

**Hacettepe Üniversitesi**

**Fen Fakültesi İstatistik Bölümü**

# STAT366 Regression Analysis

**Metin Uslu 21076757**

**Feyim Toprak 21021521**

İçindekiler

[STAT366 Regression Analysis 1](#_Toc75326627)

[A. UYGULAMANIN AMACI VE KAPSAMI 4](#_Toc75326628)

[1. Uygulamanın Amacı 4](#_Toc75326629)

[2. Uygulamanın Kapsamı (Verilerin toplanması) 4](#_Toc75326630)

[B. MODELDE KULLANILAN VERİ SETİ 4](#_Toc75326631)

[C. MODEL DENKLEMİ 5](#_Toc75326632)

[D. MODELDE KULLANILAN DEGİSKENLER 5](#_Toc75326633)

[1. Bağımlı Değişken 5](#_Toc75326634)

[2. Bağımsız Değişkenler 5](#_Toc75326635)

[E. BAGIMSIZ DEGİSKENLERİN MODELDE DENENMESİ 5](#_Toc75326636)

[a) Verilerin R’a girilmesi 5](#_Toc75326637)

[b) Bağımlı Değişkenin Normalliği Testi 6](#_Toc75326638)

[c) Doğrusallığın İncelenmesi 7](#_Toc75326639)

[d) Geçici modele ait R çıktıları 8](#_Toc75326640)

[ Belirtme Katsayısı (R2) ve Düzeltilmiş Belirtme Katsayısı (Adjusted- R2); 8](#_Toc75326641)

[ Modelin Anlamlılığı (Tümel F testi); 9](#_Toc75326642)

[ Katsayıların İncelenmesi; 9](#_Toc75326643)

[ Korelasyon Matrisi ve Gereksiz Değişkenlerin Çıkartılması; 10](#_Toc75326644)

[F. NİHAİ MODEL 10](#_Toc75326645)

[A) Değişen Varyanslılık Sorununu (Heteroskedastisite) 10](#_Toc75326646)

[B) Öz ilişki sorununu 11](#_Toc75326647)

[C) Çoklu Bağlantı Sorunu 11](#_Toc75326648)

[D) Nihai Modelin Anlamlılığı 12](#_Toc75326649)

[ Belirtme Katsayısı (R2) ve Düzeltilmiş Belirtme Katsayısı (Adjusted- R2); 12](#_Toc75326650)

[ Modelin Anlamlılığı (Tümel F testi); 13](#_Toc75326651)

[ Katsayıların İncelenmesi; 13](#_Toc75326652)

[ Bağlanım Katsayıları İçin %90 Güven Aralığı; 13](#_Toc75326653)

[ Uyum Kestirimi; 14](#_Toc75326654)

[ Ön Kestirim; 14](#_Toc75326655)

[ ve için güven aralıkları 14](#_Toc75326656)

[G. BAĞIMSIZ DEĞİŞKENLER HAKKINDA YORUM 15](#_Toc75326657)

[a) Akciğer Kapasitesi Hakkındaki Yorum 15](#_Toc75326658)

[b) Hemoglobin Düzeyi Hakkındaki Yorum 15](#_Toc75326659)

[c) Boy-Kilo Oranı Hakkındaki Yorum 15](#_Toc75326660)

[d) Spor Yapma Süresi Hakkındaki Yorum 15](#_Toc75326661)

[e) Dalış Saati Hakkındaki Yorum 15](#_Toc75326662)

[Kaynakça 15](#_Toc75326663)

**DALGIÇLARIN HAVA TÜKETİMİNİ ETKİLEYEN**

**DEGİSKENLERİN REGRESYON ANALİZİ**

# A. UYGULAMANIN AMACI VE KAPSAMI

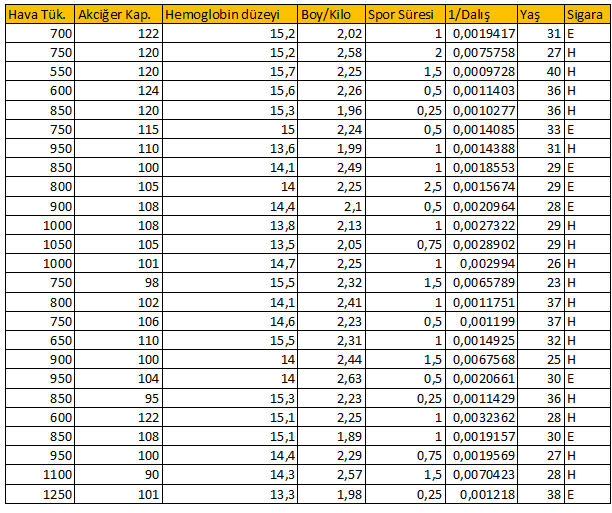
## 1. Uygulamanın Amacı

Bu uygulamanın amacı, hava tüketimi ile hava tüketimi üzerinde anlamlı etkisi olduğu düşünülen faktörlere ait çoklu regresyon modeli oluşturmaktır.

## 2. Uygulamanın Kapsamı (Verilerin toplanması)

Araştırma, İstanbul Beykoz’da bulunan bir sualtı eğitim merkezinde görevli dalgıçlar üzerinde yapılmış ve toplam 25 kişilik bir dalgıç gurubu örneklem olarak seçilmiştir. Modelde kullanılacak değişkenlerle ilgili bilgiler bir senelik gözlem süreci sonunda elde edilmiştir.

# B. MODELDE KULLANILAN VERİ SETİ

  
(Tablo-1)

# C. MODEL DENKLEMİ



# D. MODELDE KULLANILAN DEGİSKENLER

## Bağımlı Değişken

**y:** tüketilen hava miktarı bağımlı değişken olarak kabul edilmiştir.

## Bağımsız Değişkenler

***ak\_kap:*** Dalgıcın akciğer kapasitesi (%)

***hemog:*** Dalgıcın kanındaki hemoglobin düzeyi (gr/dl)

***boy\_kilo:*** Dalgıcın boy-kilo oranı (cm/kg)

***spor:*** Dalgıcın günlük ortalama spor yapma süresi (saat)

***ters\_d:*** Dalgıcın meslek hayatındaki toplam dalış saatinin tersi (1/saat)

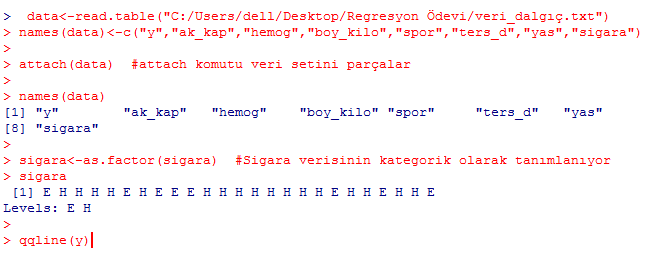
**yas:** Dalgıcın yası (yıl)

***sigara:*** Dalgıcın sigara kullanımı (Evet/ Hayır)

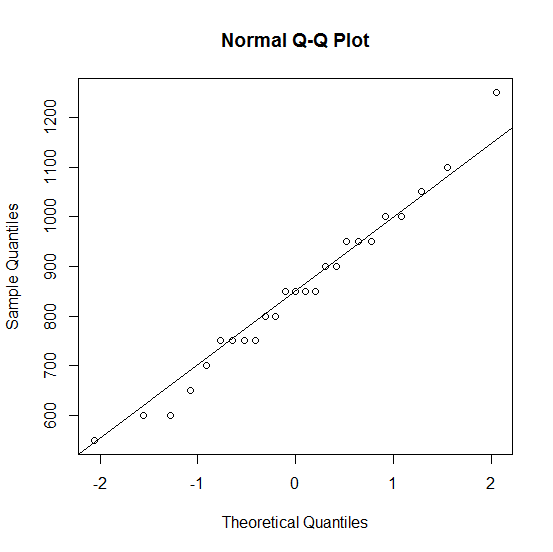
# E. BAGIMSIZ DEGİSKENLERİN MODELDE DENENMESİ

Önce bütün bağımsız değişkenleri modelde deneyeceğiz daha sonra anlamsız olan ve aralarında ilişki olan değişkenleri çıkarıp nihai modeli kuracağız.

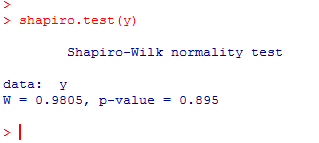
## Verilerin R’a girilmesi

  
(Tablo-2)

## Bağımlı Değişkenin Normalliği Testi

  
(Tablo-3)

Tablo-3 deki grafiğe bakıldığında noktaların doğru etrafında yer aldıkları gözleniyor buradan verinin normal dağılım göstermesini bekleriz

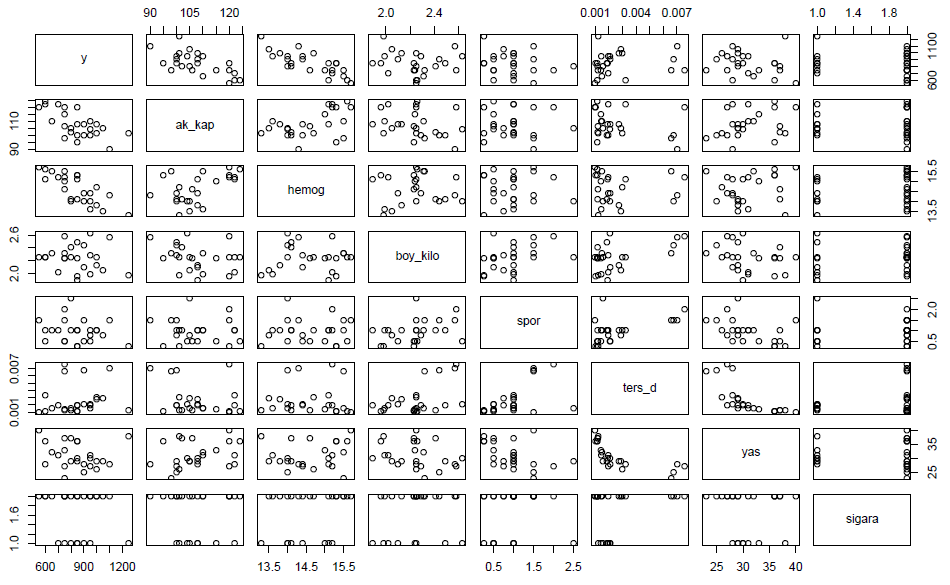


Ho: Hava tüketimi (y) değerlerinin dağılımı ile normal dağılım arasında fark yoktur.

Hs: Hava tüketimi (y) değerlerinin dağılımı ile normal dağılım arasında fark vardır.

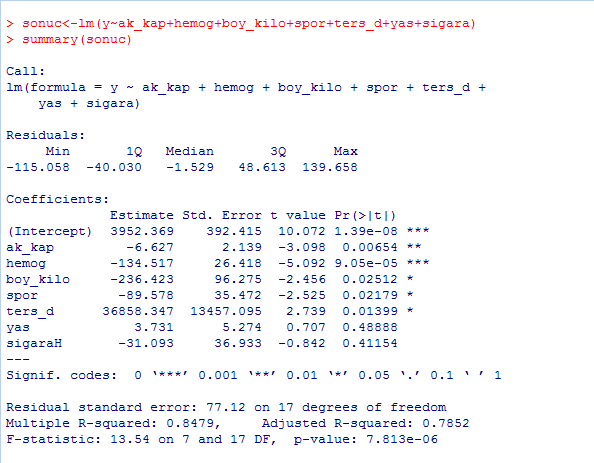
Teste ait p-value değeri 0,895 > 0,05 olduğu için H0 hipotezi kabul edilir. Bağımlı değişkenin normal dağılım gösterdiğini %95 güven düzeyi ile söylenebilir.

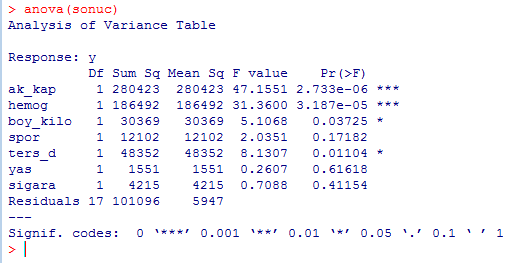
## Doğrusallığın İncelenmesi

  
(Tablo-4)

Tablo-4 deki grafikte değişkenler ikili olarak incelendiğinde aralarında çok açık doğrusallık olan değişken gözükmemektedir anca grafik biraz daha ayrıntılı incelendiğinde ters\_d (1/dalış tecrübesi) ve yaş değişkeni arasında doğrusallık olabileceği söylenebilir.

## Geçici modele ait R çıktıları

  
 (Tablo 5)

  
 (Tablo 6)

### Belirtme Katsayısı (R2) ve Düzeltilmiş Belirtme Katsayısı (Adjusted- R2);

Tablo-5 de Belirtme katsayısı R2=0,8479 ve düzeltilmiş belirtme katsayısı Adjuested-R2=0,7852 gibi yüksek değerler almıştır, bağımsız değişken sayısı çok olan modellerde Adjuested-R2’yi kullanmak daha doğrudur; bağımsız değişkenler bağımlı değişkenin %78 ini açıklar.  
“Bağımsız değişkenlerle bağımlı değişken arasında kuvvetli, doğrusal bir ilişki vardır.”

### Modelin Anlamlılığı (Tümel F testi);

Aşağıdaki hipotezler kurulup iki farklı şekilde modelin anlamlılığı test edilebilir;

H0: Kurulan regresyon modeli anlamsızdır

Hs: Kurulan regresyon modeli anlamlıdır.

***1.Yol;*** Tablo-5 deki p-valu= 7,813e-07 < 0,05 olduğundan H0 hipotezi red edilir. Kurulan regresyon modelinin anlamlı olduğu %95 güvenle söylenebilir.

***2.Yol;*** Tablo-6 daki ANOVA tablosundan;  
  
 AKT=101096 ve BKT= 280423+186492+30369+12102+48352+1551+4215=563504

Kurulan regresyon modelinin anlamlı olduğu %95 güvenle söylenebilir.

### Katsayıların İncelenmesi;

Tablo-5 deki katsayılar için;  
  
 H0: βj = 0 ( j. katsayı anlamsızdır)  
 HS: βj ≠ 0 ( j. katsayı anlamlıdır)  
  
şeklinde hipotezler kurulup Pr(> I t I) anlamlılık değerleri 0,05 den küçük olanlar için H0 hipotezleri red edilir.

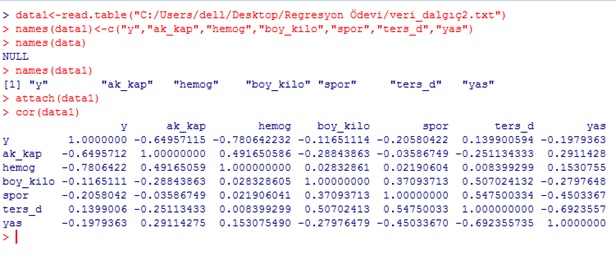
***ak\_kap, hemog, boy\_kilo, ters\_d, spor*** değişkenlerinin **katsayılarını *anlamlı****, yaş, sigara* değişkenlerinin katsayılarının *anlamsız* olduğu söylenebilir

Bu durumda geçici model şöyle olur;



Ancak bağımsız değişkenlerin aralarında ilişki olup olmadığını anlamak için korelasyon matrisine bakmamız gerekir.

### Korelasyon Matrisi ve Gereksiz Değişkenlerin Çıkartılması;

(Tablo-7)

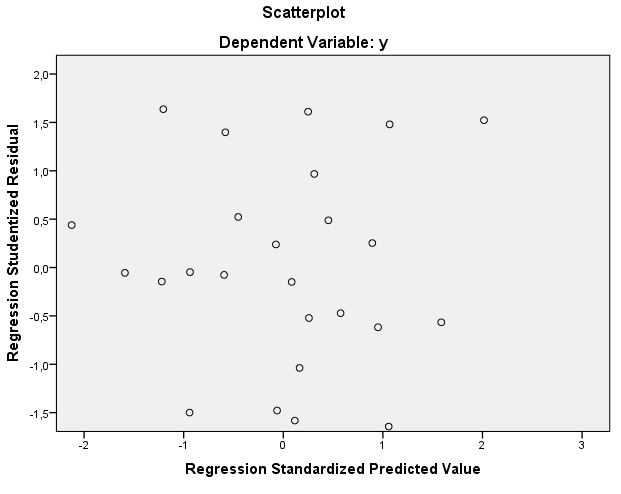
Yaş artıkça dalış tecrübesinin artacağı beklenir. Korelasyon matrisine bakıldığında da yaş ve ters\_d (1/dalış tecrübesi) arasında ters yönlü (1/dalış olduğundan) ve yaklaşık %70 lik bir ilşki vardır, ayrıca yaş ile spor yapma süresi arasındada ters yönlü %45 lik bir ilişki vardır.

*Bu ilişkiler modeli olumsuz etkileyeceğinden yaş değişkeni modelden çıkartılır.*

# F. NİHAİ MODEL

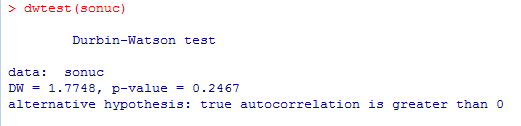
Anlamsız olan sigara ve ters\_d ile arasında yaklaşık %70 lik bağımlılık olan yaş değişkenini modelde çıkartıp nihai modeli oluşturuyoruz.

## Değişen Varyanslılık Sorununu (Heteroskedastisite)

  
 (Tablo-8)

Tahmin değerlerinin yatay eksende ve hata terimlerinin dikey eksende gösterildiği bir “hata terimleri serpme grafiği” nde, noktalar yatay huni seklinde dağılmış iseler, sabit varyans varsayımı ihlal edilmiş demektir. Eğer hata terimlerinin varyansı sabitse, noktaların dikey yöndeki yayılması, serpme grafiğinin her yönünde aynı olur.  
Tablo-8’ deki grafikte, noktaların rastgele dağılarak herhangi bir huni sekli oluşturmadığı görüldüğünden, kullanılan veriler için sabit varyans varsayımı geçerlidir.

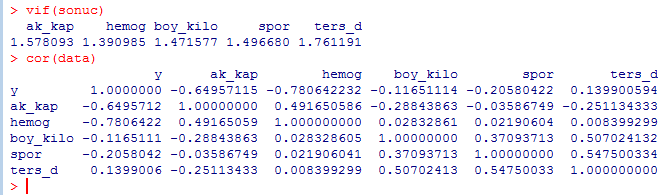
## Öz ilişki sorununu

  
(Tablo-9)

H0: Hata terimleri arasında özilişki yoktur.  
Hs: Hata terimleri arasında özilişki vardır.  
 Tablo-9 da Durbin-Watson testi sonucu bulunan p-valu=0,2467 > 0,05 olduğundan H0 hipotezi kabul edilir. Hata terimleri arasında özilişki yoktur.

## Çoklu Bağlantı Sorunu

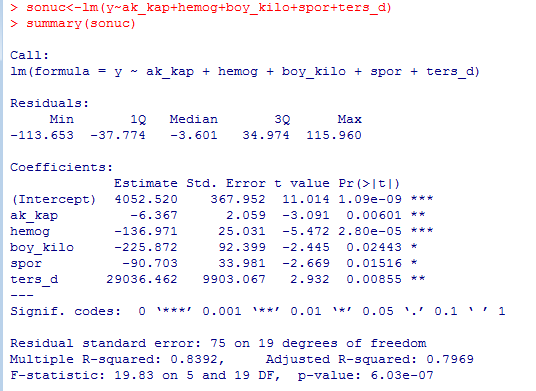
Çoklu bağlantı, doğrusal regresyon modelinden bir sapmayı ifade eder. Yani, çoklu regresyonun bir diğer varsayımı da, bağımsız değişkenler arasında çoklu bağlantı olmaması gerektiğidir.

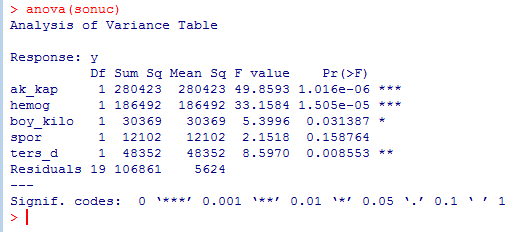
  
(Tablo-10)

Çoklu bağlantı için VIF değerlerine bakılır. Eğer VIF değeri eşik değerden küçükse, (genellikle eşik değer “10” kabul edilir) söz konusu bağımsız değişken ile, diğer bağımsız değişkenler arasında çoklu doğrusal bağlantının olmadığı kabul edilir.   
Modelimizdeki VIF değerinin tümü 10 dan küçük olduğundan çoklu doğrusal bağlantı yoktur.

İkinci yol ise bağımsız değişkenler arasındaki korelasyon katsayılarına bakıldığında bağımsız değişkenler arasında çoklu doğrusal bağımlılık olmadığı söylenebilir.

## Nihai Modelin Anlamlılığı

**** (Tablo-11)

**** (Tablo-12)

### Belirtme Katsayısı (R2) ve Düzeltilmiş Belirtme Katsayısı (Adjusted- R2);

Anlamsız olan yaş ve sigara değişkenlerini çıkardığımızda 0,0087 lik küçük bir azalmayla R2=0,8392 ve düzeltilmiş belirtme katsayısı 0,1 lik bir artmayla Adjuested-R2=0,7969 gibi yüksek değerler alır. (bu artma yaş ve sigaranın çıkartılmasının doğruluğunu gösterir)   
   
 Bağımsız değişken sayısı çok olan modellerde Adjuested-R2’yi kullanmak daha doğrudur; bağımsız değişkenler bağımlı değişkenin %80 ini açıklar.  
“Bağımsız değişkenlerle bağımlı değişken arasında kuvvetli, doğrusal bir ilişki vardır.”

### Modelin Anlamlılığı (Tümel F testi);

Aşağıdaki hipotezler kurulup iki farklı şekilde modelin anlamlılığı test edilebilir;

H0: Kurulan regresyon modeli anlamsızdır

Hs: Kurulan regresyon modeli anlamlıdır.

***1.Yol;*** Tablo-11 deki p-valu= 6,03e-07 < 0,05 olduğundan H0 hipotezi red edilir. Kurulan regresyon modelinin anlamlı olduğu %95 güvenle söylenebilir.

***2.Yol;*** Tablo-12 daki ANOVA tablosundan;  
  
 AKT=106861 ve BKT= 280423+186492+30369+12102+43352=557738

Kurulan regresyon modelinin anlamlı olduğu %95 güvenle söylenebilir.

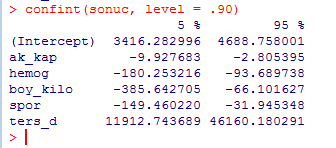
### Katsayıların İncelenmesi;

Tablo-11 deki katsayılar için;  
  
 H0: βj = 0 ( j. katsayı anlamsızdır)  
 HS: βj ≠ 0 ( j. katsayı anlamlıdır)  
  
şeklinde hipotezler kurulup Pr(> I t I) anlamlılık değerleri 0,05 den küçük olanlar için H0 hipotezleri red edilir.

***ak\_kap, hemog, boy\_kilo, ters\_d, spor*** değişkenlerinin **katsayıları anlamlıdır.**



### Bağlanım Katsayıları İçin %90 Güven Aralığı;

  
(Tablo-13)

Tablo-13 de bağlanım katsayıları için %90 düzeyinde güven aralığı verilmiştir.  
 Örneğin hemoglobin için güven aralığı; P(-9,92 ≤ β1 ≤ -2,80) = 0,90

### Uyum Kestirimi;

Birinci dalgıç için elde edilen gözlem değerleri aşağıdaki denklemde yerine yazılırsa;



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ak\_kap | hemog | boy\_kilo | spor | ters\_d | y |  |
| 122 | 15,2 | 2,02 | 1 | 0,001941748 | 700 | **702,86** |

### Ön Kestirim;

Veri kümesinde bulunmayan aşağıdaki veriler denkleme yazılırsa;



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ak\_kap | hemog | boy\_kilo | spor | ters\_d |  |
| 107 | 14 | 1,24 | 1 | 0,002617 | **1158,56** |

### ve için güven aralıkları

Birinci dalgıç için elde edilen gözlemler için uyum kestirimi güven aralığı;





Veri kümesinde bulunmayan gözlemlerin ön kestirimi güven aralığı;





# G. BAĞIMSIZ DEĞİŞKENLER HAKKINDA YORUM

1. Akciğer Kapasitesi Hakkındaki Yorum

Diğer tüm bağımsız değişkenler sabit tutulduğunda, akciğer kapasitesindeki 1 birimlik artış, beklenen hava tüketiminde 6,36 birimlik azalmaya neden olacaktır.

### Hemoglobin Düzeyi Hakkındaki Yorum

Diğer tüm bağımsız değişkenler sabit tutulduğunda, hemoglobin düzeyindeki 1 birimlik artış, beklenen hava tüketiminde 136,97 birimlik azalmaya neden olacaktır.

### Boy-Kilo Oranı Hakkındaki Yorum

Diğer tüm bağımsız değişkenler sabit tutulduğunda, boy-kilo oranındaki 1 birimlik artış, beklenen hava tüketiminde 225,87 birimlik azalmaya neden olacaktır.

### Spor Yapma Süresi Hakkındaki Yorum

Diğer tüm bağımsız değişkenler sabit tutulduğunda, spor yapma süresindeki 1 birimlik artış, beklenen hava tüketiminde 90,7 birimlik azalmaya neden olacaktır.

### Dalış Saati Hakkındaki Yorum

Diğer tüm bağımsız değişkenler sabit tutulduğunda, dalış saatinin tersindeki (1/dalış saati) 1 birimlik artış, beklenen hava tüketiminde 29036,46 birimlik artışa neden olacaktır.

# Kaynakça

Veriler Aşağıda Adı Verilen Yüksek Lisans Tezinden Alınmıştır;

SÜZENLER, O. DALGIÇLARIN HAVA TÜKETİMİNİ ETKİLEYEN DEĞİŞKENLERİN ANALİZİ VE BİR ÇOKLU REGRESYON ANALİZİ. İstanbul, 2006.