

Fen Fakültesi Aktüerya Bilimleri Bölümü

İST359 Regresyon Çözümlemesi

Regresyon Süreci Ödevi

Prof. Dr. Duru KARASOY

Hakkı KONDAK – 2220381067

17/12/2024

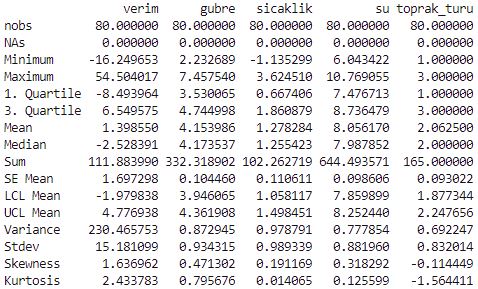
# İndeks

1. [Senaryo](#_Senaryo)
2. [Tanımlayıcı İstatistikler](#_Tanımlayıcı_İstatistikler)
3. [Varsayım Kontrolleri](#_Varsayım_Kontrolleri)
   1. [Normallik Varsayımı](#_Normallik_Varsayımı)
   2. [Doğrusallık Varsayımı](#_Doğrusallık_Varsayımı)
4. [Doğrusal Modelin Kuruluşu](#_Doğrusallık_Modelinin_Kuruluşu)
5. [Artıkların İncelenmesi](#_Artıkların_İncelenmesi)
   1. [Hat Uzaklığı](#_Hat_Uzaklığı_1)
   2. [Student Türü Artıklar](#_Student_Türü_Artıklar)
   3. [Cook Uzaklığı](#_Standart_Türü_Artıklar)
   4. [Standart Türü Artıklar](#_Standart_Türü_Artıklar_1)
6. [Yeni Veri Setinde Tanımlayıcı İstatistikler](#_Yeni_Veri_Setinde_2)
7. [Yeni Veri Setinde Varsayım Kontrolleri](#_Yeni_Veri_Setinde_1)
   1. [Normallik Varsayımı](#_Normallik_Varsayımı_1)
   2. [Doğrusallık Varsayımı](#_Doğrusallık_Varsayımı_1)

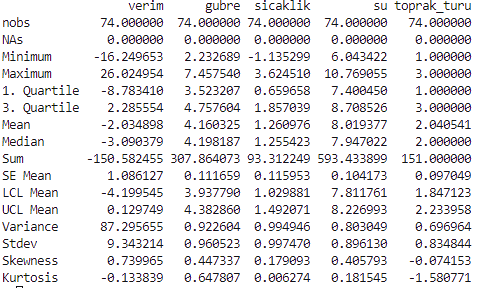
# Senaryo

# Tanımlayıcı İstatistikler

İlk datamızın tamamlayıcı istatistikleri



Yeni datamızın tamamlayıcı istatistikleri

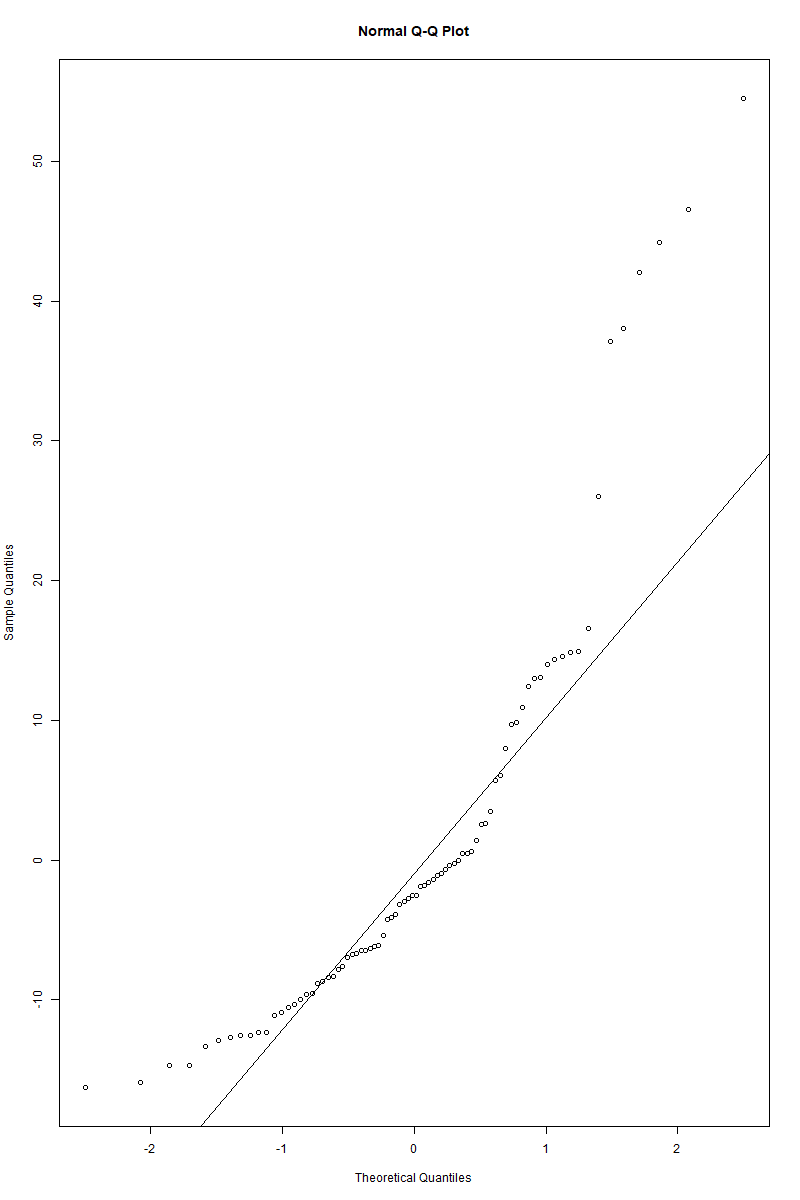


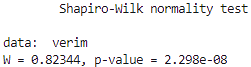
Bağımlı ve bağımsız değişkenlerimize ait en büyük, en küçük, ortalama, ortanca, varyans, standart sapma, basıklık ve çarpıklık değerleri gibi tanımlayıcı istatistikler tabloda görülmektedir.

# Varsayım Kontrolleri

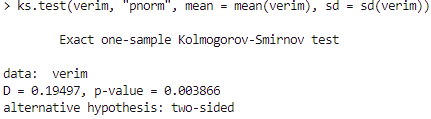
## Normallik Varsayımı

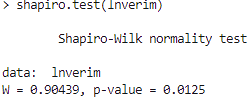
Shapiro-wilk testimize bakıyoruz





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.

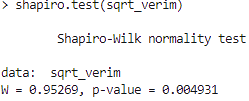
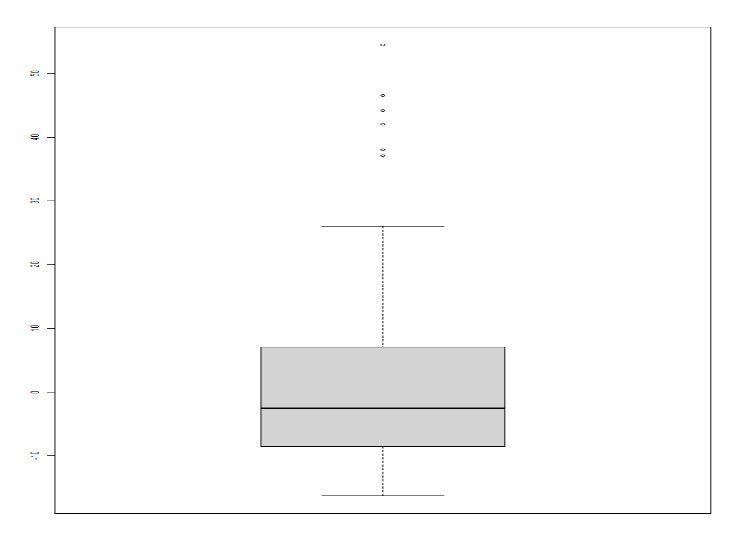
log dönüşümü yapıp bakıyoruz normalliğe. p-value 0.05’den küçük olduğu için normallik sağlanmadığını görüyoruz.

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

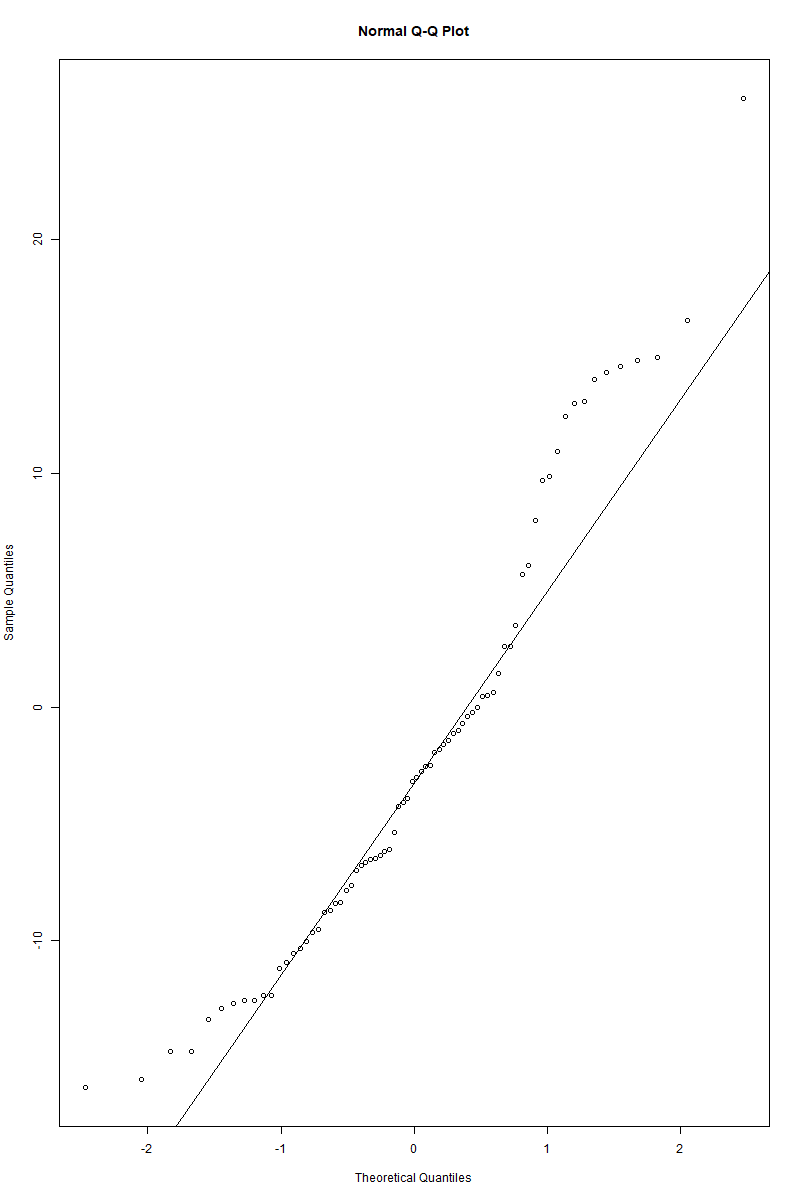
**Normallik sağlamazdığını görüyoruz ve Boxplot grafiğinden artık değerlerini bulup çıkarıyoruz**



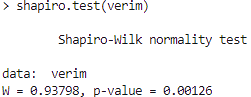
Görüldüğü üzere 17, 33, 46, ,64, 69 ve 72 artık değerlerimizdir. Bunları silip tekrardan datamızı oluşturacaz.



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



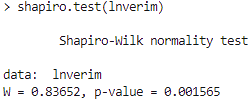
Grafikte normal dağılıyor gibi gözükse de shapiro normallik testi yapınca p-value 0.05’den küçük olduğu için H0 reddedilir. Modelimiz normal dağılıma uygun değildir.

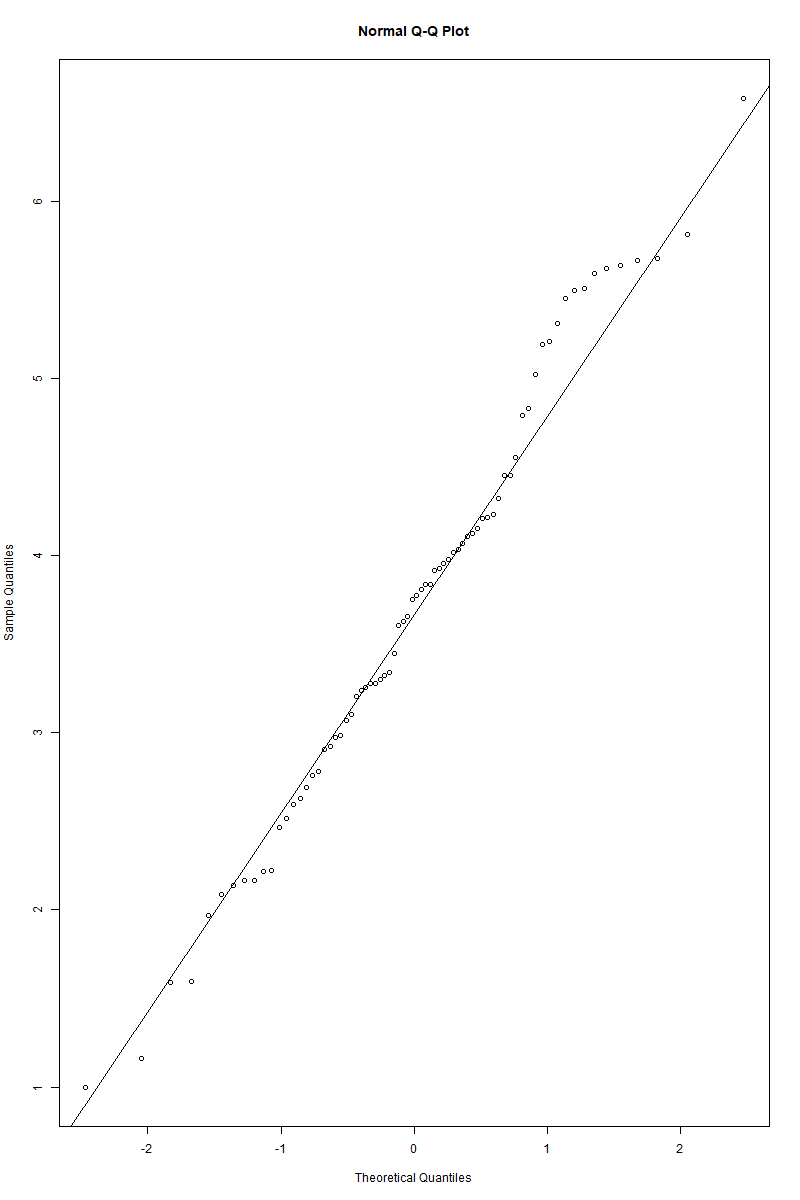


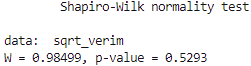
p-value 0.05’den küçük olduğu için normallik sağlanmaz.

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**.





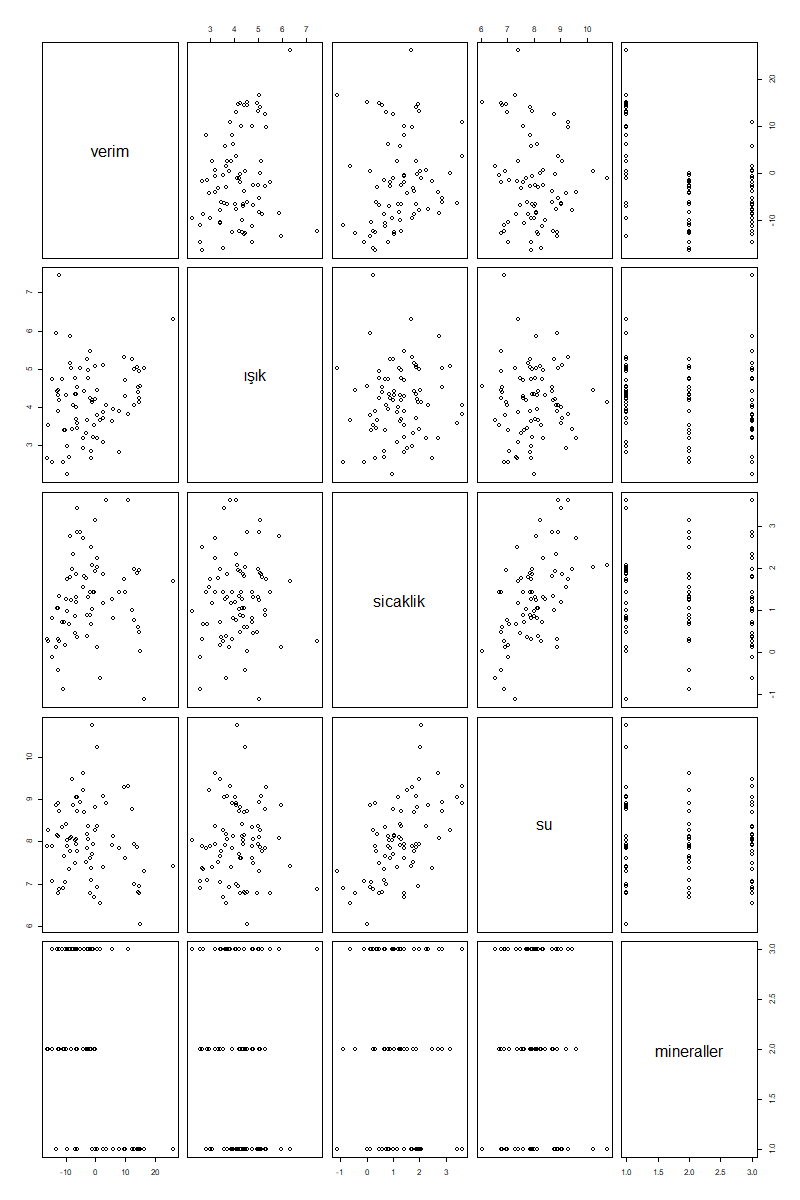
Hipotez testimizi kuruyoruz.

H0: Verilerimiz ile normal dağılım arasında fark yoktur.

H1: Verilerimiz ile normal dağılım arasında fark vardır.

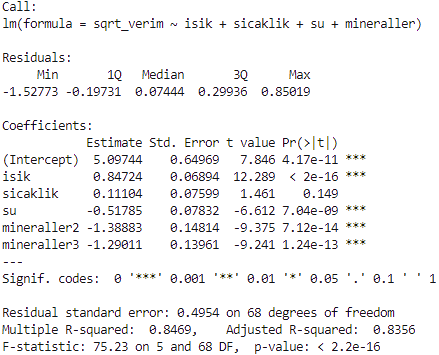
p değerimiz anlamlılık düzeyimiz olan %5’ten büyüktür. Bu durumda H0 hipotezimizi reddedemeyiz. Bağımlı değişkenimizin %95 güven düzeyinde anlamlı olduğunu söyleyebiliriz.

## 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.

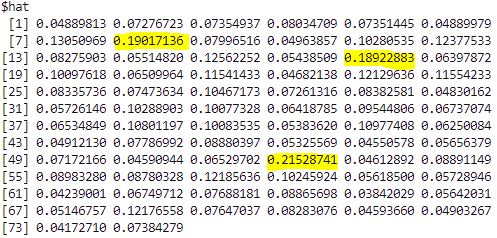
H0: β1=β2=β3=β4=β5=0 (Model anlamlı değildir.)

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

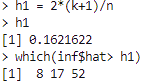
Modelimizdeki p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. H0 hipotezimizi reddederiz. Modelimiz %95 güven düzeyinde anlamlıdır. Toprak\_turu

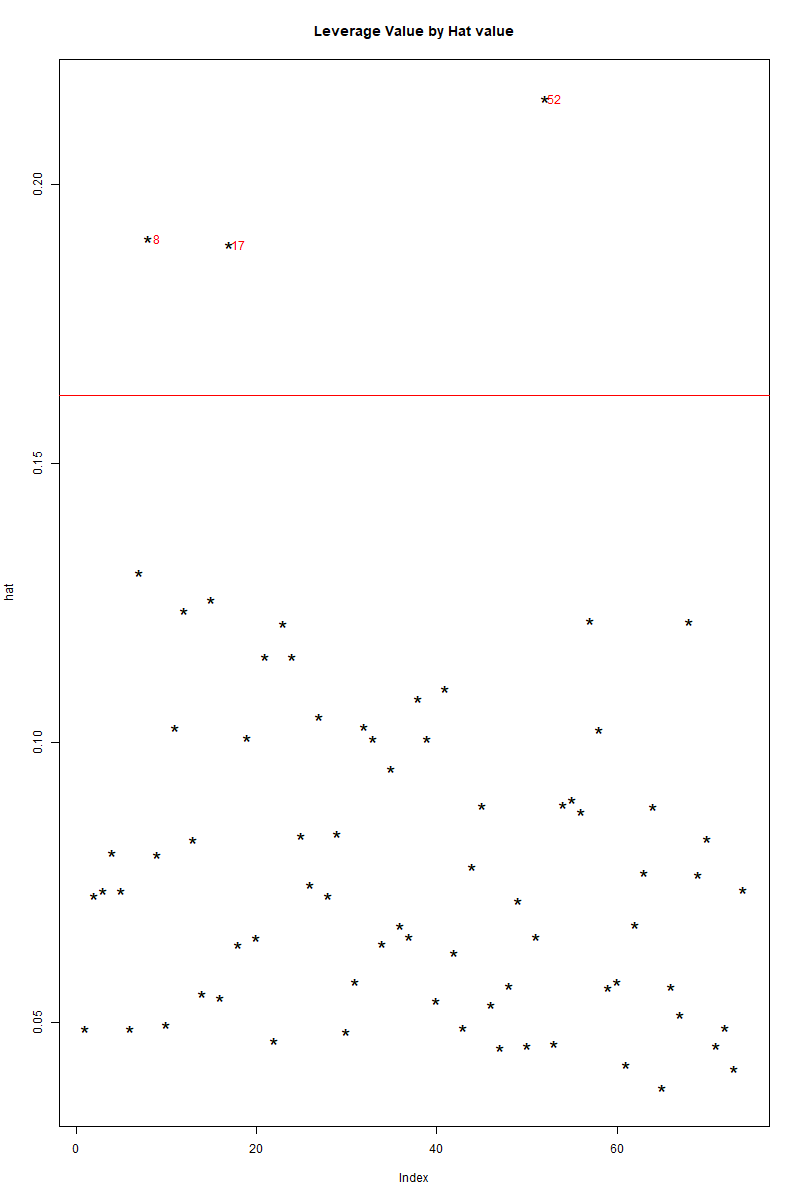
# Artıkların İncelenmesi

## Hat Uzaklığı



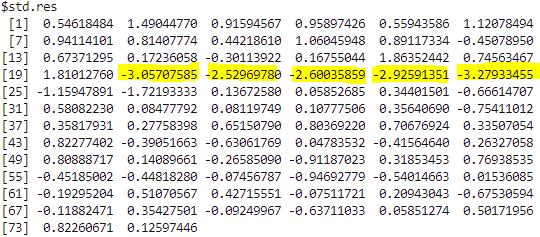
Hat Uzaklığı artık incelemesine göre değerinden büyük olan gözlemlerimiz uç değerlerdir.



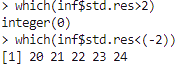


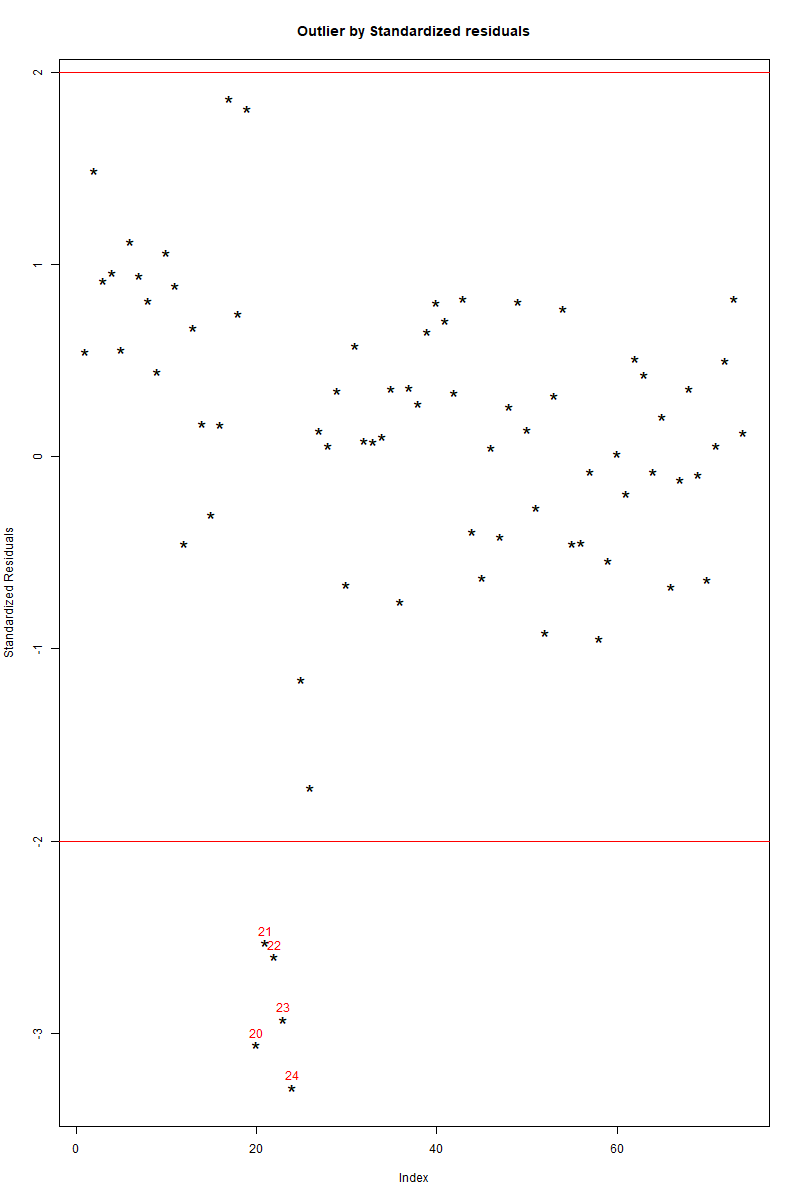
Kod üzerinde ve grafikte de görüldüğü üzere Hat Uzaklığı artık incelmesine göre 8, 17 ve 52. gözlem uç değerdir.

## Standart Türü Artıklar



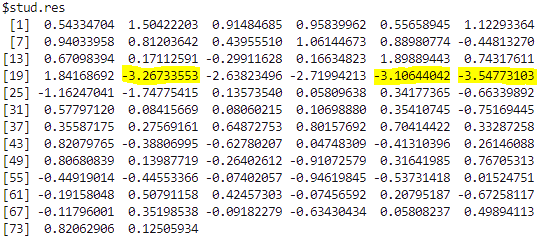
Standart Türü artıklara göre değerlerimiz (-2,2) aralığında olmalıdır. Aralık dışındaki değerler Standart Türü artıklar tablomuza göre artık değerlerdir.



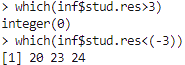


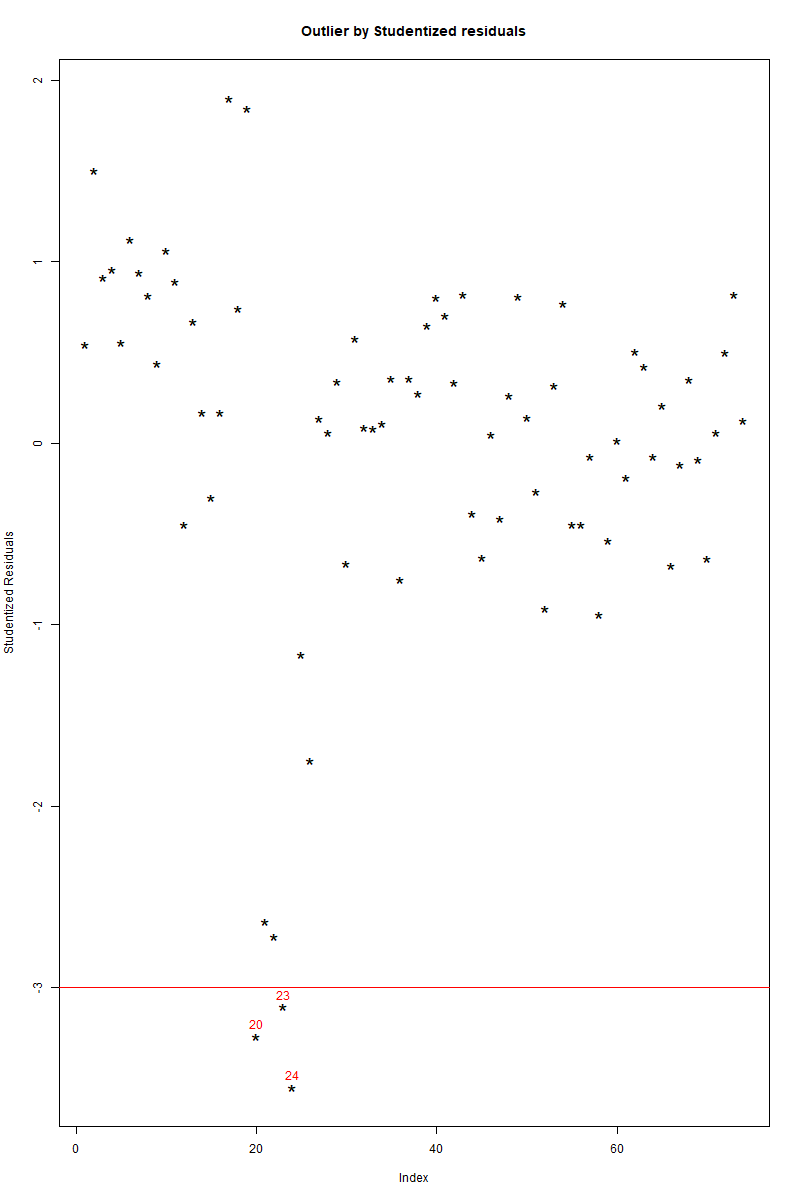
Grafikte ve kod üzerinde de görülüğü üzere 20, 21, 22, 23 ve 24. gözlemlerimiz Standart Türü artık incelmesine göre artık değerlerdir.

## Student Türü Artıklar



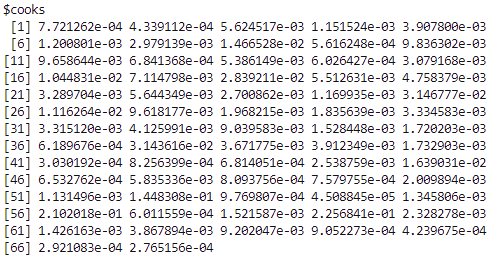
Student Türü artık incelemesine göre (-3,3) aralığı dışında kalan değerler aykırı değerlerdir.



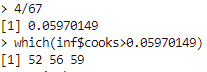


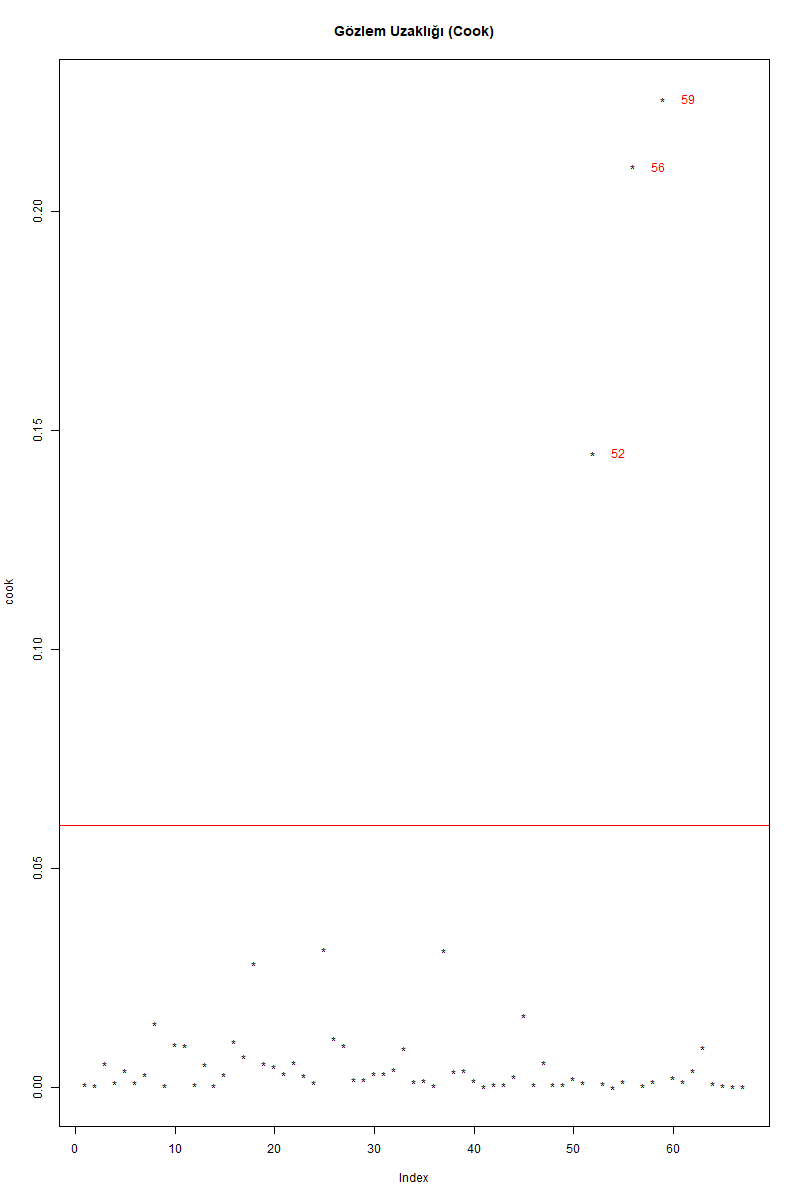
Grafikte ve kod üzerinde de görüldüğü üzere 20, 23 ve 24. Gözlemlerimiz Student Türü artık incelemesine göre artık değerlerdir.

## Cook Uzaklığı



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

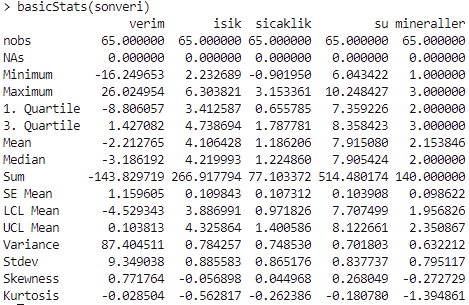




R kodu ve grafik üzerinde de görüldüğü üzere 52, 56 ve 59. gözlemlerimiz Cook Uzaklığı artık incelemesine göre artık değerlerdir.

Yeni verilerimiz üzerinde yaptığımız artık incelemesi sonucunda 19, 28, 38, 45, 51, 54, 62, 64, 84 ve 93. gözlemlerimiz atık değerlerimizdir.

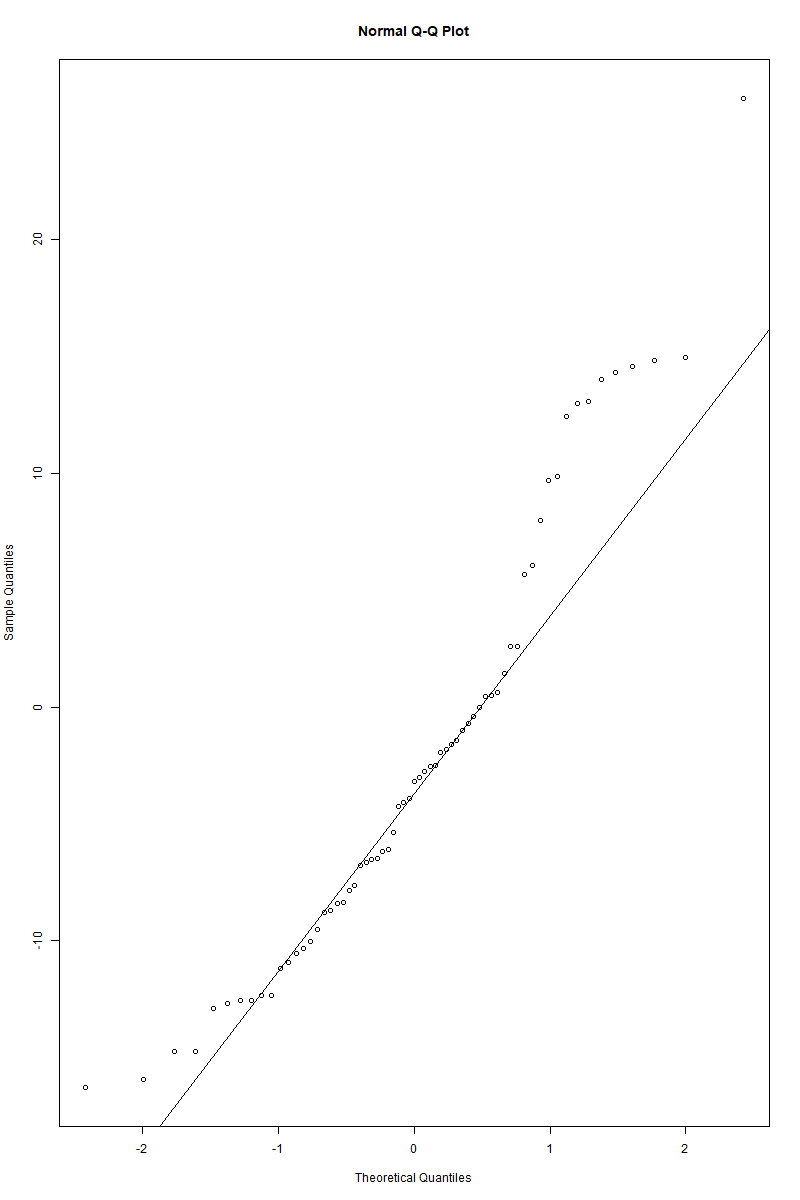
# Yeni Veri Setinde Tanımlayıcı İstatistikler



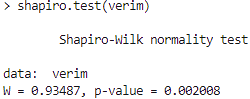
Artıkları çıkartılmış yeni veri setimize ait tanımlayıcı istatistikler tabloda görülmektedir.

# Yeni Veri Setinde Varsayım Kontrolleri

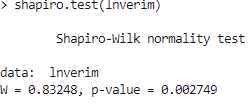
## Normallik Varsayımı

****

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

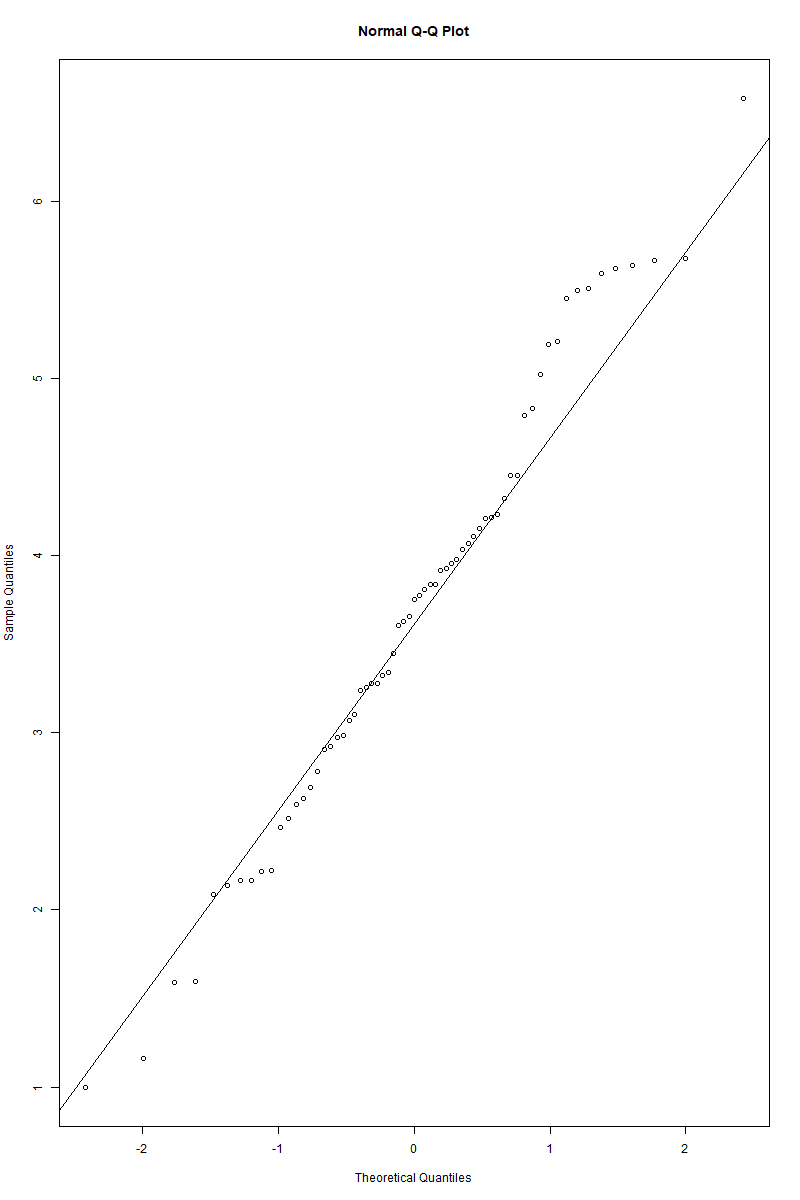


p-value 0.05’den küçük olduğu için normallik sağlanmaz

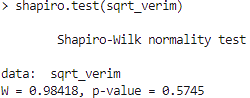


log dönüşümü yapıp bakıyoruz normalliğe. p-value 0.05’den küçük olduğu için normallik sağlanmadığını görüyoruz.

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



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



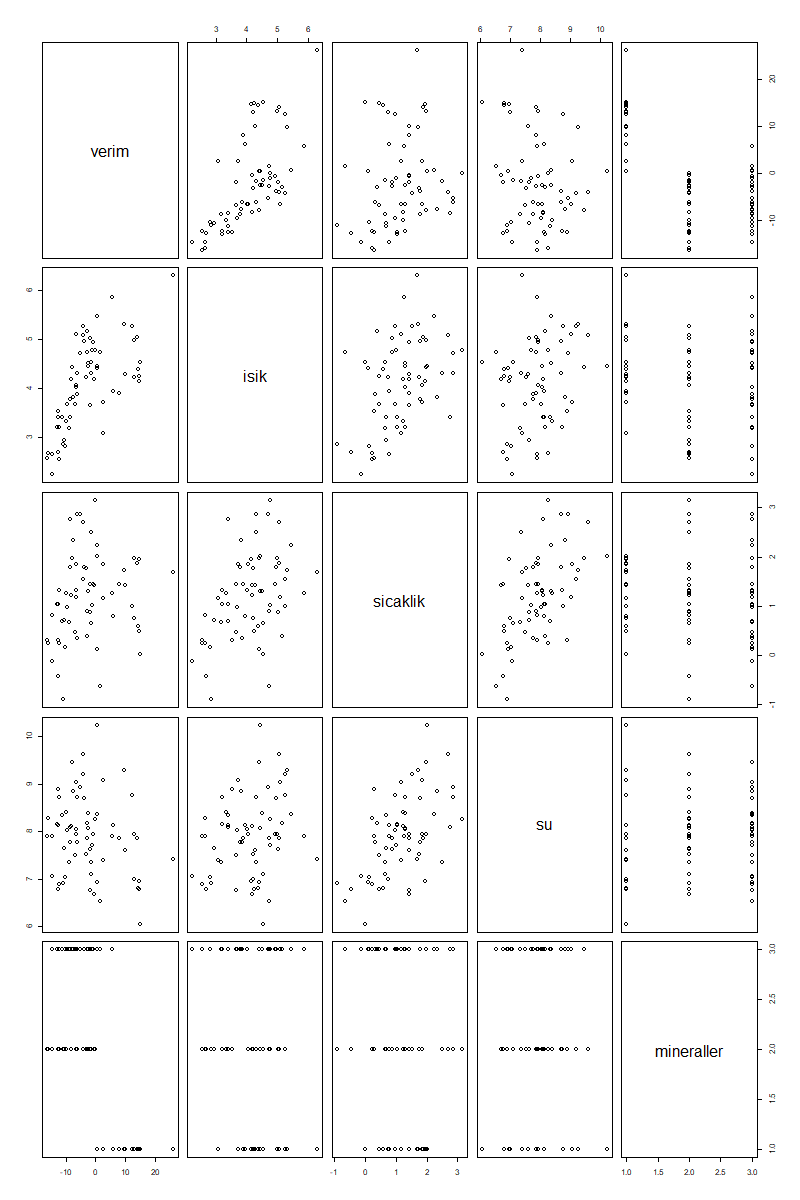
Hipotez testimizi kuruyoruz.

H0: Verilerimiz ile normal dağılım arasında fark yoktur.

H1: Verilerimiz ile normal dağılım arasında fark vardır.

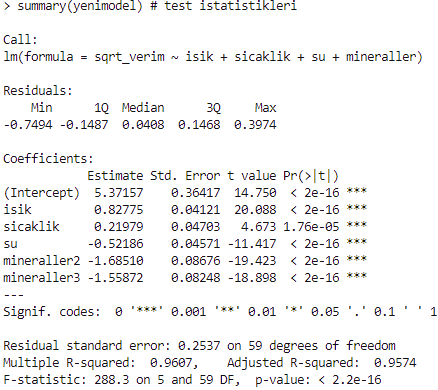
p değerimiz anlamlılık düzeyimiz olan %5’ten büyüktür. Bu durumda H0 hipotezimizi reddedemeyiz. Bağımlı değişkenimizin %95 güven düzeyinde anlamlı olduğunu söyleyebiliriz.

## 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.

Nakit değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β1=0 (Nakit değişkeni anlamlı değildir)

H1: β1≠0 (Nakit değişkeni anlamlıdır)

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

Borç değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β2=0 (Borç değişkeni anlamlı değildir)

H1: β2≠0 (Borç değişkeni anlamlıdır)

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

Maaş değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

H0: β3=0 (Maaş değişkeni anlamlı değildir)

H1: β3≠0 (Maaş değişkeni anlamlıdır)

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

Konum değişkenimizin anlamlılığını test etmek için hipotez testimizi kuruyoruz.

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

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

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

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

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

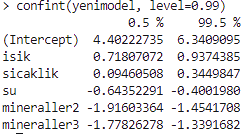
Konum değişkenimizin 3 (perakendeci) olduğu durumun p değeri anlamlılık düzeyimiz olan %5’ten küçüktür. Bu durumda H0 hipotezimizi reddederiz. Konum değişkenimizin 3 durumunda olması modelimiz için %95 güven düzeyinde anlamlıdır diyebiliriz.

**Katsayı Yorumları**

Kestirim denklemimizin sabiti, modelimizdeki diğer değişkenlerimiz sabitken bir şirketin karlılığının ortalama -0.700 olduğunu gösterir. Yani şirketler zarar durumundadır. Nakit değişkenimizin katsayısı, diğer değişkenlerin sabit olduğu durumda kasadaki nakit miktarının 1 birimlik artışının şirket karlılığında ortalama 8.858 birimlik artışa sebep olduğunu gösterir. Borç değişkeninin katsayısı bize, diğer değişkenlerin sabit olduğu durumda şirketin borcundaki 1 birimlik artışın şirket karlılığında ortalama 4.413 birimlik artışa sebep olduğunu gösterir. Maaş değişkeni katsayısı bize, diğer değişkenlerin sabit olduğu durumda maaş ödemelerindeki 1 birimlik artışın şirket karlılığında -0.710 birimlik azalışa sebep olduğunu gösterir. Konum değişkenimizin 1 olması durumunun (şirketin üretici konumda olması) şirket karlılığına bir etkisi yoktur. Konum değişkenimizin 2 olması durumunun (şirketin toptancı olması durumu) şirket karlılığına -6.904 birimlik bir etkisi vardır. Konum değişkenimizin 3 olması durumunun (şirketin perakendeci konumda olması) şirket karlılığına -6.960 birimlik bir etkisi vardır.

Belirtme katsayımız 0.9998’dir. Buna göre bağımlı değişkenimizdeki (şirket karlılığı) değişimin %99.98’i bağımsız değişkenlerimizce (Kasadaki nakit miktarı, borç miktarı, maaş ödemesi, şirketin sektördeki konumu) açıklanabilmektedir.

## Güven Aralıkları Ve Yorumları



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

P(4.402 ≤ β0 ≤6.340)=0.99 Denklem sabitimizin 4.402 ile 6.340arasında değer alması olasılığı %99’dur.

P(0.718 ≤ β1 ≤0.937)=0.99 Işık değişkenimizin katsayısının 0.718 ile 0.937aralığında bir değer alması olasılığı %99’dur.

P(0.094 ≤ β2 ≤0.344)=0.99 Sıcaklık değişkenimizin katsayısının 0.094 ile 0.344aralığında bir değer alması olasılığı %99’dur.

P(-0.643≤ β3 ≤-0.400)=0.99 Su değişkenimizin katsayısının -0.643 ile -0.400aralığında bir değer alması olasılığı %99’dur.

P(-1.916 ≤ β4 ≤-1.454)=0.99 Mineral değişkenimizin 2 olduğu durumda şirket karlılığına etkisinin

-1.916 ile -1.454aralığında olması olasılığı %99’dur.

P(-1.778 ≤ β5 ≤-1.339)=0.99 Konum değişkenimizin 3 olduğu durumda şirket karlılığına etkisinin

-1.778 ile -1.339aralığında olması olasılığı %99’dur.

# Değişen varyanslılık