında olduğunda log olabilirlik değerleri sifıra eşit veya sifirdan küçük olur. Eğer bir model çok düşük bir olabilirliğe sahipse daha yüksek log olabilirliğe sahip olur. McFadden sahte  $R^2$  değeri 0 ve 1 arasında bir

değer olup 0'a yakın değerler zayıf uyumu, 1'e yakın değerler iyi uyumu gösterir. Lojistik regresyonda çok nadiren yüksek değerlere ulaşılır. *MFR2* değerinin 0.40 ve üstünde olması iyi bir uyum olarak değerlendirilir (Cebeci, 2019).

### Cox ve Snell'in Sahte R<sup>2</sup> İstatistiği

Bir modelin olabirliği  $L(Model)$ , bağımsız değişkenler verildiğinde bağımlı değişkenin koşullu olasılığıdır. Veride  $n$  gözlem varsa,  $L(Model)$  bu tür  $n$  olasılığın çarpımıdır. Bu çarpımın  $n$ 'inci kökünü almak, her  $Y$  değerinin olasılığının bir tahminini sağlar. Cox ve Snell'in sahte R<sup>2</sup> ölçütü (MLR2) lojistik regresyon modellerinin yakınsamasını belirlemek için kullanılan  $-2\ln\left(\frac{L(model_0)}{L(model_i)}\right)$  oranına dayanan bir uyarlamadır (Cox ve Snell,1989).

$$CSR2 = 1 - \left(\frac{L(model_0)}{L(model_i)}\right)^{2/n} \quad (14)$$

Bir model sonucu mükemmel bir şekilde tahmin etse bile  $1 - (L(model_0))^{2/n}$  her zaman 1'den küçük olacağından Cox ve Snell'in sahte R<sup>2</sup> ölçütünün maksimum değeri 1 olamaz.

### Cragg ve Uhler'in Sahte R<sup>2</sup> İstatistiği

Cragg ve Uhler'in Sahte R<sup>2</sup> İstatistiği (Nagelkerke Sahte R<sup>2</sup> İstatistiği olarak da bilinir) Cox ve Snell'in sahte R<sup>2</sup> değerlerinin aralığını 1'e kadar ulaşacak şekilde ayarlar. Eşitlik 15'te görüldüğü gibi, Cox ve Snell sahte R<sup>2</sup> değeri mümkün olan maksimum değer olan  $1 - (L(model_0))^{2/n}$ 'ye bölünür (Nagelkerke, 1991). Bu nedenle mükemmel tahmin yapan bir modelin sahte R<sup>2</sup> değeri 1 olarak saptanabilir.

$$CUR2 = \frac{1 - \left(\frac{L(model_0)}{L(model_i)}\right)^{2/n}}{1 - (LL(model_0))^{2/n}} = \frac{MLR2}{1 - (LL(model_0))^{2/n}} \quad (15)$$

**3.2.2.1.(2). Karışıklık Matrisinden Hesaplanan Başarım Ölçütleri**

Tablo 3'te yapısı görülen karışıklık matrisi (confusion matrix), bağımlı değişkenin her bir sınıfı için gerçek ve tahmin değerlerinin sayılarını gösteren çapraz bir tablodur. Karışıklık matrisinden model başarımına ilişkin doğruluk (accuracy), duyarlılık (sensitivity) ve özgüllük (specificity) başta olmak üzere çok sayıda ölçüt hesaplanabilmektedir. Aşağıda bunlardan bazıları açıklanmaktadır.

Tablo 3. Karışıklık matrisi yapısı

		<b>Gerçek</b>	
		<i>Pozitif</i>	<i>Negatif</i>
<b>Tahmin</b>	<i>Pozitif</i>	<i>Doğru Pozitif (TP)</i>	<i>Yanlış Negatif (FP)</i>
	<i>Negatif</i>	<i>Yanlış Pozitif (FN)</i>	<i>Doğru Negatif (TN)</i>

Doğru Pozitif Oranı (TPR: True Positive Rate), gerçekte pozitif olup model tarafından pozitif olarak tahmin edilen yanıtların oranını gösterir.

$$TPR = TP / (TP + FN) \quad (16)$$

Doğru Pozitif Oranı, duyarlılık (sensitivity) ve bazı kaynaklarda recall olarak da adlandırılmaktadır. Doğru olarak tahminlenen gerçek pozitiflerin oranını gösterir.

Yanlış Pozitif Oranı (FPR: False Positive Rate) gerçekte negatif olduğu pozitif olarak tahmin edilenlerin oranını gösteren bir değerdir.

$$FPR = FP / (FP + TN) \quad (17)$$

Doğru Negatif Oranı (TNR: True Negative Rate), gerçekte negatif olup yine negatif olarak tahmin edilen yanıtların oranını gösteren bir değerdir. TNR, aynı zamanda *özgüllük (specifity)* olarak da adlandırılır.

$$TNR = TN/(FP + TN) \quad (18)$$

Yanlış Negatif Oranı (FNR: False Negative Rate), gerçekte pozitif olduğu halde negatif olarak tahmin edilen yanıtların oranını gösteren bir değerdir.

$$FNR = FN/(TP + FN) \quad (19)$$

Model tahmin doğruluğu (ACC: Accuracy) doğru sınıflanmış ya da doğru tahmin edilmiş pozitif ve negatif yanıt sayısının toplam yanıt sayısına bölünmesiyle hesaplanan bir oran değeridir. Bu nedenle doğru sınıflama oranını gösteren bir ölçüttür.

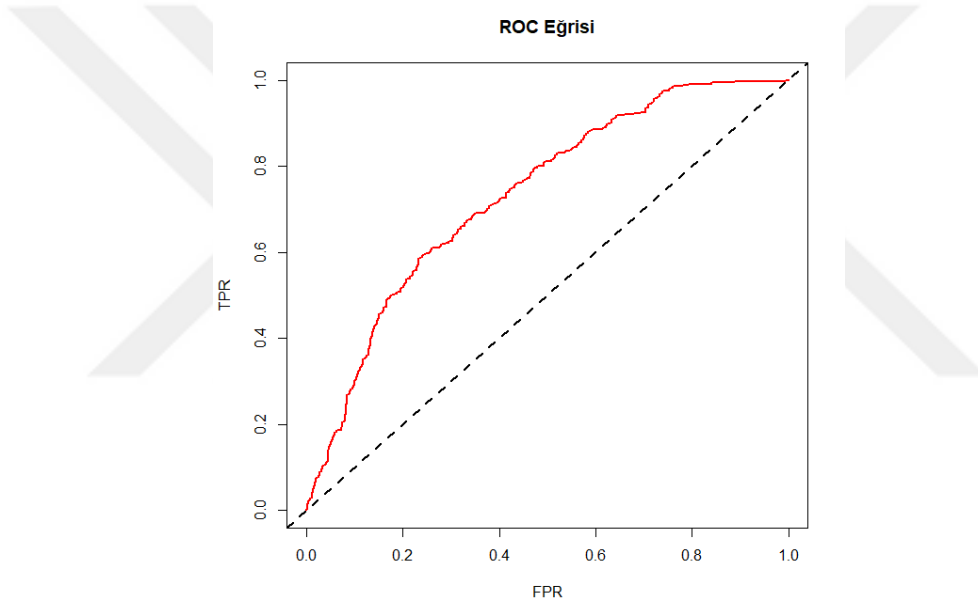
$$ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{N} \quad (20)$$

Sınıflama hatası (CE: Classification error) yanlış sınıflanmış ya da yanlış tahmin edilmiş pozitif ve negatif yanıt sayısının toplam yanıt sayısına bölünmesiyle hesaplanan bir oran değeridir.

$$CE = MCR = \frac{FP+FN}{TP+TN+FP+FN} = 1 - ACC \quad (21)$$

Bir sinyal saptama teorisi terimi olarak *Alıcı İşletim Karakteristiği* (ROC: Receiver Operating Characteristic) 1970'lerde lojistik regresyona dayanan tıbbi tanı işlemlerinde kullanılmıştır. ROC, lojistik regresyon modellerinde sınıflama başarımını yani modelin gücünü ortaya koyan bir grafiktir. Şekil'de de görüldüğü gibi ROC eğrisi  $y$  ekseninde doğru pozitif oranı (TPR) veya duyarlılık;  $x$  ekseninde yanlış pozitif oranı (FPR) veya *1-özgüllük* değerleri kullanılarak çizilir. Bir başka

ifadeyle farklı optimal eşik değerlerinde (Optimal cutoff values) modelin çalıştırılması ile hesaplanan duyarlılık ve 1-özgüllük değerleri ile çizilen bir grafiktir. Eşik değeri ya da kesme değeri bağımlı değişken sınıfları arasındaki karar sınırınıdır. Bu değerin altında ve üstünde olanlar bağımlı değişkenin iki ayrı sınıfını oluştururlar. ROC eğrisi ile doğru sınıflama ile yanlış sınıflama arasındaki değişimi görsel olarak izlemek mümkün olur. ROC eğrisinin sol üst köşeye yakın olması yani eğri altındaki alanın büyük olması modelin ayırma kabiliyetinin daha iyi olduğu anlamına gelir.



Şekil 6: ROC eğrisi örneği

ROC eğrisi altındaki alan kısaca eğrinin altındaki alan (AUC: Area Under Curve) model uyum iyiliği kullanılan bir ölçüt olup modelin sınıflama başarımı derecesini gösterir. AUC değerleri  $[0,1]$  aralığındadır. Örneğin AUC'nin 0.75 olması modelin pozitif ve negatifleri %75 olasılıkla ayırabileceğini ifade etmektedir. Yüksek AUC değerleri modelin bağımlı değişken sınıflarını ayırmada daha iyi olduğu anlamına gelir. TPR 1'e eşit ve FPR 0'a eşit olduğunda ROC eğrisi grafiğinin

sol üst köşesine değer ve AUC 1 olur. Bu durumda modelin duyarlılığı ve özgüllüğü %100 olur. Yani AUC'nin 1'e eşit olması pozitif ve negatiflerin hatasız ayrılabilirdiği anlamında gelir ki model uyumunun mükemmel olduğunu gösterir. AUC'nin 0 olması ise herhangi bir doğru sınıflama yapılamadığını anlamına gelir ki tüm pozitifler negatif, tüm ise negatifler pozitif olarak sınıflanabilirler. Bu durumda modelin tam olarak başarısız olduğu söylenebilir. Diğer yandan ROC çizgisi köşegen çizgisi ile çakıştığında yani AUC değeri 0.5 olduğunda modelin pozitif ve negatif sınıfları ayırma gücünün olmadığı anlaşılır. Bu durumda ROC eğrisi sol alt köşeden sağ üst köşeye uzanan köşegen boyunca düz bir doğru olarak görünür. Yukarıda yapılan açıklamalardan da anlaşılacağı gibi AUC değerlerinin 0.5-1.0 aralığında olması gerekir. Bu değerlerin 0.8 ve üstünde olması modelin yüksek başarılı olduğunu gösterir. Bazı kaynaklarda AUC'a göre model başarımını sözel olarak ifade etmek için Tablo 4'de verilen değerlendirme tablosu kullanılmaktadır (Cebeci, 2019, 2020b).

Tablo 4. AUC değerlerinin yorumlanması

<b>AUC</b>	<b>Yorum</b>
<b>0.91-1.00</b>	<i>Mükemmel</i>
<b>0.81-0.90</b>	<i>İyi</i>
<b>0.71-0.80</b>	<i>Orta</i>
<b>0.61-0.70</b>	<i>Zayıf</i>
<b>0.50-0.60</b>	<i>Kötü (Kullanışsız)</i>

### 3.2.2.1.(3). Bilgi Teorisine Dayanan Başarım Ölçütleri

Akaike Bilgi Kriteri (AIC: Akaike Information Criterion), bir modelde terim sayısı arttıkça modeli cezalandıran bir ölçüttür (Akaike, 1974).

$$AIC = -2 \ln L(\hat{\theta}) + 2k \quad (22)$$

Eşitlik (22)'de:

$L(\hat{\theta})$  : İncelenen model için maksimum olabilirlik fonksiyonu,

$k$ : Modeldeki parametre sayısıdır.

Çünkü modelin olabilirlik fonksiyonu ( $L(\hat{\theta})$ ) olarak genelde hata kareler ortalaması ( $MSE$ ) kullanılır.

$$AIC = n \ln \left( \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \right) + 2k \quad (23)$$

Model değerlendirmede kullanılan bir başka bilgi ölçütü, Schwarz'ın Bayesçi Bilgi Ölçütü (Schwarz, 1978) veya kısaca Bayesçi Bilgi Ölçütü (BIC) olup aslında AIC'in bir uyarlamasıdır (Cebeci, 2020a). Logit modelleri için BIC Eşitlik (24)'teki gibi hesaplanırken OLS modelleri için Eşitlik (25)'deki gibi hesaplanır.

$$BIC = -2 \ln L(\hat{\theta}) + \ln(n) k \quad (24)$$

$$BIC = n \ln \left( \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) + \ln(n) k = n \ln(MSE) + \ln(n) k \quad (25)$$

BIC parametre sayıları için daha fazla ceza uyguladığından gözlem sayısı arttıkça AIC ve BIC birbirine uymayan sonuçlar üretirler.

### 3.2.3. Veri Analizi ve Hesaplama Araçları

Veri analizleri R istatistiksel hesaplama ve grafik analiz ortamında (R Core Team, 2021) yapılmıştır. Çalışmada aşağıdaki analiz adımları izlenmiştir.

- Verisetini okuma

- Sınıf yanlılığını kontrol etme
- Değişken seçme
- Logit modelleri oluşturma
- Model diyagnostiklerini yapma
- Varsayımların kontrolü
- Modelleri karşılaştırma ve başarılı modeli önerme

Analizlerin yapılmasında R'nin temel paketlerindeki çeşitli fonksiyonlar hesaplama ve görselleştirme fonksiyonları kullanılmıştır. Lojistik regresyon analizleri *stats* paketindeki Genelleştirilmiş Doğrusal Modeller (Generalized Linear Models kısaca GLM) için kullanılan *glm* fonksiyonu *logit* modunda çalıştırılmıştır.

Modellere girecek bağımsız değişkenlerin seçiminde bilgi değerlerini hesaplamak ve değişken önem değerlerini hesaplamak üzere *caret* paketindeki (Kuhn, 2020) *varImp* fonksiyonu kullanılmış, ayrıca R'nin *step* fonksiyonu ile iki yönde kademeli lojistik regresyon analizi yapılmıştır.

Çoklu bağımlılık kontrolü (multi-colinearity) için *car* paketindeki (Fox ve Weisberg, 2019) *vif* fonksiyonu kullanılmıştır. Varyans Şişme Faktörü (VIF: Variance Inflation Factor) değeri yüksek, genel olarak 5 veya 10'un üzerinde, olan değişkenler çoklu bağımlılık gösterdiklerinden modelden çıkarılmalıdır. VIF'i yüksek değişkenler modelin katsayı tahmininde sorun yaşadığını gösterir. Ancak bu sorun genel olarak tahminlerin kalitesini düşürmez. Öte yandan model başarımından fazlaca düşme olmaksızın daha basit bir modelle çalışma avantajı da ortaya çıkmaktadır. Çalışmada logit ve bağımsız değişken arasındaki ilişkilerin kontrolü grafik üzerinde görsel olarak incelenmiş ve ayrıca *car* paketindeki *boxTidwell* fonksiyonu kullanılmıştır. Yine aynı paketteki *durbinWatsonTest* fonksiyonu ile otokorelasyon araştırılmıştır.

Modellerin tahmin başarımlarını AUC ölçütüne göre değerlendirmede ve ROC eğrilerinin çiziminde *ROCR* paketi (Sing ve ark., 2005), diğer görseller için

*corrplot* paketi (Wei ve Simko, 2017) ve *ggplot2* paketi (Wickham, 2009) ile çalışılmıştır.

Model başarımlarını karşılaştırmak için karışıklık matrisleri oluşturularak doğruluk değerleri hesaplanmıştır. Sahte  $R^2$  değerlerini hesaplamak için *DescTools* paketinden (Signorell ve ark., 2021) *PseudoR2* fonksiyonu, *pscl* paketinden (Jackman, 2020) *pR2* fonksiyonu, *modEvA* paketinden (Barbosa ve ark., 2015) *RsqGLM* fonksiyonu, *ModelMetrics* paketinden *auc* fonksiyonu (Hunt, 2020) yararlanılmıştır. Çalışmada gerçekleştirilen analizlerde kullanılan program EK 1'de verilmiştir.



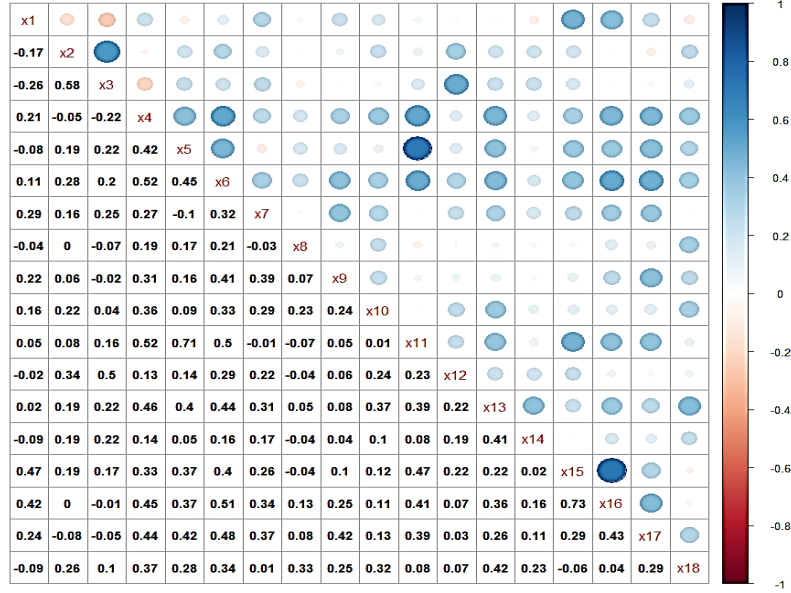
#### 4. BULGULAR VE TARTIŞMA

##### 4.1. Varsayımların Kontrolü

Lojistik regresyon modellerini karşılaştırmaya geçmede önce varsayım kontrolleri gerçekleştirilir. Oluşturulan bir modele girecek olan değişkenlerin seçilmesine *değişken seçme* denilir. Değişken seçme model eğitimi öncesinde ve sonrasında yapılır. Analiz öncesi yapılan değişken seçme genelde otokorelasyon ve çoklu bağımlılık gösteren değişkenlerin elenmesi için gerçekleştirilir.

Araştırmada çalışılan bağımsız değişkenler ( $x_1, x_2, x_3, \dots, x_{18}$ ) arasındaki korelasyonlar incelenmiştir. Böylece modeller oluşturulurken aralarında yüksek korelasyon bulunan değişkenlerin belirlenmesi ve bu değişkenlerden birinin seçilerek modele eklenmesi amaçlanmıştır. Bu tür işlemler yüksek boyutlu ve hacimli verilerde değişken indirgeme yoluyla hesaplama maliyetinin düşürülmesine katkı sağlamaktadır.

R'nin *corrplot* paketinde (Wei ve Simko, 2017), *corrplot* fonksiyonu kullanılarak değişkenler arasındaki Pearson korelasyonları için Şekil 7'de gösterilen korelasyonlar matrisi oluşturulmuştur. Şekildeki korelasyon matrisinde mavi renkler pozitif ve kırmızı renkler negatif korelasyonu göstermektedir. Dairelerin çapı ise korelasyon ölçüsünün büyüklüğünü temsil etmektedir. Matrisin diyagonal altında ise korelasyonların değerleri görülmektedir.



Şekil 7: Bağımsız değişkenler arasındaki korelasyonlar

Şekil 7 incelendiğinde x2 ile x3; x5 ile x11 ve x15 ile x16 değişkenleri arasında ilişkinin yüksek olduğu görülmektedir. Bu durumda modelleme esnasında otokorelasyona yol açabilecek bu değişken çiftlerinden birinin modele dâhil edilmesi önerilmektedir. Bu çalışmada söz konusu korelasyonların yüksek olmaması nedeniyle modellemede sorun yaratmayacağı düşünülmüştür. Nitekim *car* paketindeki *durbinWatsonTest* fonksiyonu ile yapılan otokorelasyon testinde tüm modeller için değişkenler arasında otokorelasyon saptanmamıştır ( $p>0.01$ ).

*Varyans Şişme Faktörü (VIF)* çoklu bağımlılık kontrolünde sıkça kullanılan bir ölçüttür (Cebeci, 2020b).

$$VIF_j = 1/(1 - R_j^2) \quad (26)$$

VIF, modeldeki çoklu bağımlılık nedeniyle bir regresyon katsayısı varyansının ne kadar şişirildiğini ölçer. Eğer VIF eşik değeri olan 5'ten (veya daha

az tutucu şekilde 10'dan) büyükse değişkenin çoklu bağımlılığını söz konusu olabilir ve şişmeye yol açan değişkenlerin modele eklenmemesi gerekir. Bu çalışmada tüm modeller içerdikleri değişkenler için *car* paketindeki (Fox ve Weisberg, 2019) *vif* fonksiyonuyla kontrol edilmiş ve çoklu bağımlılık saptanmamıştır.

Çalışmada logit ve bağımsız değişkenler arasındaki doğrusallığın kontrolünde Box-Tidwell testi uygulanmıştır (Box ve Tidwell, 1962). Box-Tidwell testi doğrusal bir modeldeki bağımsız değişkenlerin güç dönüşümlerinin maksimum olasılık tahminlerinin, genellikle genel bir doğrusal olmayan en küçük kareler rutini kullanmaktan daha verimli olarak yapılmasını sağlamaktadır. Modellere göre yapılan testlerde Model S1, Model S2 ve Model S3'teki bağımsız değişkenlerle bağımlı değişken logiti arasında doğrusallık ihlali saptanmamıştır ( $p > 0.05$ ). Ancak Model S4'te  $x_3$  ( $p = 0.0034$ ),  $x_5$  ( $p = 0.0002$ ),  $x_6$  ( $1e-07$ ) ve  $x_{18}$  ( $p = 0.0005$ ) değişkenleri bakımından doğrusallık ihlali olduğu saptanmıştır. Lojistik regresyon için logitin doğrusallığı ihlal edildiğinde orijinal sürekli bağımsız yerine doğal logaritmasının eklenmesi halinde sorunun çözebileceği belirtilmiştir (Schreiber-Gregory ve Bader, 2018). Bu nedenle Model S4'ün bir alt versiyonu olarak söz konusu değişkenlerin doğal logaritmalarının bağımsız değişken olarak kullanıldığı Model S4a oluşturulmuştur.

## 4.2. Lojistik Regresyon Modelleri

### 4.2.1. Tek Değişkenli Modeller

Araştırmada öncelikle Tablo 5'te değişkenlerden her biri ( $x_1, x_2, \dots, x_{18}$ ) tek başına bağımsız değişken olarak alınarak tek değişkenli lojistik regresyon modelleri oluşturulmuş ve başarımları incelenmiştir. Bunun için Eşitlik (27)'de verilen model oluşturulmuş ve analiz edilmiştir.

$$E(y_{ij}) = \beta_{0i} + \beta_{1i}x_i + e_{ij} \quad (27)$$

Eşitlik (27) 'de:

$y_{ij}$  :  $j$ . bağımlı değişken değeri (safılık)

$x_i$  :  $i$ . Bağımsız değişken ( $i=1, \dots, 18$ )

$e_{ij}$ :  $j$ . gözleme ilişkin hata.

Tek değişkenli basit lojistik regresyon modelleri için yapılan analizde saptanan regresyon katsayıları Tablo 5'te gösterilmiştir. Elde edilen sonuçlara göre  $x_1$ ,  $x_2$  ve  $x_{10}$  değişkenleri ile kurulan tek değişkenli modellerde regresyon katsayıları önemli bulunmuştur ( $p < 0.05$ ). Bununla birlikte  $x_5$  ve  $x_{18}$  değişkenlerine ait regresyon katsayılarının da önemli olma eğilimi gösterdikleri ( $p < 0.1$ ) anlaşılmaktadır.

Tablo 5. Tek değişkenli modeller için regresyon katsayıları

<b>BD</b>	<b>Katsayı</b>	<b>Kestirim</b>	<b>Std.Hata</b>	<b>z</b>	<b>Pr(&gt; z )</b>
<b>x1</b>	$\beta_0$	8.4540	2.7048	3.126	0.00177**
	$\beta_1$	-0.9006	0.3004	-2.998	0.00271**
<b>x2</b>	$\beta_0$	11.0755	3.5870	3.088	0.00202**
	$\beta_1$	-0.6006	0.2004	-2.996	0.00273**
<b>x3</b>	$\beta_0$	6.6320	2.7799	2.386	0.0170*
	$\beta_1$	-0.1673	0.0742	-2.255	0.0241*
<b>x4</b>	$\beta_0$	-1.4474	3.5586	-0.407	0.684
	$\beta_1$	0.0914	0.1753	0.522	0.602
<b>x5</b>	$\beta_0$	-15.3896	4.5425	-3.388	0.0007***
	$\beta_1$	0.9529	0.2735	3.485	0.0005***
<b>x6</b>	$\beta_0$	-4.7565	2.5718	-1.850	0.0644
	$\beta_1$	0.3835	0.1905	2.013	0.0441*
<b>x7</b>	$\beta_0$	-2.0120	2.7565	-0.730	0.465
	$\beta_1$	0.06368	0.0725	0.878	0.380
<b>x8</b>	$\beta_0$	-6.9032	3.1587	-2.185	0.0289*
	$\beta_1$	0.2241	0.0970	2.310	0.0209*
<b>x9</b>	$\beta_0$	1.44195	2.44815	0.589	0.556
	$\beta_1$	-0.02030	0.04767	-0.426	0.670
<b>x10</b>	$\beta_0$	-3.54718	3.44307	-1.030	0.303
	$\beta_1$	0.06109	0.05320	1.148	0.251
<b>x11</b>	$\beta_0$	-2.73606	2.50355	-1.093	0.274
	$\beta_1$	0.05428	0.04293	1.264	0.206

<b>x12</b>	$\beta_0$	1.5727	2.8130	0.559	0.576
	$\beta_1$	-0.1052	0.2523	-0.417	0.677
<b>x13</b>	$\beta_0$	0.90349	3.12163	0.289	0.772
	$\beta_1$	-0.04733	0.29565	-0.160	0.873
<b>x14</b>	$\beta_0$	-2.4657	2.9258	-0.843	0.399
	$\beta_1$	0.2747	0.2804	0.980	0.327
<b>x15</b>	$\beta_0$	2.60366	3.77122	0.690	0.490
	$\beta_1$	-0.04083	0.06984	-0.585	0.559
<b>x16</b>	$\beta_0$	0.827064	3.074871	0.269	0.788
	$\beta_1$	-0.008096	0.058843	-0.138	0.891
<b>x17</b>	$\beta_0$	-2.5958	2.8863	-0.899	0.368
	$\beta_1$	0.1823	0.1751	1.041	0.298
<b>x18</b>	$\beta_0$	3.6778	3.6219	1.015	0.310
	$\beta_1$	-0.3009	0.3318	-0.907	0.365

Tek deęişkenli modellere ait başarımlar karşılaştırmaları Tablo 5'te gösterilmiştir. Tablo 6'da her bir başarımlar ölçütü için ilk beş modele ait sonuçlar koyu olarak işaretlenmiştir. Lojistik regresyon analizinde AUC ölçütünün %80 ve üstünde olması iyi olarak değerlendirilebilmektedir. Elde edilen sonuçlara göre bu isteęi karşılayan modelin 0.819 AUC deęeri ile x5 deęişkeniyle oluşturulan model olduęu görülmektedir. Bununla birlikte x1 ve x2 deęişkenleri de 0.70 civarında AUC deęerleri verdięi için orta düzeyde sonuçlar vermişlerdir. Dięer deęişkenlerle oluşturulan modeller 0.50-0.65 arasındaki AUC deęerleriyle çok başarısız tahmin gücüne sahiptirler.

AIC ve BIC ölçütleri bakımından yine x5 deęişkeniyle oluşturulan model dięerlerinden daha iyi olup %80 sınıflama doęruluęuna sahiptir. Sonuç olarak iki şakak arası genişlik (x5) deęişkenin basit tek deęişkenli lojistik modellerde sınıflama başarımlarının dięerlerine daha iyi olduęu ve saflık tayininde bir ön sınıflayıcı olarak kullanılabilceęi anlaşılmaktadır.

Tablo 6. Tek değişkenli modellerin analiz sonuçları

	<b>AIC</b>	<b>BIC</b>	<b>AUC</b>	<b>ACC</b>	<b>MFR2</b>	<b>CSR2</b>	<b>CUR2</b>
<b>Model x1</b>	<b>87.663</b>	<b>92.160</b>	<b>0.706</b>	<b>0.686</b>	<b>0.112</b>	<b>0.140</b>	<b>0.189</b>
<b>Model x2</b>	<b>87.165</b>	<b>91.662</b>	<b>0.706</b>	<b>0.714</b>	<b>0.117</b>	<b>0.146</b>	<b>0.198</b>
<b>Model x3</b>	<b>92.661</b>	<b>97.158</b>	<b>0.629</b>	<b>0.671</b>	<b>0.059</b>	<b>0.076</b>	<b>0.103</b>
<b>Model x4</b>	97.949	102.446	0.542	0.600	0.003	0.004	0.005
<b>Model x5</b>	<b>79.407</b>	<b>83.904</b>	<b>0.819</b>	<b>0.800</b>	<b>0.200</b>	<b>0.236</b>	<b>0.319</b>
<b>Model x6</b>	93.740	98.237	0.625	0.571	0.048	0.062	0.084
<b>Model x7</b>	97.392	101.889	0.525	0.614	0.009	0.012	0.016
<b>Model x8</b>	<b>91.366</b>	<b>95.863</b>	<b>0.682</b>	<b>0.657</b>	<b>0.073</b>	<b>0.093</b>	<b>0.126</b>
<b>Model x9</b>	98.040	102.537	0.516	0.600	0.002	0.003	0.004
<b>Model x10</b>	96.851	101.348	0.570	0.614	0.015	0.019	0.026
<b>Model x11</b>	96.369	100.866	0.637	0.614	0.020	0.026	0.035
<b>Model x12</b>	98.047	102.544	0.546	0.600	0.002	0.002	0.003
<b>Model x13</b>	98.196	102.693	0.445	0.600	0.000	0.000	0.000
<b>Model x14</b>	97.014	101.511	0.590	0.586	0.013	0.017	0.023
<b>Model x15</b>	97.878	102.375	0.529	0.600	0.004	0.005	0.007
<b>Model x16</b>	98.203	102.700	0.518	0.600	0.000	0.000	0.000
<b>Model x17</b>	97.114	101.611	0.571	0.586	0.012	0.016	0.021
<b>Model x18</b>	97.388	101.885	0.540	0.629	0.009	0.012	0.016

#### 4.2.2. Çok Değişkenli Modeller

##### 4.2.2.1 Tam Model

İlgilenilen tüm bağımsız değişkenleri içeren modele *tam model* denilir. Çatalburun köpeklerinin saflığına ilişkin tam model Eşitlik 28'de gösterildiği gibi tüm bağımsız değişkenleri içeren bir model olarak oluşturulmuştur.

$$g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{17} X_{17} + \beta_{18} X_{18} + \varepsilon \quad (28)$$

Oluşturulan tam modele ait katsayılar Tablo 7'de gösterildiği gibi saptanmıştır. Sonuçlara göre x1, x2 ve x10 değişkenlerinin katsayılarının önemli olduğu ( $p < 0.05$ ); x18 değişkenin önemli olma eğiliminde olduğu ( $p < 0.1$ ); diğer tüm değişkenlere ait katsayıların ise önemsiz oldukları ( $p > 0.05$ ) saptanmıştır. Bu sonuçlar değişken sayısını azaltmak için kullanılarak yeni bir model oluşturulabileceğine işaret etmektedir.

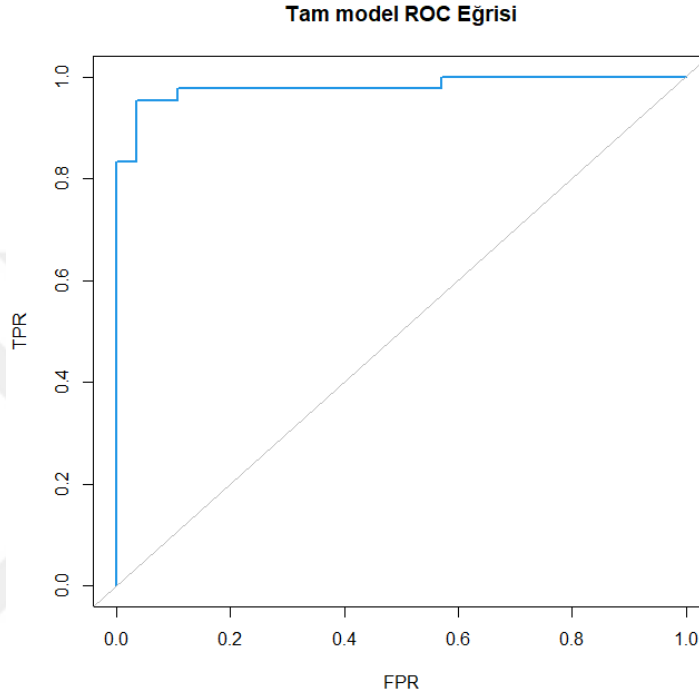
Tablo 7. Tam modele ait analiz sonuçları

<b>Katsayı</b>	<b>Kestirim</b>	<b>SH</b>	<b>z</b>	<b>Pr(&gt; z )</b>
$\beta_0$	7.883617	25.160186	0.313	0.7540
$\beta_1$	-5.276553	2.670429	-1.976	0.0482*
$\beta_2$	-3.904393	1.876712	-2.080	0.0375*
$\beta_3$	-0.300268	0.355993	-0.843	0.3990
$\beta_4$	0.277235	1.130266	0.245	0.8062
$\beta_5$	2.407795	1.357449	1.774	0.0761
$\beta_6$	2.012099	1.259079	1.598	0.1100
$\beta_7$	-0.090407	0.374947	-0.241	0.8095
$\beta_8$	-0.101778	0.252915	-0.402	0.6874
$\beta_9$	0.006635	0.301650	0.022	0.9825
$\beta_{10}$	1.107403	0.544041	2.036	0.0418*
$\beta_{11}$	-0.335718	0.249084	-1.348	0.1777
$\beta_{12}$	-1.057023	1.220005	-0.866	0.3863
$\beta_{13}$	5.724171	3.603025	1.589	0.1121
$\beta_{14}$	2.965025	3.550632	0.835	0.4037
$\beta_{15}$	0.015449	0.603538	0.026	0.9796
$\beta_{16}$	0.083137	0.376586	0.221	0.8253
$\beta_{17}$	-0.051293	1.588367	-0.032	0.9742
$\beta_{18}$	-7.353248	3.829326	-1.920	0.0548

Tam modele ilişkin model başarımlı ölçütleri Tablo 8'de verilmiştir. Sonuçlara göre AUC ölçütünün 0.98 ile mükemmel olduğu ve ayrıca Şekil 8'de görülen ROC eğrisinin de ideale yakın bir davranış gösterildiği anlaşılmaktadır. Tablo 25'teki sonuçlara göre model doğruluğunun (ACC) yaklaşık %93 düzeyinde saptanmış olup oldukça Çatalburun köpeklerinde saflık tayininin tüm bağımsız değişkenlerin kullanıldığı model ile başarılı bir şekilde yapılabileceğini göstermektedir. Aynı şekilde MacFadden sahte R<sup>2</sup> değerinin %75 civarında yüksek saptanması da bu tespiti doğrulamaktadır.

Tablo 8. Tam modele ait başarımlar ölçütleri

<i>AIC</i>	<i>BIC</i>	<i>AUC</i>	<i>ACC</i>	<i>MFR2</i>	<i>CSR2</i>	<i>CUR2</i>
61.769	104.490	0.980	0.929	0.748	0.634	0.858



Şekil 8: Tam modele ait ROC eğrisi

#### 4.2.2.2. Seçilen Değişkenlerle Oluşturulan Modeller

Yukarıda saptandığı gibi tam modelin başarımlarını yüksek olmakla birlikte bu modelde 18 bağımsız değişkenin yer alması çeşitli dezavantajlara yol açar. Bu kadar çok sayıda bağımsız değişkenin pratikte ölçülmesi tahmin yapılmasında zaman ve maliyet gücünü yaratır. Diğer yandan bir modele ne kadar çok bağımsız değişken eklenirse modelin başarımlarının yükselebileceği bilinen bir gerçektir. Böyle bir durumda aşırı uyum sorununun da söz etmek mümkündür. Bu sorunlardan kaçınmanın çeşitli yolları mevcuttur. Bunlardan biri yanıt değişkenini açıklamada katkısı yüksek olan değişkenleri saptamaktır. Bu değişkenlerden bazıları yanıt değişkenindeki

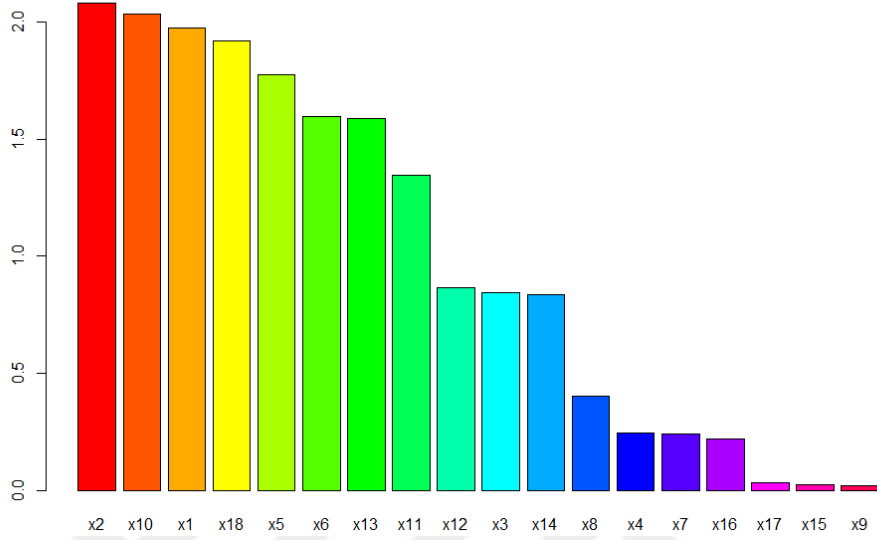
varyasyonun çok büyük bölümünü açıklayacağından tam modele yakın başarımlı düzeylerine çok daha az değişkenle erişilebilmesi olasıdır. Bu tür indirgenmiş modelleri oluşturmak bilgisayarda işlem süresini de düşüreceğinden büyük hacimli verilerde işlem maliyetini de düşürecektir. Değişken seçiminde uygulanabilecek ayırma analizi, değişken önemliliği saptama ve kademeli regresyon analizi gibi çeşitli yöntemler söz konusudur. Bu çalışmada değişken önemliliğine ve kademeli lojistik analizine göre değişken seçimi olmak üzere farklı iki yöntem uygulanmaktadır.

#### 4.2.2.2.(1). Değişken Önemliliği

R'nin *caret* paketindeki (Kuhn, 2020) *varImp* fonksiyonu kullanılarak tam modeldeki terimlerin yani bağımsız değişkenlerin her biri için *önem dereceleri* (*VIMP: Variable Importance*) ve sıra sayıları (rank) saptanarak Tablo 9'da listelenmiş ve Şekil 9'da grafik halinde verilmiştir.

Tablo 9. Bağımsız değişkenlerin önem değerleri ve sıraları

<b>BD</b>	<b>VIMP</b>	<b>Sıra</b>
<b>x1</b>	1.97591997	3
<b>x2</b>	2.08044288	1
<b>x3</b>	0.84346435	10
<b>x4</b>	0.24528257	13
<b>x5</b>	1.77376478	5
<b>x6</b>	1.59807159	6
<b>x7</b>	0.24112038	14
<b>x8</b>	0.40241878	12
<b>x9</b>	0.02199688	18
<b>x10</b>	2.03551372	2
<b>x11</b>	1.34780974	8
<b>x12</b>	0.86640869	9
<b>x13</b>	1.58871262	7
<b>x14</b>	0.83506956	11
<b>x15</b>	0.02559696	17
<b>x16</b>	0.22076430	15
<b>x17</b>	0.03229260	16
<b>x18</b>	1.92024615	4



Şekil 9: Değişkenlerin önemleri çubuk grafiği

Elde edilen sonuçlara göre değişkenlerin önem değerleri en düşük 0.02199688 ile x9 değişkeni ve en yüksek 2.08044288 ile x2 değişkeni için saptanmıştır. Tablo 9'daki sıralamaya göre en önemli değişkenlerin sırasıyla burun bir kanat uzunluğu (x2) olduğu; bunu göğüs çevresi (x10), burun çapı (x1), iki şakak arası genişlik (x5), iki kulak arası uzunluk (x6) izlediği görülmektedir. Burada ortaya çıkan sorun hangi değişkenlerin seçileceği ve modele alınması gerektiğidir. Bu konuda deterministik bir yöntem bulunmamakta; yani seçimler sezgisel yaklaşıma dayanmaktadır. Öte yandan Cebeci (2020a) değişkenlerin önem değerlerine göre sıralanmasından sonra 75. yüzdellik ( $P_{75}$ ) ve 90. yüzdellik ( $P_{90}$ ) değerlerine karşılık gelen önem değerlerinin eşik ya da kesim noktası olarak kullanılabilceğini; bu eşığe eşit veya büyük önem derecesine sahip değişkenlerin seçilebileceğini bildirmişlerdir. Bu öneri doğrultusunda Tablo 9'daki önem değerleri için  $P_{75} = 1.729841$  ve  $P_{90} = 1.993798$  olarak saptanmıştır. Buna göre  $P_{90}$  yüzdeliğinden büyük önem değerine

sahip değişkenler  $x_2$  ve  $x_{10}$  değişkenleri olmuştur. Seçilen bu iki değişken için model Eşitlik 29'daki gibi oluşturulmuş ve Model S1 olarak adlandırılmıştır.

*Model S1:*

$$g(Y) = \beta_0 + \beta_1 X_2 + \beta_2 X_{10} + \varepsilon \quad (29)$$

Benzer şekilde  $P_{75}$  yüzdeliğinden büyük önem derecesine sahip değişkenlerin  $x_1$ ,  $x_2$ ,  $x_5$ ,  $x_{10}$  ve  $x_{18}$  değişkenleri olduğu belirlenmiştir. Seçilen bu iki değişken için model Eşitlik 30'daki gibi oluşturulmuş ve Model S2 olarak adlandırılmıştır.

*Model S2:*

$$g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_5 + \beta_4 X_{10} + \beta_5 X_{18} + \varepsilon \quad (30)$$

Değişken önemliliğine göre değişken seçerken model başarımı için yeterli olabileceği düşünülen değişkenler de keyfi olarak seçilebilir. Eşitlik 31'de verilen Model S3 önemlilik derecesi en yüksek  $x_1$ ,  $x_2$  ve  $x_{10}$  değişkenlerinden oluşturulmuştur.

*Model S3:*

$$g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{10} + \varepsilon \quad (31)$$

#### 4.2.2.2.(2). Kademeli Lojistik Regresyon İle Değişken Seçimi

Lojistik regresyonda yaygın şekilde uygulanan değişken seçimi yöntemlerinden biri kademeli lojistik regresyon (stepwise logistic regression) yöntemini kullanmaktır. Bu regresyon ileriye doğru eklemeli, geriye doğru elemeli veya her iki yönde eklemeli ve elemeli olarak uygulanabilmektedir. Bu çalışmada her iki yönde kademeli lojistik regresyon analizi sonucunda  $x_1$ ,  $x_2$ ,  $x_5$ ,  $x_6$ ,  $x_{10}$ ,

x14, x18 değişkenleri önemli değişkenler olarak saptanmıştır. Bu değişkenlerden oluşturulan Model S4 Eşitlik 32'de verilmiştir.

*Model S4:*

$$g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_5 + \beta_4 X_6 + \beta_5 X_{10} + \beta_6 X_{14} + \beta_7 X_{18} + \varepsilon \quad (32)$$

Daha önce Bölüm 4.1'de yapılan varsayım kontrolünde Model S4'teki x2, x5, x6 ve x18 değişkenleri ile logit arasındaki ilişki doğrusal bulunmadığından Model S4 bu değişkenlerin logaritmaları alınarak Eşitlik 33'teki model oluşturulmuştur.

*Model S4a:*

$$g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_5 + \beta_4 \ln X_6 + \beta_5 X_{10} + \beta_6 X_{14} + \beta_7 \ln X_{18} + \varepsilon \quad (33)$$

#### 4.2.2.2.(3). Seçilen Değişkenlerle Oluşturulan Modellere Ait Analizler

Eşitlik 20'deki Model S1 modeli için VIF değerleri x2 ve x10 için sırasıyla 2.030561 ve 2.030561 olarak saptanmış olup 5'ten küçük oldukları için çoklu-bağlanımlılık olmadığı sonucuna varılmıştır. Modele ait analiz sonuçları Tablo 10'da görülmektedir. Tablodaki sonuçlara göre kesme yüksekliği ( $\beta_0$ ) dışındaki tüm katsayılar önemli bulunmuşlardır ( $p < 0.01$ )

Tablo 10. Model S1 analizi sonuçları

BD	Katsayı	Kestirim	SH	z	Pr(> z )
	$\beta_0$	3.78848	4.28616	0.884	0.376757
x2	$\beta_1$	-1.11198	0.30583	-3.636	0.000277 ***
x10	$\beta_2$	0.25520	0.08966	2.846	0.004422 **

Eşitlik 30'da verilen Model S2'ye ait VIF değerleri x1, x2, x5, x10 ve x18 için sırasıyla 1.511402, 2.120757, 1.310981, 3.062369 ve 1.722682 olarak saptanmış olup çoklu-bağlanımlılık olmadığı görülmüştür. Tablo 11'de gösterilen analiz

sonuçlarına göre x18 değişkeni hariç diğer değişkenlere ait regresyonların sıfırdan önemli derecede farklı oldukları ( $p < 0.01$ ) görülmektedir.

Tablo 11. Model S2 analizi sonuçları

<b>BD</b>	<b>Katsayı</b>	<b>Kestirim</b>	<b>SH</b>	<b>z</b>	<b>Pr(&gt; z )</b>
	$\beta_0$	-2.4359	7.3521	-0.331	0.74041
<b>x1</b>	$\beta_1$	-1.4420	0.5413	-2.664	0.00773 **
<b>x2</b>	$\beta_2$	-1.3798	0.4342	-3.178	0.00149 **
<b>x5</b>	$\beta_3$	1.0619	0.3334	3.185	0.00145 **
<b>x10</b>	$\beta_4$	0.5147	0.1570	3.278	0.00105 **
<b>x18</b>	$\beta_5$	-0.9548	0.6069	-1.573	0.11568

Eşitlik 31'deki Model S3 modelinde VIF değerleri x1,x2 ve x10 için sırasıyla 1.416300, 2.629437 ve 3.141374 saptanmış olup çoklu-bağımlılık olmadığı görülmüştür. Model ait analiz sonuçları Tablo 12'de görülmektedir. Elde edilen sonuçlara göre modelde yer alan x1, x2 ve x10 değişkenlerine ait regresyonlar sıfırdan önemli derecede farklı bulunmuşlardır ( $p < 0.01$ ).

Tablo 12. Model S3 analizi sonuçları

<b>BD</b>	<b>Katsayı</b>	<b>Kestirim</b>	<b>SH</b>	<b>z</b>	<b>Pr(&gt; z )</b>
	$\beta_0$	9.3281	5.2372	1.781	0.074893.
<b>x1</b>	$\beta_1$	-1.4277	0.4554	-3.135	0.001720 **
<b>x2</b>	$\beta_2$	-1.3060	0.3764	-3.470	0.000521 ***
<b>x10</b>	$\beta_3$	0.4214	0.1265	3.332	0.000862 ***

Eşitlik 32'de görülen Model S4 modelinde VIF değerleri x1, x2, x5 x6, x10, x14 ve x18 değişkenleri için sırasıyla 3.023204, 3.628704, 1.558069, 2.798935, 5.438208, 2.754153 ve 4.394348 olarak saptanmış ve çoklu bağımlılık olmadığı saptanmıştır. Model S4'e ait analiz sonuçları Tablo 13'te verilmiştir. Modelde yer alan x6 ve x10 değişkeni hariç tüm diğer tüm değişkenlerin regresyonları önemli bulunmuştur ( $p < 0.05$ ).

Tablo 13. Model S4 analizi sonuçları

<b>BD</b>	<b>Katsayı</b>	<b>Kestirim</b>	<b>SH</b>	<b>z</b>	<b>Pr(&gt; z )</b>
	$\beta_0$	-8.1090	9.9077	-0.818	0.41310
<b>x1</b>	$\beta_1$	-2.5417	0.8730	-2.912	0.00360 **
<b>x2</b>	$\beta_2$	-2.3803	0.7508	-3.170	0.00152 **
<b>x5</b>	$\beta_3$	0.9611	0.4078	2.3570	0.01842 *
<b>x6</b>	$\beta_4$	0.9365	0.5541	1.6900	0.09100 .
<b>x10</b>	$\beta_5$	0.7964	0.2622	3.0370	0.00239 **
<b>x14</b>	$\beta_6$	2.4582	1.1066	2.2210	0.02632 *
<b>x18</b>	$\beta_7$	-2.8503	1.1649	-2.447	0.01441 *

Tablo 14'te Eşitlik 33'teki Model S4a'nın analiz sonuçları görülmektedir. Sonuçlara göre ln(x6) terimi dışında regresyon katsayıları önemli bulunmuştur.

Tablo 14. Model S4a analizi sonuçları

<b>BD</b>	<b>Katsayı</b>	<b>Kestirim</b>	<b>SH</b>	<b>z</b>	<b>Pr(&gt; z )</b>
	$\beta_0$	74.9433	31.2996	2.394	0.01665 *
<b>x1</b>	$\beta_1$	-2.5934	0.8855	-2.929	0.00340 **
<b>ln x2</b>	$\beta_2$	-43.9167	13.8210	-3.178	0.00149 **
<b>ln x5</b>	$\beta_3$	14.1240	6.1332	2.303	0.02129 *
<b>ln x6</b>	$\beta_4$	12.8487	7.3436	1.750	0.08018 .
<b>x10</b>	$\beta_5$	0.8289	0.2726	3.041	0.00236 **
<b>x14</b>	$\beta_6$	2.4662	1.0990	2.244	0.02483 *
<b>ln x18</b>	$\beta_7$	-31.9604	12.8259	-2.492	0.01271 *

#### 4.2.2.2.(4). Seçilen Değişkenlerle Oluşturulan Modellerin Başarımları

Seçilen değişkenlerle oluşturulan Model S1, Model S2, Model S3 ve Model S4 ve Model S4a çeşitli başarımlar ölçütlerine göre karşılaştırılmış ve elde edilen sonuçlar Tablo 15'te listelenmiştir.

Tablo 15. Model başarımlarının karşılaştırılması

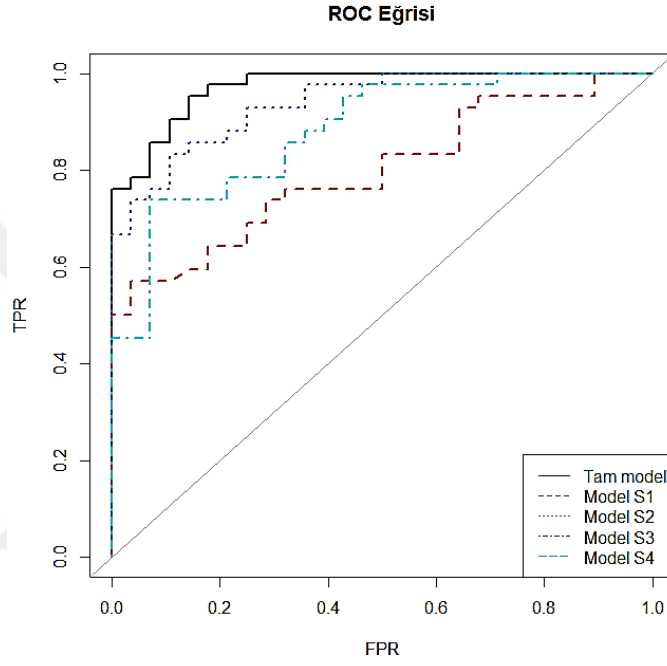
<b>Model</b>	<b>AIC</b>	<b>BIC</b>	<b>AUC</b>	<b>ACC</b>	<b>MFR2</b>	<b>CSR2</b>	<b>CUR2</b>
<b>Model S1</b>	78.081	84.826	0.797	0.714	0.235	0.271	0.367
<b>Model S2</b>	55.448	68.939	0.939	0.829	0.539	0.516	0.697
<b>Model S3</b>	65.888	74.882	0.880	0.786	0.386	0.405	0.547
<b>Model S4</b>	47.194	65.182	0.972	0.886	0.669	0.594	0.802
<b>Model S4a</b>	47.841	65.828	0.970	0.885	0.669	0.594	0.802
<b>Tam Model</b>	61.769	104.490	0.980	0.929	0.748	0.634	0.858

Tablo 15'teki sonuçlara göre AUC değerleri Model S1, Model S2, Model S3, Model S4, Model S4a ve Tam model için sırasıyla 0.797, 0.939, 0.880, 0.972, 0.970 ve 0.980 olarak saptanmış olup en yüksek AUC değerini Tam modelin verdiği, bunu çok yakın bir değerle Model S4'ün izlediği Şekil 10'dan da görülebilmektedir. Beklendiği üzere modele eklenen her bir terim performans artışına katkıda bulunmuştur. AUC ölçütlerine göre Model S4 ile Tam Model karşılaştırıldığında tam modele fazladan değişken eklemenin model başarımına katkısının ancak %1'den çok daha düşük bir düzeyde olduğu anlaşılmaktadır. Benzer şekilde yalnızca dört değişkenin ( $x_1$ ,  $x_2$ ,  $x_5$  ve  $x_{10}$ ) yer aldığı Model S2 başarımının AUC ölçütünün 0.939 gibi mükemmel yakın bulunması da önemli bir sonuç olup bu modelin de saflık tayininde karşılaştırılabilir bir başarımlı gösterebileceğine işaret etmektedir. Çünkü saflık tayinini daha kısa sürede ve daha kolay yapılabilmesi mümkün olacaktır.

Tablo 15 incelendiğinde en başarılı modelin sırasıyla 47.194 AIC ve 65.182 BIC değeri ile Model S4 olduğu görülmektedir. Bu modeli 47.841 AIC ve 65.828 BIC değeriyle Model S4a izlemiştir. Gerek bu ölçütler ve gerekse diğer ölçütler bakımından Model S4 ve Model S4a arasında farklılığın önemli olmadığı, dolayısıyla logaritmik dönüşümle yapılan dönüştürmenin sonuçları farklılaştırmadığı anlaşılmakta, kaldı ki ek hesaplama maliyeti getirmesi nedeniyle Model S4'ün kullanılması önerilmektedir. Bunun alternatifi olarak logit-bağımsız değişken arası ilişkinin doğrusallığını ihlali halinde bağımsız değişkeni kesikleştirme yoluyla kategorik değişkene dönüştürmek ve modele böylece eklemek de mümkündür. Ancak sorunun bu çalışma yerine çalışmanın gelecekte yürütülecek safhalarında geniş olarak incelenmesi planlanmıştır. Yapılan karşılaştırmalarda Model S4'ü 55.448 AIC ve 68.939 BIC değeriyle Model S2 izlemiştir. Tam modelin AIC ve BIC ölçütlerine göre başarımlı düzeyi ise Model S4 ve Model S2'nin gerisinde kalmaktadır.

Modellerin doğruluk değerlerine sıralamasına gelince en yüksek doğruluk %92.9 değeriyle tam modele ait iken bunu %88.6 doğruluk değeriyle Model S4

izlemiştir. Model S1'nin doğruluğu %71.4, Model S2'nin doğruluğu %82.9 ve Model S3'ün doğruluğu ise %78.6 olarak saptanmıştır. Saptanan doğruluk dereceleri itibariyle tam modelin daha başarılı olduğu ancak Model S4'ün makul sayılabilecek bir doğruluk gösterdiği anlaşılmaktadır.



Şekil 10: Karşılaştırılan modellere ait ROC eğrileri grafiği

Tablo 14'de MFR2, CSR2 ve CUR2 ölçütleri bakımından yapılan başarımların sıralaması *Tam Model* > *Model S4* > *Model S2* > *Model S3* > *Model S1* şeklinde saptanmıştır. Bu ölçütlere göre değerlendirme yapıldığında da tam modelden sonra en başarılı modelin Model S4 olduğu anlaşılmaktadır.



## 5. SONUÇ VE ÖNERİLER

Bu tez çalışmasında basitliği ve uygulanabilirliği yüksek sınıflayıcılardan biri olan lojistik regresyon ile Çatalburun köpek ırkında çeşitlik morfolojik özellikler kullanılarak saflık tayini başarımı çalışılmıştır.

Morfolojik özelliklerin her biri için ayrı ayrı oluşturulan tek değişkenli modellerle elde edilen bulgulara göre burun çapı (x1), burun bir kanat uzunluğu (x2), burun iki kanat uzunluğu (x3), iki şakak arası genişlik (x5), iki kulak arası uzunluk x(6), iki şakak ortası ile iki scapula arası uzunluk (x8) değişkenleri için oluşturulan modellerde regresyonlar önemli bulunmasına karşın modellerin hiçbiri kabul edilebilir bir başarıma düzeyine erişilememiştir. Bunlar arasında iki şakak arası genişlik (x5) değişkeni, 0.819 düzeyinde AUC ölçütü değeriyle en yüksek başarımlı modeli vermektedir. Bu değişkeni 0.706 gibi daha düşük düzeyli AUC değerleriyle burun çapı (x1) ve burun bir kanat uzunluğu (x2) değişkenleri takip etmektedir. AIC, BIC, ACC ve diğer başarıma ölçütleri açısından da sıralama benzerdir. Bu sonuçlar Çatalburun köpeğinin söz konusu değişkenlerle karakterize edilen bir ırk olduğunu teyit etmektedir. Sonuç olarak basit lojistik regresyon ile saflık tayini için kullanılacak tek değişkenli modelin ancak x5 değişkeni için uygulanabilir olduğu, ancak doğruluk düzeyinin 0.801 gibi düşük olması nedeniyle saflık tayininde tek değişkenli modellerin iyi çalışmadığı görülmektedir.

Bağımsız değişkenlerin tümü alınarak oluşturulan tam modelin AUC ölçütünün 0.98 ile mükemmel olduğu ve ayrıca ROC eğrisinin de ideale yakın bir davranış gösterdiği saptanmıştır. Tam modelin sınıflama doğruluğunun da yaklaşık %93 düzeyinde saptanmış olup oldukça Çatalburun köpeklerinde saflık tayininin tüm bağımsız değişkenlerin kullanıldığı modelle başarılı bir şekilde yapılabileceği ortaya göstermektedir.

Tam modelin başarımı yüksek olmakla birlikte bu modeli kullanarak tahmin yapmak için çok sayıda değişkenin ölçülmesi gerekli olmaktadır. Bu nedenle

değişkenlerden önemli olanların seçilerek tam modele yakın bir başarımlı verebilecek modeller tercih edilir. Değişken önem derecesi ve kademeli lojistik regresyonla yapılan değişken seçimine göre x1, x2, x5, x6, x10, x14 ve x18 değişkenlerini içeren indirgenmiş modelin 0.972 AUC değeri ve %88.6 sınıflama doğruluğu ile tam modele çok yakın bir başarımlı göstermesi sahada yapılacak saflık çalışmalarında bu modelin başarımlı kullanılabileceğini göstermektedir. Öte yandan bu modelin AIC ve BIC değerlerinin tam modele göre küçük olması da saflık tayininde bu modelin kullanılabileceğini ortaya koymaktadır. Bu sonuçlar aynı zamanda burun çapı (x1), burun bir kanat uzunluğu (x2), iki şakak arası genişlik (x5), iki kulak arası uzunluk (x6), göğüs çevresi (x10), arka pati bilek kalınlığı (x14) ve kulak genişliği (x18) özelliklerinin ırkın saflığını belirleyici temel özellikler olduğuna işaret etmektedir.

Bu çalışmada lojistik regresyon ile Çatalburun köpek ırkında sürekli morfolojik özellikleri kullanarak başarımlı sınıflama yapılabileceği ortaya konulmuştur. Gelecek yıllarda ırka ait daha fazla veri elde edilmesiyle daha yüksek başarımlı modellerin elde edilmesi mümkün olabilecektir. Ancak şu anda ortaya konulan model ile sahadan koruma amacıyla uzman olmayan kişilerce toplanacak saf köpekleri saptamak için kullanılabilecek bir kolaylık sağlanmış olmaktadır. Modele göre seçilen hayvanların sayısı daha az olacağından merkezde uzman değerlendirmesi için gerekli zaman ve toplama maliyeti de düşürülmüş olacaktır.

## KAYNAKLAR

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6): 716–723.
- Akın, H. B. & Şentürk, E. (2012). Bireylerin Mutluluk Düzeylerinin Ordinal Lojistik Regresyon Analizi ile İncelenmesi. *Öneri Dergisi*, 10 (37) , 183-193 .
- Aktaş, Y. A. (2009). Okul Öncesi Dönemde Matematik Eğitimi. Nobel Kitabevi.
- Arı, A. & Onder, H. (2013). Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri. *Anadolu Tarım Bilimleri Dergisi*, 28 (3), 168-174.
- Aygün, T. ve Çelikyürek, H. (2017). Büyükbaş Hayvancılıkta Kayıt Tutma Sistemine Yönelik Bir Bilgisayar Paket Programının Hazırlanması. *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi*, 27(3): 453-459.
- Box, G. E. P. ve Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics* 4, 531-550.
- Barbosa, A.M., Real, R., Munoz, A.R. & Brown, J.A. (2015). New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Diversity and Distributions*, 19(10):1333-1338.
- Bircan, H. (2004). Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Dergisi*, 2: 185-208.
- Burmaoğlu, S., Oktay, E. Ve Özen, Ü. (2009). Birleşmiş Milletler Kalkınma Programı Beşerî Kalkınma Endeksi Verilerini Kullanarak Diskriminant Analizi ve Lojistik Regresyon Analizinin Sınıflandırma Performanslarının Karşılaştırılması. *Savunma Bilimleri Dergisi*, 8(2), 23-49.
- Cebeci, Z. (2019). R ile Parametrik Olmayan İstatistik Analiz.1.Baskı, Abaküs Yayınevi, İstanbul. 516 s.
- Cebeci, Z. (2020a). Ver, Bilimde R İle Ver Önışlem.1.Baskı, Nobel Akademik Yayıncılık, Ankara. 534 s.

- Cebeci, Z. (2020b). R Uygulamalı Yeniden Örnekleme Teknikleri.1.Baskı, Pegem Akademi, Ankara. 632 s.
- Compton, B. W., Rhymer, J. M. ve McCollough, M. (2002). Habitat selection by wood turtles (*Clemmys insculpta*): an application of paired logistic regression. *Ecology*, 83(3): 833-843.
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2): 215-232.
- Cox, D.R. ve Snell, E.J. (1989) The Analysis of Binary Data 2nd ed. London: Chapman and Hall.
- Çokluk, Ö. (2010). Lojistik Regresyon Analizi: Kavram ve Uygulama. Ankara Üniversitesi Eğitim Fakültesi Ölçme ve Değerlendirme Bölümü. 1359-1407.
- Dinçel, D. ve Petek, M. (2012). Çatalburun Köpekleri ve Irk Özellikleri. *Uludağ Üniversitesi Veteriner Fakültesi Dergisi*, 1: 25-28.
- Dinçer, R. Ü. (2006). Tarsus Çatalburun Av Köpekleri (Türk Pointer'i). Tarsus Belediyesi Kültür Yayınları, Mersin. 107s.
- Dinse, G. E. ve Haseman, J. K. (1986). Logistic regression analysis of incidental-tumor data from animal carcinogenicity experiments. *Toxicological Sciences*, 6(1):44-52.
- Drobnjak, D. Oğrak, Y. Z. ve Urosevic, M. (2012). Tarsus Çatalburun Breed of Turkish Hunting Dog (Turkish Pointer). *Dicle Üniversitesi Veteriner Fakültesi Dergisi*, 2(5): 71-74.
- Ertuğrul, M. ve Yılmaz, O. (2012). Türkiye yerli köpek ırk ve tipleri. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2(1): 99-106.
- Eyduran, E., Ozdemir, T., Çak, B. ve Alarslan, E. (2005). Using of logistic regression in Animal Science. *Journal of Applied Sciences*, 5: 1753-1756.
- Farhangfar, H. Rowlinson, P. Hassani, S. ve Nasri, M. F. (2010). Logistic regression analysis of some environmental factors affecting days open in Iranian primiparous Holstein cows. *Advances in Animal Biosciences*, 1(1), 276.

- Fox, J. ve Weisberg, S. (2019). *An R Companion to Applied Regression*, 3rd Edn. Thousand Oaks CA: Sage.  
(<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>)
- Girginer, N., Cankuş, B., “Eskişehir’de belediye hizmetlerine ilişkin algı ve memnuniyetin ikili lojistik regresyon analizi ile ölçülmesi”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, Vol. 11, No. 1, Haziran 2010, s.297-323.
- Gregory, D. ve Bader, K. (2018). Logistic and linear regression assumptions: Violation recognition and control. In *Proc. of Midwest SAS User Group 2018 Conference, Indiana, AA-091*. pp. 1-22.
- Hunt, T. (2020). *ModelMetrics: Rapid Calculation of Model Metrics*. R package version 1.2.2.2. URL: <https://CRAN.R-project.org/package=ModelMetrics>
- Jackman, S. (2020). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. United States Studies Centre, University of Sydney. Sydney, New South Wales, Australia. R package version 1.5.5. URL <https://github.com/atahk/pscl/>
- Keshavarzi, H. et al (2017). Abortion studies in Iranian dairy herds: I. Risk factors for abortion. *Livestock Science*, 195: 45-52.
- Kuhn, M. (2020). *Caret: Classification and Regression Training*. R package version 6.0-86. URL <https://CRAN.R-project.org/package=caret>
- Kneib, T. Knauer, F. ve Küchenhoff, H. (2011). A general approach to the analysis of habitat selection. *Environmental and Ecological Statistics*, 18(1):1-25.
- Köksal, B. (2011). Regresyon analizinde ROC eğrisi kestrimi ile model seçimi. M. Ü. Yüksek Lisans Tezi, İstanbul, 172 s.
- Lepper, M., Kass, P. H., ve Hart, L. A. (2002). Prediction of adoption versus euthanasia among dogs and cats in a California animal shelter. *Journal of Applied Animal Welfare Science*, 5(1), 29-42.

- Long, J. Scott (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.
- McFadden, D. (1979). Quantitative methods for analysing travel behavior of individuals: Some recent developments. In D. A. Hensher & P. R. Stopher (Eds.), Behavioural travel modelling (pp. 279-318). London: Croom Helm.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3): 691–692.
- Nelder, J. A. ve Wedderburn, R. W. M (1972). Generalized Lineal Model. *Journal of the Royal Statistical Society*, 370-384.
- Nemes, S. ve Hartel, T. (2010). Summary measures for binary classification systems in animal ecology. *North-Western Journal of Zoology*, 6(2): 323-330.
- North, M. P. ve Reynolds, J. H. (1996). Microhabitat analysis using radiotelemetry locations and polytomous logistic regression. *The Journal of Wildlife Management*, 639-653.
- Oğuzlar, A. (2010). Lojistik Regresyon Analizi Yardımıyla Suçlu Profiline Belirlenmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 19 (1), 21-35.
- Pereira, D. F., Oliveira, S. C. D. ve Penha, N. L. (2011). Logistic regression to estimate the welfare of broiler breeders in relation to environmental and behavioral variables. *Engenharia Agrícola*, 31(1):33-40.
- R Development Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6 (2), 461–464.
- Signorell, A. et mult. al. (2021). DescTools: Tools for descriptive statistics. R package version 0.99.44. URL: <https://cran.r-project.org/package=DescTools>

- Smith, T. J. ve McKenna, C. M. (2013). A comparison of logistic regression pseudo  $R^2$  indices. *Multiple Linear Regression Viewpoints*, 39(2):17-26.
- Tomaszyk, K., Dobek, A., Moliński, K., Gut, A., ve Szwaczkowski, T. (2014). Synergy factors in the analysis of lamb survival. *Italian Journal of Animal Science*, 13(4): 3337.
- Topal, M., Emsen, E., Yağanoğlu, A. M., ve Unit, A. R. (2017). Chaid And Logistic Regression Approaches For Assessing The Effects Of Non-Genetic Factors On Lamb Mortality. *Journal of Animal and Plant Sciences*, 27(1): 40-47.
- Wickham, H (2009). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.
- Wei, T., Simko, W. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from <https://github.com/taiyun/corrplot>



## ÖZGEÇMİŞ

Mustafa Ersen Bebek, ilk ve orta öğrenimini Mersin il merkezindeki İleri İlkokulu'nda (2001), Mezitli Ortaokulu'nda (2004), Dumlupınar Lisesi'nde (2007) tamamlamıştır. Yükseköğrenimini Van Yüzüncü Yıl Üniversitesi Veteriner Fakültesinde 2015 yılında tamamlamıştır. Ayrıca Erasmus Programı kapsamında 2013-2014 ilk yarıyılı Romanya Iasi'daki Ion Ionescu de la Brad University of Agricultural Sciences and Veterinary Medicine'de öğrenim görmüştür. Stajını Romanya Iasi Ion Ionescu de la Brad University of Agricultural Sciences and Veterinary Medicine 'de 2013 yaz ayı boyunca üç ay süre ile pet hayvan hastalıkları üzerine yapmıştır. Anadolu Üniversitesi Sağlık Yönetimi Bölümü'nde öğrenimine 4. sınıfta devam etmektedir. Mersin Büyükşehir Belediyesi Tarımsal ve Veteriner Hizmetleri Daire Başkanlığı'na bağlı Evcil ve Sahipsiz Hayvanlar Şube Müdürlüğü'nde 2015 yılından itibaren veteriner hekim olarak çalışmaktadır.



# **EKLER**



## EK 1. Analizlerde Kullanılan R betiği

```
# Verisetinin okunması
> setwd("D:/LisansüstüÖğrenciler/Mustafa Bebek")
> if(!require(readxl)){
+   install.packages("readxl",
+   repo="https://cloud.r-project.org");
+   library(readxl)
+ }

> catalburun = read_excel("catalburun.xlsx", sheet=1, col_names=TRUE)
> catalburun <- data.frame(catalburun[,-c(1:5)])
> catalburun$y <- as.factor(catalburun$y)

> str(catalburun)
> head(catalburun)
> table(catalburun$y)

#TEK DEĞİŞKENLİ MODELLER
> cbx1 = glm(y~x1, data=catalburun, family="binomial")
> summary(cbx1)
> cbx2 = glm(y~x2, data=catalburun, family="binomial")
> summary(cbx2)
> cbx3 = glm(y~x3, data=catalburun, family="binomial")
> summary(cbx3)
> cbx4 = glm(y~x4, data=catalburun, family="binomial")
> summary(cbx4)
> cbx5 = glm(y~x5, data=catalburun, family="binomial")
> summary(cbx5)
> cbx6 = glm(y~x6, data=catalburun, family="binomial")
> summary(cbx6)
> cbx7 = glm(y~x7, data=catalburun, family="binomial")
> summary(cbx7)
> cbx8 = glm(y~x8, data=catalburun, family="binomial")
> summary(cbx8)
> cbx9 = glm(y~x9, data=catalburun, family="binomial")
> summary(cbx9)
> cbx10 = glm(y~x10, data=catalburun, family="binomial")
> summary(cbx10)
> cbx11 = glm(y~x11, data=catalburun, family="binomial")
> summary(cbx11)
> cbx12 = glm(y~x12, data=catalburun, family="binomial")
> summary(cbx12)
> cbx13 = glm(y~x13, data=catalburun, family="binomial")
> summary(cbx13)
> cbx14 = glm(y~x14, data=catalburun, family="binomial")
> summary(cbx14)
> cbx15 = glm(y~x15, data=catalburun, family="binomial")
> summary(cbx15)
```

```

> cbx16 = glm(y~x16, data=catalburun, family="binomial")
> summary(cbx16)
> cbx17 = glm(y~x17, data=catalburun, family="binomial")
> summary(cbx17)
> cbx18 = glm(y~x18, data=catalburun, family="binomial")
> summary(cbx18)

#TEK DEĞİŞKENLİ MODELLER İÇİN BAŞARIM ÖLÇÜTLERİ

# Paketleri yükle
> if(!require(pscl))
+ {install.packages("pscl",
+   repo="https://cloud.r-project.org");
+   library(pscl)}

> if(!require(ModelMetrics)){
+   install.packages("ModelMetrics",
+     repo="https://cloud.r-project.org");
+     library(ModelMetrics)}

> if(!require(DescTools)){
+   install.packages("DescTools",
+     repo="https://cloud.r-project.org");
+     library(DescTools)}

> performance <- c()
> actual <- catalburun$y
> pR2 <- c("McFadden", "McFaddenAdj", "CoxSnell", "Nagelkerke",
+   "AldrichNelson", "VeallZimmermann", "Efron", "McKelveyZavoina",
+   "Tjur", "all")

#x1
> preds = predict(cbx1, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx1)
> aic = AIC(cbx1)
> bic = BIC(cbx1)
> other = pR2(cbx1)[4:6] #veya alternatif olarak aşağıdaki komut
> other = PseudoR2(cbx1, which="all")[c(1,3,4)]
> Model.1 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.1)

#x2
> preds = predict(cbx2, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx1)
> aic = AIC(cbx2)
> bic = BIC(cbx2)

```

```

> other = pR2(cbx2)[4:6]
> Model.2 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.2)

#x3
> preds = predict(cbx3, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx3)
> aic = AIC(cbx3)
> bic = BIC(cbx3)
> other = pR2(cbx3)[4:6]
> Model.3 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.3)

#x4
> preds = predict(cbx4, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx4)
> aic = AIC(cbx4)
> bic = BIC(cbx4)
> other = pR2(cbx4)[4:6]
> Model.4 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.4)

#x5
> preds = predict(cbx5, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx5)
> aic = AIC(cbx5)
> bic = BIC(cbx5)
> other = pR2(cbx5)[4:6]
> Model.5 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.5)

#x6
> preds = predict(cbx6, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx6)
> aic = AIC(cbx6)
> bic = BIC(cbx6)
> other = pR2(cbx6)[4:6]
> Model.6 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.6)

```

```

#x7
> preds = predict(cbx7, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx7)
> aic = AIC(cbx7)
> bic = BIC(cbx7)
> other = pR2(cbx7)[4:6]
> Model.7 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.7)

#x8
> preds = predict(cbx8, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx8)
> aic = AIC(cbx8)
> bic = BIC(cbx8)
> other = pR2(cbx8)[4:6]
> Model.8 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.8)

#x9
> preds = predict(cbx9, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx9)
> aic = AIC(cbx9)
> bic = BIC(cbx9)
> other = pR2(cbx9)[4:6]
> Model.9 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.9)

#x10
> preds = predict(cbx10, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx10)
> aic = AIC(cbx10)
> bic = BIC(cbx10)
> other = pR2(cbx10)[4:6]
> Model.10 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.10)

#x11
> preds = predict(cbx11, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx11)
> aic = AIC(cbx11)

```

```

> bic = BIC(cbx11)
> other = pR2(cbx11)[4:6]
> Model.11 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.11)

#x12
> preds = predict(cbx12, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx12)
> aic = AIC(cbx12)
> bic = BIC(cbx12)
> other = pR2(cbx12)[4:6]
> Model.12 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.12)

#x13
> preds = predict(cbx13, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx13)
> aic = AIC(cbx13)
> bic = BIC(cbx13)
> other = pR2(cbx13)[4:6]
> Model.13 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.13)

#x14
> preds = predict(cbx14, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx14)
> aic = AIC(cbx14)
> bic = BIC(cbx14)
> other = pR2(cbx14)[4:6]
> Model.14 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.14)

#x15
> preds = predict(cbx15, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx15)
> aic = AIC(cbx15)
> bic = BIC(cbx15)
> other = pR2(cbx15)[4:6]
> Model.15 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.15)

```

```

#x16
> preds = predict(cbx16, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx16)
> aic = AIC(cbx16)
> bic = BIC(cbx16)
> other = pR2(cbx16)[4:6]
> Model.16 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.16)

#x17
> preds = predict(cbx17, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx17)
> aic = AIC(cbx17)
> bic = BIC(cbx17)
> other = pR2(cbx17)[4:6]
> Model.17 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.17)

#x18
> preds = predict(cbx18, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbx18)
> aic = AIC(cbx18)
> bic = BIC(cbx18)
> other = pR2(cbx18)[4:6]
> Model.18 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> performance <- rbind(performance, Model.18)

> round(performance, 3)

# TAM MODEL
> cbfull = glm(y~., data=catalburun, family="binomial")
> summary(cbfull)
> durbinWatsonTest(cbfull) # Otokorelasyon testi

# ROC eğrisi ve AUC ölçütü
> if(!require(ROCR))
+ {install.packages("ROCR",
+ repo="https://cloud.r-project.org");
+ library(ROCR)}
> preds = predict(cbfull, new=catalburun, type="response")
> ROCpred = ROCR::prediction(preds, catalburun$y)
> auc = performance(ROCpred, measure = "auc")
> auc = auc@y.values[[1]]
> auc

```

```

# Tam model başarımı
> preds = predict(cbfull, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbfull)
> aic = AIC(cbfull)
> bic = BIC(cbfull)
> other = pR2(cbfull)[4:6]
> Model.Tam <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> Model.Tam

# DEĞİŞKEN SEÇİMİ

# Kademeli Lojistik regresyonla değişken seçimi
> cblognull = glm(formula=y~1, data=catalburun, family="binomial")
> cblogfull = glm(formula=y~., data=catalburun, family="binomial")
> cblogstep = step(cblognull, direction="both",
scope=list(lower=cblognull, upper=cblogfull))
> summary(cblogstep)
# Seçilen değişkenleri listele
> selvars = names(unlist(cblogstep[[1]]))
> selvars = selvars[-("(Intercept)" %in% selvars)]
> selvars

# caret ile değişken önemliliği saptama
> if(!require(caret))
+ {install.packages("caret",
+ repo="https://cloud.r-project.org");
+ library(caret)}
> cbfull <- glm(y~., data=catalburun, family="binomial")
> xnames <- paste0("x",1:18)
> varimp <- caret::varImp(cbfull)$Overall
> names(varimp) <- xnames
> cutoff <- quantile(varimp, c(.75, .90))
> xnames[varimp>=cutoff[1]]
> xnames[varimp>=cutoff[2]]
> varimp <- sort(varimp, decreasing=TRUE)
> barplot(varimp, names=names(varimp), col=rainbow(18))

# SEÇİLEN DEĞİŞKENLERLE OLUŞTURULAN MODELLER
# Seçilen model 1 (Model S1)
> cbmodels1= glm(y~x2+x10, data=catalburun, family="binomial")
> summary(cbmodels1)
> car::vif(cbmodels1) # Çoklu bağlantımlılık kontrolü

# ROC eğrisi ve AUC ölçütü
> if(!require(ROCR))
+ {install.packages("ROCR",

```

```

+ repo="https://cloud.r-project.org");
+ library(ROCR)}
> preds = predict(cbmodels1, new=catalburun, type="response")
> ROCpred = ROCR::prediction(preds, catalburun$y)
> auc = performance(ROCpred, measure = "auc")
> auc = auc@y.values[[1]]
> auc

> ROCperf = ROCR::performance(ROCpred,
+ measure = "tpr", x.measure = "fpr")
> plot(ROCperf, col=1, lty=1, lwd=3, xlab="FPR", ylab="TPR")
> title("ROC Eğrisi")
> abline(a=0, b=1, col="gray")
> legend("bottomright", legend=c("Tam model", "Model S1",
+ "Model S2", "Model S3", "Model S4"),
+ col=c(1,2,3,4,5), lty=c(1,2,3,4,5), lwd=c(2,2,2,2,2))

> ROCperf = ROCR::performance(ROCpred,
+ measure = "tpr", x.measure = "fpr")
> plot(ROCperf, col=2, lty=2, lwd=2, xlab="FPR", ylab="TPR", add=TRUE)

# Model S1 başarımı
> preds = predict(cbmodels1, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbmodels1)
> aic = AIC(cbmodels1)
> bic = BIC(cbmodels1)
> other = pR2(cbmodels1)[4:6]
> Model.S1 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> Model.S1

# Seçilen model 2 (Model S2)
> cbmodels2= glm(y~x1+x2+x5+x10+x18, data=catalburun, family="binomial")
> summary(cbmodels2)
> car::vif(cbmodels2) # Çoklu bağımlılık kontrolü

# ROC eğrisi ve AUC ölçütü
> preds = predict(cbmodels2, new=catalburun, type="response")
> ROCpred = ROCR::prediction(preds, catalburun$y)
> auc = performance(ROCpred, measure = "auc")
> auc = auc@y.values[[1]]
> auc

> ROCperf = ROCR::performance(ROCpred,
+ measure = "tpr", x.measure = "fpr")
> plot(ROCperf, col=3, lwd=2, lty=3, xlab="FPR", ylab="TPR", add=TRUE)

```

```

# Model S2 başarımı
> preds = predict(cbmodels2, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbmodels2)
> aic = AIC(cbmodels2)
> bic = BIC(cbmodels2)
> other = pR2(cbmodels2)[4:6]
> Model.S2 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> Model.S2

# Seçilen model 3 (Model S3)
> cbmodels3= glm(y~x1+x2+x10, data=catalburun, family="binomial")
> summary(cbmodels3)
> car::vif(cbmodels3) # Çoklu bağlantımlılık kontrolü

# ROC eğrisi ve AUC ölçütü
> preds = predict(cbmodels3, new=catalburun, type="response")
> ROCpred = ROCR::prediction(preds, catalburun$y)
> auc = performance(ROCpred, measure = "auc")
> auc = auc@y.values[[1]]
> auc

> ROCperf = ROCR::performance(ROCpred,
+   measure = "tpr", x.measure = "fpr")
> plot(ROCperf, col=4, lwd=2, lty=4, xlab="FPR", ylab="TPR", add=TRUE)

# Model S3 başarımı
> preds = predict(cbmodels3, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbmodels3)
> aic = AIC(cbmodels3)
> bic = BIC(cbmodels3)
> other = pR2(cbmodels3)[4:6]
> Model.S3 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> Model.S3

# Seçilen model 4 (Model S4)
> cbmodels4= glm(y~x1+x2+x5+x6+x10+x14+x18, data=catalburun,
family="binomial")
> summary(cbmodels4)
> car::vif(cbmodels4) # Çoklu bağlantımlılık kontrolü

> preds = predict(cbmodels4, new=catalburun, type="response")
> ROCpred = ROCR::prediction(preds, catalburun$y)
> auc = performance(ROCpred, measure = "auc")
> auc = auc@y.values[[1]]
> auc

```

```

> ROCperf = ROCR::performance(ROCpred,
+   measure = "tpr", x.measure = "fpr")
> plot(ROCperf, col=5, lwd=2, lty=5, xlab="FPR", ylab="TPR", add=TRUE)

# Model S4 başarımı
> preds = predict(cbmodels4, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbmodels4)
> aic = AIC(cbmodels4)
> bic = BIC(cbmodels4)
> other = pR2(cbmodels4)[4:6]
> Model.S4 <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> Model.S4

# Logit ve bağımsız değişkenlerin doğrusallığının testi
> logodds1 <- cbmodels1$linear.predictors
> boxTidwell(logodds1 ~ x2+x10, data=catalburun)
> plot(logodds1 ~ x2+x10, data=catalburun)

> logodds2 <- cbmodels2$linear.predictors
> boxTidwell(logodds2 ~ x1+x2+x5+x10+x18, data=catalburun)
> plot(logodds2 ~ x1+x2+x5+x10+x18, data=catalburun)

> logodds3 <- cbmodels3$linear.predictors
> boxTidwell(logodds3 ~ x1+x2+x10, data=catalburun)
> plot(logodds3 ~ x1+x2+x10, data=catalburun)

> logodds4 <- cbmodels4$linear.predictors
> boxTidwell(logodds4 ~ x1+x2+x5+x6+x10+x14+x18, data=catalburun, tol=1e-
06, max.iter=10)
> plot(logodds4 ~ x1+x2+x5+x6+x10+x14+x18, data=catalburun)

# Model S4a - Alternatif Model S4
> cbmodels4a = glm(y~x1+log(x2)+log(x5)+log(x6)+x10+x14+log(x18),
data=catalburun, family="binomial")
> summary(cbmodels4a)
> car::vif(cbmodels4a) # Çoklu bağlantımlılık kontrolü

> preds = predict(cbmodels4a, new=catalburun, type="response")
> ROCpred = ROCR::prediction(preds, catalburun$y)
> auc = performance(ROCpred, measure = "auc")
> auc = auc@y.values[[1]]
> auc

> ROCperf = ROCR::performance(ROCpred,
+   measure = "tpr", x.measure = "fpr")
> plot(ROCperf, col=6, lwd=2, lty=5, xlab="FPR", ylab="TPR", add=TRUE)

```

```
# Model S4a başarımı
> preds = predict(cbmodels4a, new=catalburun, type="response")
> preds = ifelse(preds >= 0.5, 1, 0)
> acc = sum(actual==preds)/length(actual)
> auc = ModelMetrics::auc(cbmodels4a)
> aic = AIC(cbmodels4a)
> bic = BIC(cbmodels4a)
> other = pR2(cbmodels4)[4:6]
> Model.S4a <- c(AIC=aic, BIC=bic, AUC=auc, ACC=acc, other)
> Model.S4a

# Model başarımları tablosu
> performance <- rbind(Model.S1, Model.S2, Model.S3,
+   Model.S4, Model.S4a, Model.Tam)
> round(performance, 3)
```

## EK 2. Çatalburun köpeklere ait veri yapısı (catalburun.xlsx)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	T	D	AKS	bal	5	10.50	20.60	42.40	22.10	15.30	9.60	30.00	30.00	59.30	73.50	67.00	12.40	10.80	11.70	61.10	61.30	18.40	12.20	0	
2	T	D	K	bal	2	8.30	15.30	30.90	19.60	17.00	13.60	34.90	31.40	48.00	61.40	59.40	16.60	10.40	10.40	51.40	49.20	19.00	10.60	1	
3	T	D	KGB	bal	2	8.30	17.30	35.00	20.40	18.50	14.40	33.80	34.90	46.00	65.50	58.00	10.50	9.40	9.50	56.50	54.00	17.20	11.30	1	
4	T	D	KGB	bal	3	9.70	17.30	35.70	21.20	17.40	13.60	37.90	33.20	58.20	65.80	58.50	10.80	10.30	10.00	54.80	51.20	18.60	11.30	0	
5	T	D	AKGB	bal	2	8.80	16.40	33.40	21.60	17.30	13.80	34.90	32.30	43.20	65.80	61.80	10.80	10.10	9.90	58.40	54.80	16.20	9.80	1	
6	T	D	K	bal	2	9.90	16.20	32.60	21.50	17.60	13.50	37.40	33.20	52.10	62.10	59.50	10.10	10.50	10.50	55.40	55.70	17.10	11.60	1	
7	T	D	BK	bal	4	8.40	16.60	34.70	19.40	17.30	12.10	34.50	32.30	46.00	56.70	55.30	10.40	10.00	9.90	49.10	51.10	16.10	9.70	1	
8	T	D	S	bal	4	7.70	16.60	36.40	19.90	17.60	12.20	38.10	32.10	52.20	62.10	57.40	10.80	10.00	9.80	51.90	51.80	17.10	10.30	1	
9	T	D	BSP	bal	4	10.00	19.00	38.20	21.30	17.20	13.00	35.00	32.20	51.50	62.10	63.80	11.30	10.40	10.60	57.20	59.50	17.10	10.10	0	
10	T	D	BSP	bal	5	9.40	17.60	37.50	22.50	17.40	16.20	44.90	33.60	49.20	81.20	59.60	12.90	11.50	12.10	55.10	57.40	18.20	10.70	1	
11	T	D	BSP	bal	3	8.30	17.30	40.10	20.10	16.40	13.60	34.50	36.50	47.10	58.00	61.00	16.90	10.00	9.80	54.40	56.10	14.60	10.50	1	
12	T	D	S	bal	3	7.70	17.20	39.70	20.60	16.50	13.50	39.80	40.00	42.60	66.00	58.50	12.80	10.40	9.90	52.50	47.50	16.20	11.10	1	
13	T	D	BKP	bal	5	9.30	17.30	38.30	21.00	18.60	13.10	44.80	34.10	49.20	71.80	63.50	11.70	11.80	11.20	58.50	56.50	16.60	11.00	1	
14	T	D	BSP	bal	5	8.60	16.10	35.00	21.60	17.60	14.10	42.00	33.60	50.00	69.00	63.50	10.80	11.00	11.00	58.00	58.50	17.00	10.50	1	
15	T	D	BSP	bal	2	8.50	16.90	39.80	19.50	17.60	13.60	35.20	32.50	49.50	69.50	61.00	9.60	9.50	10.00	53.00	59.80	15.50	10.60	0	
16	T	D	KGB	bal	3	9.40	16.70	33.90	20.20	16.80	14.20	38.30	31.60	56.10	65.20	63.10	11.20	11.10	11.20	54.20	58.70	17.50	10.90	1	
17	T	D	S	bal	3	8.10	18.20	36.70	21.20	17.50	14.20	36.50	32.00	48.00	61.80	56.50	12.70	11.20	10.90	50.50	50.40	15.20	12.00	1	
18	T	D	BSP	bal	3	8.40	13.90	37.90	20.60	17.30	13.80	36.50	32.60	55.50	64.00	59.50	11.70	10.80	10.80	54.00	53.00	18.00	11.50	1	
19	T	D	S	bal	3	9.80	17.90	36.90	21.50	17.00	15.10	41.50	31.00	53.50	69.50	62.00	11.20	10.90	11.00	57.50	57.80	17.60	10.60	1	
20	T	D	K	bal	3	8.80	16.10	32.50	19.50	15.00	14.00	35.60	31.50	47.50	59.10	58.30	11.80	9.50	9.00	51.00	52.50	15.70	11.00	0	
21	T	D	K	bal	3	9.80	16.60	33.90	21.00	17.50	14.50	39.80	36.00	55.50	65.00	65.10	11.50	10.00	10.40	57.20	58.60	17.50	10.70	1	
22	T	D	K	bal	3	9.80	16.60	33.90	21.00	17.50	14.50	39.80	36.00	55.50	65.00	65.10	11.50	10.00	10.40	57.20	58.60	17.50	10.70	1	

KN:Köpek Numarası

Y:safalık (1:saf ,0:saf değil)

x1:burun çapı

x2:burun bir kanat uzunluğu

x3:burun iki kanat uzunluğu

x4:burun ucu alın uzunluğu

x5:iki şakak arası genişlik

x6:iki kulak arası uzunluk

x7:boyun kalınlığı

x8:iki şakak ortası ile iki scapula arası uzunluk

x9:iki scapula ile kuyruk arası uzunluk

x10:göğüs çevresi

x11:cidago yüksekliği (ön bacak üstü)

x12:kuyruk kalınlığı (1. boğum)

x13:ön pati bilek kalınlığı

x14:arka pati bilek kalınlığı

x15:ön bacak uzunluğu

x16:arka bacak uzunluğu

x17:kulak uzunluğu

x18:kulak genişliği