#1

library(MASS)

# 50 kişilik bir veri seti oluşturma,Rastgele sayılar

set.seed(123)

n <- 50

# Bağımsız değişkenlerin oluşturulması,bağımsız değişkenlere farklı ranjlarda değer giriyoruz.

degisken1 <- runif(n, min = 0, max = 100) # 0-100 arasında rastgele değerler

degisken2 <- runif(n, min = 500, max = 750) # 500-750 arasında rastgele değerler

degisken3 <- runif(n, min = 250, max = 500) # 250-500 arasında rastgele değerler

degisken4 <- runif(n, min = 500, max = 1000) # 500-1000 arasında rastgele değerler

degisken5 <- runif(n, min = 50, max = 200) # 50-200 arasında rastgele değerler

# Bağımlı değişkenin oluşturulması

bagimli\_degisken <- rnorm(n, mean = 0, sd = 1)

# Veri setinin oluşturma

veri <- data.frame(degisken1, degisken2, degisken3, degisken4, degisken5, bagimli\_degisken)

# Oluşturduğumuz veri setini gözlemleme

head(veri)

summary(veri)

# Veri setinin yapısını gözlemleme

str(veri)

# Veri setinin özet istatistiklerini inceleme

summary(veri)

# Eksik değerleri kontrol etme ve tamamlama

is.na(veri)

# Aykırı değerleri belirleme ve işleme

outliers <- boxplot.stats(veri$degisken1)$out

outliers <- boxplot.stats(veri$degisken2)$out

outliers <- boxplot.stats(veri$degisken3)$out

outliers <- boxplot.stats(veri$degisken4)$out

outliers <- boxplot.stats(veri$degisken5)$out

# Histogramlarla değişkenlerin dağılımlarını görselleştirme

hist(veri$degisken1, main = "Degisken1 Dağılımı", xlab = "Degisken1 Değerleri")

hist(veri$degisken2, main = "Degisken2 Dağılımı", xlab = "Degisken2 Değerleri")

# Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi scatter plotlar ile görselleştirme

plot(veri$bagimli\_degisken ~ veri$degisken1, main = "Degisken1 ve Bagimli Degisken Arasındaki İlişki", xlab = "Degisken1", ylab = "Bagimli Degisken")

# Diğer değişkenler için de aynı şekilde devam edebilirsiniz.

# Korelasyon analizi yapalım

cor(veri$degisken1, veri$bagimli\_degisken) #Bu sonuca göre, "degisken1" ile "bagimli\_degisken" arasında zayıf bir pozitif korelasyon bulunmaktadır. Korelasyon katsayısı 0.07975679'dur, bu da değişkenler arasında çok düşük bir ilişki olduğunu gösterir.

cor(veri$degisken2,veri$bagimli\_degisken) #Sonuca göre, "degisken2" ile "bagimli\_degisken" arasında negatif bir korelasyon bulunmaktadır. Korelasyon katsayısı -0.08175974'tür, yani değişkenler arasında çok düşük bir negatif ilişki olduğunu gösterir. Bu, bir değişken artarken diğerinin azaldığına işaret eder, ancak ilişki çok zayıftır.

cor(veri$degisken3,veri$bagimli\_degisken) #Sonuca göre, "degisken3" ile "bagimli\_degisken" arasında bir pozitif korelasyon bulunmaktadır. Korelasyon katsayısı 0.1133065'tir, yani değişkenler arasında düşük bir pozitif ilişki olduğunu gösterir.

cor(veri$degisken4,veri$bagimli\_degisken) #Sonuca göre, "degisken4" ile "bagimli\_degisken" arasında pozitif bir korelasyon bulunmaktadır. Korelasyon katsayısı 0.1635945'tir, yani değişkenler arasında düşük bir pozitif ilişki olduğunu gösterir.

cor(veri$degisken5,veri$bagimli\_degisken) #Sonuca göre, "degisken5" ile "bagimli\_degisken" arasında pozitif bir korelasyon bulunmaktadır. Korelasyon katsayısı 0.08379184'tür, yani değişkenler arasında çok düşük bir pozitif ilişki olduğunu gösterir.

# T-test yapalım (örnek olarak)

t.test(veri$degisken2, mu = 550)

# (Ho): Veri setinin popülasyon ortalaması 550'dir.

(H1): Veri setinin popülasyon ortalaması 550'den farklıdır.

# Yorum: p-değeri oldukça küçüktür (4.987e-09), bu da H0 hipotezi reddetmek için güçlü bir kanıt olduğunu gösterir. Bu nedenle, veri setinin popülasyon ortalamasının 550'den farklı olduğunu söyleyebiliriz.

|  |
| --- |
| library(MASS)  > # 50 kişilik bir veri seti oluşturma,Rastgele sayılar  > set.seed(123)  > n <- 50  > degisken1 <- runif(n, min = 0, max = 100) # 0-100 arasında rastgele değerler  > degisken2 <- runif(n, min = 500, max = 750) # 500-750 arasında rastgele değerler  > degisken3 <- runif(n, min = 250, max = 500) # 250-500 arasında rastgele değerler  > degisken4 <- runif(n, min = 500, max = 1000) # 500-1000 arasında rastgele değerler  > degisken5 <- runif(n, min = 50, max = 200) # 50-200 arasında rastgele değerler  > # Bağımlı değişkenin oluşturulması  > bagimli\_degisken <- rnorm(n, mean = 0, sd = 1)  > # Veri setinin oluşturma  > veri <- data.frame(degisken1, degisken2, degisken3, degisken4, degisken5, bagimli\_degisken)  > # Oluşturduğumuz veri setini gözlemleme  > head(veri)  degisken1 degisken2 degisken3 degisken4 degisken5 bagimli\_degisken  1 28.75775 511.4578 399.9972 923.7266 85.8089 -0.65194990  2 78.83051 610.5500 333.2059 748.7636 194.3538 0.23538657  3 40.89769 699.7312 372.1533 693.9545 140.2049 0.07796085  4 88.30174 530.4748 488.6185 623.2245 127.2545 -0.96185663  5 94.04673 640.2370 370.7256 555.5482 110.3860 -0.07130809  6 4.55565 551.6328 472.5876 694.9972 182.0370 1.44455086  > summary(veri)  degisken1 degisken2 degisken3 degisken4 degisken5  Min. : 2.461 Min. :500.2 Min. :252.6 Min. :555.5 Min. : 57.15  1st Qu.:27.137 1st Qu.:556.5 1st Qu.:314.9 1st Qu.:664.8 1st Qu.: 92.59  Median :50.295 Median :611.3 Median :376.2 Median :750.0 Median :124.78  Mean :52.009 Mean :619.3 Mean :378.8 Mean :756.5 Mean :126.90  3rd Qu.:78.084 3rd Qu.:685.6 3rd Qu.:434.0 3rd Qu.:828.5 3rd Qu.:157.79  Max. :99.427 Max. :746.2 Max. :496.1 Max. :992.8 Max. :195.40  bagimli\_degisken  Min. :-2.053247  1st Qu.:-0.718990  Median :-0.052688  Mean :-0.007875  3rd Qu.: 0.675080  Max. : 3.241040  > # Veri setinin yapısını gözlemleme  > str(veri)  'data.frame': 50 obs. of 6 variables:  $ degisken1 : num 28.8 78.8 40.9 88.3 94 ...  $ degisken2 : num 511 611 700 530 640 ...  $ degisken3 : num 400 333 372 489 371 ...  $ degisken4 : num 924 749 694 623 556 ...  $ degisken5 : num 85.8 194.4 140.2 127.3 110.4 ...  $ bagimli\_degisken: num -0.6519 0.2354 0.078 -0.9619 -0.0713 ...  > # Veri setinin özet istatistiklerini inceleme  > summary(veri)  degisken1 degisken2 degisken3 degisken4 degisken5  Min. : 2.461 Min. :500.2 Min. :252.6 Min. :555.5 Min. : 57.15  1st Qu.:27.137 1st Qu.:556.5 1st Qu.:314.9 1st Qu.:664.8 1st Qu.: 92.59  Median :50.295 Median :611.3 Median :376.2 Median :750.0 Median :124.78  Mean :52.009 Mean :619.3 Mean :378.8 Mean :756.5 Mean :126.90  3rd Qu.:78.084 3rd Qu.:685.6 3rd Qu.:434.0 3rd Qu.:828.5 3rd Qu.:157.79  Max. :99.427 Max. :746.2 Max. :496.1 Max. :992.8 Max. :195.40  bagimli\_degisken  Min. :-2.053247  1st Qu.:-0.718990  Median :-0.052688  Mean :-0.007875  3rd Qu.: 0.675080  Max. : 3.241040  > # Eksik değerleri kontrol etme ve tamamlama  > is.na(veri)  degisken1 degisken2 degisken3 degisken4 degisken5 bagimli\_degisken  [1,] FALSE FALSE FALSE FALSE FALSE FALSE  [2,] FALSE FALSE FALSE FALSE FALSE FALSE  [3,] FALSE FALSE FALSE FALSE FALSE FALSE  [4,] FALSE FALSE FALSE FALSE FALSE FALSE  [5,] FALSE FALSE FALSE FALSE FALSE FALSE  [6,] FALSE FALSE FALSE FALSE FALSE FALSE  [7,] FALSE FALSE FALSE FALSE FALSE FALSE  [8,] FALSE FALSE FALSE FALSE FALSE FALSE  [9,] FALSE FALSE FALSE FALSE FALSE FALSE  [10,] FALSE FALSE FALSE FALSE FALSE FALSE  [11,] FALSE FALSE FALSE FALSE FALSE FALSE  [12,] FALSE FALSE FALSE FALSE FALSE FALSE  [13,] FALSE FALSE FALSE FALSE FALSE FALSE  [14,] FALSE FALSE FALSE FALSE FALSE FALSE  [15,] FALSE FALSE FALSE FALSE FALSE FALSE  [16,] FALSE FALSE FALSE FALSE FALSE FALSE  [17,] FALSE FALSE FALSE FALSE FALSE FALSE  [18,] FALSE FALSE FALSE FALSE FALSE FALSE  [19,] FALSE FALSE FALSE FALSE FALSE FALSE  [20,] FALSE FALSE FALSE FALSE FALSE FALSE  [21,] FALSE FALSE FALSE FALSE FALSE FALSE  [22,] FALSE FALSE FALSE FALSE FALSE FALSE  [23,] FALSE FALSE FALSE FALSE FALSE FALSE  [24,] FALSE FALSE FALSE FALSE FALSE FALSE  [25,] FALSE FALSE FALSE FALSE FALSE FALSE  [26,] FALSE FALSE FALSE FALSE FALSE FALSE  [27,] FALSE FALSE FALSE FALSE FALSE FALSE  [28,] FALSE FALSE FALSE FALSE FALSE FALSE  [29,] FALSE FALSE FALSE FALSE FALSE FALSE  [30,] FALSE FALSE FALSE FALSE FALSE FALSE  [31,] FALSE FALSE FALSE FALSE FALSE FALSE  [32,] FALSE FALSE FALSE FALSE FALSE FALSE  [33,] FALSE FALSE FALSE FALSE FALSE FALSE  [34,] FALSE FALSE FALSE FALSE FALSE FALSE  [35,] FALSE FALSE FALSE FALSE FALSE FALSE  [36,] FALSE FALSE FALSE FALSE FALSE FALSE  [37,] FALSE FALSE FALSE FALSE FALSE FALSE  [38,] FALSE FALSE FALSE FALSE FALSE FALSE  [39,] FALSE FALSE FALSE FALSE FALSE FALSE  [40,] FALSE FALSE FALSE FALSE FALSE FALSE  [41,] FALSE FALSE FALSE FALSE FALSE FALSE  [42,] FALSE FALSE FALSE FALSE FALSE FALSE  [43,] FALSE FALSE FALSE FALSE FALSE FALSE  [44,] FALSE FALSE FALSE FALSE FALSE FALSE  [45,] FALSE FALSE FALSE FALSE FALSE FALSE  [46,] FALSE FALSE FALSE FALSE FALSE FALSE  [47,] FALSE FALSE FALSE FALSE FALSE FALSE  [48,] FALSE FALSE FALSE FALSE FALSE FALSE  [49,] FALSE FALSE FALSE FALSE FALSE FALSE  [50,] FALSE FALSE FALSE FALSE FALSE FALSE  #Tüm değerlerin FALSE olarak dönmesi, veri setinde hiçbir eksik (NA) değer olmadığını gösterir. Yani, tüm hücreler dolu ve eksiksiz veri içerir. Bu durumda, herhangi bir eksik değerle başa çıkmak için özel bir işlem yapmamıza gerek yoktur. Veri setimiz eksiksizdir ve direkt olarak analiz yapabiliriz.  outliers <- boxplot.stats(veri$degisken1)$out  > outliers <- boxplot.stats(veri$degisken2)$out  > outliers <- boxplot.stats(veri$degisken3)$out  > outliers <- boxplot.stats(veri$degisken4)$out  > outliers <- boxplot.stats(veri$degisken5)$out  > # Histogramlarla değişkenlerin dağılımlarını görselleştirme  > hist(veri$degisken1, main = "Degisken1 Dağılımı", xlab = "Degisken1 Değerleri")  > hist(veri$degisken2, main = "Degisken2 Dağılımı", xlab = "Degisken2 Değerleri")  > plot(veri$bagimli\_degisken ~ veri$degisken1, main = "Degisken1 ve Bagimli Degisken Arasındaki İlişki", xlab = "Degisken1", ylab = "Bagimli Degisken")  > # Diğer değişkenler için de aynı şekilde devam edebilirsiniz.  > # Korelasyon analizi yapalım  > cor(veri$degisken1, veri$bagimli\_degisken) #Bu sonuca göre, "degisken1" ile "bagimli\_degisken" arasında zayıf bir pozitif korelasyon bulunmaktadır. Korelasyon katsayısı 0.07975679'dur, bu da değişkenler arasında çok düşük bir ilişki olduğunu gösterir.  [1] 0.07975679  > cor(veri$degisken2,veri$bagimli\_degisken) #Sonuca göre, "degisken2" ile "bagimli\_degisken" arasında negatif bir korelasyon bulunmaktadır. Korelasyon katsayısı -0.08175974'tür, yani değişkenler arasında çok düşük bir negatif ilişki olduğunu gösterir. Bu, bir değişken artarken diğerinin azaldığına işaret eder, ancak ilişki çok zayıftır.  [1] -0.08175974  > cor(veri$degisken3,veri$bagimli\_degisken) #Sonuca göre, "degisken3" ile "bagimli\_degisken" arasında bir pozitif korelasyon bulunmaktadır. Korelasyon katsayısı 0.1133065'tir, yani değişkenler arasında düşük bir pozitif ilişki olduğunu gösterir.  [1] 0.1133065  > cor(veri$degisken4,veri$bagimli\_degisken) #Sonuca göre, "degisken4" ile "bagimli\_degisken" arasında pozitif bir korelasyon bulunmaktadır. Korelasyon katsayısı 0.1635945'tir, yani değişkenler arasında düşük bir pozitif ilişki olduğunu gösterir.  [1] 0.1635945  > cor(veri$degisken5,veri$bagimli\_degisken) #Sonuca göre, "degisken5" ile "bagimli\_degisken" arasında pozitif bir korelasyon bulunmaktadır. Korelasyon katsayısı 0.08379184'tür, yani değişkenler arasında çok düşük bir pozitif ilişki olduğunu gösterir.  [1] 0.08379184  > t.test(veri$degisken2, mu = 550)  One Sample t-test  data: veri$degisken2  t = 7.08, df = 49, p-value = 4.987e-09  alternative hypothesis: true mean is not equal to 550  95 percent confidence interval:  599.5990 638.9145  sample estimates:  mean of x  619.2568  # (Ho): Veri setinin popülasyon ortalaması 550'dir.  (H1): Veri setinin popülasyon ortalaması 550'den farklıdır.  # yorum: p-değeri (p-value): 4.987e-09 p-değeri çok küçüktür, bu nedenle H0 hipotezi güçlü bir şekilde reddeder. Bu, veri setinin popülasyon ortalamasının 550'den farklı olduğu sonucuna yol açar. H1 kabul edilir. |
|  | |
| |  | | --- | |  | | |

#2. veri seti için keşfsel veri analizi

library(readxl)

> # Veri setinin yapısını görüntüler#

> str(eda\_veri\_set)

tibble [163 × 9] (S3: tbl\_df/tbl/data.frame)

$ sıra : num [1:163] 1 2 3 4 5 6 7 8 9 10 ...

$ yas : num [1:163] 25 22 24 22 23 23 22 24 22 21 ...

$ AGNO : num [1:163] 3 2 4 3 3 4 2 4 3 3 ...

$ lisepuanı: num [1:163] 80 70 85 69 53 79 68 82 58 58 ...

$ S.Kaygı : num [1:163] 30 50 36 55 53 41 51 49 45 35 ...

$ turkce : num [1:163] 72 60 42 52 32 54 50 38 44 32 ...

$ Biyoloji : num [1:163] 66 66 64 46 40 56 62 60 46 42 ...

$ Felsefe : num [1:163] 56 56 74 60 24 56 82 78 34 24 ...

$ Tarih : num [1:163] 58 58 60 82 36 74 46 64 30 30 ...

> # Değişkenlerin özet istatistiklerini verir#

> summary(eda\_veri\_set)

sıra yas AGNO lisepuanı S.Kaygı

Min. : 1.0 Min. :19.00 Min. :2.000 Min. :53.00 Min. :29.00

1st Qu.: 41.5 1st Qu.:21.00 1st Qu.:3.000 1st Qu.:59.00 1st Qu.:38.00

Median : 82.0 Median :23.00 Median :3.000 Median :69.00 Median :44.00

Mean : 82.0 Mean :22.53 Mean :3.074 Mean :69.68 Mean :44.35

3rd Qu.:122.5 3rd Qu.:23.50 3rd Qu.:4.000 3rd Qu.:79.50 3rd Qu.:50.00

Max. :163.0 Max. :25.00 Max. :4.000 Max. :85.00 Max. :66.00

turkce Biyoloji Felsefe Tarih

Min. :20.00 Min. :22.00 Min. : 22.00 Min. :24.00

1st Qu.:40.00 1st Qu.:51.00 1st Qu.: 60.00 1st Qu.:54.00

Median :46.00 Median :62.00 Median : 72.00 Median :62.00

Mean :47.28 Mean :62.37 Mean : 68.52 Mean :62.18

3rd Qu.:54.00 3rd Qu.:72.00 3rd Qu.: 80.00 3rd Qu.:70.00

Max. :90.00 Max. :98.00 Max. :100.00 Max. :98.00

> # Eksik değerleri kontrol etme ve doldurma#

> is.na(eda\_veri\_set)

sıra yas AGNO lisepuanı S.Kaygı turkce Biyoloji Felsefe Tarih

[1,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[2,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[3,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[4,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[5,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[7,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[8,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[9,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[10,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[11,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[12,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[13,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[14,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[15,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[16,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[17,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[18,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[19,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[20,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[21,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[22,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[23,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[24,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[25,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[26,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[27,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[28,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[29,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[30,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[31,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[32,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[33,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[34,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[35,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[36,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[37,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[38,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[39,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[40,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[41,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[42,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[43,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[44,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[45,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[46,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[47,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[48,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[49,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[50,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[51,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[52,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[53,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[54,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[55,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[56,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[57,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[58,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[59,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[60,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[61,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[62,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[63,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[64,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[65,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[66,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[67,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[68,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[69,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[70,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[71,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[72,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[73,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[74,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[75,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[76,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[77,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[78,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[79,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[80,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[81,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[82,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[83,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[84,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[85,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[86,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[87,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[88,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[89,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[90,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[91,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[92,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[93,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[94,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[95,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[96,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[97,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[98,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[99,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[100,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[101,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[102,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[103,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[104,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[105,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[106,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[107,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[108,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[109,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[110,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[111,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[ reached getOption("max.print") -- omitted 52 rows ]>

#Tüm değerlerin FALSE olarak dönmesi, eda\_veri\_set veri setinde hiçbir eksik (NA) değer olmadığını gösterir. Yani, tüm hücreler dolu ve eksiksiz veri içerir. Bu durumda, herhangi bir eksik değerle başa çıkmak için özel bir işlem yapmamıza gerek yoktur. Veri setimiz eksiksizdir ve direkt olarak analiz yapabiliriz.

# Eksik değerleri kontrol eder#

> eda\_veri\_set[is.na(eda\_veri\_set)] <- 0

> # Eksik değerleri 0 ile doldurur#

> # Aykırı değerleri kontrol etme ve işleme#

> outlier <- boxplot.stats(eda\_veri\_set$yas)$out # Aykırı değerleri bulur#

> eda\_veri\_set <- eda\_veri\_set[!eda\_veri\_set$yas %in% outlier, ] # Aykırı değerleri çıkarır#

> # Histogram

> hist(eda\_veri\_set$AGNO, main="Değişkenin Dağılımı", xlab="AGNO")

> # Kutu Grafiği

> boxplot(yas ~ AGNO, data=eda\_veri\_set, main="yas vs AGNO Kutu Grafiği")

> # Tek Örneklem t-testi#

> t.test(eda\_veri\_set$AGNO, mu=2)

One Sample t-test

data: eda\_veri\_set$AGNO

t = 18.692, df = 162, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 2

95 percent confidence interval:

2.960196 3.187043

sample estimates:

mean of x

3.07362

#(Ho): Veri setinin popülasyon ortalaması 2'dir.

(H1): Veri setinin popülasyon ortalaması 2'den farklıdır.

# Yorum:p-değeri çok küçüktür, ortalama 3.07362 olarak bulunmuştur.bu nedenle H0 hipotezi güçlü bir şekilde reddeder. Bu, veri setinin popülasyon ortalamasının 2'den farklı olduğu sonucuna ulaşılır.

> # Bağımsız İki Örneklem t-testi#

> t.test(eda\_veri\_set$yas, eda\_veri\_set$AGNO)

Welch Two Sample t-test

data: eda\_veri\_set$yas and eda\_veri\_set$AGNO

t = 155.1, df = 242.26, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

19.21298 19.70727

sample estimates:

mean of x mean of y

22.53374 3.07362

(Ho): İki grup arasındaki ortalamalar arasında fark yoktur.

(H1): İki grup arasındaki ortalamalar arasında fark vardır.

# p-değeri çok küçüktür(p-value < 2.2e-16). Bu, iki grup arasında anlamlı bir fark olduğu sonucuna yol açar. Bu durumda, yaş ve AGNO arasında anlamlı bir ilişki olduğunu gösterir(H0 red,H1 kabul).

> # Korelasyon Analizi#

> cor(eda\_veri\_set$yas, eda\_veri\_set$AGNO)

[1] -0.05558965

# Yorum: Elde edilen korelasyon katsayısı -0.05558965'tir. Bu değer, "yas" ile "AGNO" arasında çok düşük bir negatif ilişki olduğunu gösterir.