# 



HACETTEPE ÜNİVERSİTESİ- İSTATİSTİK BÖLÜMÜ

İSTATİSTİKSEL ANALİZ UYGULAMALARI

Ders Sorumlusu : Dr. Onur Toka

Hazırlayan : Şükriye Nur Şencan

İçindekiler

[GİRİŞ 1](#_Toc30362022)

[GEREKLİ PAKETLER 1](#_Toc30362023)

[VERİ HAZIRLAMA 2](#_Toc30362024)

[VERİ ÖN İŞLEME VE ÖZETLEYİCİ BİLGİLER 3](#_Toc30362025)

[Veri İmportu 3](#_Toc30362026)

[Veri Ön işleme 3](#_Toc30362027)

[Veri için Özetleyici Bilgiler 3](#_Toc30362028)

[Değişkenlere Genel Bakış 4](#_Toc30362029)

[ANALİZ 4](#_Toc30362030)

[DEĞİŞKENLER VE GRAFİK YORUMLARI 4](#_Toc30362031)

[GRE Skoru Değişkeni 4](#_Toc30362032)

[TOEFL Skoru Değişkeni 5](#_Toc30362033)

[Scatter Plot 6](#_Toc30362034)

[CGPA (Akademik Ortalama) Değişkeni 8](#_Toc30362035)

[Change\_ofAdmit Değişkeni 9](#_Toc30362036)

[Kategorik değişkenler için Grafikler 10](#_Toc30362037)

[University\_Rating(Üniversite Sınıflandırması) ve Reserch(Araştırma Deneyimi) Değişkenleri 10](#_Toc30362038)

[Lollipop Grafiği 18](#_Toc30362039)

[ANOVA 20](#_Toc30362040)

[WİLCOX TESTİ 22](#_Toc30362041)

[REGRESYON 23](#_Toc30362042)

[Çoklu Doğrusal Regresyon 23](#_Toc30362043)

[Bağımlı Değişkene ait Q-Q Plot Grafiği 23](#_Toc30362044)

[Doğrusallık 23](#_Toc30362045)

[Katsayı Yorumları 25](#_Toc30362046)

[Regresyon Katsayıları için Güven Aralığı 26](#_Toc30362047)

[Ön Kestirim 27](#_Toc30362048)

[K EN YAKIN KOMŞU SINIFLANDIRMA ALGORİTMASI 27](#_Toc30362049)

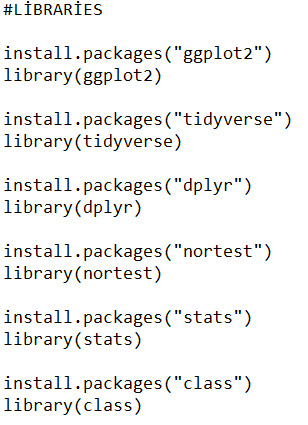
[ÖZET 29](#_Toc30362050)

# GİRİŞ

Bu veriyi gördüğümde değişkenlerin master programına kabul olasılığını nasıl etkilediğinin araştırıldığını düşündüm. Hangi değişkenlerin kabul olasılığını olumlu etkilediği hangi değişkenlerin bu olasılığı olumsuz etkilediği araştırılıyor olabilir.

Bu probleme yapılacak istatistiksel analiz , bir kişinin master programına kabul edilme olasılığının kaç olduğu ve master programına katılmak isteyen birinin hangi koşulları yerine getirmesi gerektiği hakkında bilgi sahibi olmada yardımcı olur.

# GEREKLİ PAKETLER



**Ggplot2:** kütüphanesi R’daki en iyi veri görselleştirme araçlarındandır. Grafik Gramerine dayalı olarak kullanılan bir kütüphanedir. Grafiklerle, grafiğe dayalı estetikle ve ayrıntılarla ilgilenir.

**Tidyverse :** düzenlemelerin, görselleştirmelerin, modellemelerin kolay bir şekilde yapılmasını sağlayan R’ın birçok paketini içinde bulunduran büyükçe bir pakettir. Veri bilimi ile uğraşanların sistematik bir şekilde ilerlemesine, işlemler arasında bağlantıların kolay kurulup anlaşılmasına yardımcı olur.

**Dplyr :** 5 ana fiilden oluşmaktadır. Bunlar Filter, select, mutate, arrange, summarise’dır. Bunların hepsi bir araya gelerek “grup bazında” herhangi bir işlemi gerçekleştirmenizi sağlar. dplyr, verilerin nasıl saklandığını özetlemek için tasarlanmıştır.

**Nortest :** Normallik hipotezlerini test eden bir kütüphanedir.

**Stats :** Bu paket istatistiksel hesaplamalar ve rastgele sayı üretimi için işlevler içerir.

**Class :** Sınıflandırma için k en yakın komşu dahil olmak üzere çeşitli fonksiyonları içeren bir kütüphanedir.

# VERİ HAZIRLAMA

Veri kümesi master programına kabul durumuna ait 4 nicel (GRE skoru, TOEFL skoru, niyet ve referans mektuplarını gücü, akademik ortalama) ,2 kategorik (Üniversite Sınıflandırması, Araştırma Deneyimi) ve 1 bağımlı (Kabul edilme olasılığı) değişken ve 200 gözlemden oluşmaktadır.

Değişkenler:

1. GRE Skoru (Maksimum 340)

2. TOEFL Skoru (Maksimum 120)

3. Universite Sınıflandırması (1 En kötü 5 En iyi)

4. Niyet Mektubu ve Referans Mektuplarının Gücü (En iyisi 5 puan)

5. Lisans Akademik Ortalaması (10 üzerinden)

6. Araştırma Deneyimi (0: Yok 1 : Var)

7. Kabul Edilme Olasılığı (Chance\_ofAdmit)

## VERİ ÖN İŞLEME VE ÖZETLEYİCİ BİLGİLER

### Veri İmportu

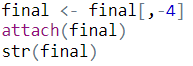


Veri import edilmiştir.

### Veri Ön işleme

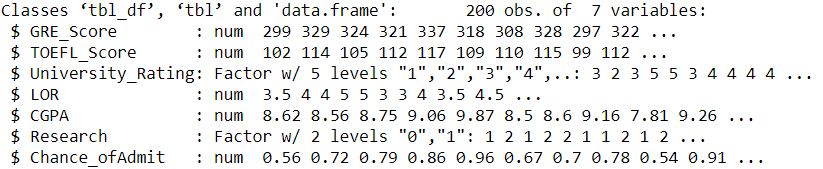


Veride kayıp gözlem olup olmadığına sum(is.na()) kodu ile bakılmıştır ve kayıp gözlem olmadığı görülmüştür.



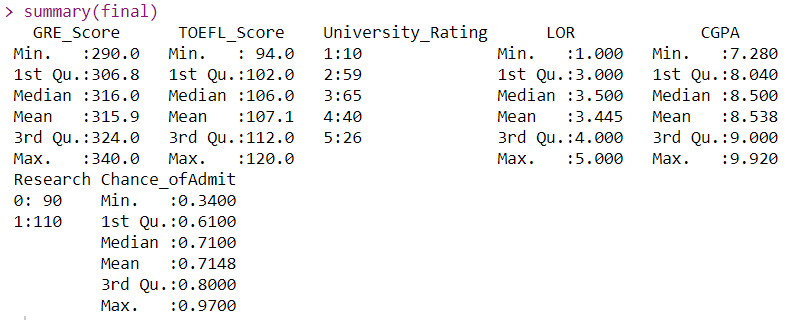
Analize dahil edilmeyeceği için “SOP“ değişkeni veri seti içerisinden kaldırılmıştır. str() fonksiyonu ile değişkenlerin tiplerine bakılmıştır. Kategorik değişkenler numeric olarak gözüktüğü için bu değişkenlerin tipleri değiştirilip faktör haline getirilmiştir.





Tabloya bakıldığında Üniversity\_Rating ve Reserch değişkenlerinin tipinin faktör olarak değiştiği görülmektedir.

### Veri için Özetleyici Bilgiler

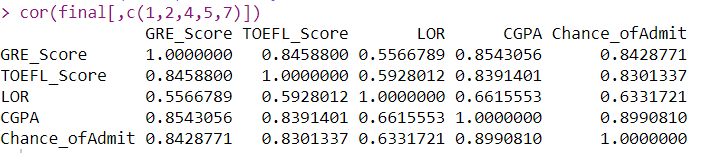


Summary() fonksiyonu ile veri setindeki numerik değişkenlere ait minimum, maksimum değerler , çeyreklikler kısacası özetleyici bilgiler verilmiştir. Örneğin GRE\_Score değişkenine ait maksimum değerin 340, minimum değerin de 290 olduğu görülmektedir. Ortalama skor değerinin 315.9 ve medyan değeri ise 316’dır. 1. Çeyreklik değerinin 306.8 ve 3. Çeyrek değerinin de 324 olduğu görülmektedir.

Üniversite sınıflandırması değişkeninde (1)en kötü üniversiteden 10 tane ve (5) en iyi üniversiteden de 26 tane olduğu görülmektedir.

Araştırma değişkenine bakıldığında araştırma deneyimi olmayan 90 , araştırma deneyimi olan 110 kişinin olduğu görülmektedir.

### Değişkenlere Genel Bakış



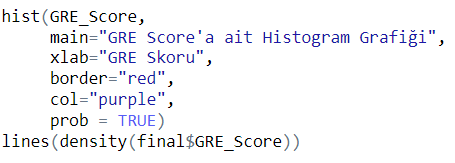
Cor() fonksiyonu ile değişkenler arasındaki korelasyon değerlerine bakılmıştır. Buna göre Chace\_ofAdmit(Kabul edilme olasılığı) bağımlı değişkeni ile GRE\_Score ,TOEFL\_Score ve CGPA(akademik ortalama) bağımsız değişkenleri arasında pozitif yönlü yüksek bir ilişki olduğu görülmektedir.

GRE skoru ile TOEFL skoru değişkenleri arasında da 0.84’lük pozitif ve yüksek bir ilişki olduğu söylenebilir. GRE skoru ve TOEFL skoru ile akademik ortalama(CGPA) değişkenleri arasında da pozitif yönlü yüksek bir ilişki vardır.

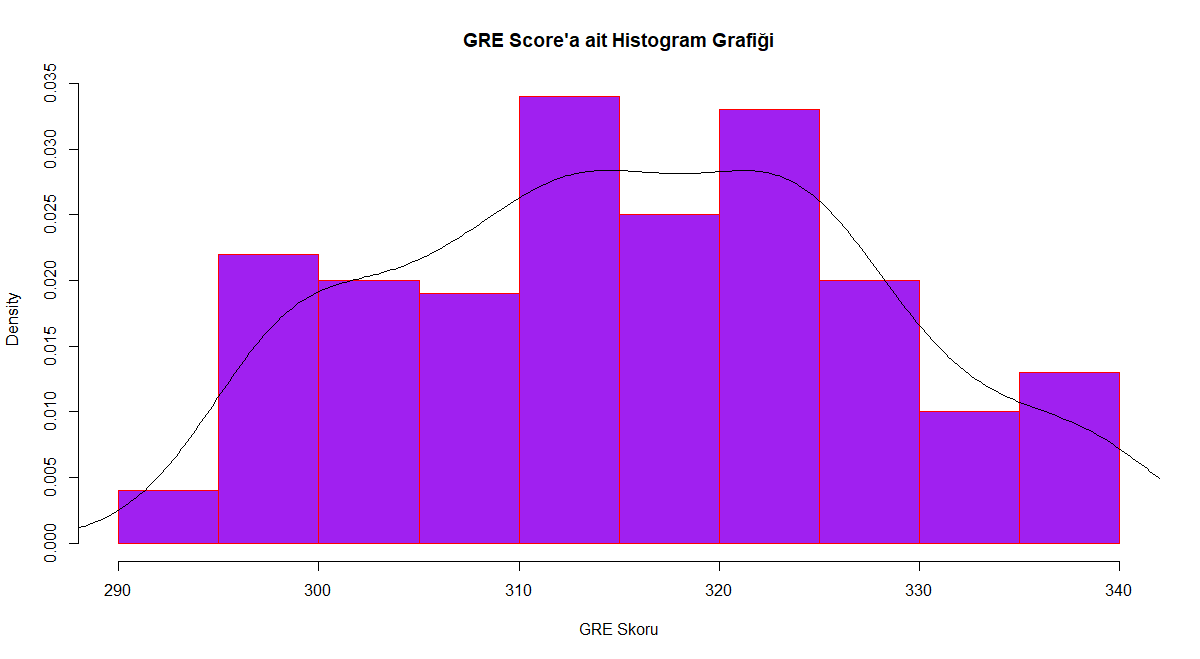
# ANALİZ

## DEĞİŞKENLER VE GRAFİK YORUMLARI

### GRE Skoru Değişkeni

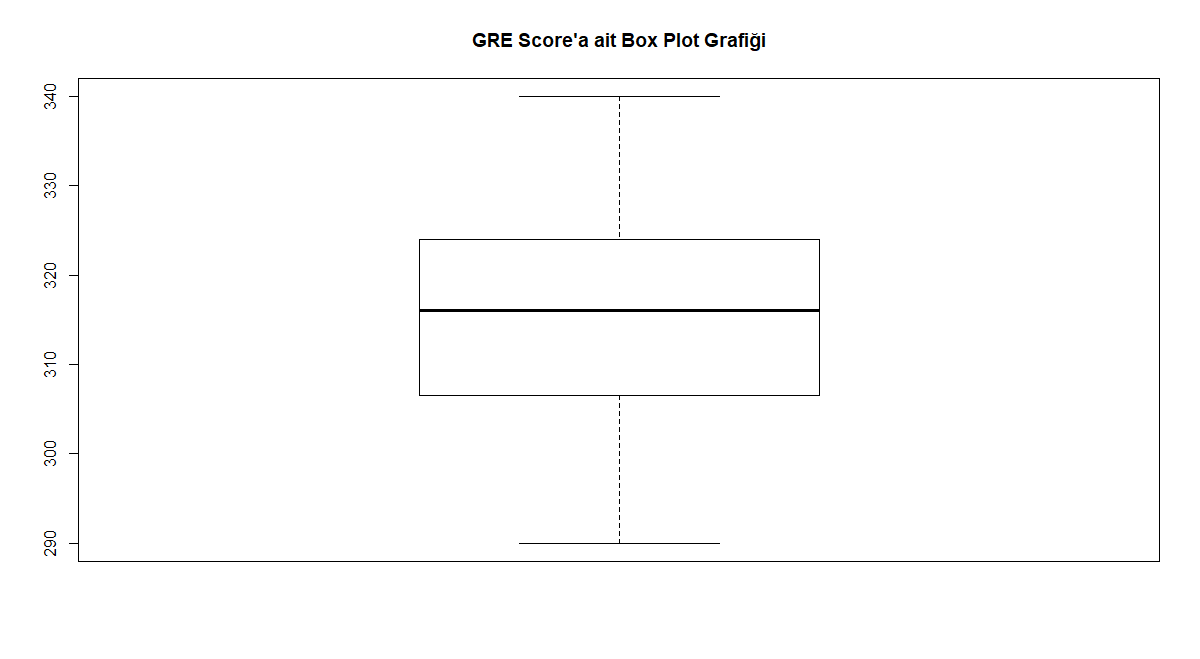


Hist() kodu ile GRE\_Score değişkenine ait histogram grafiği çizdirilmiştir. Main kodu ile grafiğe başlık verilmiştir ve x ekseni GRE Skoru olarak adlandırılmıştır. Col ve border ile grafiğe ait renklendirmeler yapılmıştır. Prop kodu ile histogram grafiğine yoğunluk tahmini eklenmiştir. Lines komutu ile yoğunluk tahmini için çizgi çekilmesi sağlanmıştır.



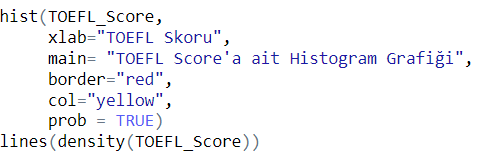
GRE skor değeri 310 ve 325 arasında daha fazla yoğunluk gösterirken 290 ve 295 skorları aralığında ise en az yoğunluk göstermektedir.

* Box plot grafiği dağılımın çarpıklığına bakmak ve aykırı değerlerin tespit edilmesi için kullanılır.

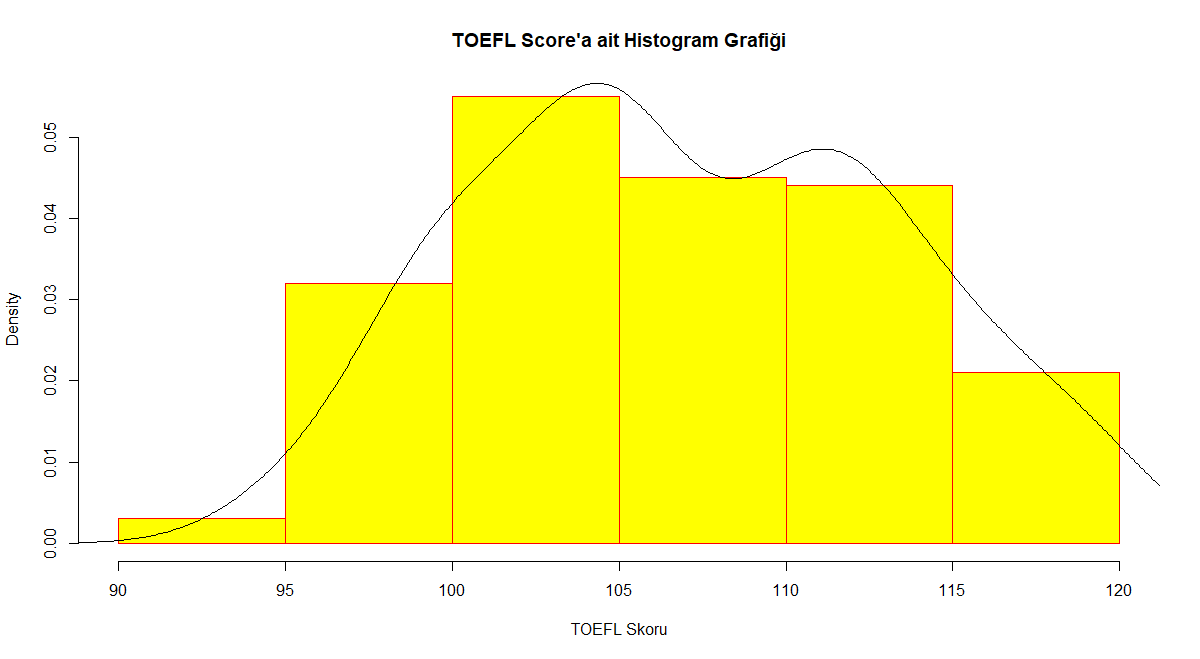
 

GRE\_Skore değişkeni için dağılım merkezinin 315 olduğu görülmektedir. Box plot grafiğine göre skor değişkenine ait dağılımın simetrik olduğu söylenebilir. Aykırı değer gözlenmemektedir.

### TOEFL Skoru Değişkeni

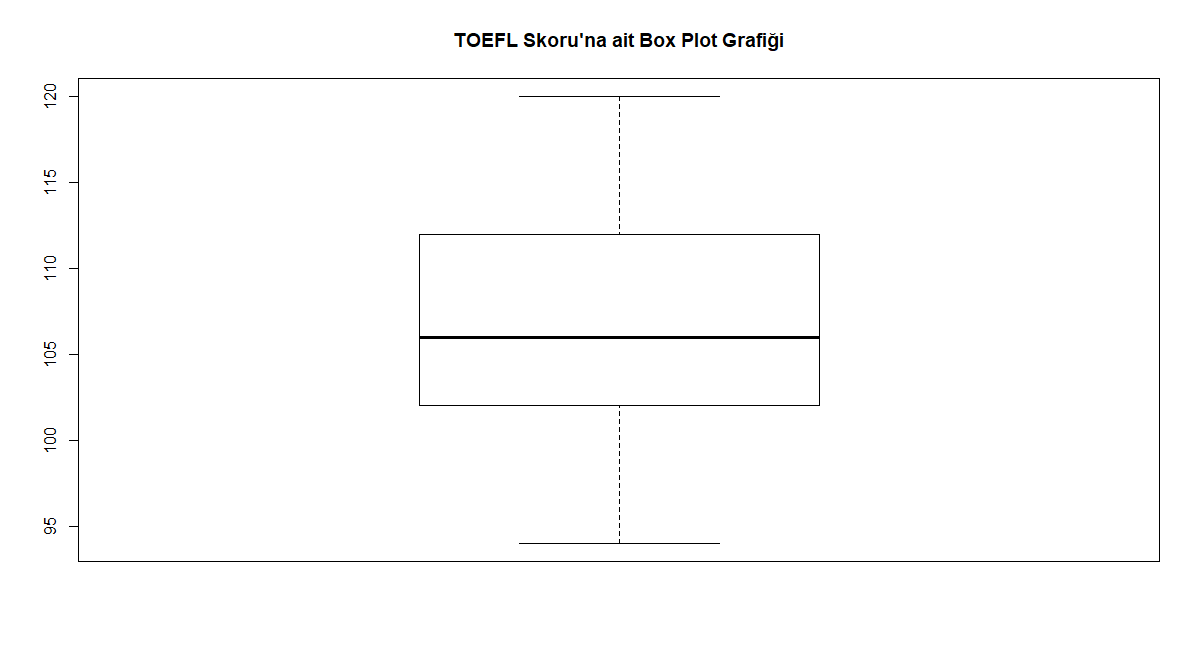


Hist() kodu ile TOEFL\_Score değişkenine ait histogram grafiği çizdirilmiştir. Main kodu ile grafiğe başlık verilmiştir ve x ekseni TOEFL Skoru olarak adlandırılmıştır. Prop kodu ile histogram grafiğine yoğunluk tahmini eklenmiştir. Lines komutu ile yoğunluk tahmini için çizgi çekilmesi sağlanmıştır.



TOEFL skoru 100 ve 115 arasında daha fazla yoğunluk gösterirken 90 ve 95 skorları aralığında ise en az yoğunluk göstermektedir.

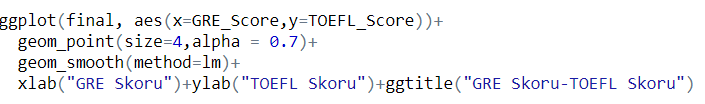




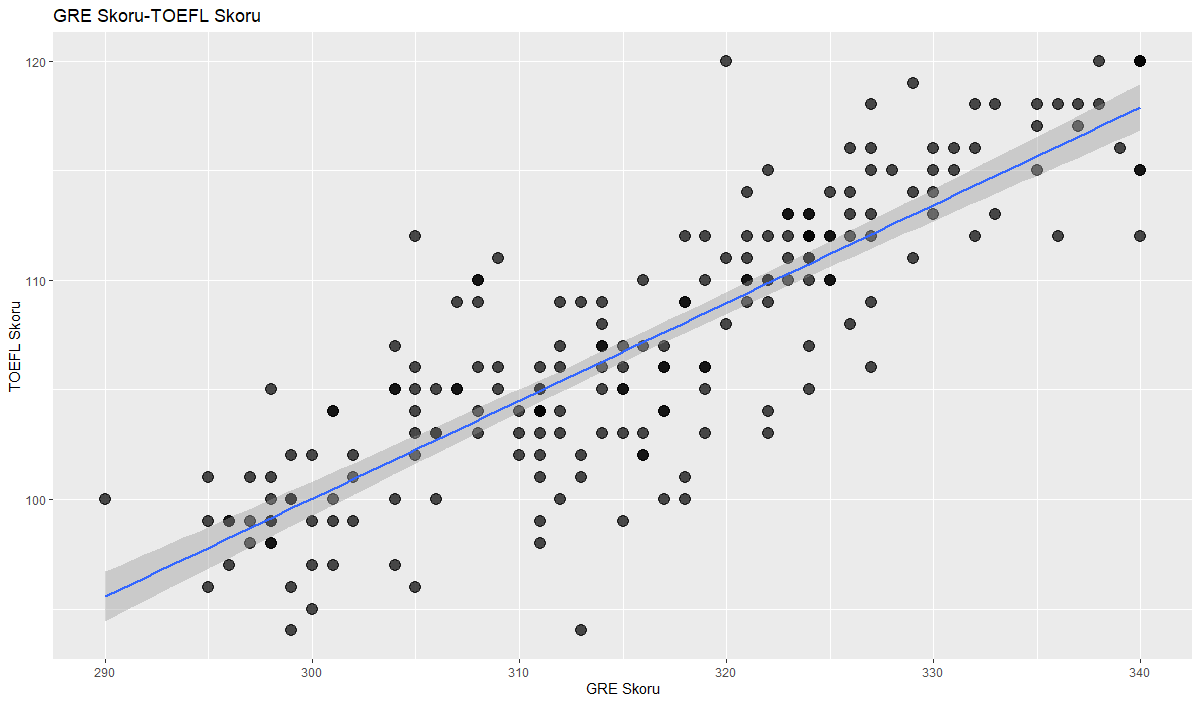
TOEFL\_Skore değişkeni için dağılım merkezinin 105 olduğu görülmektedir. Box plot grafiğine göre skor değişkenine ait dağılımın simetrik olduğu söylenebilir. Aykırı değer gözlenmemektedir.

### Scatter Plot

* Scatter plot grafiği verilen 2 tane sürekli değişkeninin birbirlerine göre durumlarına inceler.

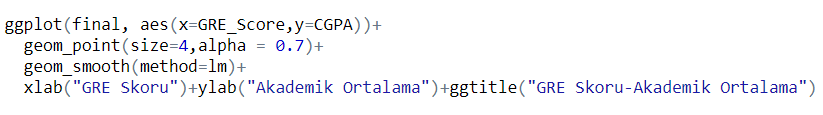


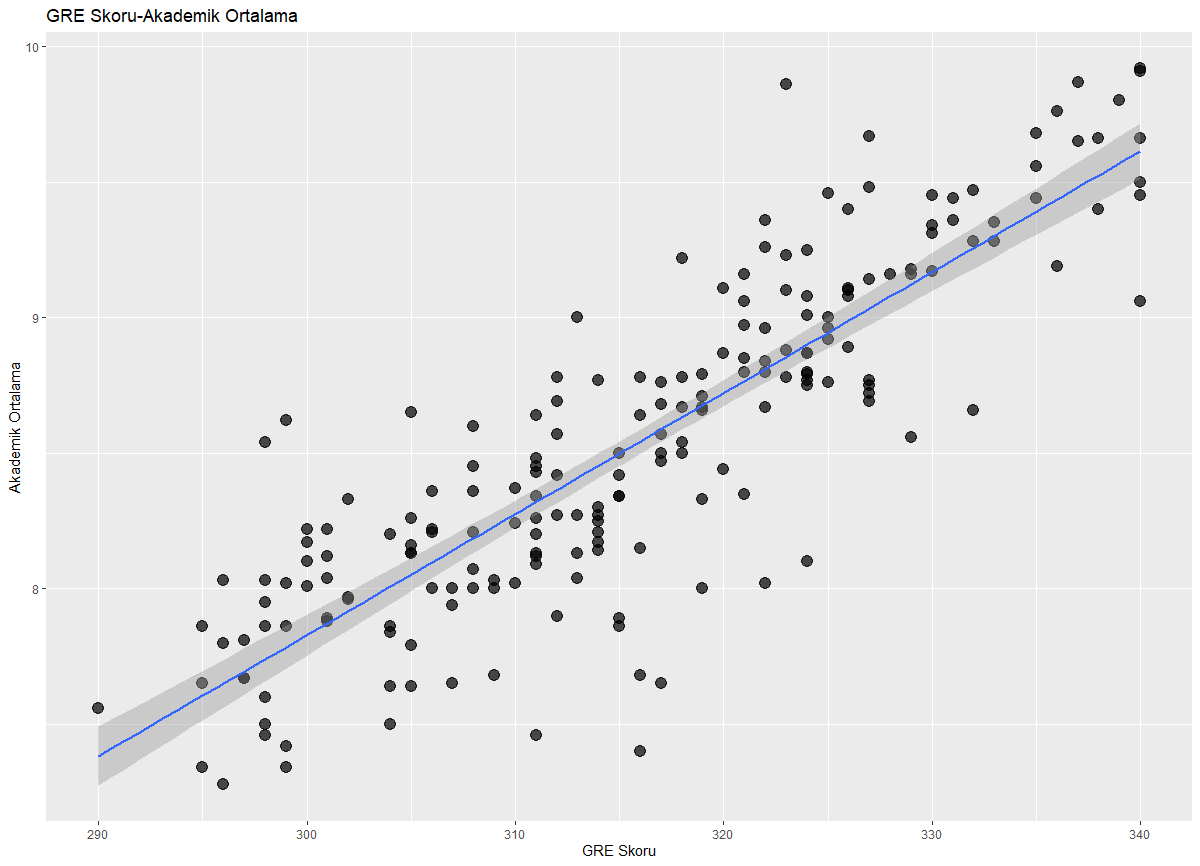
GRE Skor ve TOEFL Skor değişkenlerine ait scatter plot grafiği çizilmiştir. Grafiği daha anlaşılır hale getirmek için ”method=lm” kodu ile regresyon eğrisi çizdirilmiştir.



Grafiğe bakıldığında genel olarak GRE Skoru arttıkça TOEFL Skorunun da arttığı görülmektedir. GRE skoru ile TOEFL skoru arasında pozitif doğrusal bir ilişkinin olduğu söylenebilir.

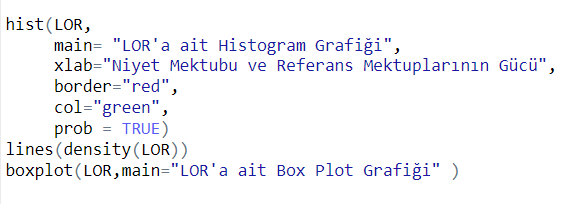
* GRE skoru ve Akademik ortalamaya ilişkin scatter plot grafiği çizdirilmiştir.

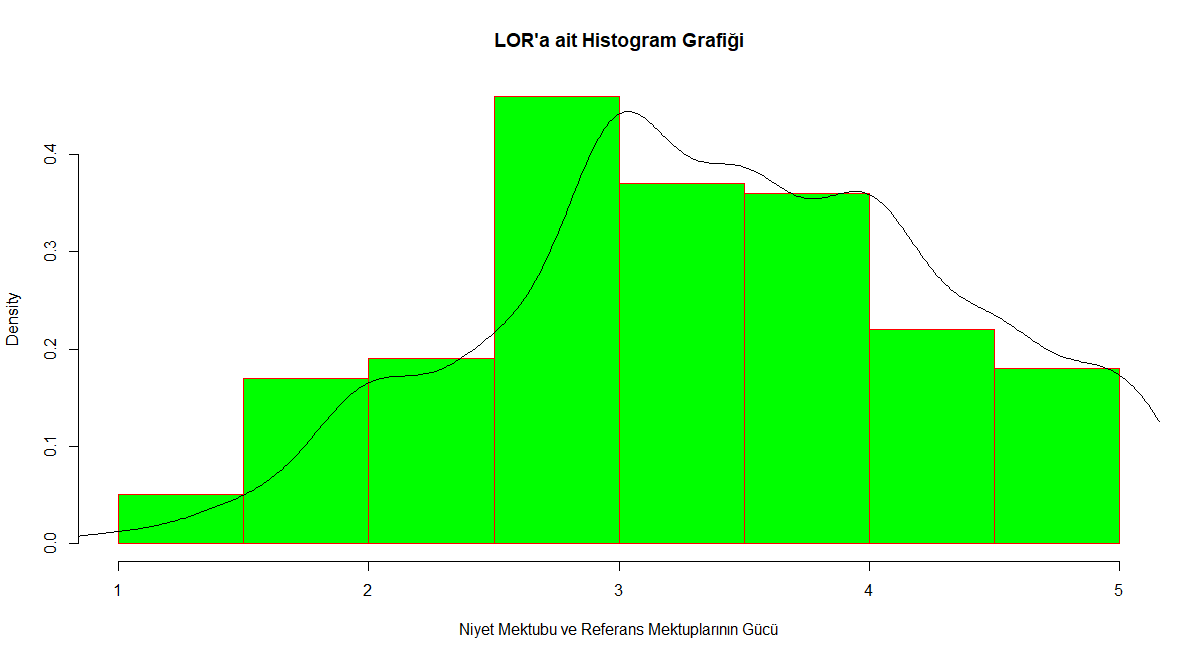




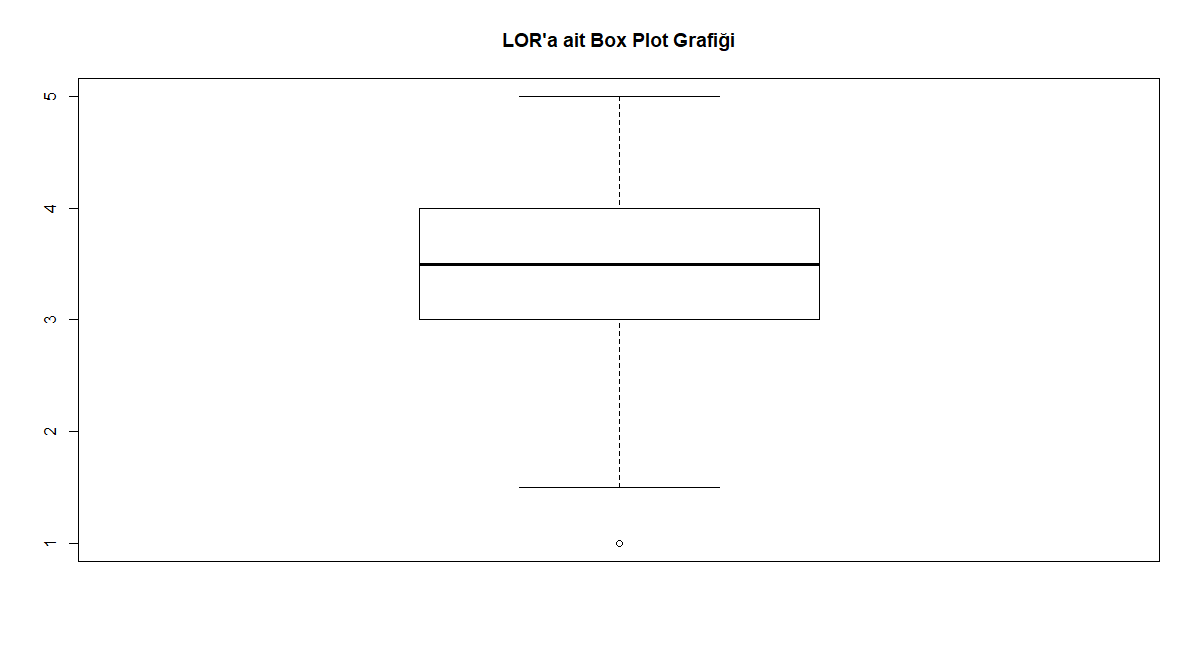
GRE skoru yüksek olan kişilerin akademik ortalamalarının da yüksek olduğu görülmektedir.

#### LOR (Niyet ve Referens Mektuparının Gücü) Değişkeni



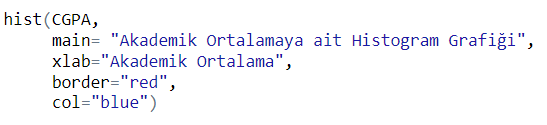


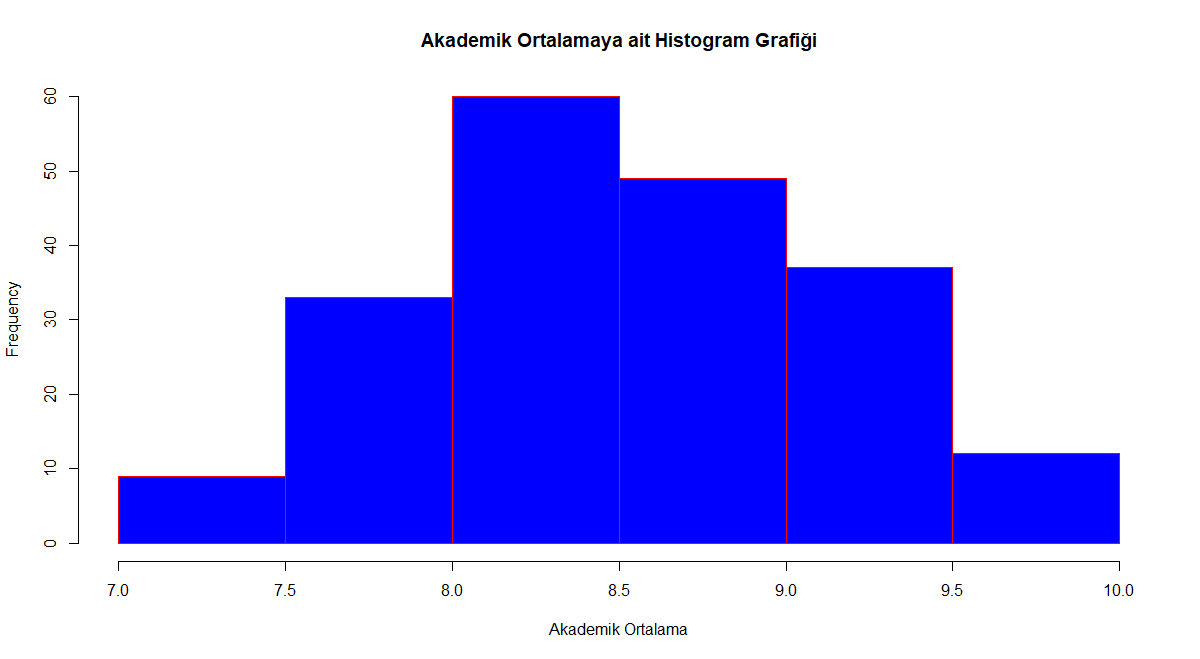
Niyet ve referans mektupları değişkeninde 2.5 ile 4 arasında alınan puanların yoğunluğunun daha fazla olduğu görülmektedir. 1 ile 1.5 arasındaki puanlar ise oldukça azdır. Histogram grafiğine bakıldığında LOR değişkenine ait dağılımın sola çarpık olduğu görülmektedir.



LOR değişkeni için dağılım merkezinin 3.5 olduğu görülmektedir. Box plot grafiğine göre LOR değişkenine ait dağılımın simetrik olduğu söylenebilir. Bir tane aykırı değer gözlenmektedir.

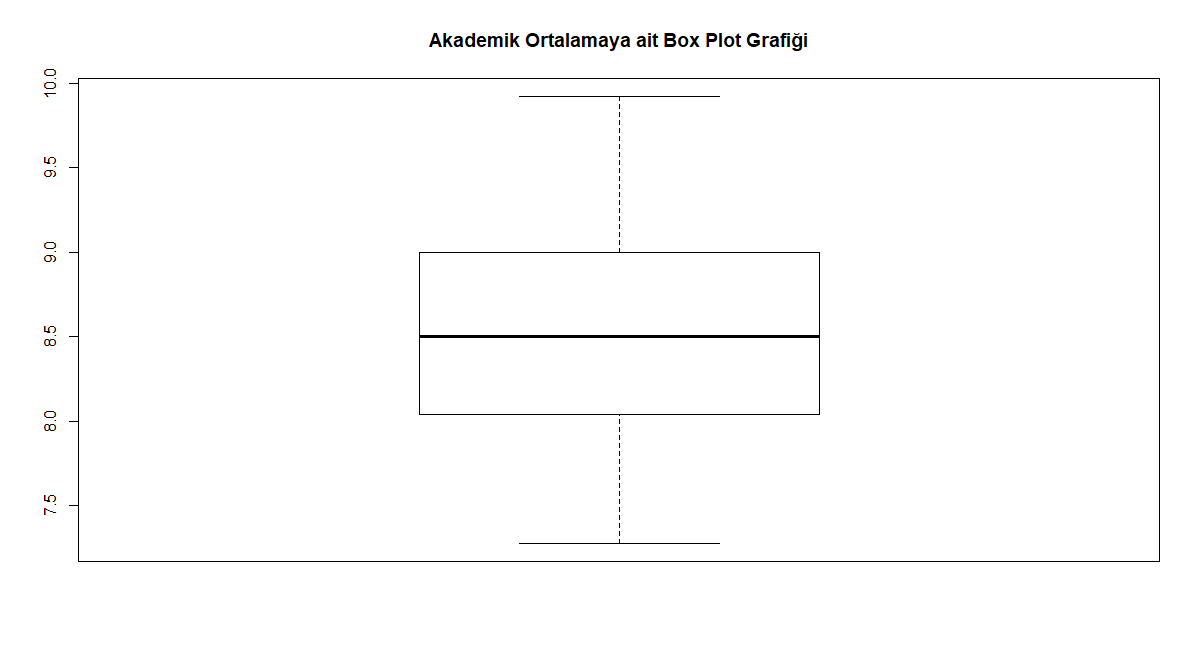
### CGPA (Akademik Ortalama) Değişkeni





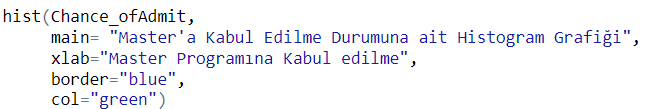
Histogram grafiğine bakıldığında akademik ortalama için sıklığın en fazla 8 ile 8.5 arasında olduğu görülmektedir .Grafiğe bakıldığında simetrik bir dağılım gösterdiğini söyleyebiliriz.

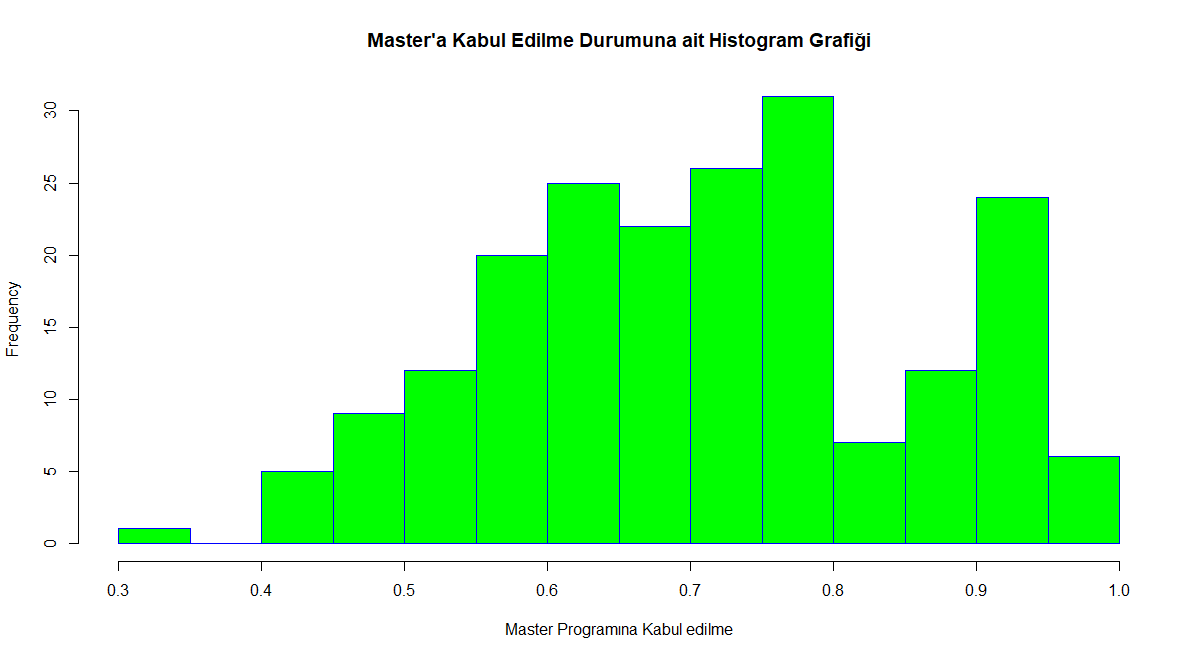




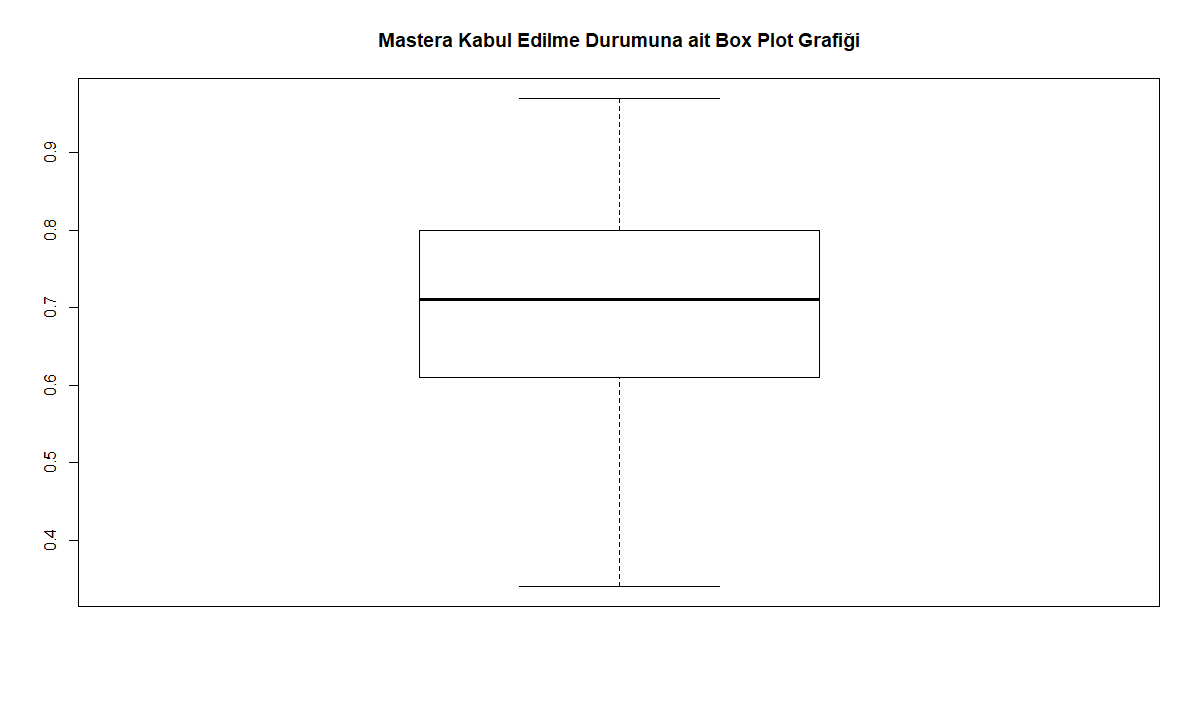
CGPA değişkeni için dağılım merkezinin 8.5 olduğu görülmektedir. Box plot grafiğine göre akademik ortalama değişkenine ait dağılımın simetrik olduğu söylenebilir. Aykırı değer gözlenmemektedir.

### Change\_ofAdmit Değişkeni





Histogram grafiğine bakıldığında mastera kabul edilme olasılığı için sıklığın en fazla 0.6 ile 0.8 arasında olduğu görülmektedir .



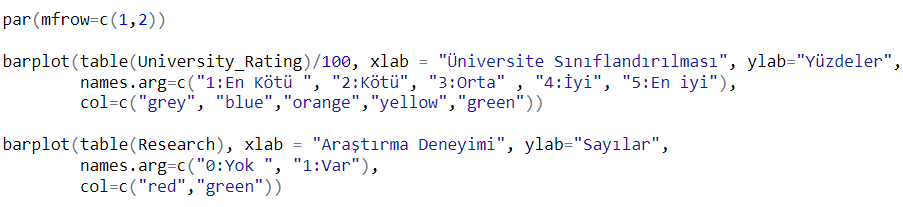
Change\_ofAdmit değişkeni için dağılım merkezinin 0.7 olduğu görülmektedir. Aykırı değer gözlenmemiştir.

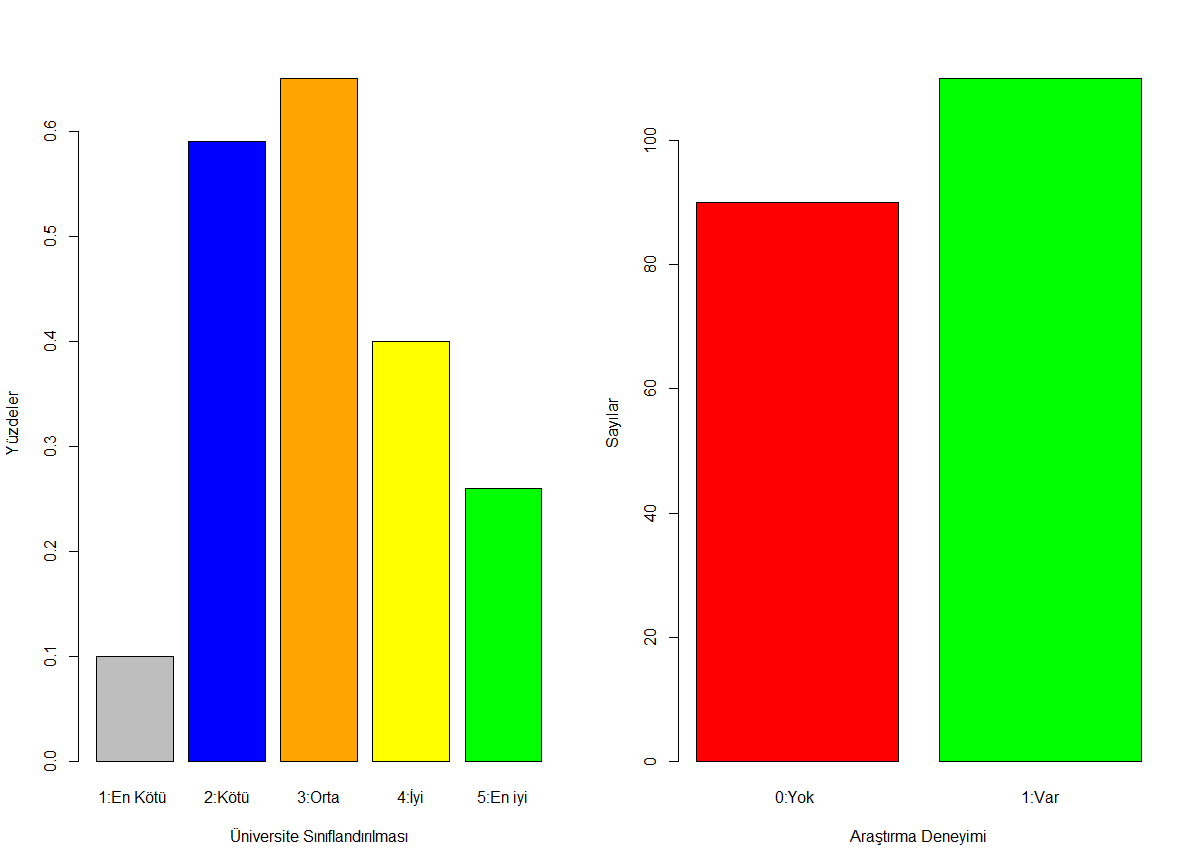
### Kategorik değişkenler için Grafikler

### University\_Rating(Üniversite Sınıflandırması) ve Reserch(Araştırma Deneyimi) Değişkenleri

#### Sütun Grafikleri

Kategorik değişkenler için sütun grafiği çizdirilmiştir.

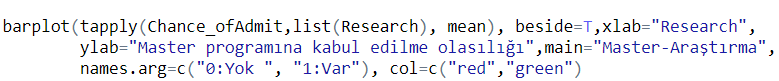


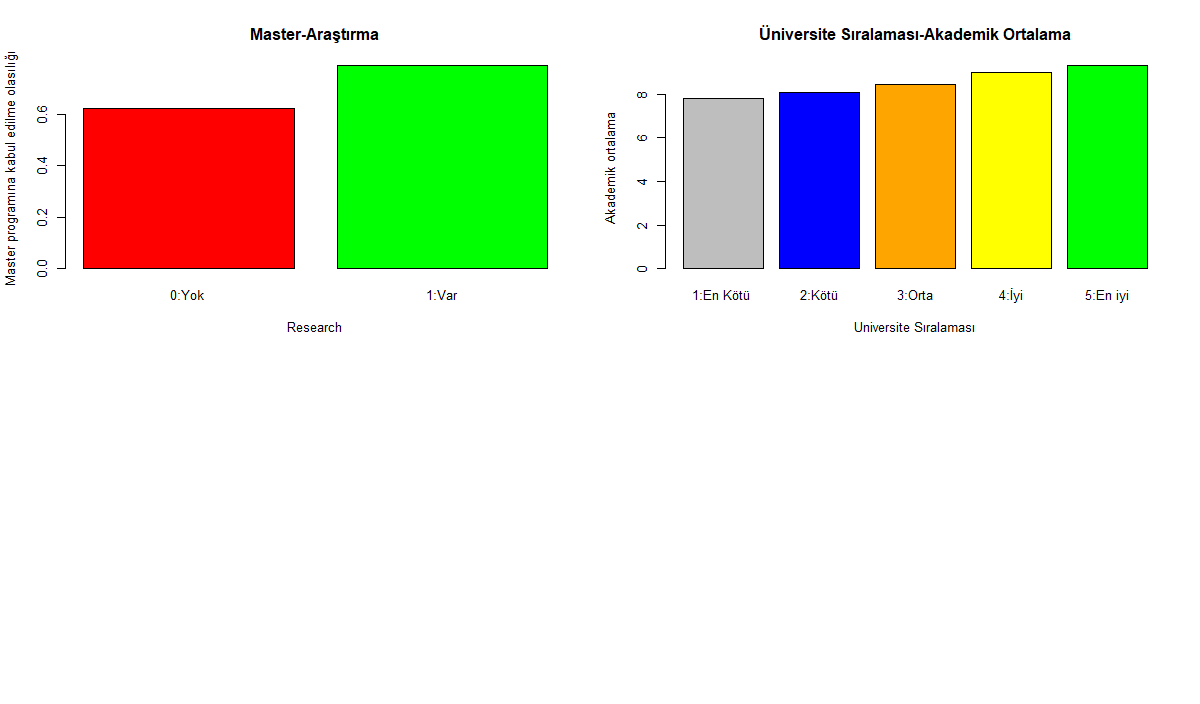


Üniversite sınıflandırması değişkeni için grafiğe bakıldığında en fazla orta seviye üniversitelerde yoğunlaşma olduğu görülmektedir orta seviye üniversiteler , üniversite sınıflandırılmasının %60 ı kapsamaktadır. En kötü üniversite sayısı azdır ve %10 luk bir kısımda yer almaktadır.

Araştırma deneyimi olanların sayısının daha fazla olduğu görülmektedir.

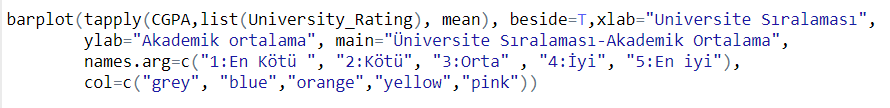
* Mastera kabul edilme olasılığı (sürekli) değişkeni ile araştırma deneyimi değişkeni(kategorik) sütun grafiği çizdirilmiştir.

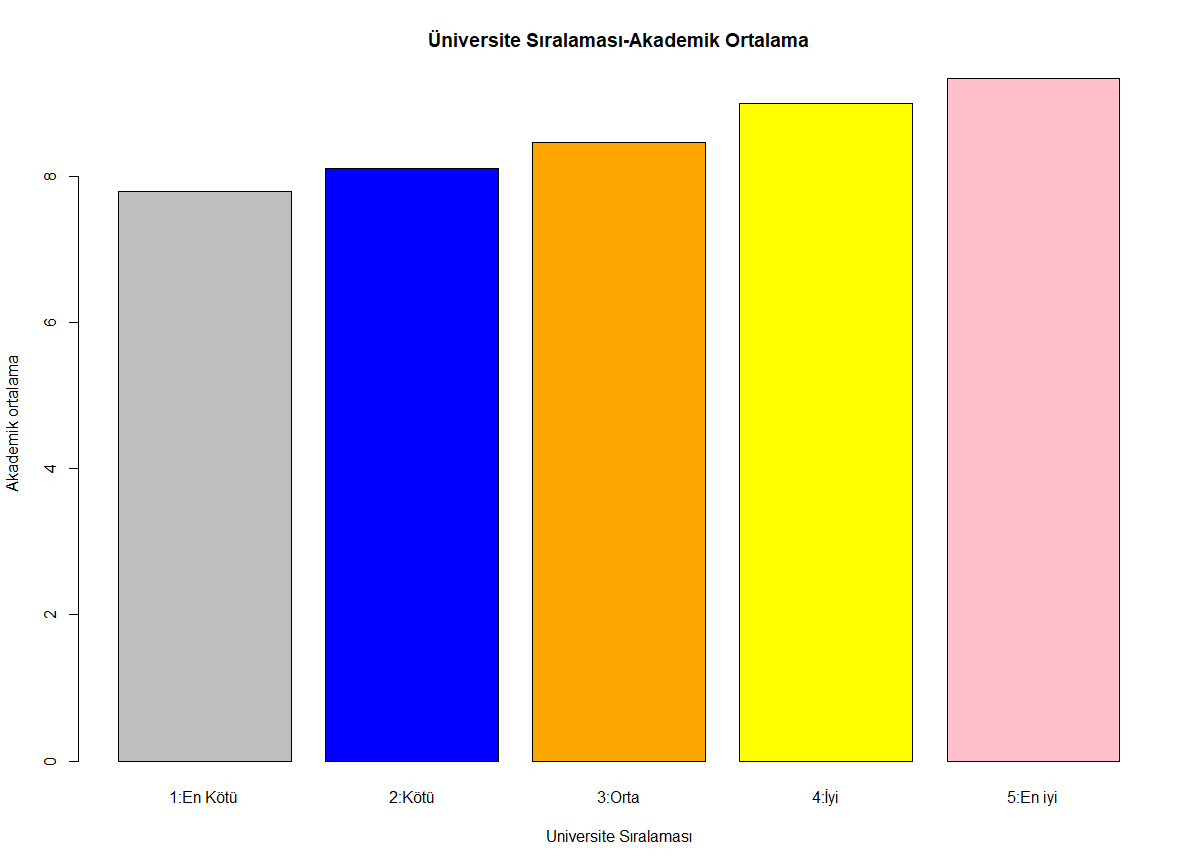




Araştırma deneyimi olup mastera kabul edilenlerin sayısının daha fazla olduğu görülmektedir.

* Akademik ortalama (sürekli) değişkeni ile üniversite sınıflandırması değişkeni(kategorik) sütun grafiği çizdirilmiştir.





Akademik ortalama arttıkça üniversite kalitesinin az da olsa arttığı görülmektedir.

#### Pasta Grafiği

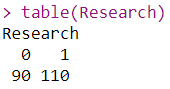
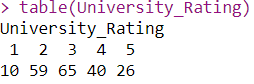
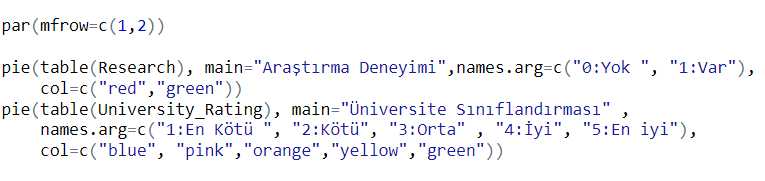
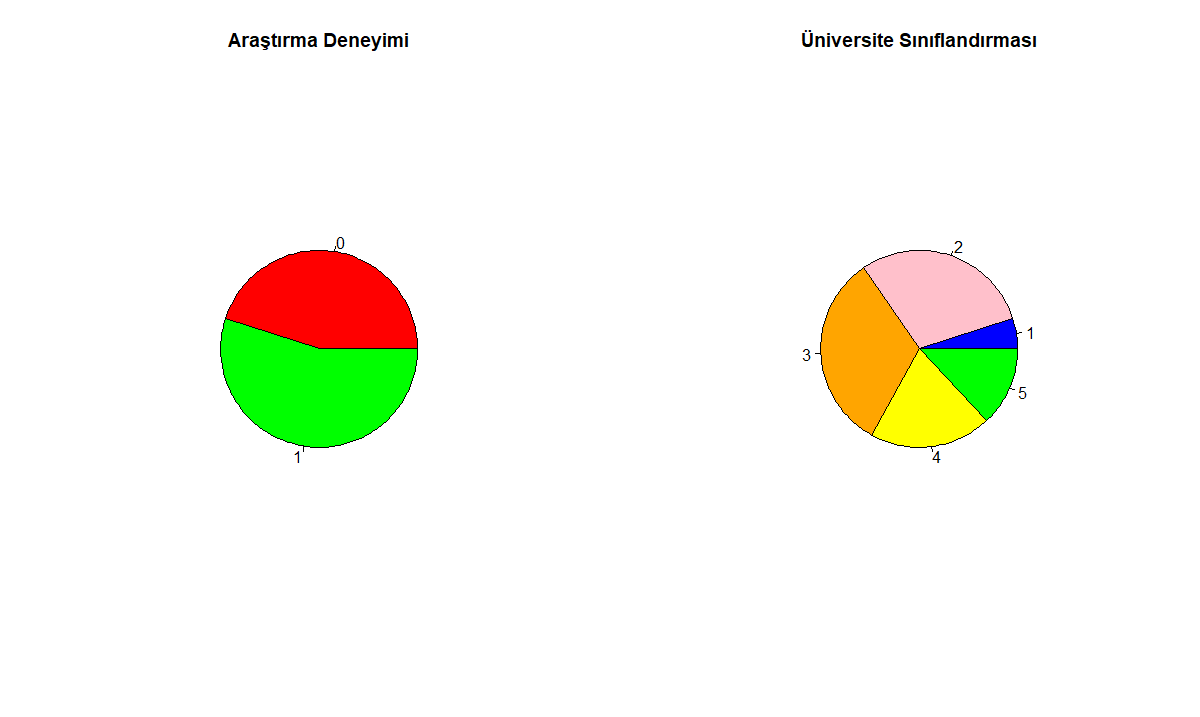
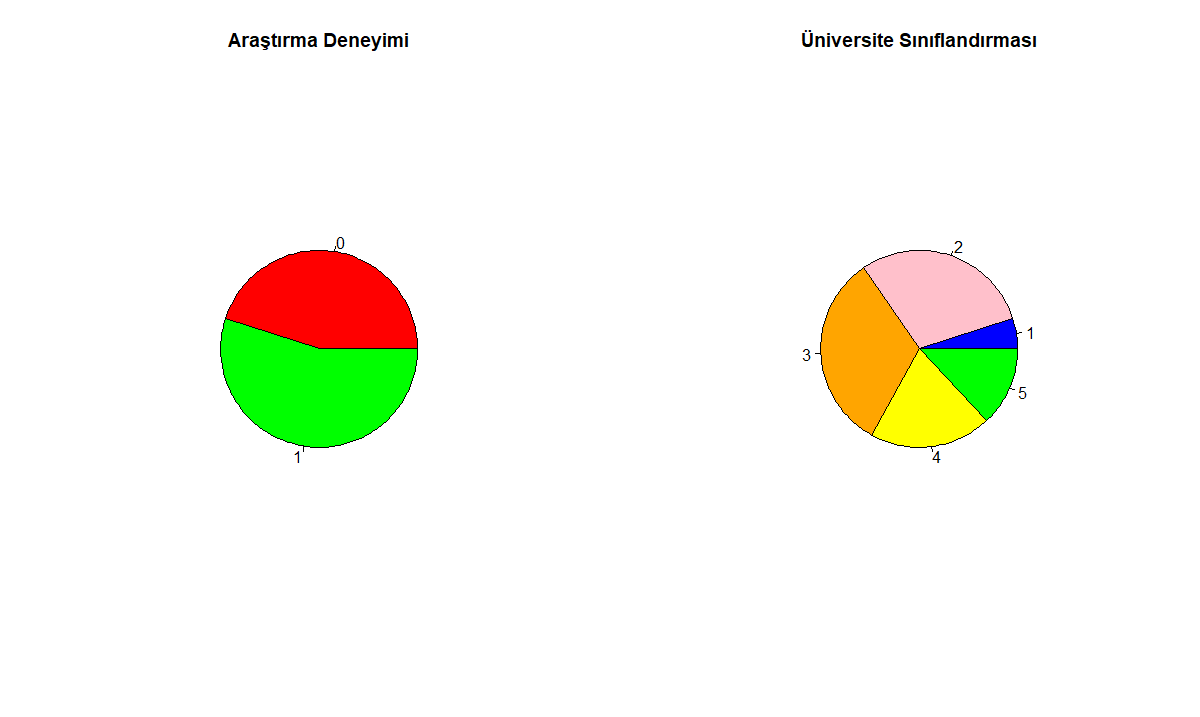
 

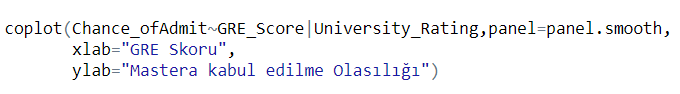
Table fonksiyonu ile kategorik değişkenlere ait kaçar değer olduğu gösterilmiştir.

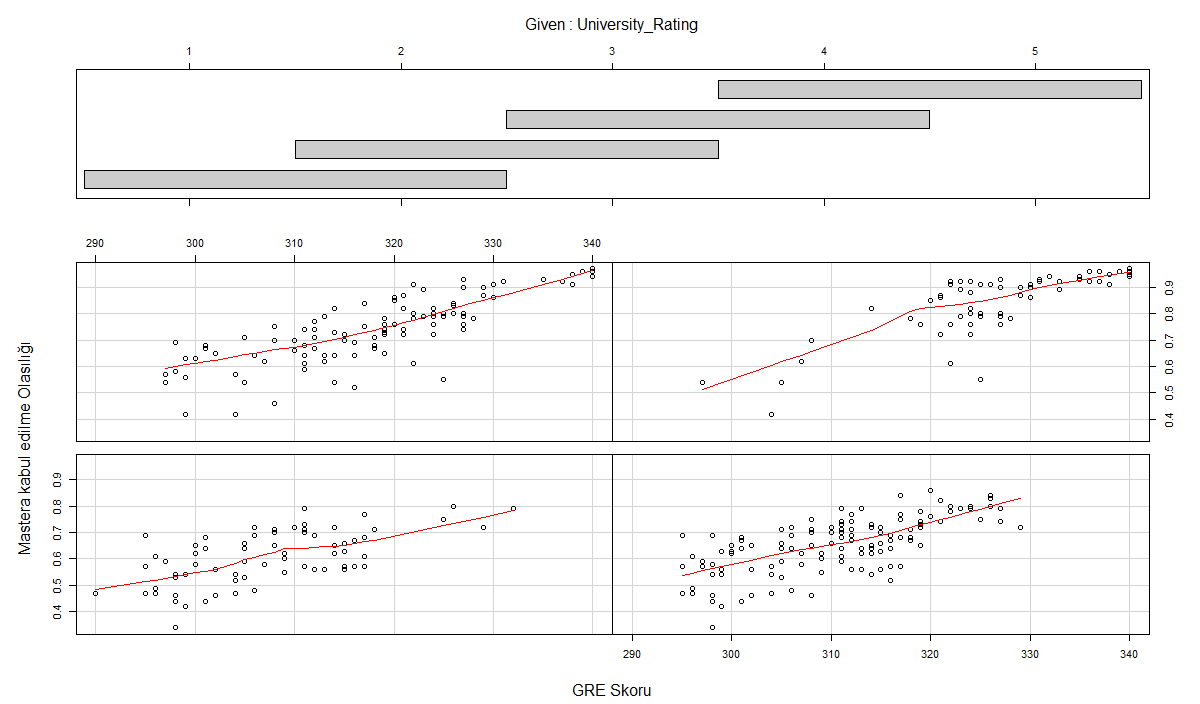




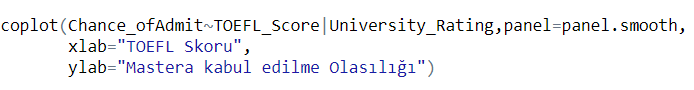
#### Coplot Grafikleri

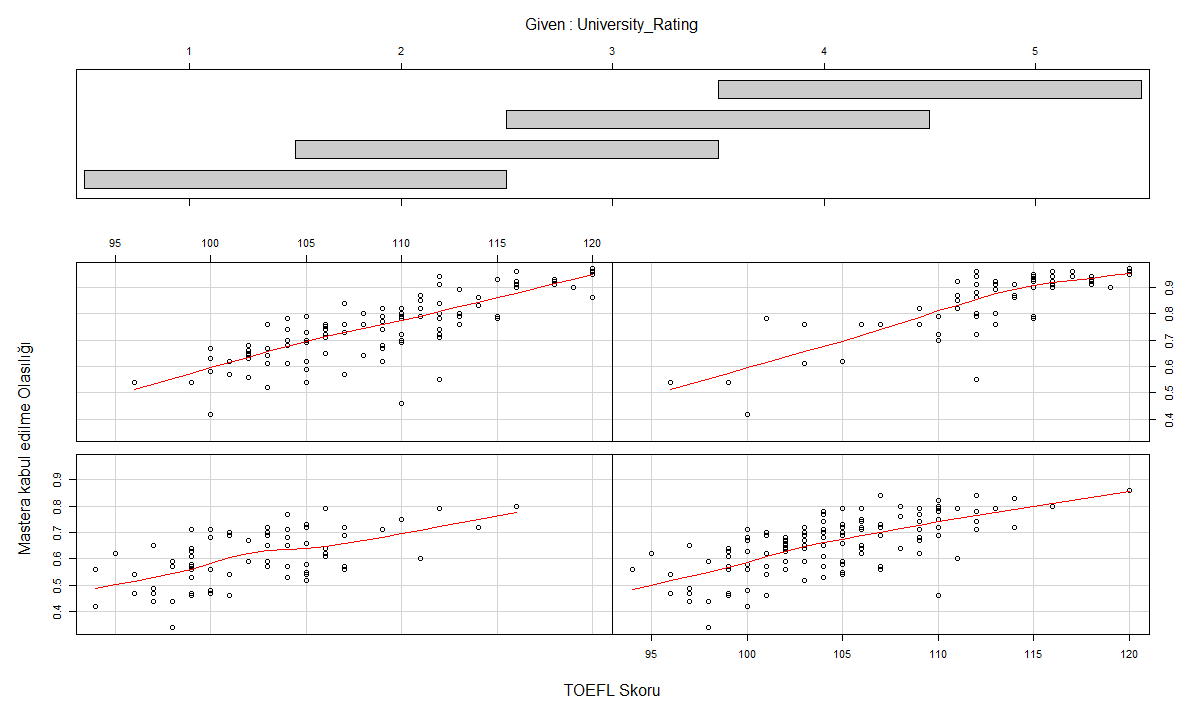
Değişkenlere göre üniversite sınıflandırmasına ait coplot grafikleri ve kodları aşağıdaki gibi verilmiştir.



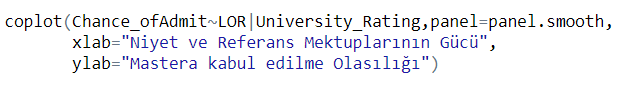


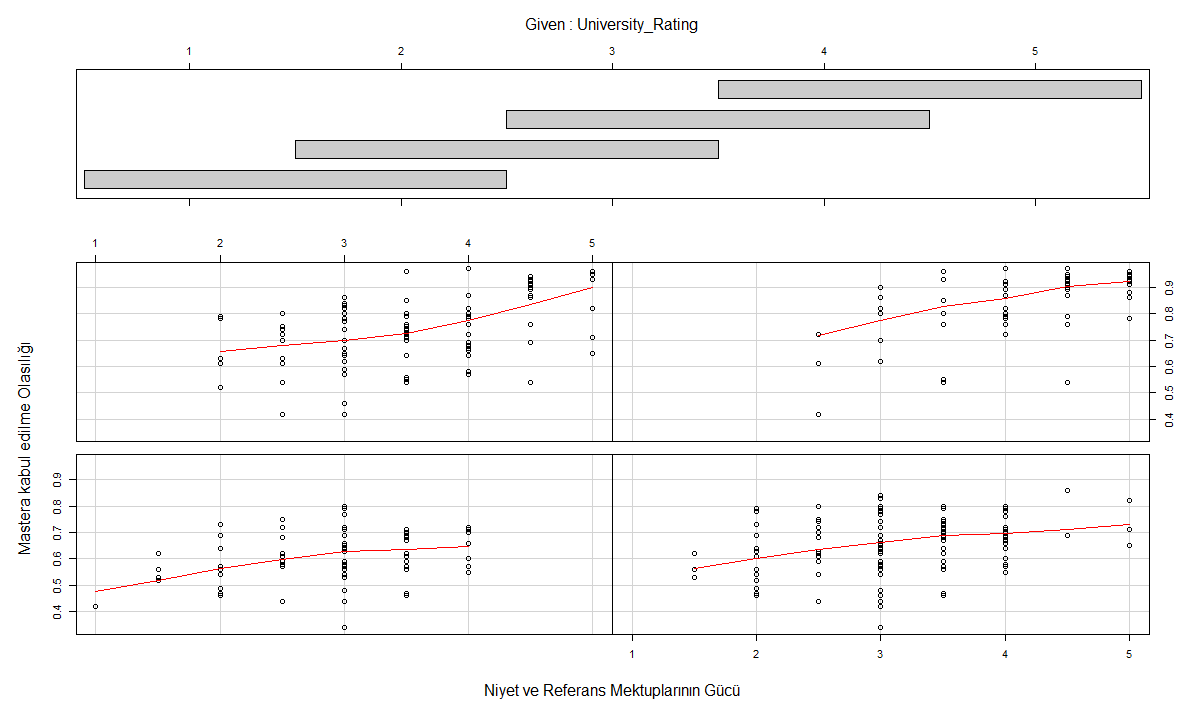
Üniversite sınıflandırmalarına göre GRE Skorunun master programına katılma olasılığı üzerine dağılışı grafikte görüldüğü gibidir. 4 ve 5. Üniversiteki kişilerin GRE skorlarının daha yüksek ve üniversiteye kabul olasılıklarının daha fazla olduğu görülmektedir.



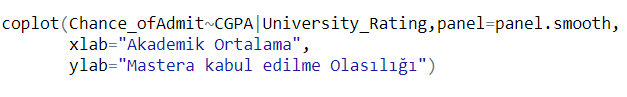


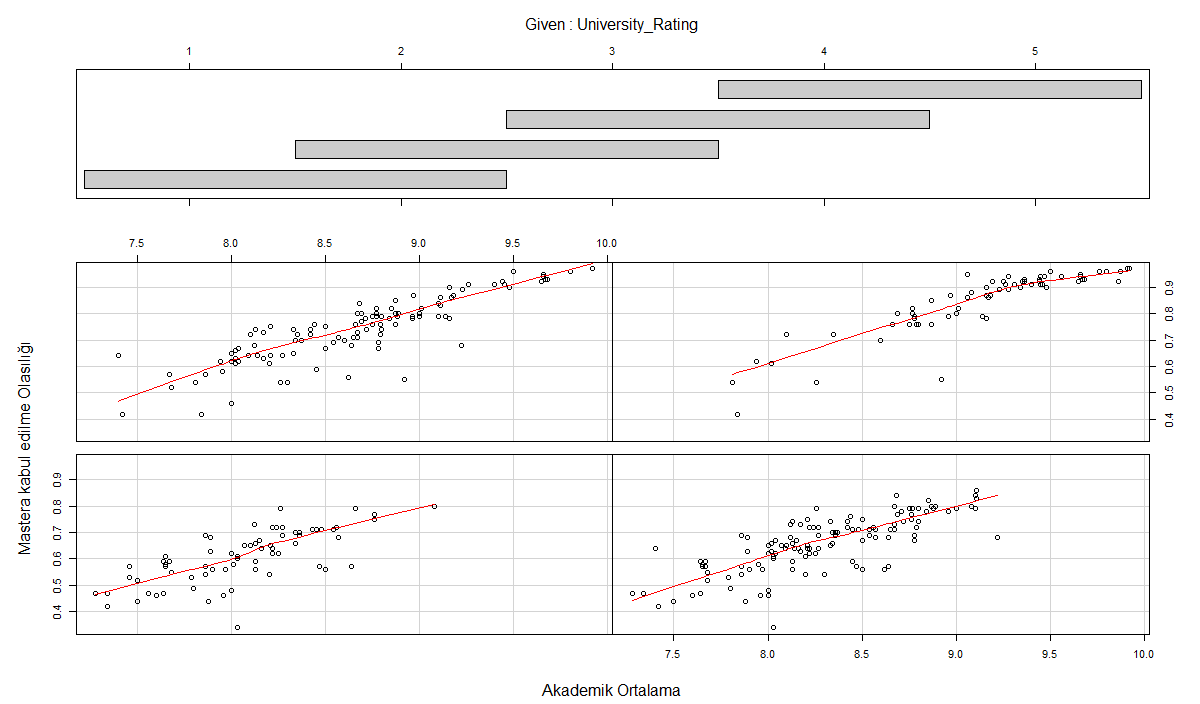
Üniversite sınıflandırmalarına göre TOEFL Skorunun master programına katılma olasılığı üzerine dağılışı grafikte görüldüğü gibidir.





Üniversite sınıflandırmalarına göre niyet ve referans mektupları gücünün master programına katılma olasılığı üzerine dağılışı grafikte görüldüğü gibidir. En iyi üniversitelerde (4,5) Niyet ve referans mektuplarının gücü ile mastera kabul edilme olasılığının daha yüksek olduğu görülmektedir.

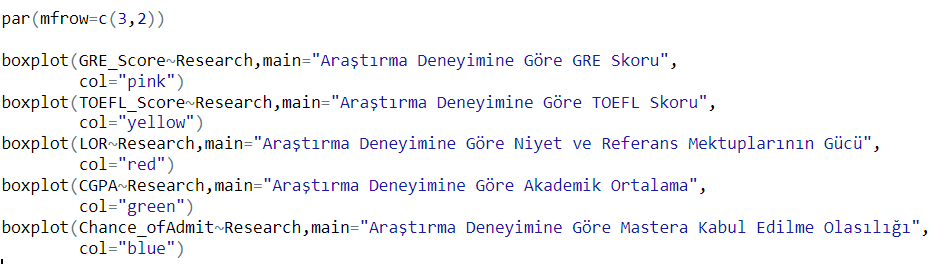


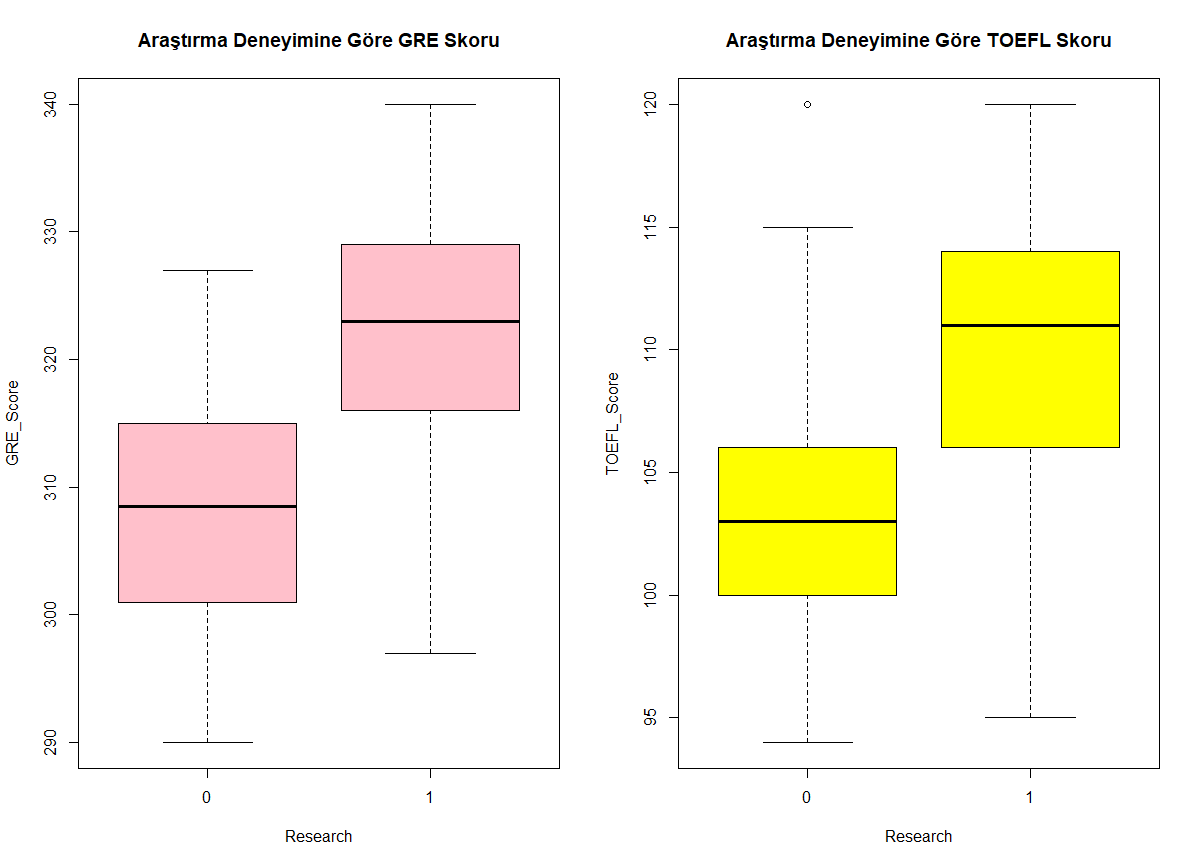


Üniversite sınıflandırmalarına göre akademik ortalamanın master programına katılma olasılığı üzerine dağılışı grafikte görüldüğü gibidir. Akademik ortalama arttıkça master programına katılma olasılığı artar. 4 ve 5. Üniversitelerde mastera kabul durumunun daha yüksek olduğu görülmektedir.

#### Boxplot Grafiği

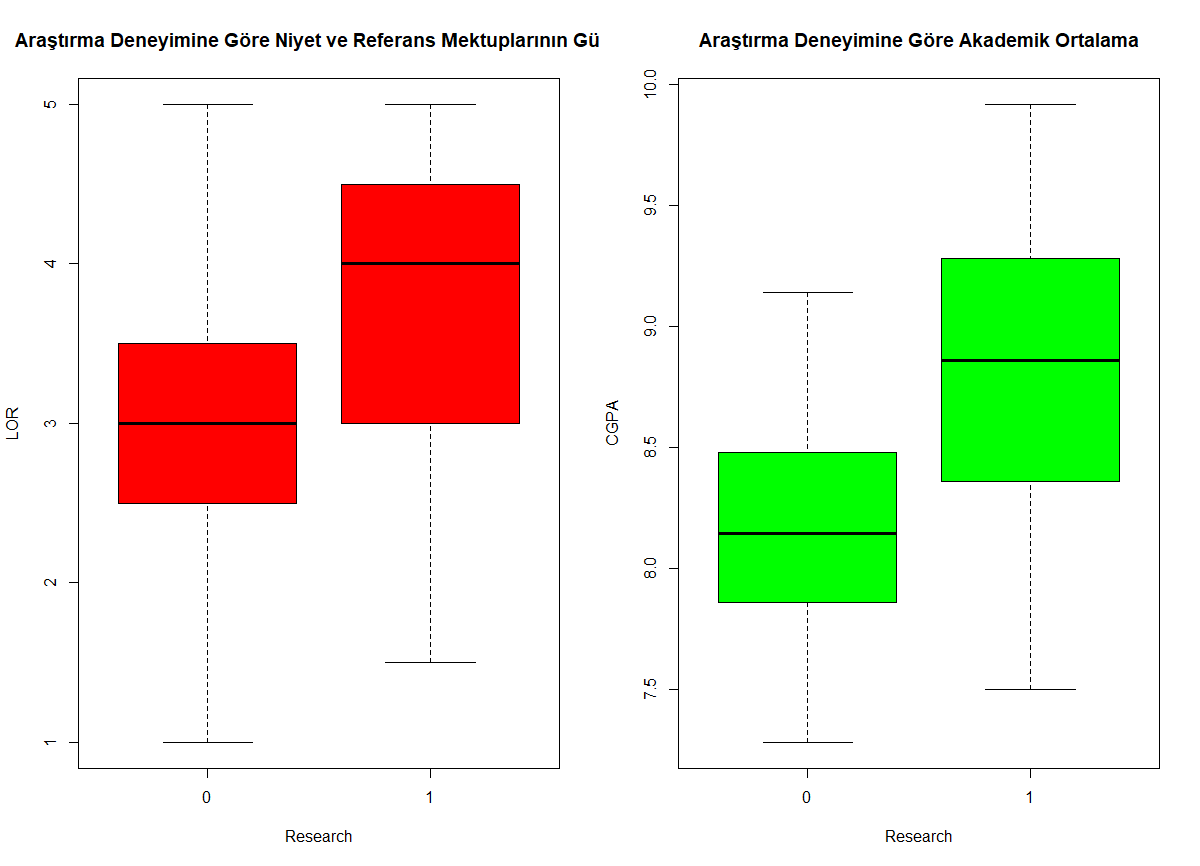
Araştırma Deneyimi değişkeni ile diğer değişkenler arasında box plot grafiği çizdirilmiştir.



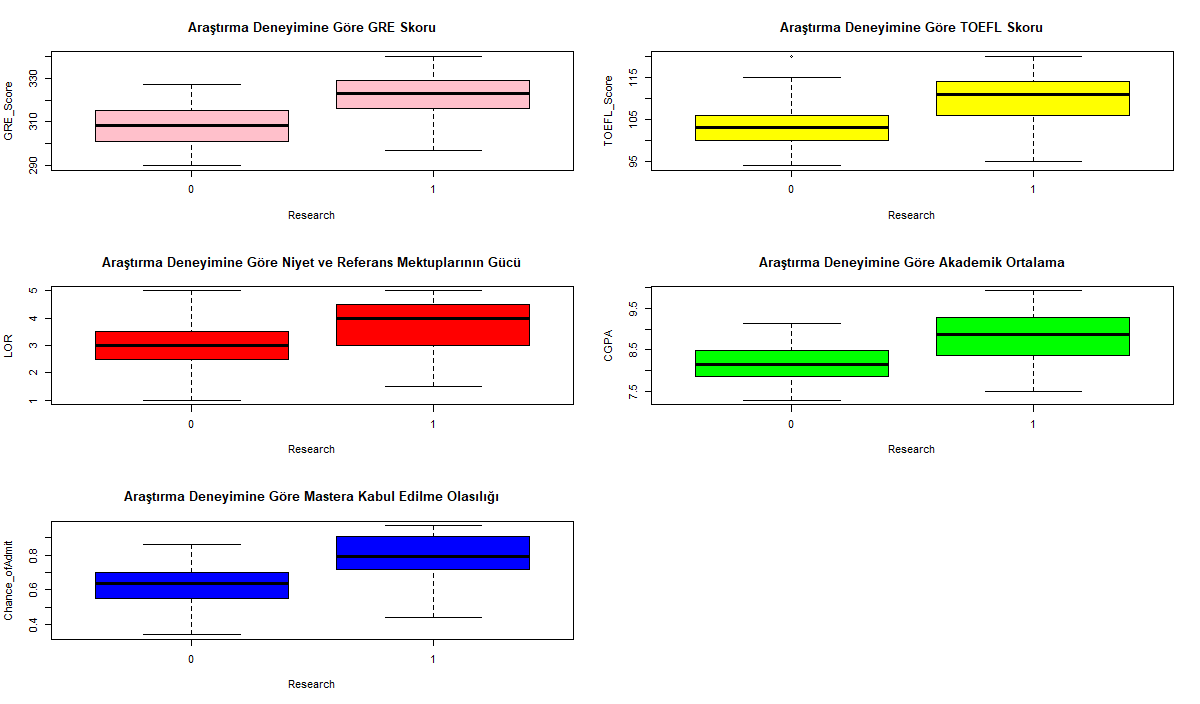


GRE Skor değişkeni için Araştırma deneyimi olmadığı durumda dağılımın ortalaması 310 civarındadır. Araştırma deneyimi olduğu durumda ise dağılımın ortalaması 320 civarındadır. Aykırı değer ise gözlenmemiştir.

TOEFL skor değişkeni için araştırma deneyiminin olmadığı durumda bir tane aykırı değer gözlenmiştir.

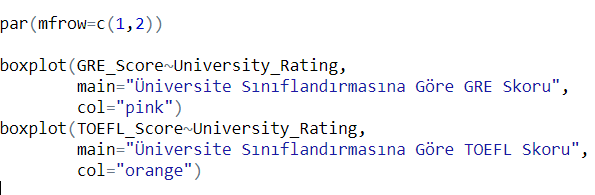


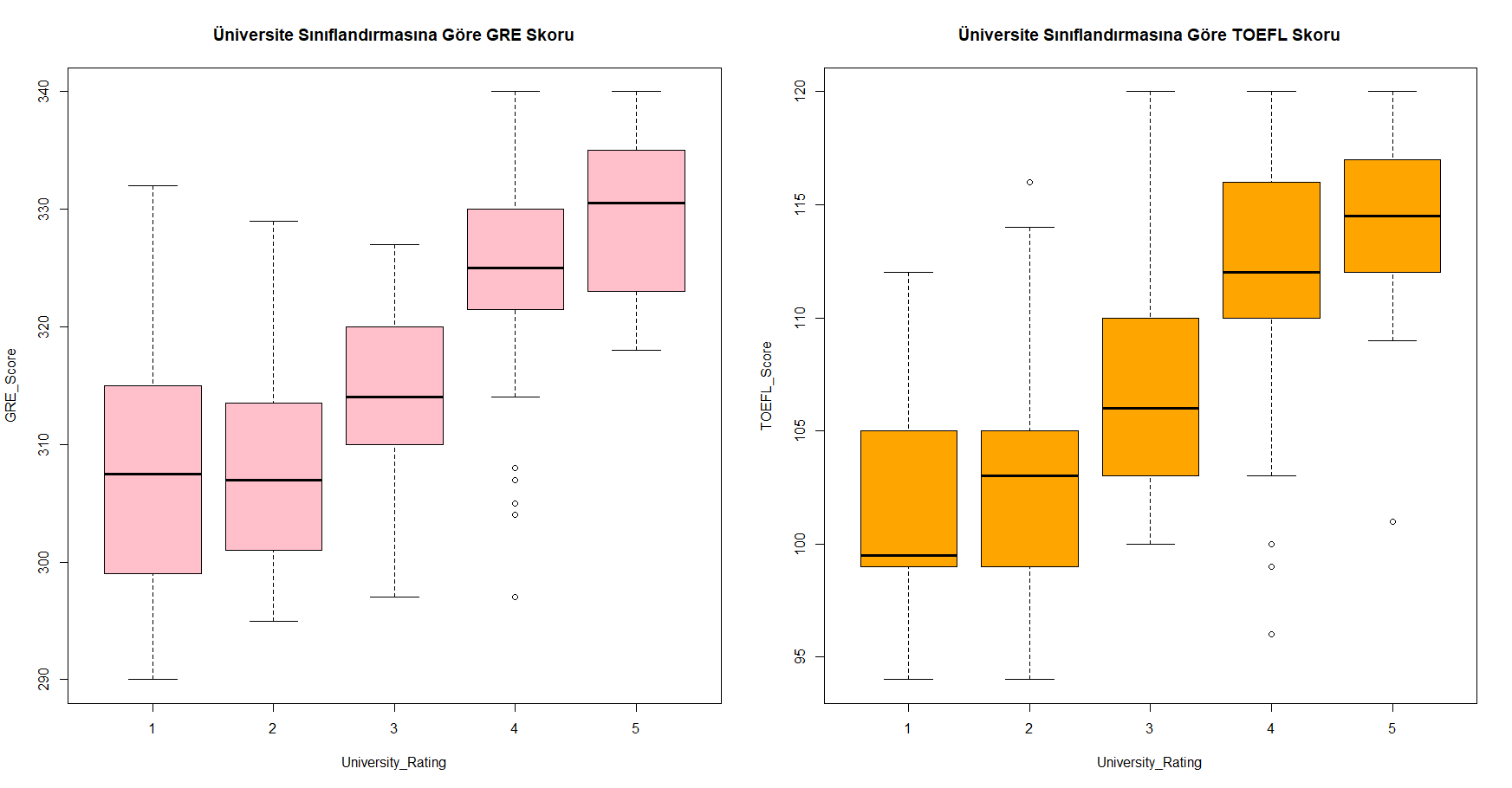
Niyet ve Referans Mektuplarının Gücü değişkeni için araştırma deneyimi olduğu durumda dağılımın merkezi 4’ten yüksek olduğu ve dağılımın sola çarpık olduğu görülmektedir.



Araştırma deneyimine sahip olanların, mastera kabul edilme olasılığının daha yüksek olduğu söylenebilir.

Üniversite Sınıflandırma değişkeni ile diğer değişkenler için box plot grafikleri çizdirilmiştir.



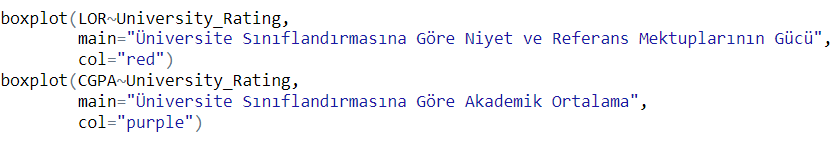


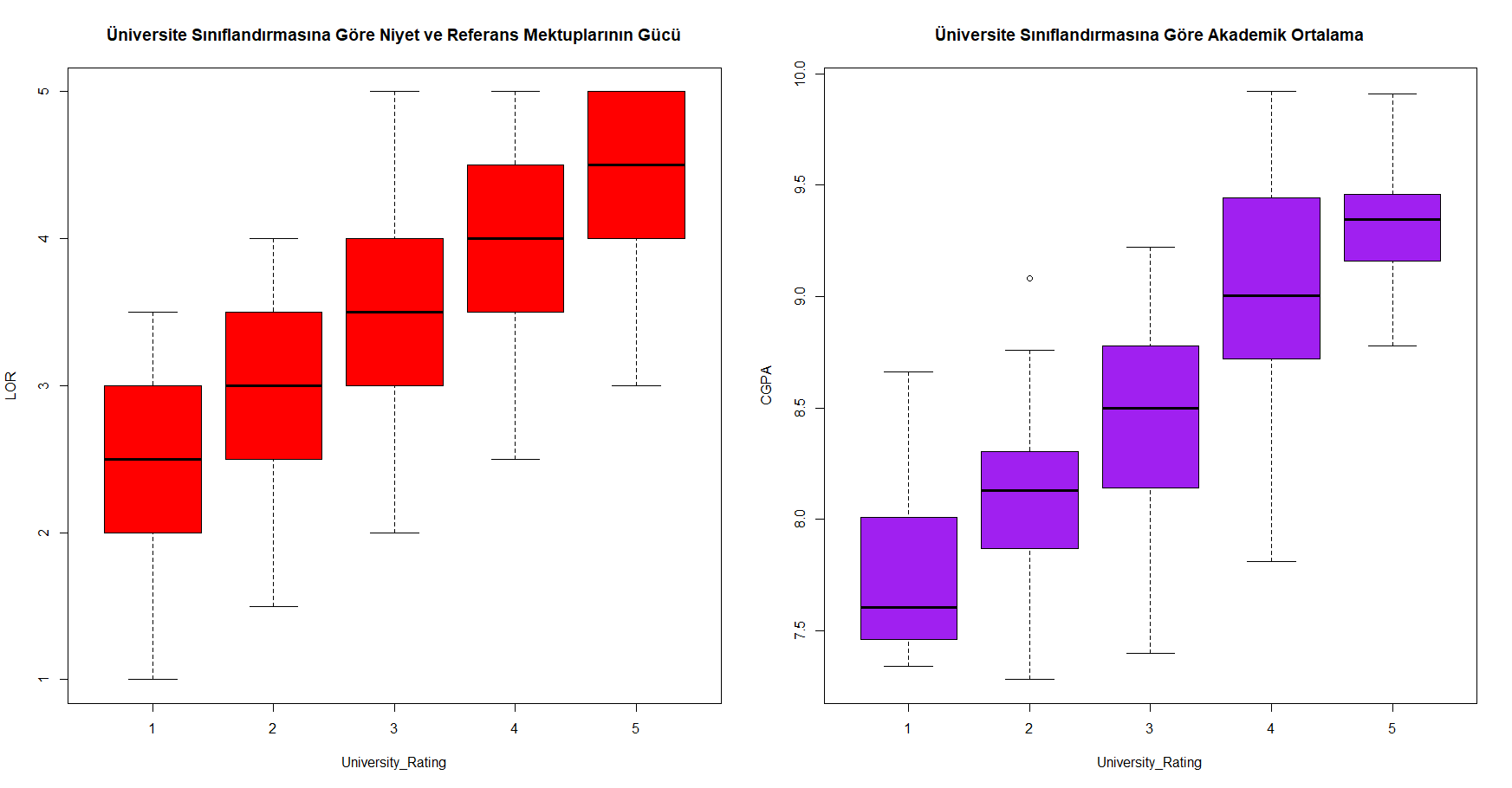
GRE Skoru değişkeni için Üniversite sınıflandırmasının (4) iyi olduğu durumda aykırı değerler gözlemlenmiştir. Dağılım sağa çarpıktır, aykırı değerler gözlenmektedir. 1. Ve 2. Üniversitelerde GRE skorları arasında fark olmadığı söylenebilir.

TOEFL Skoru değişkeni için Üniversite sınıflandırmasının (1) en kötü olduğu durumda dağılımın sağa çarpık olduğu görülmektedir.

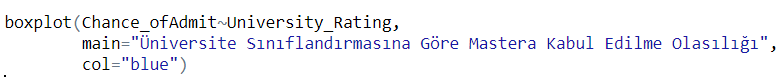
TOEFL Skoru değişkeni için Üniversite sınıflandırmasının (4) iyi olduğu durumda dağılımın sağa çarpık olduğu görülmektedir aynı zamanda aykırı değerler gözlemlenmiştir.

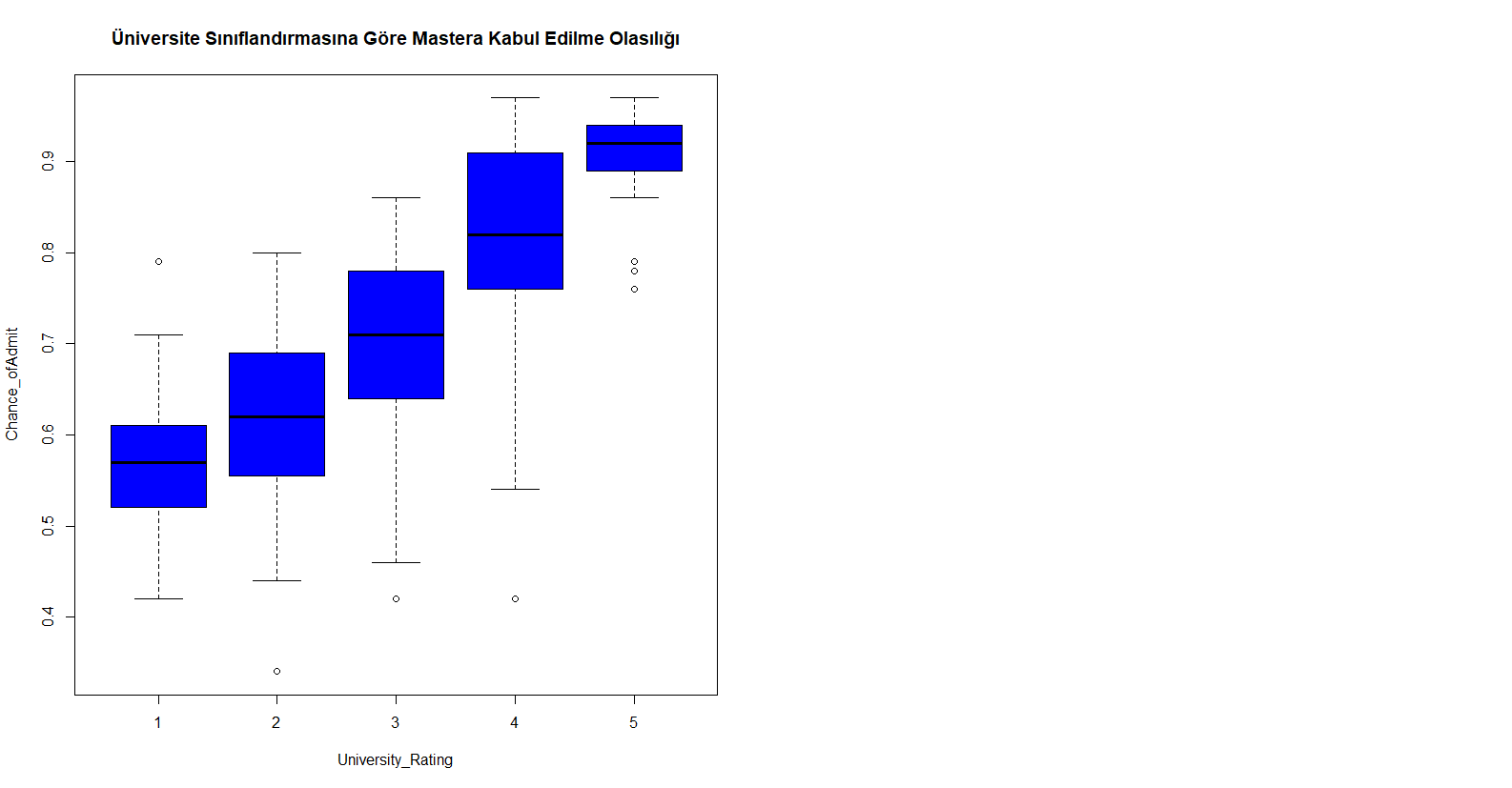
TOEFL Skoru değişkeni için Üniversite sınıflandırmasının (5) en iyi olduğu durumda da aykırı değerler gözlemlenmiştir yani örneğin TOEFL Skoru düşük olduğu halde en iyi üniversite sınıflandırmasında yer alma durumu karşımıza çıkmış olabilir.





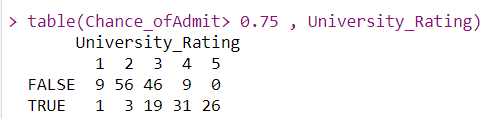
Akademik ortalama(CGPA) değişkeni için Üniversite sınıflandırmasının (1) en kötü olduğu durumda box plot grafiğinde dağılımın sağa çarpık olduğu görülmektedir.



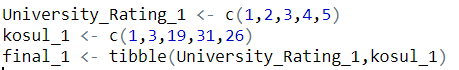


Master programına kabul edilme olasılığı değişkeni için Üniversite sınıflandırmasının 2,3,4 ve 5 olduğu durumlarda aykırı gözlemler bulunmaktadır.

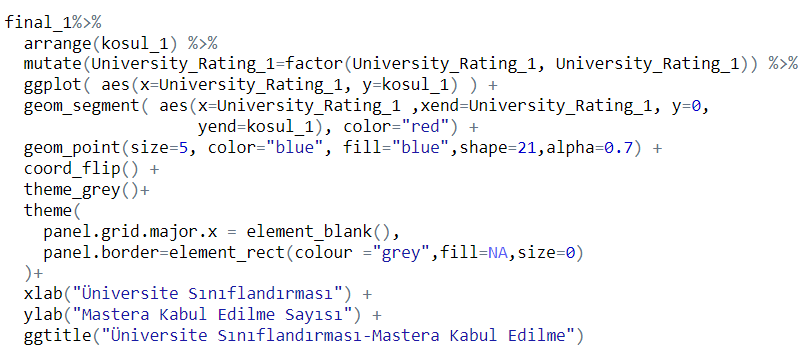
### Lollipop Grafiği



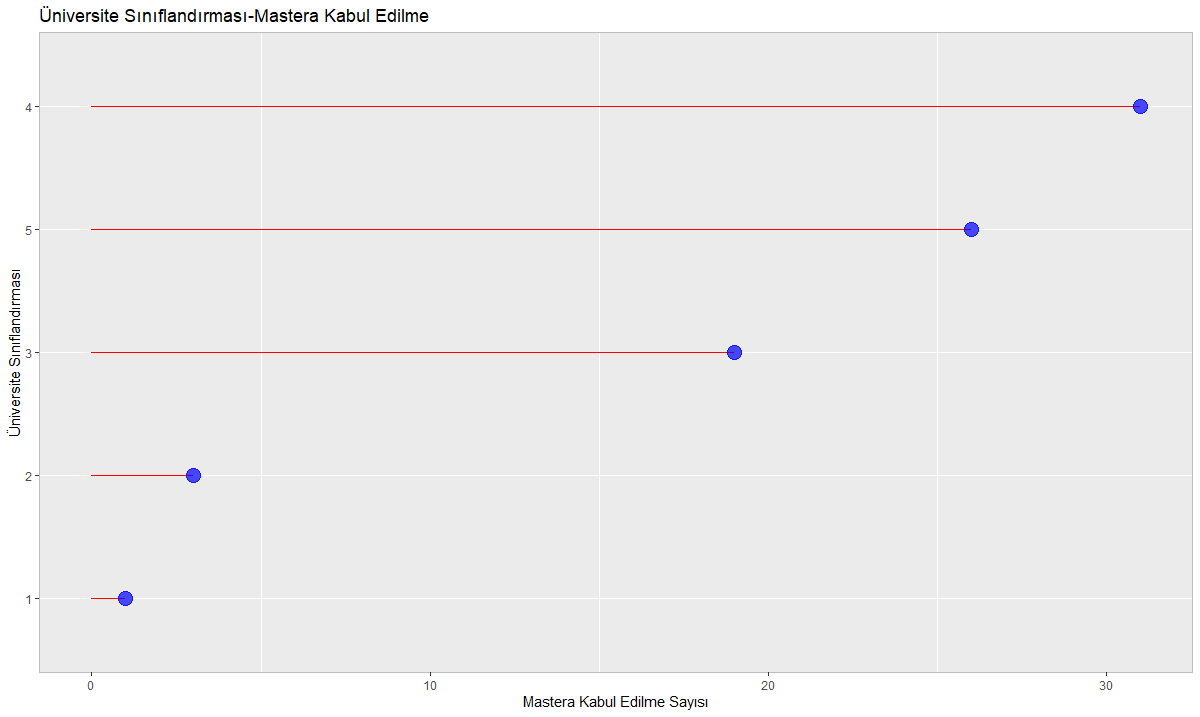
Üniversite sınıflandırmasına göre master programına kabul edilme koşulu 0.75 ten büyük olanların tablosu görülmektedir.



Kabul edilenlerin sayıları koşul\_1 olarak girilmiştir. Final\_1 adında yeni bir tibble oluşturulmuştur.

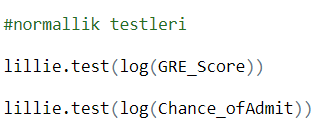


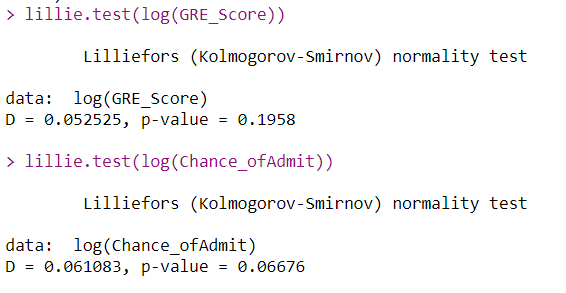
Arrange() koşuluyla koşul\_1 değişkeni sıralanmıştır. Mutate fonksiyonu ile University\_Rating\_1 değişkeni faktöre dönüştürülmüştür. Ggplot fonksiyonu ile grafik çizimine başlanmıştır x eksenine University\_Rating\_1 değişkeni , y eksenine de koşul\_1 değişkeni atanmıştır.   
geom\_segment fonksiyonu ile (x, y) ve (xend, yend) noktaları arasında düz bir çizgi çizilmiştir, color fonksiyonu ile rengi belirtilmiştir. Geom\_point fonksiyonu ile geom saçılma noktaları oluşturulup; renk ,dolgu,tip ve saydamlık özellikleri eklenmiştir. Coord\_flip fonksiyonu ile koordinat eksenleri çevrilmiştir. Theme fonksiyonları ile tema belirtilmiştir. X ekseni , y ekseni ve grafik isimlendirilmiştir.



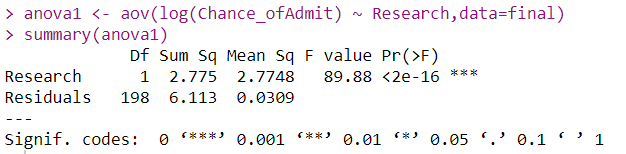
Grafiğe bakıldığında en iyi üniversitelerde(4,5) mastera kabul edilen öğrenci sayısının en fazla olduğu görülmektedir. En kötü üniversitede(1) ise mastera kabul edilen öğrenci sayısının oldukça az olduğu görülmektedir.

## ANOVA





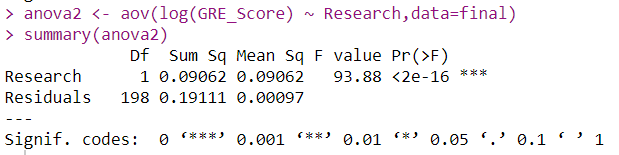
Logaritmaları alındığında GRE skoru ve mastera kabul edilme olasılığı değişkenlerinin normal dağıldığı söylenir.



Ho: Araştırma deneyimi ile mastera kabul edilme durumu değişkenlerinin ortalamaları arasında fark yoktur.

Hs: Araştırma deneyimi ile mastera kabul edilme durumu değişkenlerinin ortalamaları arasında fark vardır.

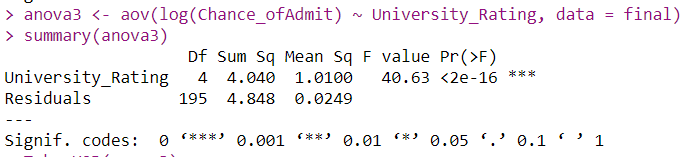
P=0.000 < α=0.05 olduğu için H0 hipotezi reddedilir. Araştırma durumu ile mastera kabul edilme durumu ortalamaları arasında fark olduğunu %95 güvenle söyleyebiliriz.



Ho: GRE skoru ile mastera kabul edilme durumu değişkenlerinin ortalamaları arasında fark yoktur.

Hs: GRE skoru ile mastera kabul edilme durumu değişkenlerinin ortalamaları arasında fark vardır.

P=0.000 < α=0.05 olduğu için H0 hipotezi reddedilir. Gruplar arasında fark olduğunu %95 güvenle söyleyebiliriz.

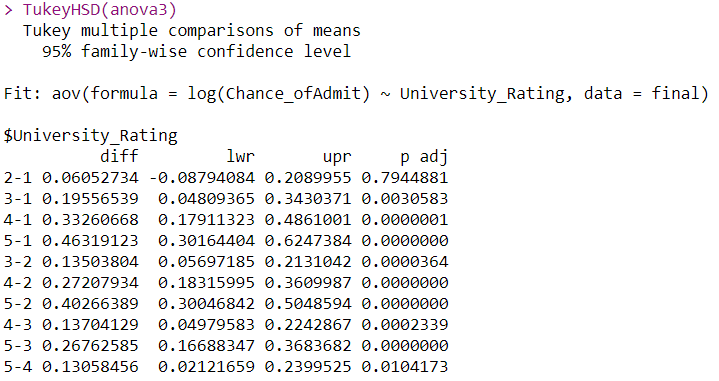


Ho: Üniversite sınıflandırmasına göre mastera kabul edilme durumu ortalamaları arasında fark yoktur.

Hs: Üniversite sınıflandırmasına göre mastera kabul edilme durumu ortalamaları arasında fark vardır.

P=0.000 < α=0.05 olduğu için H0 hipotezi reddedilir. Gruplar arasında fark olduğunu %95 güvenle söyleyebiliriz.

Üniversite sınıflandırmasına göre mastera kabul edilme durumunda farklılığı yaratan grubu bulmak için tukey testi yapılmıştır.



Ho: 1. Ve 2. Üniversite arasında fark yoktur.

Hs: 1. Ve 2. Üniversite arasında fark vardır.

P=0.7944 > α=0.05 olduğu için H0 Hipotezi reddedilemez. İki grup arasında fark olmadığını %95 güvenle söyleyebiliriz.

Ho: 3. Ve 1. Üniversite arasında fark yoktur.

Hs: 3. Ve 1. Üniversite arasında fark vardır.

P=0.0030 < α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 4. Ve 1. Üniversite arasında fark yoktur.

Hs: 4. Ve 1. Üniversite arasında fark vardır.

P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 5. Ve 1. Üniversite arasında fark yoktur.

Hs: 5. Ve 1. Üniversite arasında fark vardır.

P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 3. Ve 2. Üniversite arasında fark yoktur.

Hs: 3. Ve 2. Üniversite arasında fark vardır.

P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 4. Ve 2. Üniversite arasında fark yoktur.

Hs: 4. Ve 2. Üniversite arasında fark vardır.

P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 5. Ve 2. Üniversite arasında fark yoktur.

Hs: 5. Ve 2. Üniversite arasında fark vardır.

P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 4. Ve 3. Üniversite arasında fark yoktur.

Hs: 4. Ve 3. Üniversite arasında fark vardır.

P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 5. Ve 3. Üniversite arasında fark yoktur.

Hs: 5. Ve 3. Üniversite arasında fark vardır.

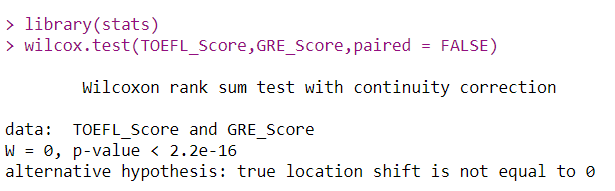
P=0.0000< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

Ho: 5. Ve 4. Üniversite arasında fark yoktur.

Hs: 5. Ve 4. Üniversite arasında fark vardır.

P=0.0104< α=0.05 olduğu için H0 Hipotezi reddedilir. İki grup arasında fark olduğunu %95 güvenle söyleyebiliriz.

## WİLCOX TESTİ



Ho: TOEFL skoru ve GRE skoru arasında konum yönünden farklılık yoktur.

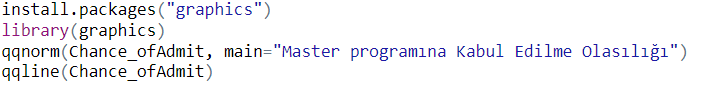
Hs : TOEFL skoru ve GRE skoru arasında konum yönünden farklılık vardır.

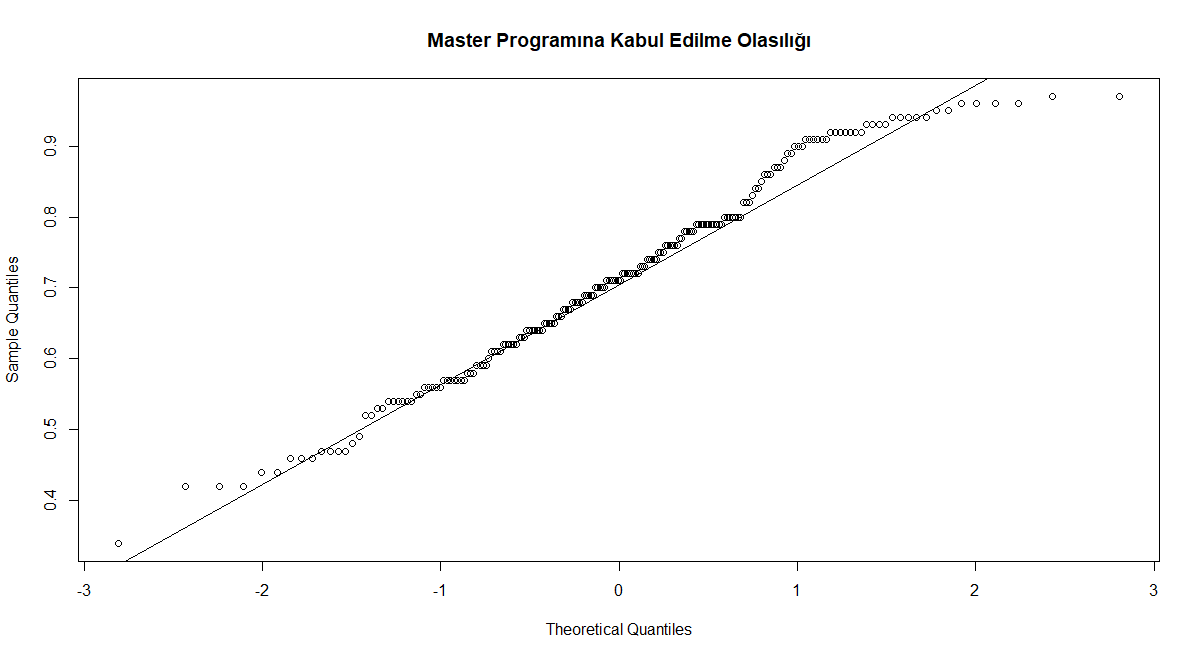
Çıkan sonuca göre p=0.000< α=0.05’tir. Dolayısıyla Ho hipotezi reddedilir yani 2 grup arasında konum yönünden farklılık olduğunu %95 güven düzeyinde söyleyebiliriz.

## REGRESYON

### Çoklu Doğrusal Regresyon

### Bağımlı Değişkene ait Q-Q Plot Grafiği



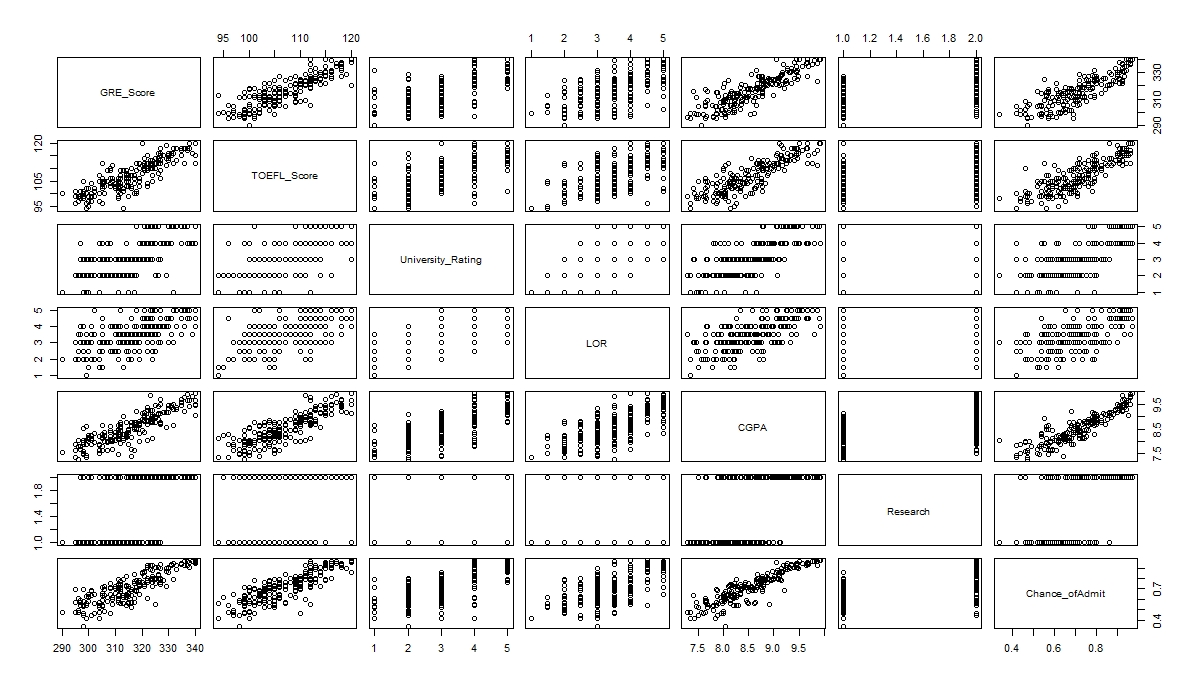


Noktalar 45 derecelik doğru etrafında dağıldığı için grafiğe göre bağımlı değişkenin normal dağılım gösterdiği söylenebilir.

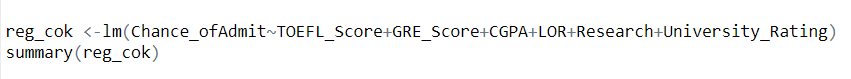
### Doğrusallık

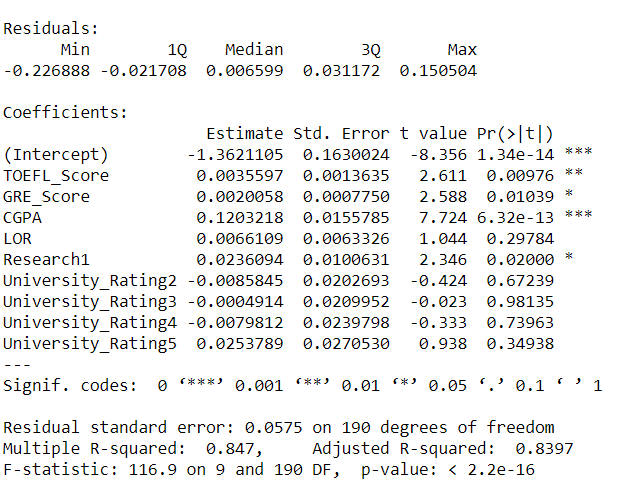


Pairs() fonksiyonu ile grafikler yardımıyla değişkenler arasındaki doğrusallığa bakılmıştır.



Y bağımlı değişkeni(Change\_ofAdmit) ile CGPA, TOEFL\_Skore, GRE\_Skore bağımsız değişkenleri arasında doğrusala yakın bir ilişki görülmektedir. Bunun yanında GRE\_Skore ve TOEFL\_Skore bağımsız değişkenleri arasında ilişki olduğu görülmektedir ayrıca CGPA bağımsız değişkeninin hem GRE\_Skore hem de TOEFL\_Skore bağımsız değişkenleri ile ilişkili olduğu görülmektedir dolayısıyla çoklu bağlantı sorunu olduğu söylenebilir.





Ho: Model istatistiksel olarak anlamlı değildir.

Hs: Model istatistiksel olarak anlamlıdır.

Tablodaki p value değeri=0.0000 < 0.05 olduğu için H0 hipotezi reddedilir. Modelin anlamlı olduğu 0.05 anlamlılık düzeyinde söylenebilir. Bağımsız değişkenlerden TOEFL skoru, GRE skoru, akademik ortalama değişkenleri anlamlıdır. Niyet ve referans mektuplarının gücü değişkeni ise anlamsızdır. Bağımlı değişkendeki toplam değişimin %84’ünü bağımsız değişkenlerce açıklamaktadır(R-suared).

Model denklemi =

- 1.36 - 0.003 + 0.002 + 0.12 + 0.006 + 0.02 - 0.008 – 0.0004 - 0.007 + 0.025 0.0575

### Katsayı Yorumları

**Sabit terim:** TOEFL\_Score, GRE\_Score, CGPA, LOR bağımsız değişkenleri 0 olduğunda bağımlı değişken ortalama -1.36’dır.

H0: Sabit terimin modele katkısı önemsizdir.

HS: Sabit terimin modele katkısı önemlidir.

Sabit terim için p value değeri=0.000 <α = 0.05 olduğu için H0 hipotezi reddedilir. Sabit terimin modele katkısı önemlidir.

**TOEFL\_Score Katsayısı :**

H0: TOEFL\_Score değişkeninin modele katkısı önemsizdir.

HS: TOEFL\_Score değişkeninin modele katkısı önemlidir.

Tablodan bakıldığında p value değeri=0.009 < α = 0.05 olduğu görülmektedir dolayısıyla H0 hipotezi reddedilir. Bağımsız değişkenin modele katkısı önemlidir. GRE\_Score, CGPA, LORdeğişkenleri sabit tutulduğunda ; TOEFL\_Score değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 0.0035 birim artırır.

**GRE\_Score Katsayısı:**

H0: GRE\_Score değişkeninin modele katkısı önemsizdir.

HS: GRE\_Score değişkeninin modele katkısı önemlidir.

Tablodan bakıldığında p value değeri=0.01 < α = 0.05 olduğu görülmektedir dolayısıyla H0 hipotezi reddedilir. GRE\_Score bağımsız değişkeninin modele katkısı önemlidir. Diğer değişkenler sabit tutulduğunda ; GRE\_Score bağımsız değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 0.0020 birim artırır.

**CGPA(Akademik Ortalama) Katsayısı :**

H0: CGPA değişkeninin modele katkısı önemsizdir.

HS: CGPA değişkeninin modele katkısı önemlidir.

Tablodan bakıldığında p value değeri=0.000 < α = 0.05 olduğu görülmektedir dolayısıyla H0 hipotezi reddedilir. Bağımsız değişkeninin modele katkısı önemlidir. GRE\_Score, TOEFL\_Score,LOR bağımsız değişkenleri sabit tutulduğunda ; akademik ortalama değişkenindeki 1 birimlik artış bağımlı değişkeni ortalama 0.12 birim artırır.

**LOR(Niyet ve Referans Mektuplarının Gücü) Katsayısı:**

H0: LOR değişkeninin modele katkısı önemsizdir.

HS: LOR değişkeninin modele katkısı önemlidir.

Tablodan bakıldığında p value değeri=0.29 > α = 0.05 olduğu görülmektedir dolayısıyla H0 hipotezi reddedilemez. LOR değişkeninin modele katkısı önemsizdir.

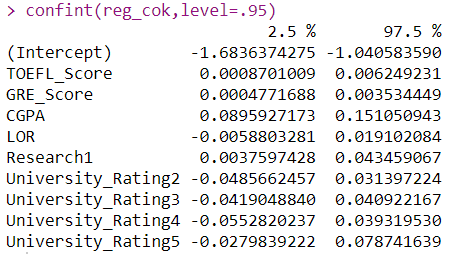
**Research(1) Katsayısı** :

H0: Araştırma deneyimi olan ile araştırma deneyimi olmayan arasında fark yoktur.

HS: Araştırma deneyimi olan ile araştırma deneyimi olmayan arasında fark vardır.

Tabloya bakıldığında p value değeri=0.02 < α = 0.05 olduğu için H0 hipotezi reddedilir. Research(1) ile Research(0) arasında fark olduğu %95 güvenle söylenebilir.

### Regresyon Katsayıları için Güven Aralığı



**Sabit Terim Güven Aralığı:** TOEFL\_Score, GRE\_Score,CGPA ,LOR bağımsız değişkenleri 0 olduğunda bağımlı değişkenin ortalama -1.6836 ile -1.0405 arasında olduğu %95 güvenle söylenir. Sabit terimin güven aralığı 0’ı içermediği için %95 güvenle anlamlıdır.

**TOEFL\_Score için Güven Aralığı :** , GRE\_Score, CGPA ,LOR bağımsız değişkenleri sabit tutulduğunda TOEFL\_Score bağımsız değişkeni 1 birim arttığında bağımlı değişkenin ortalama 0.0008 ile 0.0062 arasında olduğu %95 güvenle söylenir. TOEFL\_Score güven aralığı 0’ı içermediği için %95 güvenle anlamlıdır.

**GRE\_Score için Güven Aralığı:** Diğer bağımsız değişkenler sabit tutulduğunda GRE\_Score bağımsız değişkeni 1 birim arttığında bağımlı değişkenin ortalama 0.0004 ile 0.0035 arasında olduğu %95 güvenle söylenir. GRE\_Score güven aralığı 0’ı içermediği için %95 güvenle anlamlıdır.

**CGPA için Güven Aralığı:** Akademik ortalama bağımsız değişkeni güven aralığı 0’ı içermediği için %95 güvenle anlamlıdır.

**LOR için Güven Aralığı :** LOR ortalama bağımsız değişkeni güven aralığı 0’ı içerdiği için %95 güvenle anlamsızdır.

**Research(1) için Güven Aralığı:** Araştırma deneyimi(1) bağımsız değişkeni güven aralığı 0’ı içermediği için %95 güvenle anlamlıdır.

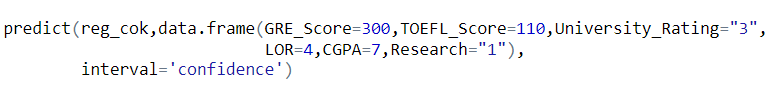
**University Rating(2) için Güven Aralığı:** Üniversite sınıflandırması(2) bağımsız değişkeni güven aralığı 0’ı içerdiği için %95 güvenle anlamsızdır.

**University Rating(3) için Güven Aralığı:** Üniversite sınıflandırması(3) bağımsız değişkeni güven aralığı 0’ı içerdiği için %95 güvenle anlamsızdır.

**University Rating(4) için Güven Aralığı:** Üniversite sınıflandırması(4) bağımsız değişkeni güven aralığı 0’ı içerdiği için %95 güvenle anlamsızdır.

**University Rating(5) için Güven Aralığı:** Üniversite sınıflandırması(5) bağımsız değişkeni güven aralığı 0’ı içediği için %95 güvenle anlamsızdır.

### Ön Kestirim

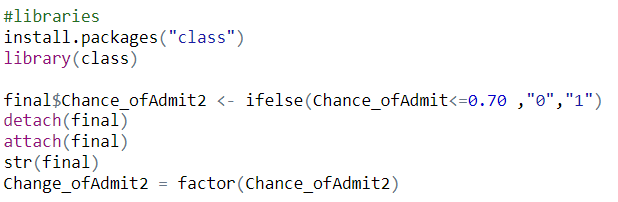


Sonuç:

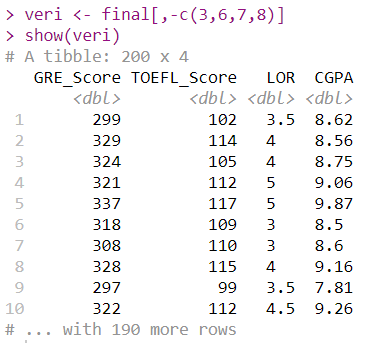


Veri setinde bulunmayan değişken değerleri girilmiştir, tahmin değeri 0.52 olarak bulunmuştur. GRE Skoru 300, TOEFL skoru 110 , Universite sınıflandırması 3 ,niyet ve referans mektupları puanı 4, akademik ortalaması 7 ve araştırma deneyimine sahip bir öğrenicinin mastera kabul edilme olasılığı 0.52’dir. Bu değişkenler için bağımlı değişkenin(mastera kabul edilme olasılığının) ortalama 0.4703 ile 0.5756 arasında olduğu söylenebilir.

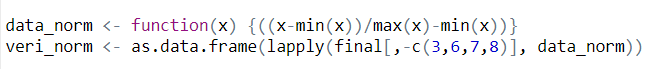
## K EN YAKIN KOMŞU SINIFLANDIRMA ALGORİTMASI



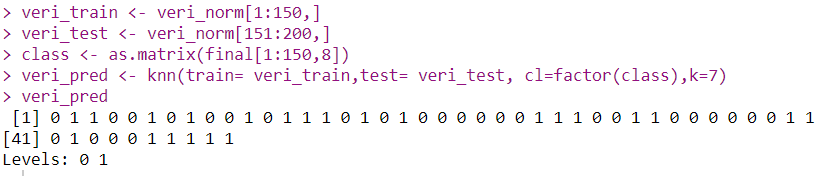
Öncelikle k en yakın komşu sınıflandırmasını yapabilmek için class kütüphanesi yüklenmiştir. Chance\_ofAdmit2 adında yeni bir değişken oluşturulmuştur. Master a kabul edilme olasılığı 0.70’ten küçük eşit olanlar 0, bunun dışındakiler 1 olacak şekilde koşul oluşturulmuştur. Bu değişken veri setine detach ve attach komutları yardımıyla eklenmiş ve factor() fonksiyonu ile kategorik hale getirilmiştir.



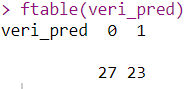
Verinin içinden GRE\_Score, TOEFL\_Score, LOR, CGPA değişkenleri seçilmiş ve veri olarak adlandırılmıştır.



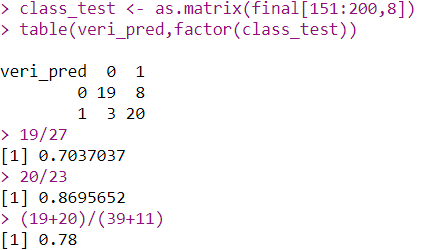
Verideki değişkenlere normalleştirme işlemi yapılmıştır.



200 gözlemin 0.75’i eğitim , 0.25’i test kümesi olarak ayrılmıştır. Class ise Chance\_ofAdmit (master programına kabul edilip edilmeme) sütununu içermektedir. Knn() fonksiyonu yardımı ile mastera kabul edilip edilmeme durumuna ait 50 tane kestirim değeri oluşturulmuştur.



Kestirim değerleri ile ilgili sonuçlar yukarıdaki gibi verilmiştir. Kabul edilmeyen 27, kabul edilen 23 gözlem vardır.



50 tane veri içerisinden 39 tanesi doğru tahmin edilmiştir. Başarı oranının 0.78 olduğu görülmektedir. Master programına kabul edilme koşulu 1 olanların 0.86 oranıyla daha iyi tahmin edildiği görülmektedir.

# ÖZET

Bu istatistiksel analiz sonucunda bir kişinin mastera kabul edilme olasılığını etkileyen faktörlerin akademik ortalama ,araştırma deneyimi, TOEFL skoru ve GRE skoru olduğu görülmüştür. Akademik ortalaması yüksek , araştırma deneyimi olan , TOEFL ve GRE skorları iyi olan kişilerin mastera kabul edilme olasılığının daha yüksek olduğu sonucuna varılmıştır. Bunun yanı sıra GRE skor ile TOEFL skor arasında doğru orantılı bir ilişki olduğu görülmüştür. Üniversite sınıflandırmasının da mastera kabul koşulunu etkilemesini beklerdim ancak öyle olmadı. Analizde daha çok sınıflandırma algoritması kullanıp en son sonuçları karşılaştırıp en iyi tahminin hangi yönteme ait olduğuna karar versem çok daha güzel olurdu ancak bazı kütüphaneler çalışmadığı için bunu yapamadım.

KAYNAKÇA:

<http://www.sthda.com/english/wiki/one-way-anova-test-in-r>

<http://geog.uoregon.edu/GeogR/topics/coplots.html>

<http://bilgisayarkavramlari.sadievrenseker.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/>

R Cookbook

Udemy-Modern R Programlama Eğitimi

<https://www.data-to-viz.com/graph/lollipop.html>

<https://docplayer.biz.tr/15448102-K-en-yakin-komsu-algoritmasi-ve-bir-uygulama-kredi-riskini-siniflandirma.html>

<https://www.newslabturkey.org/r-dplyr-paketi-ile-veri-nasil-islenir-ve-analiz-edilir/?gclid=EAIaIQobChMIqeKL9MmG5wIV0OeaCh3bgAV3EAAYAiAAEgK0MPD_BwE>

<https://towardsdatascience.com/k-nearest-neighbors-algorithm-with-examples-in-r-simply-explained-knn-1f2c88da405c>

<https://bookdown.org/burak2358/SARP-TR/betimleyici-istatistikler-ve-hipotez-testi.html>