

Fen Fakültesi Aktüerya Bilimleri Bölümü

İST359 Regresyon Çözümlemesi

Regresyon Süreci Ödevi

Prof. Dr. Duru KARASOY

Hakkı KONDAK – 2220381067

17/12/2024

# İndeks

1. [Senaryo](#_Senaryo)
2. [Tanımlayıcı İstatistikler](#_Tanımlayıcı_İstatistikler)
3. [Varsayım Kontrolleri](#_Varsayım_Kontrolleri)
   1. [Normallik Varsayımı](#_Normallik_Varsayımı)
   2. [Doğrusallık Varsayımı](#_Doğrusallık_Varsayımı)
   3. [Doğrusal Modelin Kuruluşu](#_Doğrusallık_Modelinin_Kuruluşu)
4. [Artıkların İncelenmesi](#_Artıkların_İncelenmesi)
   1. [Hat Uzaklığı](#_Hat_Uzaklığı_1)
   2. [Standart Türü Artıklar](#_Standart_Türü_Artıklar_1)
   3. [Student Türü Artıklar](#_Student_Türü_Artıklar)
   4. [Cook Uzaklığı](#_Standart_Türü_Artıklar)
5. [Yeni Veri Setinde Varsayım Kontrolleri](#_Yeni_Veri_Setinde_1)
   1. [Yeni Veri Setinde Tanımlayıcı İstatistikler](#_Yeni_Veri_Setinde)
   2. [Normallik Varsayımı](#_Normallik_Varsayımı_1)
   3. [Doğrusallık Varsayımı](#_Doğrusallık_Varsayımı_1)
6. [Regresyon katsayılarının anlamlılıkları](#_Katsayı_Anlamlılıkları)
7. [Katsayı Yorumları](#_Katsayı_Yorumları)
8. [Güven Aralıkları Ve Yorumları](#_Güven_Aralıkları_Ve)
9. [Değişken Varyanslılık Sorunu](#_Değişen_varyanslılık)
10. [Öz İlişki Sorununu](#_Öz_İlişki_Sorunu)
11. [Çoklu Bağlantı Sorunu](#_Çoklu_Bağlantı_Sorunu)
12. [Uyum Kestirimi](#_Uyum_Kestirimi)
13. [Ön Kestirim](#_Ön_Kestirim)
14. [Güven Aralıkları ve Yorumları](#_Uyum_Kestirimi)
15. [Değişken Seçimi Yöntemleri](#_Değişken_Seçimi_Yöntemleri)
    1. [İleriye Doğru Seçim Yöntemi](#_İleriye_Doğru_Seçim)
    2. [Geriye Doğru Seçim Yöntemi](#_Geriye_Doğru_Seçim)
    3. [Adımsal Seçilim Yöntemi](#_Adımsal_Seçim_Yöntemi)
16. [Ridge Regresyonu](#_Ridge_Regresyonu)
17. [Yeni Veri Setinde Artıkların İncelenmesi](#_Yeni_Veri_Setinde_3)
    1. [Hat Uzaklığı](#_Hat_Uzaklığı_2)
    2. [Standart Türü Artıklar](#_Standart_Türü_Artıklar_2)
    3. [Student Türü Artıklar](#_Student_Türü_Artıklar_1)
    4. [Cook Uzaklığı](#_Cook_Uzaklığı_1)

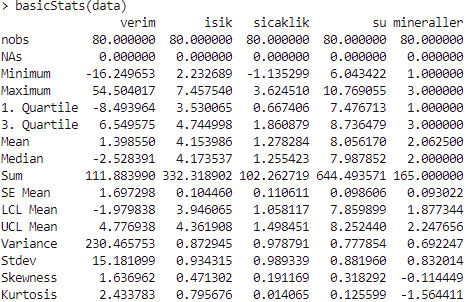
# 

# Senaryo

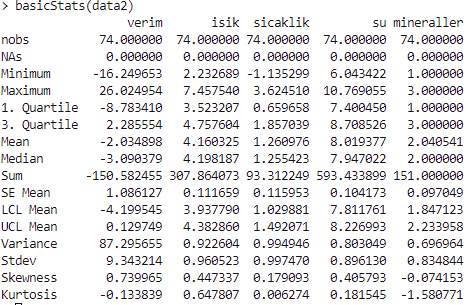
Regresyon sürecinde bir topraksız tarım yapan suda marul yetiştiren bir çiftçiyi inceleyeceğiz. Veri setimizdeki “y” değişkenimiz ürünün “Verimini”, “x1” değişkenimiz ürünün aldığı “Işık” miktarını, “x2” değişkenimiz ortamın “Sıcaklığını”, “x3” değişkenimiz “Su” miktarını, “x4” değişkenimiz ise bitkinin alığı “Mineraller” olmak üzere ifade edilir. (1 Azotu, 2 Kalsiyumu, 3 Potasyumu ifade eder.)

# Tanımlayıcı İstatistikler

İlk datamızın tamamlayıcı istatistikleri



Yeni datamızın tamamlayıcı istatistikleri

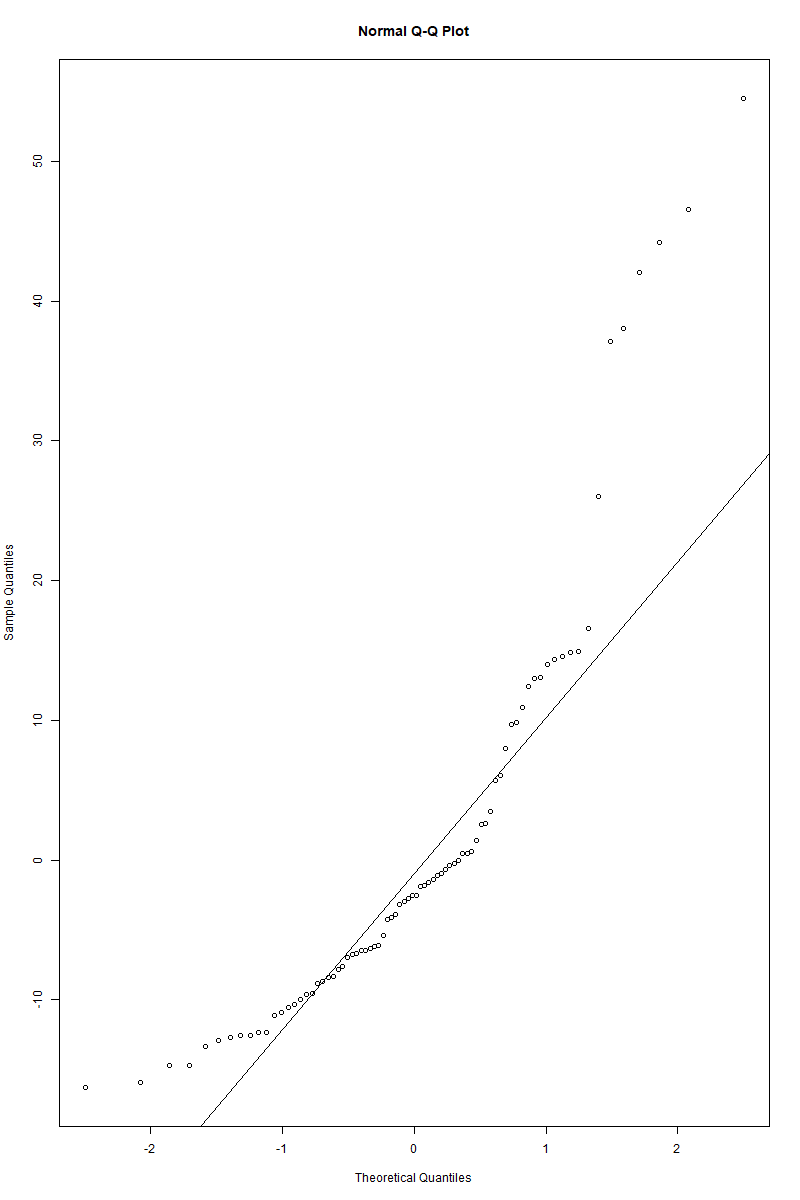


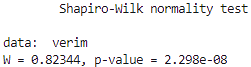
Bağımlı ve bağımsız değişkenlerimize ait en büyük, en küçük, ortalama, ortanca, varyans, standart sapma, basıklık ve çarpıklık değerleri gibi tanımlayıcı istatistikler tabloda görülmektedir.

# Varsayım Kontrolleri

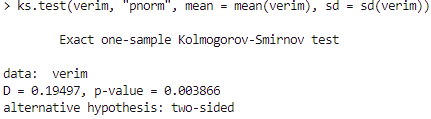
## Normallik Varsayımı

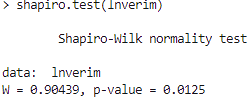
Shapiro-wilk testimize bakıyoruz





p-value 0.05’den küçük olduğu için normallik sağlanmaz.

Kolmogorov-swirnov’a göre normallik bakıyoruz ve p-value 0.05’den küçük olduğu için normallik sağlanmaz.

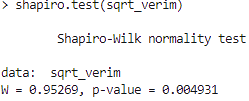
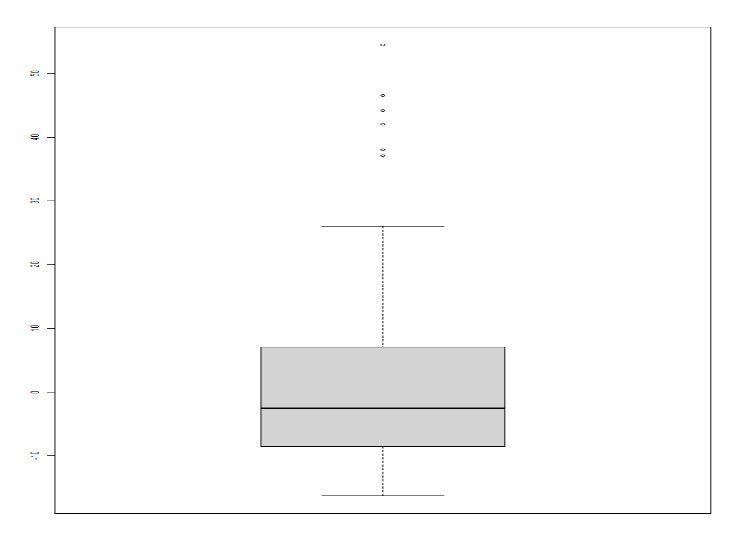
log dönüşümü yapıp bakıyoruz normalliğe. p-value 0.05’den küçük olduğu için normallik sağlanmadığını görüyoruz.

karekök dönüşümü yapıp bakıyoruz normalliğe. p-value 0.05’den küçük olduğu için normallik sağlanmadığını görüyoruz.

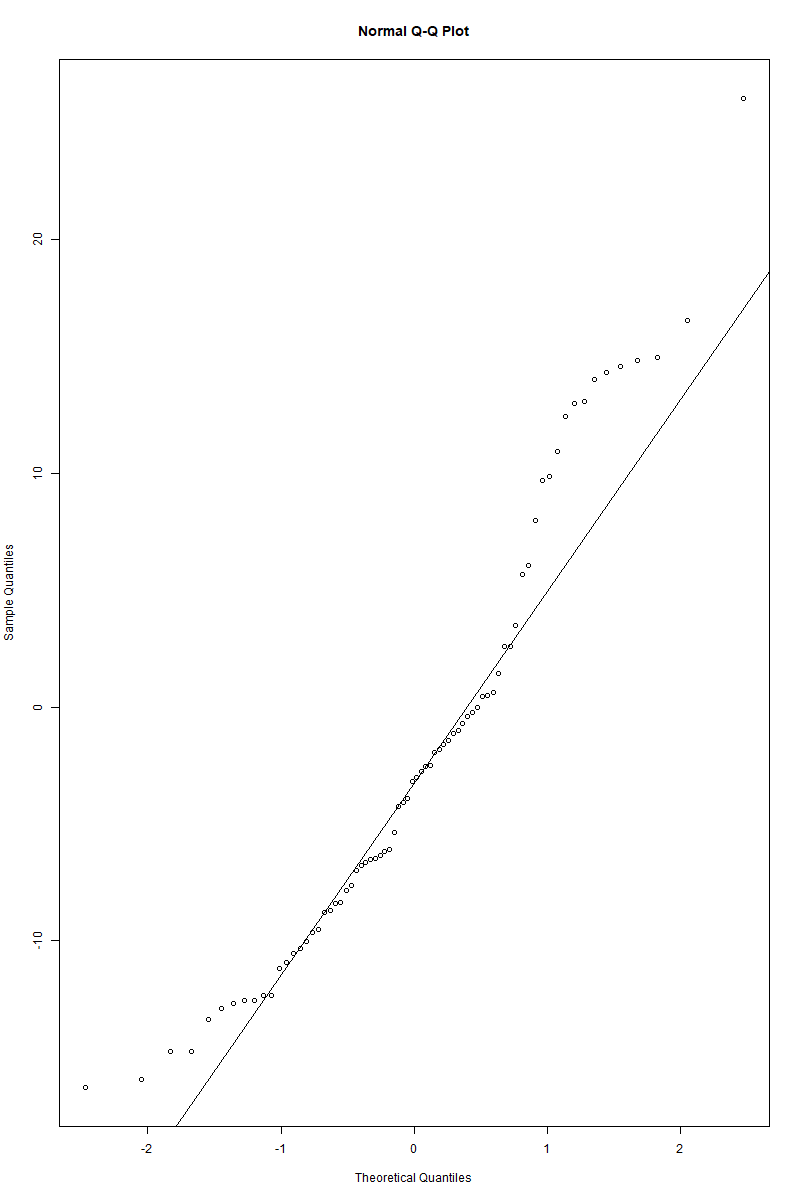
**Normallik sağlamazdığını görüyoruz ve Boxplot grafiğinden artık değerlerini bulup çıkarıyoruz**



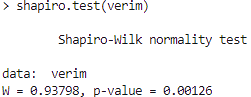
Görüldüğü üzere 17, 33, 46, ,64, 69 ve 72 artık değerlerimizdir. Bunları silip tekrardan datamızı oluşturacaz.



**Boxplot’tan sonra normallik var mı diye bakıyoruz**



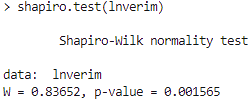
Grafikte normal dağılıyor gibi gözükse de shapiro normallik testi yapınca p-value 0.05’den küçük olduğu için H0 reddedilir. Modelimiz normal dağılıma uygun değildir.

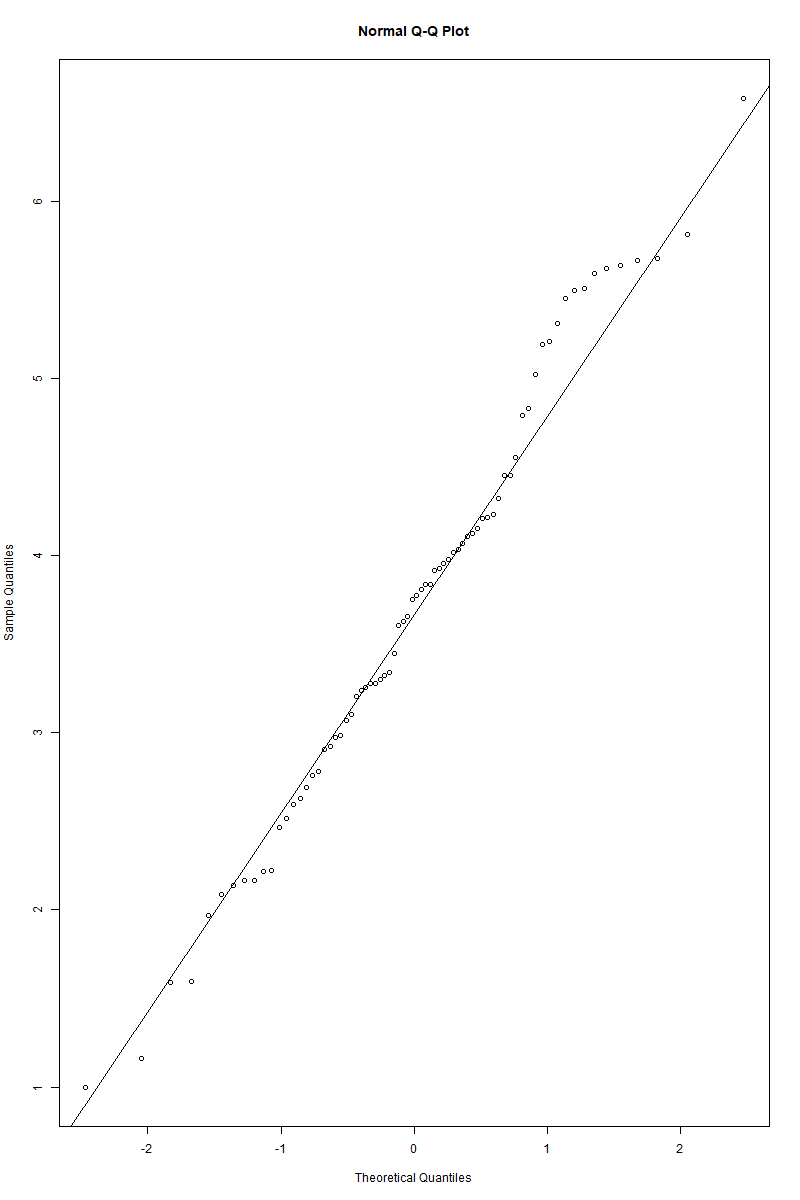


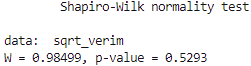
p-value 0.05’den küçük olduğu için normallik sağlanmaz.

p-value 0.05’den küçük olduğu için normallik sağlanmaz.

log dönüşümü yapıp bakıyoruz normalliğe. p-value 0.05’den küçük olduğu için normallik sağlanmadığını görüyoruz.

 **Hala normalik sağlamıyor o yüzden karekök dönüşümü uyguluyoruz**.





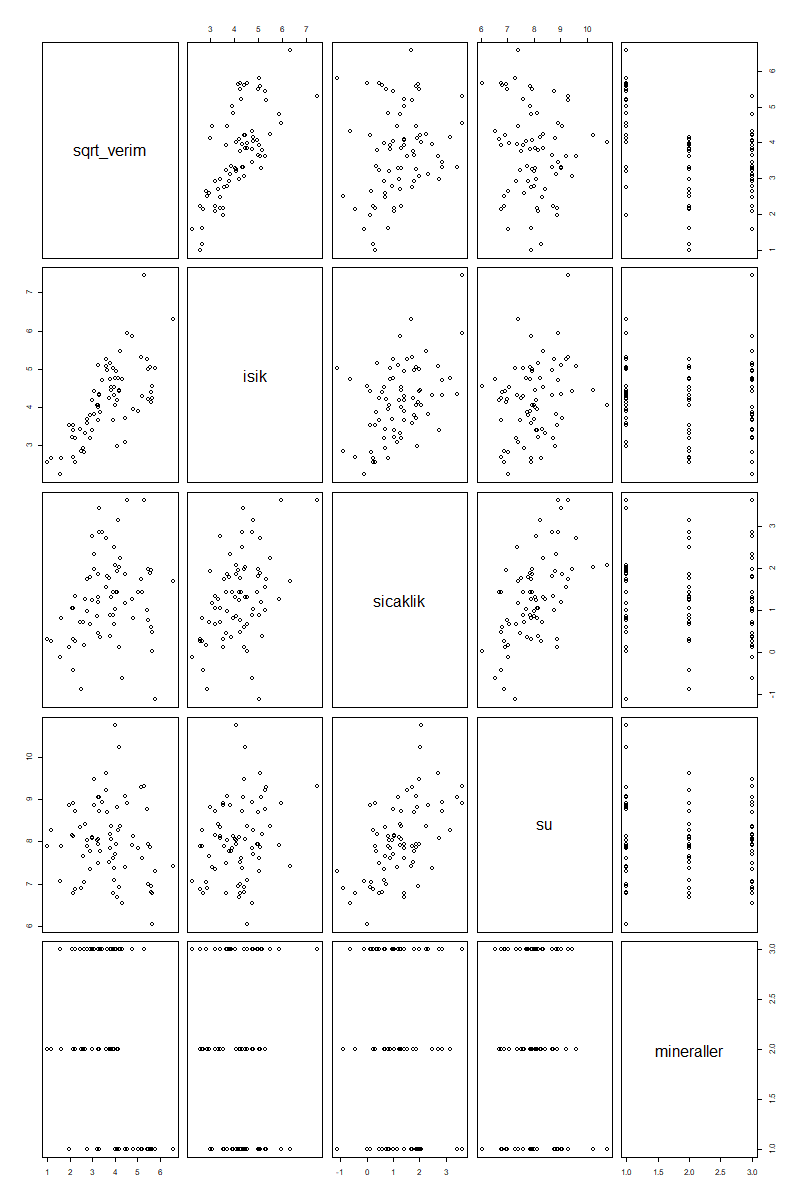
Hipotez testimizi kuruyoruz.

H0: Verilerimiz ile normal dağılım arasında fark yoktur.

H1: Verilerimiz ile normal dağılım arasında fark vardır.

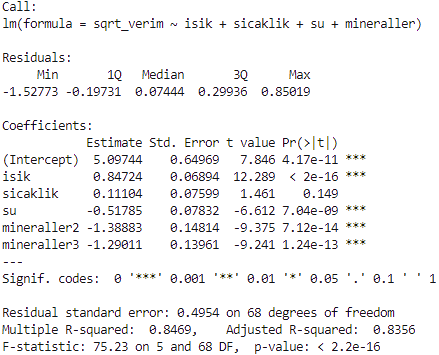
p değerimiz anlamlılık düzeyimiz olan %5’ten büyüktür. Bu durumda H0 hipotezimizi reddedemeyiz. Bağımlı değişkenimizin %95 güven düzeyinde anlamlı olduğunu söyleyebiliriz.

## Doğrusallık Varsayımı



Grafiklerde görüldüğü üzere bağımlı değişkenimiz ile bağımsız değişkenlerimiz arasında doğrusallık vardır. Doğrusal model kurulabilir. Ancak bağımsız değişkenlerimiz arasında da ilişki durumu söz konusudur.

# Doğrusallık Modelinin Kuruluşu



Model anlamlılığı hipotezimizi kuruyoruz.

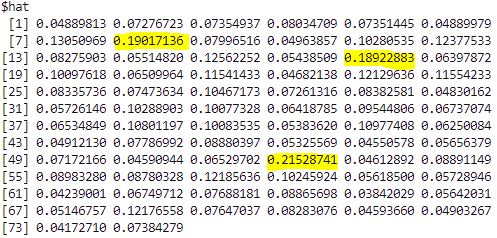
H0: β1=β2=β3=β4=β5=0 (Model anlamlı değildir.)

H1: β1,β2,β3,β4,β5’ten en az birisi sıfırdan farklıdır. (Model anlamlıdır.)

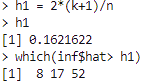
Modelimizdeki p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. H0 hipotezimizi reddederiz. Modelimiz %95 güven düzeyinde anlamlıdır. Toprak\_turu

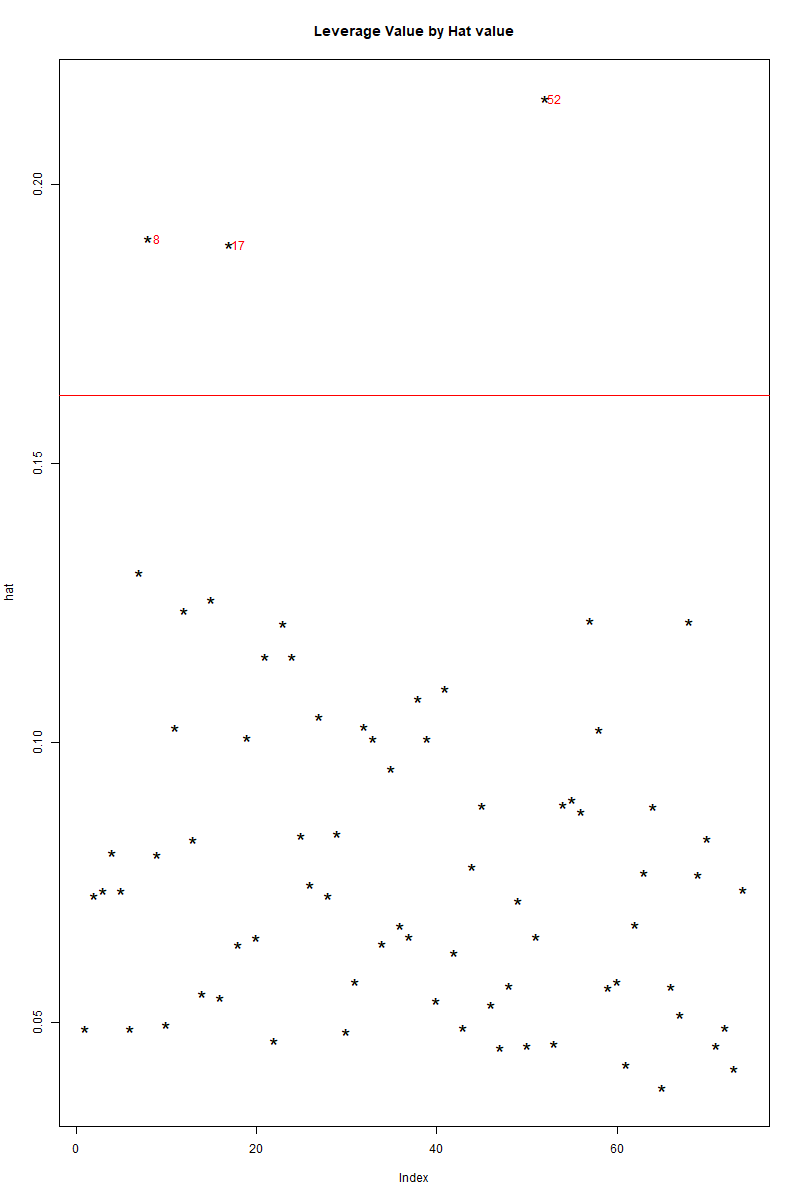
# Artıkların İncelenmesi

## Hat Uzaklığı



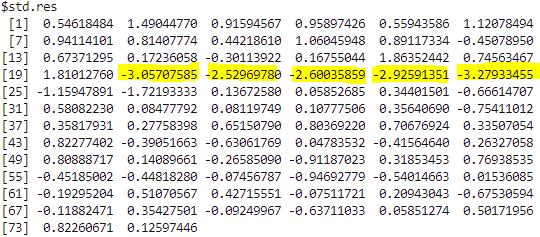
Hat Uzaklığı artık incelemesine göre değerinden büyük olan gözlemlerimiz uç değerlerdir.



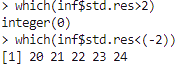


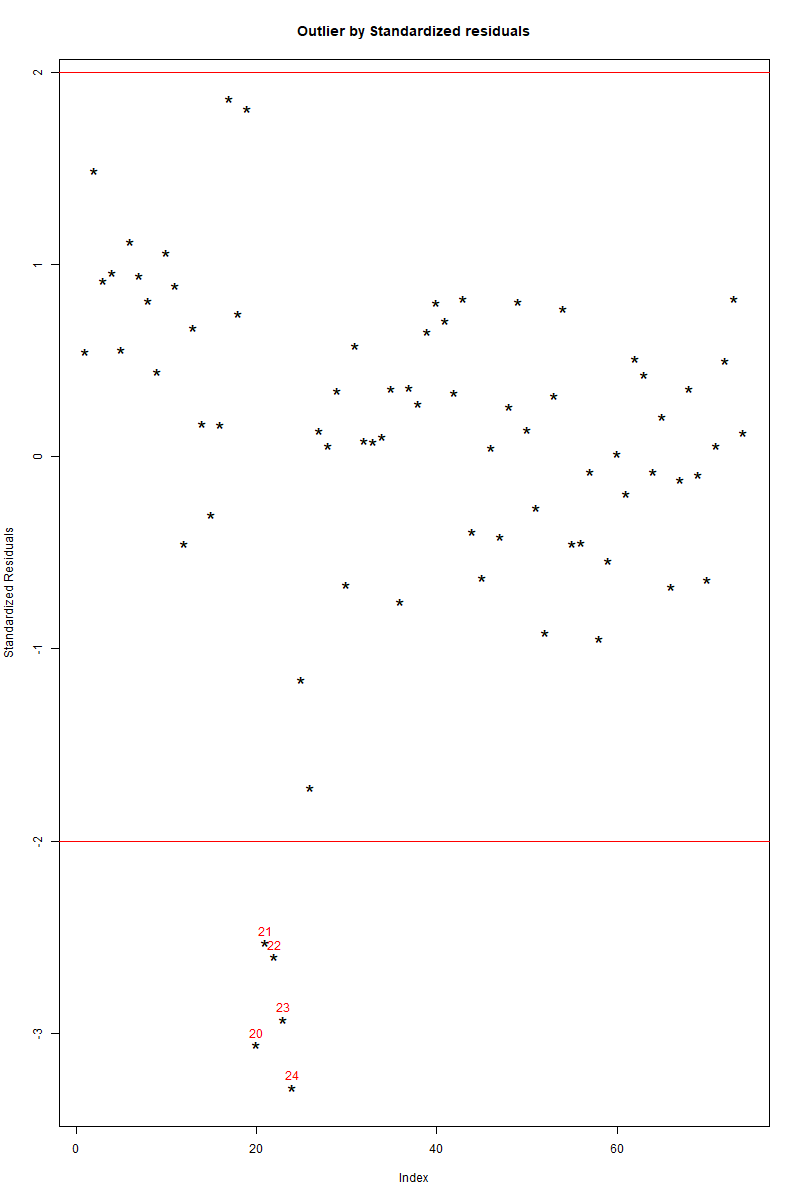
Kod üzerinde ve grafikte de görüldüğü üzere Hat Uzaklığı artık incelmesine göre 8, 17 ve 52. gözlem uç değerdir.

## Standart Türü Artıklar



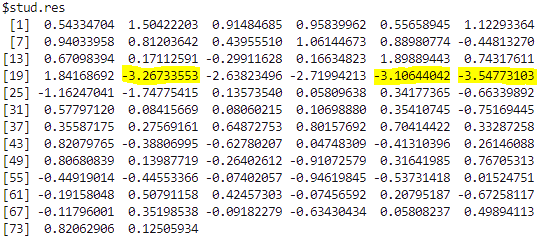
Standart Türü artıklara göre değerlerimiz (-2,2) aralığında olmalıdır. Aralık dışındaki değerler Standart Türü artıklar tablomuza göre artık değerlerdir.



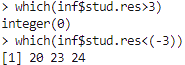


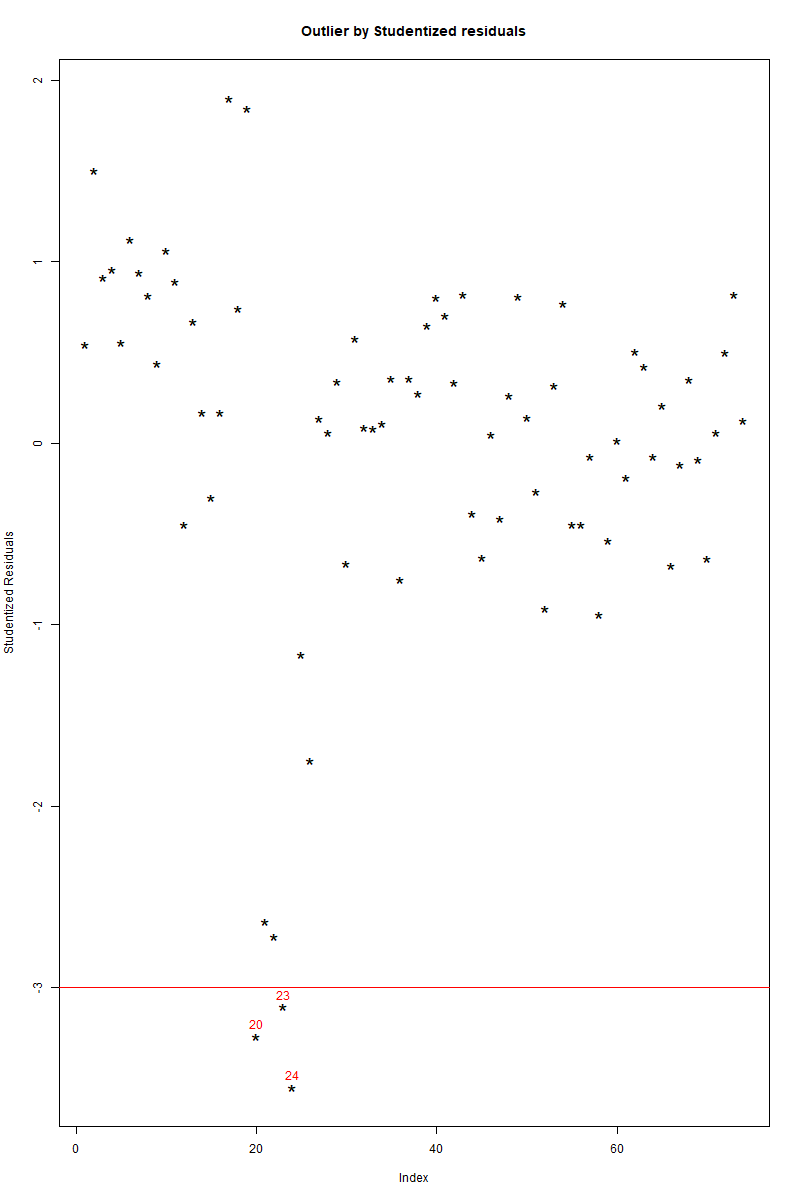
Grafikte ve kod üzerinde de görülüğü üzere 20, 21, 22, 23 ve 24. gözlemlerimiz Standart Türü artık incelmesine göre artık değerlerdir.

## Student Türü Artıklar



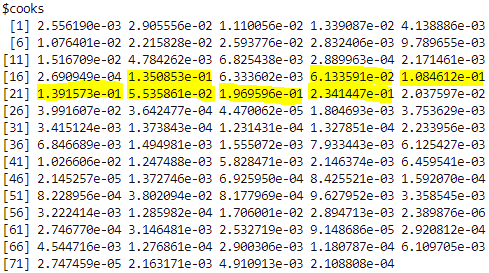
Student Türü artık incelemesine göre (-3,3) aralığı dışında kalan değerler aykırı değerlerdir.



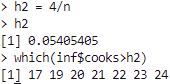


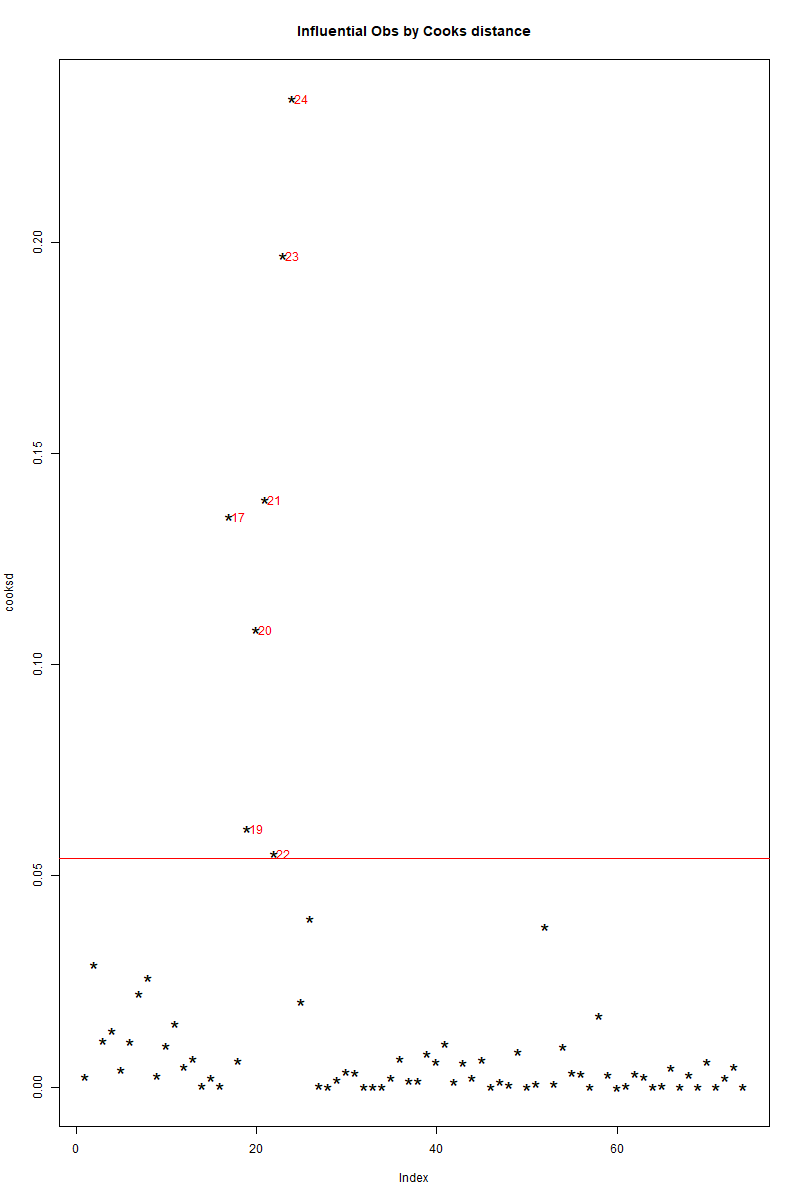
Grafikte ve kod üzerinde de görüldüğü üzere 20, 23 ve 24. Gözlemlerimiz Student Türü artık incelemesine göre artık değerlerdir.

## Cook Uzaklığı



Cook Uzaklığı artık incelemesine göre n>50 olduğu için değerinden büyük gözlemlerimiz etkin değerdir.

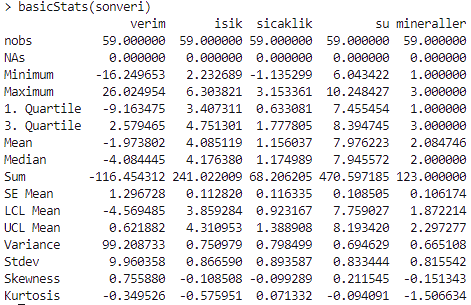




R kodu ve grafik üzerinde de görüldüğü üzere 17, 19, 20, 21, 22, 23 ve 24. gözlemlerimiz Cook Uzaklığı artık incelemesine göre artık değerlerdir.

Yeni verilerimiz üzerinde yaptığımız artık incelemesi sonucunda 17, 19, 20, 21, 22, 23, 24, 8 ve 52. gözlemlerimiz atık değerlerimizdir.

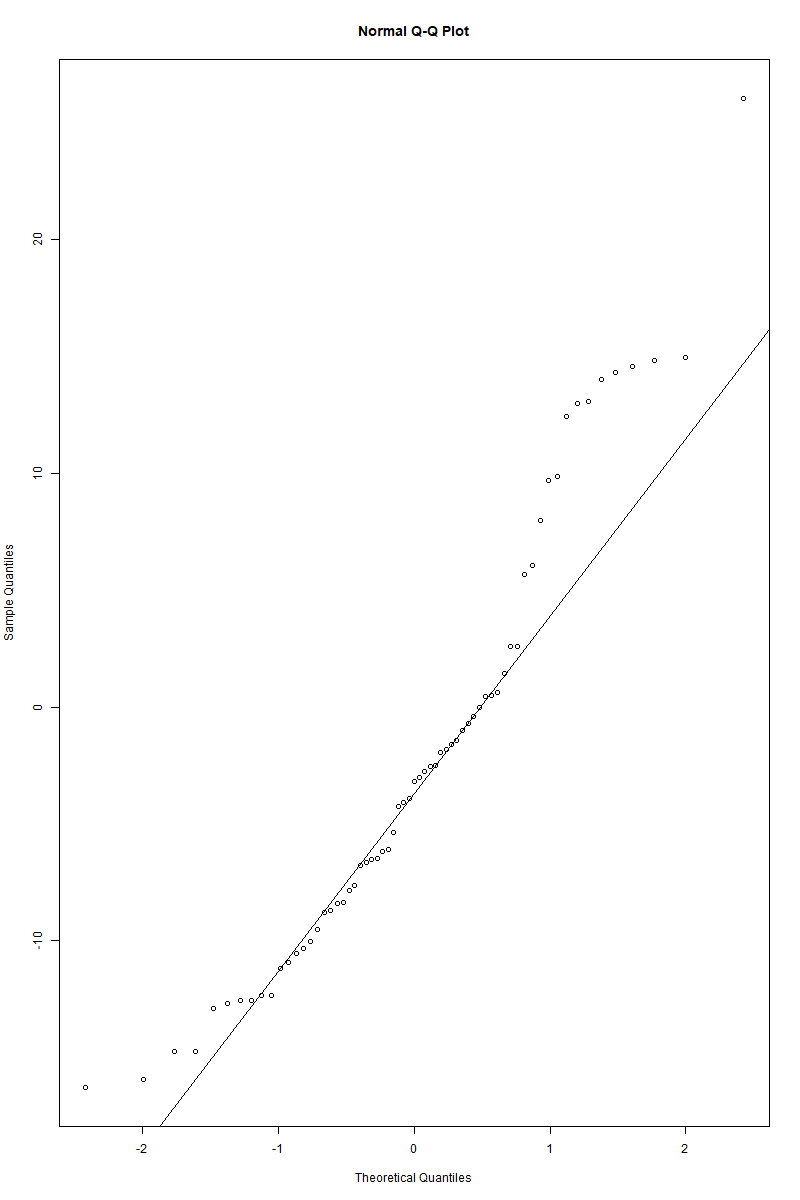
# Yeni Veri Setinde Tanımlayıcı İstatistikler



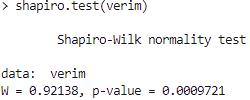
Artıkları çıkartılmış yeni veri setimize ait tanımlayıcı istatistikler tabloda görülmektedir.

# Yeni Veri Setinde Varsayım Kontrolleri

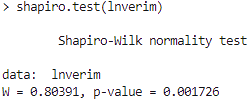
## Normallik Varsayımı

****

Grafikte görüldüğü üzere bağımlı değişkenimiz normal dağılıma uymaktadır. Emin olabilmek için Shapiro-Wilk testi uyguluyoruz.

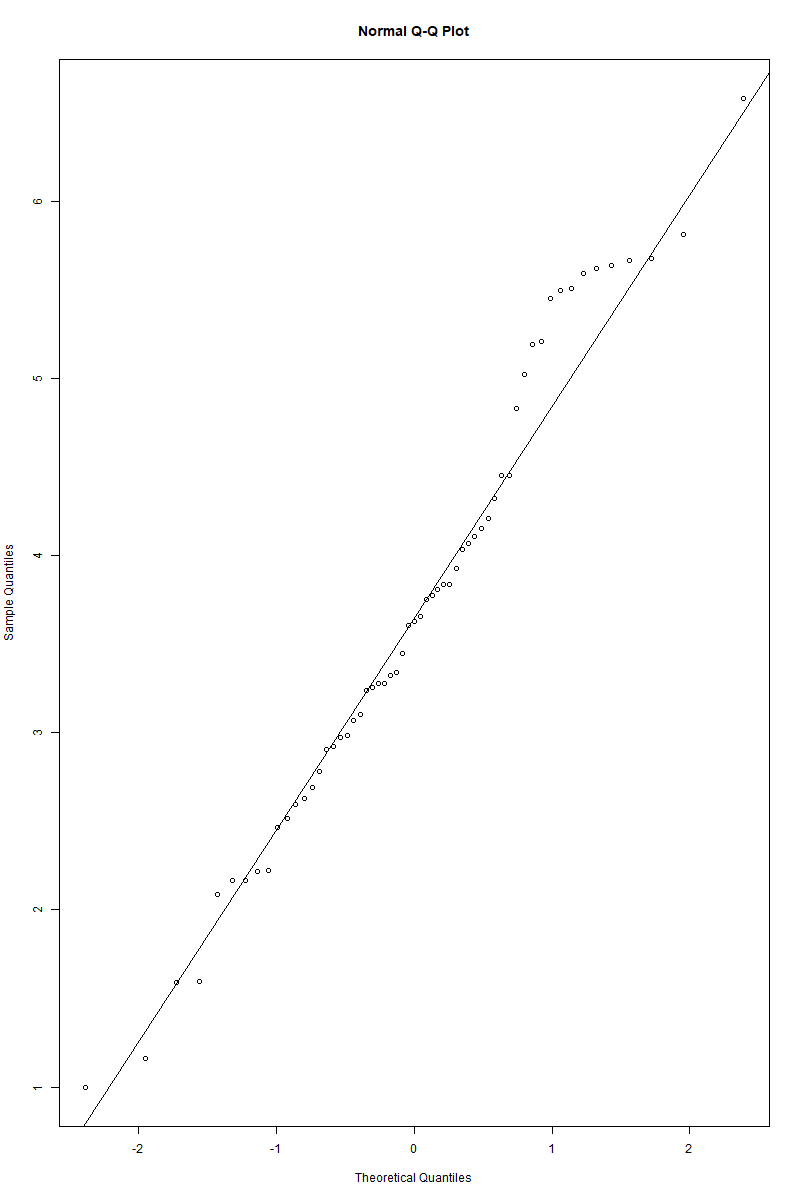


p-value 0.05’den küçük olduğu için normallik sağlanmaz

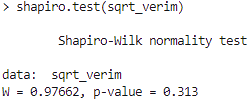


log dönüşümü yapıp bakıyoruz normalliğe. p-value 0.05’den küçük olduğu için normallik sağlanmadığını görüyoruz.

**Hala normalik sağlamıyor o yüzden karekök dönüşümü uyguluyoruz**.



Grafikte normal dağılıyor gibi gözükse de shapiro normallik testine yine de bakmamız gerekiyor



Hipotez testimizi kuruyoruz.

H0: Verilerimiz ile normal dağılım arasında fark yoktur.

H1: Verilerimiz ile normal dağılım arasında fark vardır.

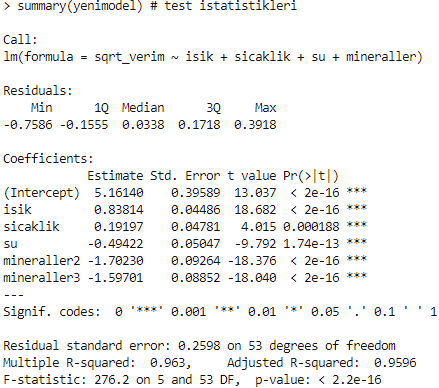
p değerimiz anlamlılık düzeyimiz olan %5’ten büyüktür. Bu durumda H0 hipotezimizi reddedemeyiz. Bağımlı değişkenimizin %95 güven düzeyinde anlamlı olduğunu söyleyebiliriz.

## Doğrusallık Varsayımı



Bağımlı değişkenimiz ile bağımsız değişkenlerimiz arasında doğrusal bir ilişki vardır. Doğrusal model kurabiliriz.

Yeni Veri Setinde Doğrusal Model Kuruluşu



Model kestirim denklemimiz şöyledir:

Model anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β1=β2=β3=β4=β5=0 (Model anlamlı değildir.)

H1: β1,β2,β3,β4,β5’ten en az birisi sıfırdan farklıdır. (Model anlamlıdır.)

p değerimiz anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddediyoruz. Yani modelimizin %95 güven düzeyinde anlamlı olduğunu söyleyebiliriz. Yani en az bir değişkenimiz sıfırdan farklıdır. Modelimizi kullanabiliriz.

# Katsayı Anlamlılıkları

Denklem sabitimizi için hipotez testimizi kuruyoruz

H0: β0=0 (Sabit terim anlamlı değildir)

H1: β ≠0 (Sabit terim anlamlıdır)

p değerimiz anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddederiz. Kestirim denklemimizdeki sabit terimin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

Işık değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β1=0 (Işık değişkeni anlamlı değildir)

H1: β1≠0 (Işık değişkeni anlamlıdır)

Işık değişkenine ait p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddediyoruz. Işık değişkenimizin modelimiz için anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

Sıcaklık değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β2=0 (Sıcaklık değişkeni anlamlı değildir)

H1: β2≠0 (Sıcaklık değişkeni anlamlıdır)

Sıcaklık değişkenine ait p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddediyoruz. Sıcaklık değişkenimizin modele katkısının anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

Su değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β3=0 (Su değişkeni anlamlı değildir)

H1: β3≠0 (Su değişkeni anlamlıdır)

Su değişkenine ait p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddediyoruz. Su değişkenimizin modele katkısının anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

Mineraller değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β4=0 (Mineraller değişkeninin 2 olduğu durum anlamlı değildir)

H1: β4≠0 (Mineraller değişkeninin 2 olduğu durum anlamlıdır)

Mineraller değişkenimizin 2 (Kalsiyum) olduğu durumun p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddederiz. Mineraller değişkenimizin 2 durumunda olması modelimiz için %95 güven düzeyinde anlamlıdır diyebiliriz.

H0: β5=0 (Mineraller değişkeninin 3 olduğu durum anlamlı değildir)

H1: β5≠0 (Mineraller değişkeninin 3 olduğu durum anlamlıdır)

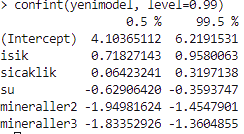
Mineraller değişkenimizin 3 (Potasyum) olduğu durumun p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddederiz. Mineraller değişkenimizin 3 durumunda olması modelimiz için %95 güven düzeyinde anlamlıdır diyebiliriz.

## Katsayı Yorumları

Kestirim denklemimizin sabiti, modelimizdeki diğer değişkenlerimiz sabitken marul verimimiz ortalama olduğunu gösterir. Işık değişkenimizin katsayısı, diğer değişkenlerin sabit olduğu durumda ışık miktarının 1 birimlik artışının marul verimliliğine ortalama birimlik artışa sebep olduğunu gösterir. Sıcaklık değişkeninin katsayısı bize, diğer değişkenlerin sabit olduğu durumda ortamdaki sıcaklık 1 birimlik marul verimine ortalama birimlik artışa sebep olduğunu gösterir. Su değişkeni katsayısı bize, diğer değişkenlerin sabit olduğu durumda su miktarı 1 birimlik artışın marul verimliliğine ortalama birimlik azalışa sebep olduğunu gösterir. Mineraller değişkenimizin 1 olması durumunun marul verimliliğine etkisi yoktur. Mineraller değişkenimizin 2 (Kalsiyum) olması durumunun marul verimliliğine - birimlik bir etkisi vardır. Mineraller değişkenimizin 3(Potasyum) olması durumunun marul verimliliğine - birimlik bir etkisi vardır.

Belirtme katsayımız 0.9607’dir. Buna göre bağımlı değişkenimizdeki (Marul Verimliliği) değişimin %96.30’i bağımsız değişkenlerimizce (Işık, Sıcaklık, Su, Mineraller) açıklanabilmektedir.

## Güven Aralıkları Ve Yorumları



Modelimize ait güven aralıkları %99 güven düzeyi ile bu şekilde hesaplanmıştır.

P(4.103 ≤ β0 ≤6.219)=0.99 Denklem sabitimizin 4.103 ile 6.219 arasında değer alması olasılığı %99’dur.

P(0.718 ≤ β1 ≤0.958)=0.99 Işık değişkenimizin katsayısının 0.718 ile 0.958 aralığında bir değer alması olasılığı %99’dur.

P(0.064 ≤ β2 ≤0.319)=0.99 Sıcaklık değişkenimizin katsayısının 0.064 ile 0.319 aralığında bir değer alması olasılığı %99’dur.

P(-0.629≤ β3 ≤-0.359)=0.99 Su değişkenimizin katsayısının -0.629 ile -0.359aralığında bir değer alması olasılığı %99’dur.

P(-1.949 ≤ β4 ≤-1.454)=0.99 Mineraller değişkenimizin 2 olduğu durumda marul verimliliğine etkisinin

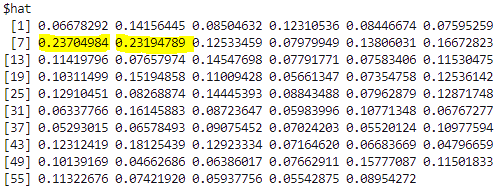
-1.949 ile -1.454 aralığında olması olasılığı %99’dur.

P(-1.833 ≤ β5 ≤-1.360)=0.99 Mineraller değişkenimizin 3 olduğu durumda marul verimliliğine etkisinin

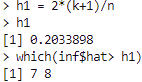
-1.833 ile -1.360 aralığında olması olasılığı %99’dur.

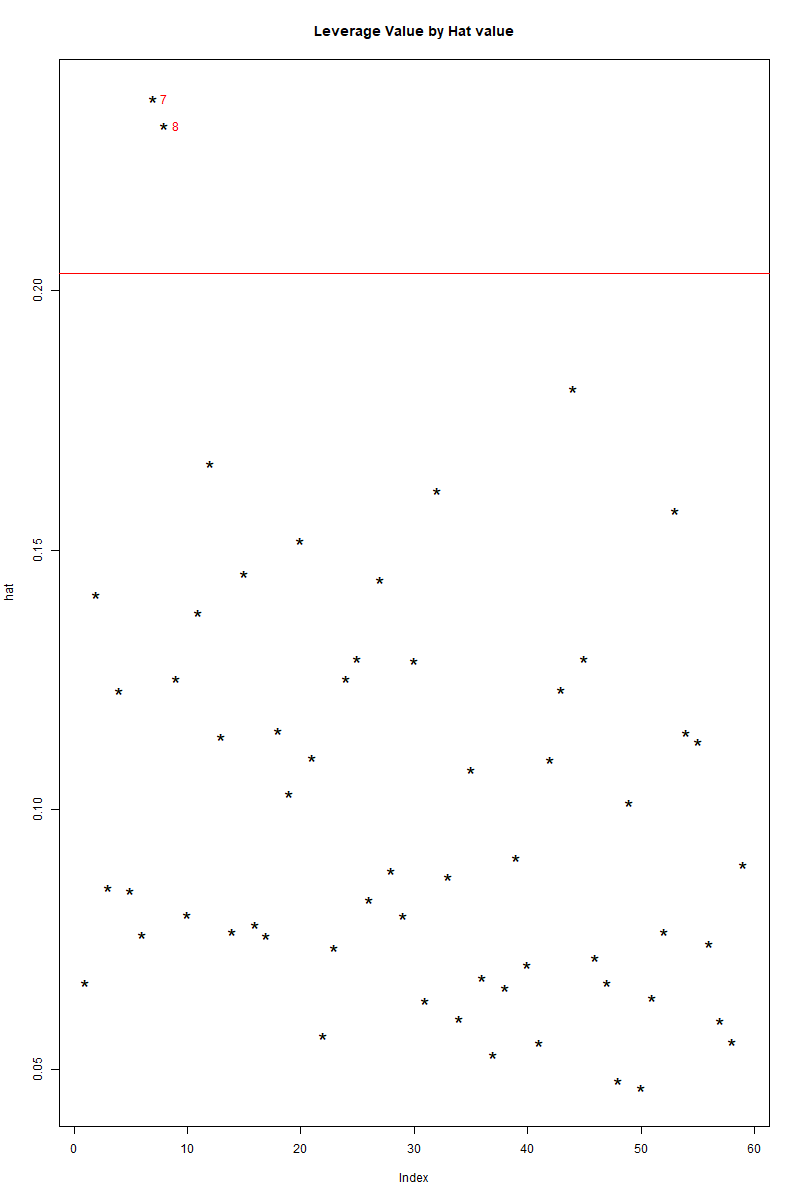
# Yeni Veri Setinde Artık İncelemesi

## Hat Uzaklığı



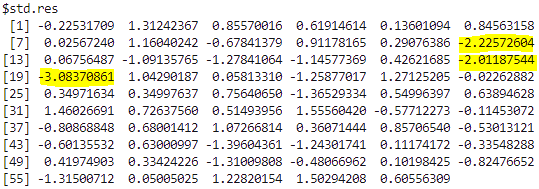
Hat Uzaklığı artık incelemesine göre değerinden büyük olan gözlemlerimiz uç değerlerdir.



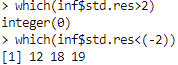


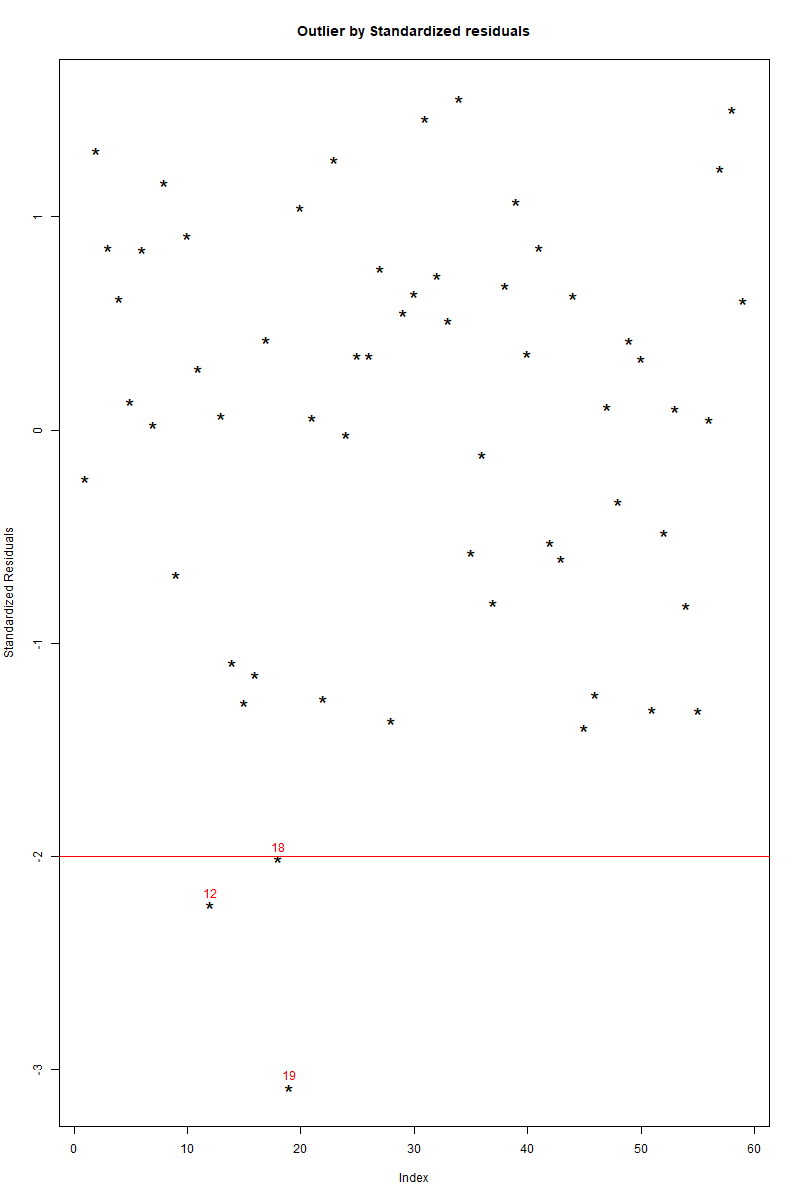
Kod üzerinde ve grafikte de görüldüğü üzere Hat Uzaklığı artık incelmesine göre 7. Ve 8. gözlem uç değerdir.

## Standart Türü Artıklar



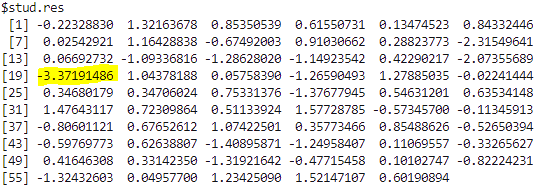
Standart Türü artıklara göre değerlerimiz (-2,2) aralığında olmalıdır. Aralık dışındaki değerler Standart Türü artıklar tablomuza göre artık değerlerdir.



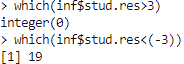


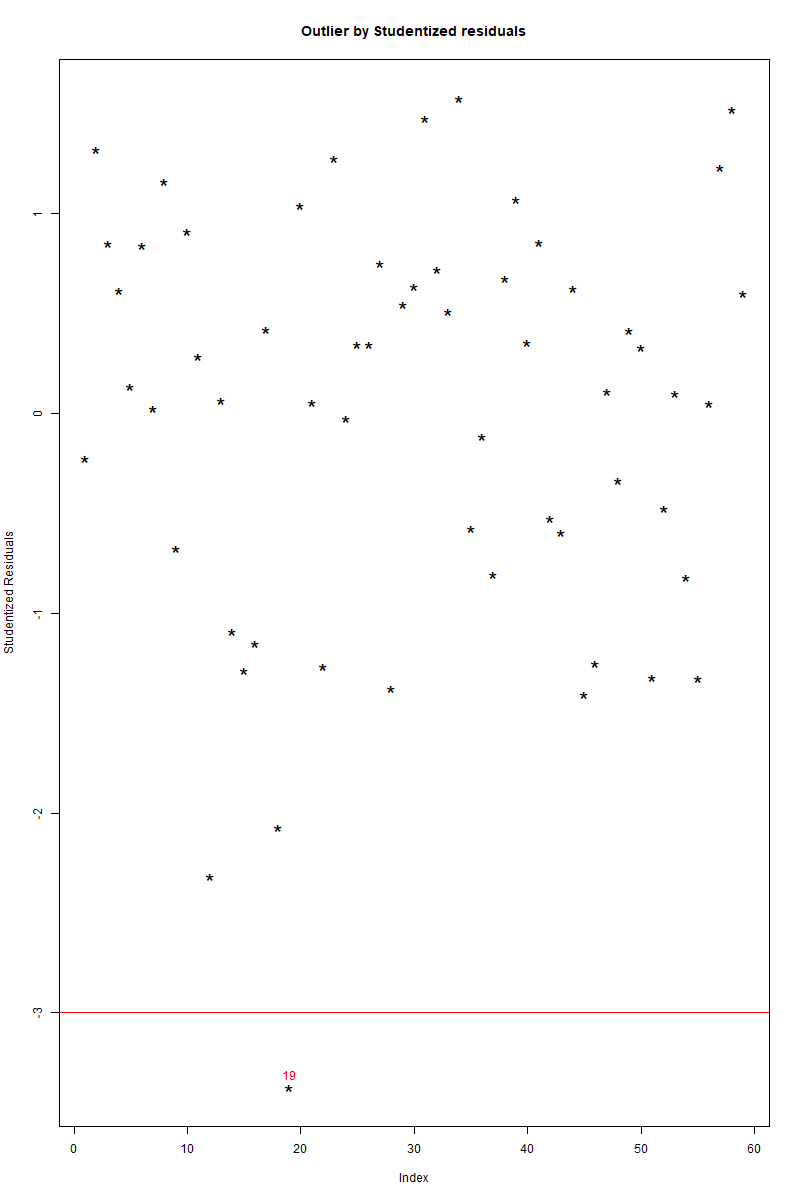
Grafikte ve kod üzerinde de görülüğü üzere 12, 19 ve 18. gözlemlerimiz Standart Türü artık incelmesine göre artık değerlerdir.

## Student Türü Artıklar



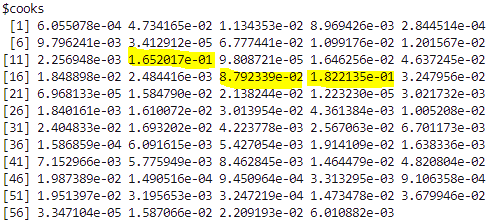
Student Türü artık incelemesine göre (-3,3) aralığı dışında kalan değerler aykırı değerlerdir.



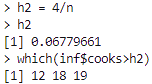


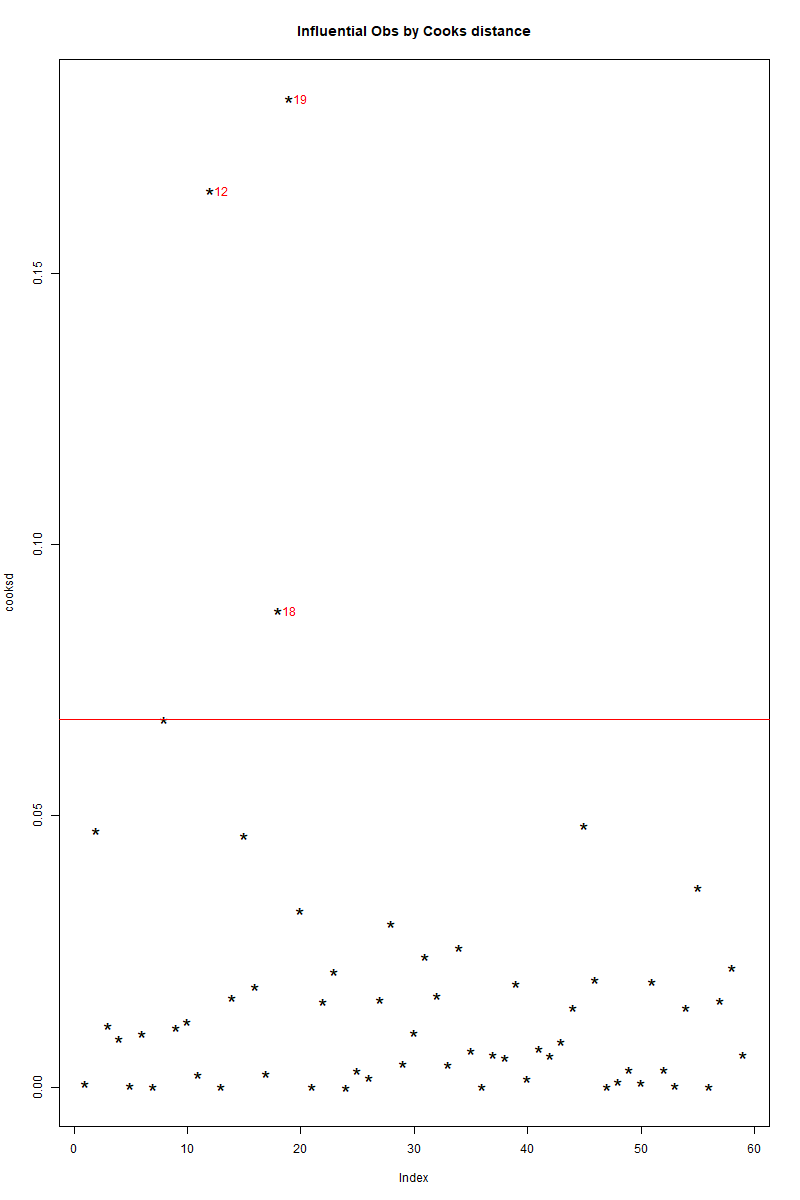
Grafikte ve kod üzerinde de görüldüğü üzere 19. Gözlemimiz Student Türü artık incelemesine göre artık değerdir.

## Cook Uzaklığı



Cook Uzaklığı artık incelemesine göre n>50 olduğu için değerinden büyük gözlemlerimiz etkin değerdir.

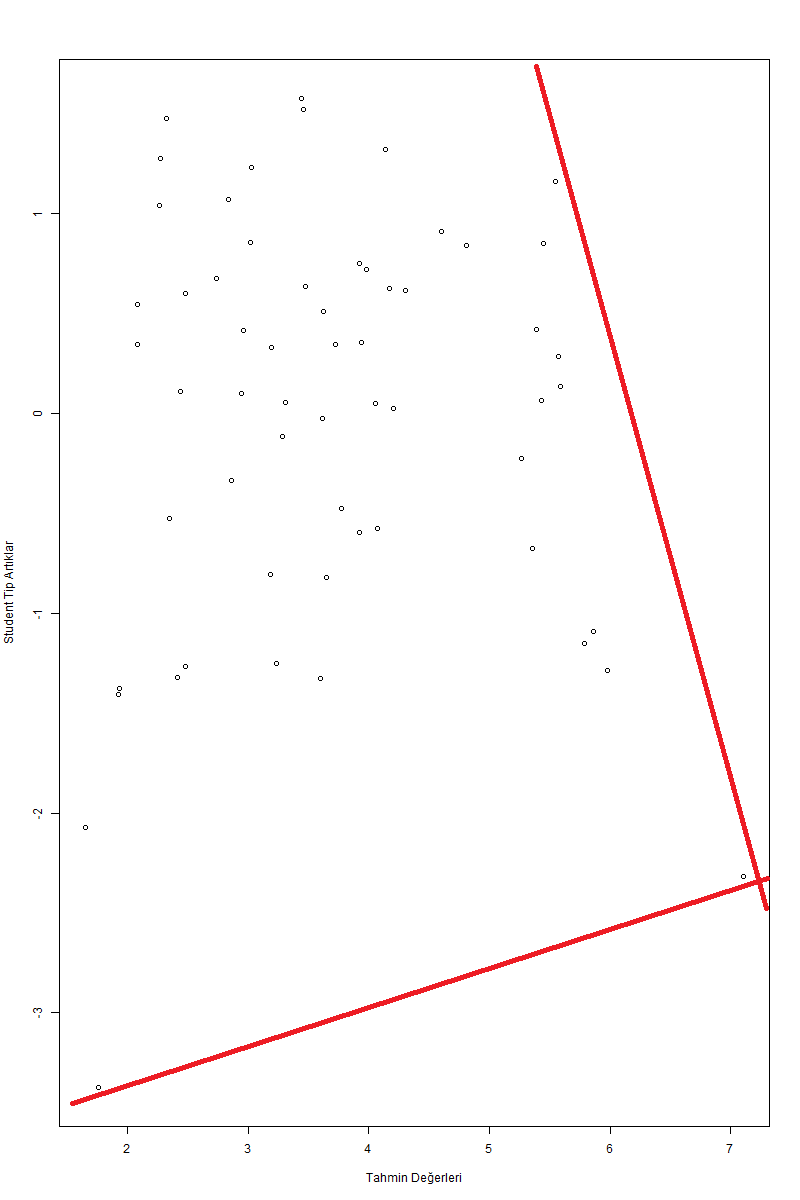




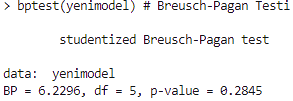
R kodu ve grafik üzerinde de görüldüğü üzere 12, 19 ve 18. gözlemlerimiz Cook Uzaklığı artık incelemesine göre artık değerlerdir.

Yeni verilerimiz üzerinde yaptığımız artık incelemesi sonucunda 7, 8, 12, 19 ve 18. gözlemlerimiz atık değerlerimizdir.

# Değişen varyanslılık



Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde yapının rasgele olmadığı, gözlemlerin nokta olarak ifade edildiği durumda sola doğru megafon oluştuğu görülmektedir. Ayrıca değişen varyans incelemesi için aşağıdaki test uygulanabilir.

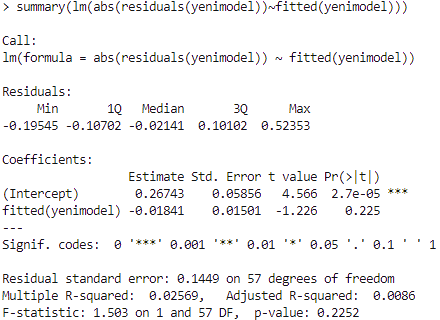
Breusch-Pagan testimizi yorumlamak için hipotez testimizi şöyle kuruyoruz:

H0: Varyanslar homojendir.

H1: Varyanslar homojen değildir.

Test sonucunda p değeri anlamlılık düzeyi olan %5’ten büyüktür. Bu durumda H0 hipotezimizi reddedemeyiz. Breusch-Pagan testine göre varyansların homojen dağıldığını %95 güven düzeyinde söyleyebiliriz.

Değişen varyans sorununun kontrolü için bir diğer yöntem artıklarla kestirim değerleri arasında bir model yazılıp yazılamadığına bakmaktır. Bunun için modelimizi kurup anlamlılığını test ediyoruz



Artıklarla kestirim değerleri arasında kurduğumuz modelin anlamlılığını test etmek için hipotez testimizi şu şekilde kuruyoruz:

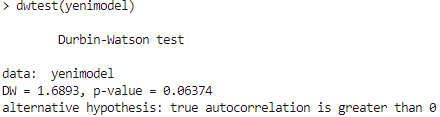
H0: Model anlamsızdır.

H1: Model anlamlıdır.

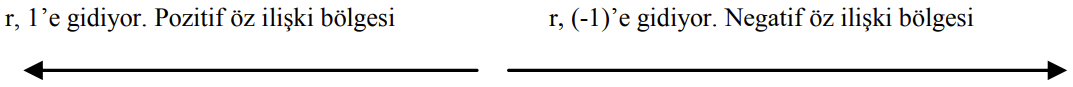
Modele ait p değeri anlamlılık düzeyimiz olan %5’ten büyüktür. Bu durumda H0 hipotezini reddedemeyiz. Yani artıklarla kestirim değerleri arasında kurulan denklem istatistiksel olarak %95 güven düzeyinde anlamsızdır.

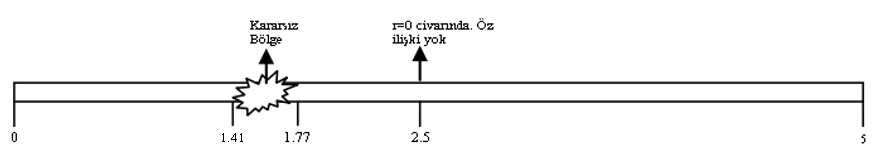
Her iki değişen varyanslılık kontrol etme yönteminin sonucuna göre değişen varyanslılık sorunumuz yoktur.

# Öz İlişki Sorunu



Durbin Watson değeri üzerinden inceleme yapılabilmektedir. Model özeti tablosunda test değeri d=1,6893 bulunmuştur.





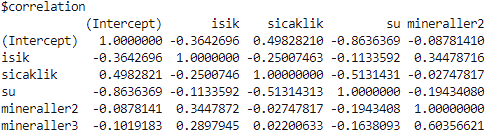
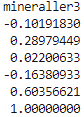
Öz ilişki olup olmadığına karar verebilmek için hipotez testimizi kurup d değerimizi Durbin-Watson tablosundaki alt ve üst değerler ile karşılaştırmamız gerekiyor. Öncelikle hipotez testimizi şöyle kuruyoruz.

H0: ρ=0 (Öz ilişki yoktur)

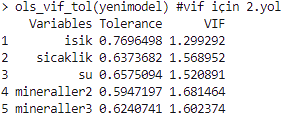
H1: ρ<0 (Negatif öz ilişki vardır)

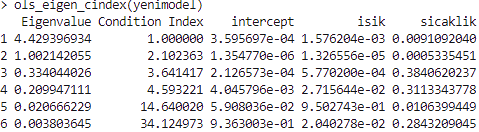
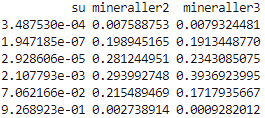
Durbin Watson değeri üzerinden inceleme yapılabilmektedir. Model özeti tablosunda test değeri d=1,6893 bulunmuştur. Alternatif hipotez kurulmasında bu değer önemlidir. Aşağıdaki tablo incelendiğinde ve d=2(1-r) yaklaşık formülü dikkate alındığında ilişkinin yönü tablo değerine göre yorumlanabilmektedir. 0 ve 5 arasındaki değerlerde eğer bir öz ilişki varsa bu öz ilişkinin pozitif olacağı biliniyor. Alternatif hipotez + öz ilişki vardır şeklinde kurulmalıdır. k=5 ve n=59 (tabloda n=59 değeri olmadığı için en yakın değer olan n=60 seçilmiştir.) değerine karşılık gelen alt ve üst sınırlar dl=1.41 du=1.77 şeklindedir. Buna göre 0.15535 değeri tablo alt değeri olan 1.41’ den küçük olduğu için H0 hipotezini reddedemeyiz. Ayrıca Durbin-Watson testindeki p değeri anlamlılık düzeyimiz olan %5’ten büyük olduğu için H0 hipotezini reddedemeyiz. Öz ilişki olmadığını %95 güven düzeyinde söyleyebiliriz.

# Çoklu Bağlantı Sorunu

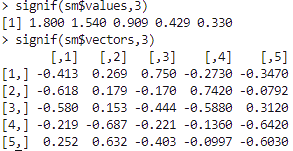
 

Korelasyon matrisine göre denklem sabiti ile su değişkeni arasında negatif yönlü güçlü bir korelasyon mevcuttur. Çoklu bağlantıdan şüphelenebilmek için VIF katsayılarına, koşul sayılarına, özdeğerlere ve özvektörlere bakmamız gerekiyor.

VIF katsayılarında çoklu bağlantıdan şüphelenebilmek için 10’dan büyük bir değere sahip olmamız gerekir ama modelimizdeki VIF katsayıları tam istediğimiz gibi 10’dan küçük ve 1’e çok yakındırlar. VIF katsayılarına bakara çoklu bağlantıdan şüphelenemeyiz.

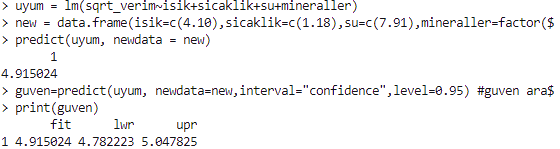


30’dan büyük koşul sayısına sahip durumlar için güçlü çoklu bağlantı yorumu yapabiliriz. 10 ile 30 arasında koşul sayısına sahip durumlarda zayıf çoklu bağlantı yorumu yapabiliriz. Buna göre sabit değerimiz ile su değişkenimiz arasında zayıf çoklu bağlantıdan şüpheleniyoruz.

Özdeğerlere baktığımızda 0’a yakın bir değer vardır. Bu durumda çoklu bağlantı vardır yorumunu yapabiliriz.

# Uyum Kestirimi

R üzerinde aşağıdaki kod ile uyum kestirimi uygulayıp güven aralıkları buluyoruz.



Uyum kestirimi veri setimizdeki bağımsız değişkenlerimizin kestirim denkleminde yerine yazılmasıyla elde edilir. Bu örnekte bağımsız değişkenlerimizi için seçilen değerler veri setimizdeki ilk değerlerdir. Bu değerleri yerine yazdığımızda denklemimiz şu şekilde olmakta.

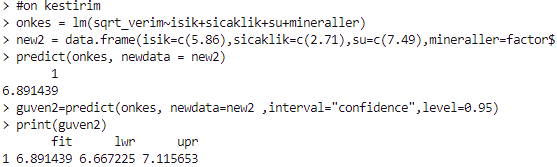
Uyum kestirimi için güven aralıklarını şu şekilde gösterebiliriz.

P(4.782 ≤ ≤ 5.047) = 0.95

Güven aralığını, yerine yazdığımız bağımsız değişkenler için bulacağımız kestirim değerinin 4.782 ile 5.047 arasında olması olasılığı %95’tir şeklinde yorumlayabiliriz.

# Ön Kestirim

R üzerinde aşağıdaki kod ile ön kestirim uygulayıp güven aralıklarını buluyoruz



Ön kestirim veri setimizde olmayan bağımsız değişkenlerin kestirim denkleminde yerine yazılmasıyla elde edilir. Bu örnek için kod üzerinde de görüldüğü gibi Işık değişkeni 5.86 Sıcaklık değişkeni 2.71 Su değişkeni 7.49 Mineraller değişkeni ise 2 durumunda alınmıştır. Bağımsız değişkenlerimizi yerine yazınca denklemimiz şu şekilde oluyor.

Ön kestirim için güven aralıklarını şu şekilde gösterebiliriz

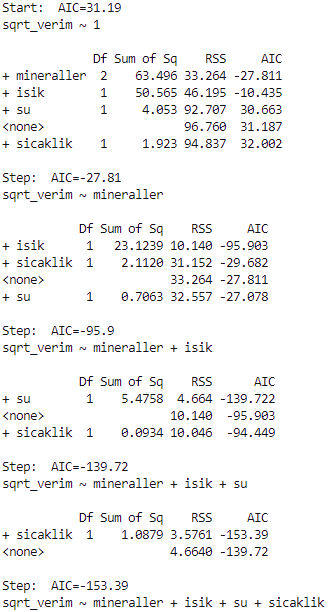
P(6.66 ≤ ≤ 6.89) = 0.95

Güven aralığını, yerine yazdığımız bağımsız değişkenler için bulacağımız kestirim değerinin

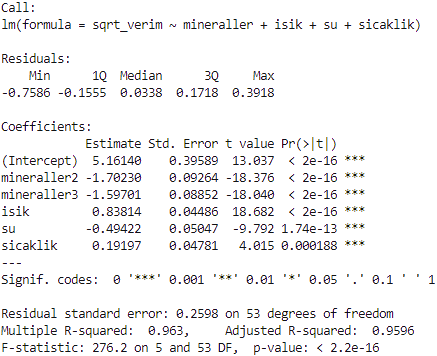
6.66 ile 6.89 arasında olması olasılığı %95’tir diye yorumlayabiliriz.

# Değişken Seçimi Yöntemleri

## İleriye Doğru Seçim Yöntemi



İleri doğru seçim yönteminde öncelikle mineraller değişkenini ekliyoruz ve anlamlı buluyoruz. Ardından ışık değişkenini ekliyoruz ve anlamlı buluyoruz. Sonra sırasıyla su ve sıcaklık değişkenlerini ekleyerek anlamlılıklarını test ediyor ve anlamlı buluyoruz



Kod üzerinde de görüldüğü gibi ileri doğru seçim yöntemine göre en iyi kestirim denklemimiz şu şekilde yazılabilir.

İleri doğru seçim yöntemine göre kurduğumuz modelin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

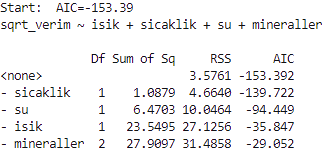
H0: β1=β2=β3=β4=β5=0 (Model anlamlı değildir.)

H1: β1,β2,β3,β4,β5’ten en az birisi sıfırdan farklıdır. (Model anlamlıdır.)

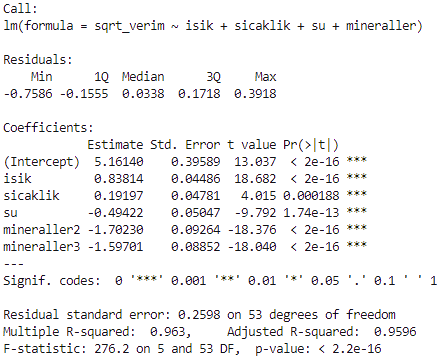
Modelimize ait p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezini reddederiz. İleri doğru seçim yöntemine göre kurduğumuz model %95 güven düzeyinde anlamlıdır.

## Geriye Doğru Seçim Yöntemi

Geriye doğru seçim yönteminde modelimize tüm bağımsız değişkenleri ekleyerek başlıyoruz ve bağımsız değişkenleri sırasıyla çıkartarak anlamlılığı test ediyoruz. Geriye doğru seçim yönteminde R çıktısını şu şekilde elde ediyoruz.



Çıktı üzerinde de görüldüğü gibi modelimizden hiçbir değişkeni çıkaramıyoruz. Geriye doğru seçim yöntemine göre en iyi kestirim denklemi şu şekilde yazılabilir



Tüm değişkenlerimizin dahil olduğu modeli kurup anlamlılığını test ediyoruz.

Modelin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β1=β2=β3=β4=β5=0 (Model anlamlı değildir.)

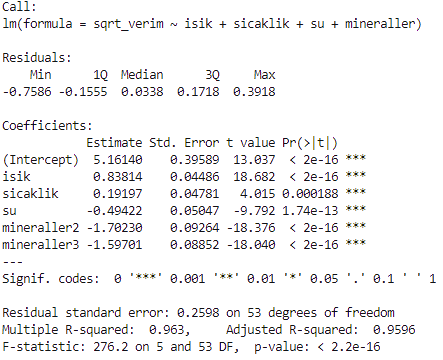
H1: β1,β2,β3,β4,β5’ten en az birisi sıfırdan farklıdır. (Model anlamlıdır.)

Modelin p değeri anlamlılık düzeyi olan %5’ten küçüktür. Bu durumda H0 hipotezini reddederiz. Geriye doğru seçim yöntemine göre kurduğumuz model %95 güven düzeyinde anlamlıdır.

İleri ve geriye doğru seçim yöntemlerimiz aynı sonucu vermiştir.

## Adımsal Seçim Yöntemi

Adımsal seçim yönteminde öncelikle ileriye doğru seçim yönteminde olduğu gibi değişkenleri sırayla ekleyerek başlarız. Farklılık olarak her değişkeni ekledikten sonra önceden eklediğimiz değişkenlerin anlamlılığını kaybedip kaybetmediğine bakarız. Adımsal seçim yönteminin R üzerindeki çıktısı şu şekildedir.



Adımsal seçim yöntemine göre sırasıyla değişkenlerimizi ekledikçe daha önce eklediğimiz değişkenler anlamlılıklarını kaybetmemişlerdir. Bu sebeple bütün bağımsız değişkenlerimiz modele dahil olmuştur. Modelimizin anlamlılığını test etmek için hipotez testini kuruyoruz.

H0: β1=β2=β3=β4=β5=0 (Model anlamlı değildir.)

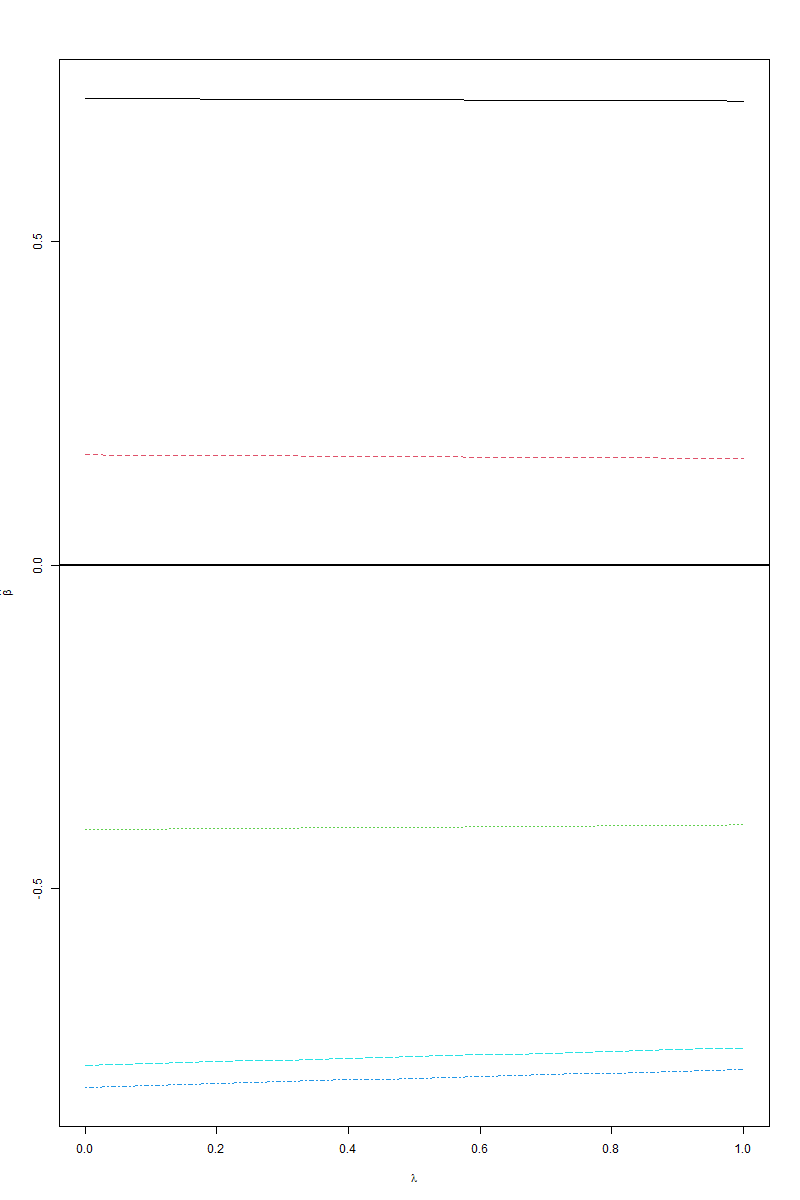
H1: β1,β2,β3,β4,β5’ten en az birisi sıfırdan farklıdır. (Model anlamlıdır.)

Modelin p değeri anlamlılık düzeyi olan %5’ten küçük olduğu için H0 hipotezini reddediyoruz. Adımsal seçim yöntemine göre oluşturduğumuz model %95 güven düzeyinde anlamlıdır.

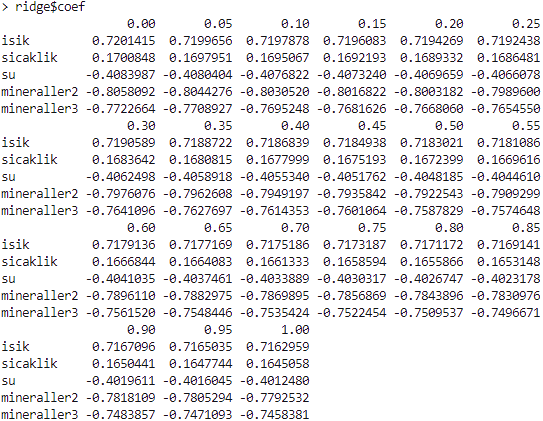
İleri doğru seçim, geriye doğru seçim ve adımsal seçim yöntemleri aynı sonucu vermiştir. Buna göre modelimiz oldukça anlamlı ve güvenilirdir.

# Ridge Regresyonu

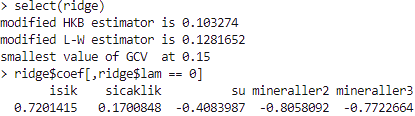
Ridge regresyonuna göre çoklu bağlantı sorunu yaşadığımız verilerimize uygun miktarda yan değeri ekleyerek çoklu bağlantı sorunundan kurtuluruz. Ridge izi grafiğine bakarak verilerimizde çoklu bağlantı olup olmadığını, çoklu bağlantı var ise ne kadarlık bir yan değerinin çoklu bağlantıyı gidereceğini görebiliriz. Ridge izi grafiği şöyledir.



Grafiğe bakıldığında λ değeri artmasına rağmen katsayılarda değişiklik yoktur. Buna göre çoklu bağlantı durumu yoktur. Sıfır civarında seyreden bir katsayı da olmadığı için anlamsız değişken yoktur yorumu yapabiliriz.



Sayısal olarak da baktığımız zaman yan değeri arttıkça katsayılar neredeyse hiç değişmemektedir. Yan değerini mümkün olduğunca küçük almak daha iyidir. Bu durumda yan değerini 0 almak en iyisi olacaktır.



Hızlı artış ya da azalış gösteren katsayılara karşılık gelen değişkenler çoklu bağlantılı değişkenlerdir. Burada mineraller2 ve mineraller3 çoklu bağlantıdan etkilenmiş değişkenlerdir.

Sıfır eksenin civarında seyreden değişkenler de modelde önemsiz değişkeni göstermektedir. Burada x2 sıfır eksenine çok yakın seyrettiği için sıcaklık değişkeninin modele katkısının anlamsız olduğu söylenebilir.