Parametrik Olmayan İstatistiksel Yöntemler Dönem Ödevi

Mert Yanık

21522134

Oluşturulan bütün kodlar, çıktılar ve Word dosyası RStudio içerisinde bulunan R Markdown ile oluşturulmuştur. Elimizde 2 farklı veri kümesi vardır. Bunlardan biri bağımsız diğeri bağımlı veri kümesidir.

# Soru 1

Birinci soruda bizden en az 3 bağımsız gruptan oluşan bir veri kümesi bulmamız ve her bir değişkeninin Normal dağılmadığını göstermemiz istenmektedir. Aşağıda bu veri kümesinin RStudio’ya okutuluşu ve ekrana basılması için gerekli kodlar yazılmıştır.

Bagimsiz <- read.delim("C:/Users/user/Desktop/NonPar/Bagimsiz.txt")  
attach(Bagimsiz)  
print(Bagimsiz)

## A.Makinesi B.Makinesi C.Makinesi  
## 1 340 294 263  
## 2 340 325 340  
## 3 356 325 356  
## 4 386 NA 371  
## 5 386 NA 371  
## 6 499 NA NA

Veri kümesi, kalite kontrolü yapmak amacı ile aynı üretimi yapan üç ayrı makineden (A, B, C) sırasıyla 6, 3 ve 5 tane olmak üzere ürünleri ele almaktadır. Burada puanların az olması dayanıklılığın daha zayıf olduğunu göstermektedir.

Bu verilere dair özetleyici istatistiklere bakalım.

summary(Bagimsiz)

## A.Makinesi B.Makinesi C.Makinesi   
## Min. :340.0 Min. :294.0 Min. :263.0   
## 1st Qu.:344.0 1st Qu.:309.5 1st Qu.:340.0   
## Median :371.0 Median :325.0 Median :356.0   
## Mean :384.5 Mean :314.7 Mean :340.2   
## 3rd Qu.:386.0 3rd Qu.:325.0 3rd Qu.:371.0   
## Max. :499.0 Max. :325.0 Max. :371.0   
## NA's :3 NA's :1

Elimizde ki 3 değişkeninde minimum - maksimum değerleri, çeyreklik değerleri ve ortalamaları görülmektedir. Bu noktada ortalama değer ile medyan değerinin birbirine olan yakınlığı veya uzaklığı ile Normal Dağılım çıkarsaması yapılabilir.

Şimdi elimizde ki 3 değişkenin tamamının Normal dağılmadığını gösterelim.

shapiro.test(A.Makinesi)

##   
## Shapiro-Wilk normality test  
##   
## data: A.Makinesi  
## W = 0.77569, p-value = 0.03513

shapiro.test(B.Makinesi)

##   
## Shapiro-Wilk normality test  
##   
## data: B.Makinesi  
## W = 0.75, p-value < 2.2e-16

shapiro.test(C.Makinesi)

##   
## Shapiro-Wilk normality test  
##   
## data: C.Makinesi  
## W = 0.77499, p-value = 0.04989

Kod çıktılarında görüldüğü üzere bütün P-Value değerleri, alfa değeri olan 0.05 den küçüktür. Bu da elimizde ki 3 değişkeninde Normal dağılmadığı anlamına gelmektedir.

# Soru 2

İkinci soruda bizden en az 3 bağımlı gruptan oluşan bir veri kümesi bulmamız ve her bir değişkeninin Normal dağılmadığını göstermemiz istenmektedir. Aşağıda kodların ekrana basımı gösterilmektedir.

Bagimli <- read.delim("C:/Users/user/Desktop/NonPar/Bagimli.txt")  
attach(Bagimli)  
print(Bagimli)

## Saat0 Saat24 Saat72  
## 1 2.5 5.9 13.9  
## 2 2.8 8.1 12.6  
## 3 2.0 8.1 14.3  
## 4 5.9 8.6 4.2  
## 5 2.5 8.1 14.5

Veri kümesi, laboratuvar farelerinin besin alma üstünlüğünün ölçülmesinde farklı hareket düzeyleri etkilerinin ölçülmesi için bir araştırmayı temel almaktadır. 5 erkek fare 0, 24 ve 72 saat aç bırakılarak yedikleri besin miktarı gram olarak ölçülmüştür.

summary(Bagimsiz)

## A.Makinesi B.Makinesi C.Makinesi   
## Min. :340.0 Min. :294.0 Min. :263.0   
## 1st Qu.:344.0 1st Qu.:309.5 1st Qu.:340.0   
## Median :371.0 Median :325.0 Median :356.0   
## Mean :384.5 Mean :314.7 Mean :340.2   
## 3rd Qu.:386.0 3rd Qu.:325.0 3rd Qu.:371.0   
## Max. :499.0 Max. :325.0 Max. :371.0   
## NA's :3 NA's :1

Soru 1’de olduğu gibi tanımlayıcı istatistikler yukarıda görülmektedir. Yine burada da medyan değer ve ortalama değerini kıyaslayarak Normal dağılım çıkarsaması yapılabilir fakat bunun için özel olarak geliştirilmiş yöntemleri kullanmak daha mantıklıdır.

Gerekli testleri kullanarak elimizde ki 3 değişkenin Normal dağılmadığını gösterelim.

shapiro.test(Saat0)

##   
## Shapiro-Wilk normality test  
##   
## data: Saat0  
## W = 0.72062, p-value = 0.01555

shapiro.test(Saat24)

##   
## Shapiro-Wilk normality test  
##   
## data: Saat24  
## W = 0.71372, p-value = 0.01326

shapiro.test(Saat72)

##   
## Shapiro-Wilk normality test  
##   
## data: Saat72  
## W = 0.69005, p-value = 0.007514

Kod çıktılarında görüldüğü üzere bütün P-Value değerleri, alfa değeri olan 0.05 den küçüktür. Bu da elimizde ki 3 değişkeninde Normal dağılmadığı anlamına gelmektedir.

# Soru 3

Bu soruda bağımsız veri kümemizden herhangi bir değişken seçilip, o değişken üzerinde İşaret ve Wilcoxon İşaret Sıra Sayıları testlerini uygulamamız istenmektedir.

Seçtiğimiz değişken A Makinesi olsun. A Makinesinin medyan değerinin 350’ye eşit olup olmadığını test edelim.

Bu durumda;

H0 Hipotezi:A Makinesinin medyan değeri 350’ye eşittir.

HS Hipotezi:A Makinesinin medyan değeri 350’ye eşit değildir.

library(DescTools)  
SignTest(A.Makinesi,mu=350)

##   
## One-sample Sign-Test  
##   
## data: A.Makinesi  
## S = 4, number of differences = 6, p-value = 0.6875  
## alternative hypothesis: true median is not equal to 350  
## 96.9 percent confidence interval:  
## 340 499  
## sample estimates:  
## median of the differences   
## 371

Yukarıda ki test işaret testidir. Bu test sonucu çıkan P-Value değeri 0.6875’tir. Yani alfa değeri olan 0.05’ten büyük çıktığı için H0 hipotezi kabul edilir. %95 güven düzeyinde A Makinesinin medyan değerinin 350’ye eşit olduğunu söyleyebiliriz.

Şimdi Wilcoxon testini uygulayalım.

library(stats)  
wilcox.test(A.Makinesi,mu=350)

##   
## Wilcoxon signed rank test with continuity correction  
##   
## data: A.Makinesi  
## V = 16, p-value = 0.2918  
## alternative hypothesis: true location is not equal to 350

Wilcoxon testinde P-Value değeri 0.2918 çıkmıştır. Aynı şekilde alfa değerinden büyük olduğu için H0 hipotezini kabul edebiliriz. İşaret testi ile Wilcoxon testlerinin aynı sonuca ulaştığı görülmektedir.

# Soru 4

Dördüncü soruda bağımsız veriye dair gerekli hipotezleri oluşturduktan sonra seçilen iki grup arasında konum yönünden farklılık olup olmadığını test etmemiz istenmektedir.

Bunun için A Makinesi ve B Makinesini ele alalım.

H0 Hipotezi:A Makinesi ve B Makinesi arasında konum yönünden farklılık yoktur.

HS Hipotezi:A Makinesi ve B Makinesi arasında konum yönünden farklılık vardır.

library(stats)  
wilcox.test(A.Makinesi,B.Makinesi,paired = FALSE)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: A.Makinesi and B.Makinesi  
## W = 18, p-value = 0.02624  
## alternative hypothesis: true location shift is not equal to 0

Çıkan sonuca göre P-Value=0.02624 < alfa=0.05’tir. Dolayısıyla H0 Hipotezi ret edilir. Yani 2 grup arasında konum yönünden farklılık olduğunu %95 güven düzeyinde söyleyebiliriz.

# Soru 5

Beşinci soruda bağımlı veriye dair gerekli hipotezleri oluşturduktan sonra seçilen iki grup arasında konum yönünden farklılık olup olmadığını test etmemiz istenmektedir.

Bunun için 24.Saat ve 72.Saati ele alalım.

H0 Hipotezi:Farelere uygulanan deneyde 24.Saat ile 72.Saat arasında konum yönünden farklılık yoktur.

HS Hipotezi:Farelere uygulanan deneyde 24.Saat ile 72.Saat arasında konum yönünden farklılık vardır.

library(stats)  
wilcox.test(Saat24,Saat72,paired = TRUE)

##   
## Wilcoxon signed rank test  
##   
## data: Saat24 and Saat72  
## V = 1, p-value = 0.125  
## alternative hypothesis: true location shift is not equal to 0

Sonuçta görüldüğü üzere p-value=0.125 > alfa=0.05’dir. Bu durumda H0 Hipotezi kabul edilebilir. Yani farelere uygulanan deneyde 24.Saat ile 72.Saat arasında konum yönünden farklılık olmadığını %95 güven düzeyinde söyleyebiliriz.

# Soru 6

Bu soruda birinci soruda hazırlamış olduğumuz bağımsız veri kümesinde ki değişkenlerin konum yönünden farklı olup olmadığı araştırılmak istenmektedir.

Bunun için Kruskal-Wallis Testini uygulayacağız.

H0 Hipotezi : Kalite kontrolü yapmak amacı ile üretilen makineler arasında fark yoktur.

HS Hipotezi : Kalite kontrolü yapmak amacı ile üretilen makinelerden en az biri farklıdır.

library(stats)  
library(PMCMRplus)  
attach(Bagimsiz)  
kruskal.test(Bagimsiz)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Bagimsiz  
## Kruskal-Wallis chi-squared = 5.5834, df = 2, p-value = 0.06132

Sonuçta görüldüğü üzere P-Value değeri 0.06 > alfa = 0.05 olduğu için H0 Hipotezi kabul edilir yani makieneler arasında fark olmadığını %95 güven düzeyinde söyleyebiliriz. Bu durumda farklılığı yaratan grup olmadığı için Post-Hoc testlerini uygulamaya gerek yoktur.

# Soru 7

Bu soruda ikinci soruda hazırlamış olduğumuz bağımlı veri kümesinde ki değişkenlerin konum yönünden farklı olup olmadığı araştırılmak istenmektedir.

Bunun için Friedman Testini kullanacağız.

R, Friedman testini çalıştırmamız için bir matris kullanmamızı istemektedir.

class(Bagimli)

## [1] "data.frame"

Görüldüğü üzere bizim veri kümemiz bir data frame’dir. Bu durumda bu veri kümesini matris haline getirmemiz ve testi bu şekilde yapmamız gerekmektedir. Aşağıda bunun nasıl yapıldığı ve daha sonrasında Friedman testinin sonucu görülmektedir.

H0 Hipotezi : Farelerin hareket düzeylerinde zamanın bir etkisi yoktur.

HS Hipotezi : Farelerin hareket düzeylerinde zamanın etkisi vardır.

attach(Bagimli)  
forFriedman=matrix(c(Bagimli[,1],Bagimli[,2],Bagimli[,3]),nrow = 5,  
 dimnames = list(1:5,c("0.Saat","24.Saat","72.Saat")))  
friedman.test(forFriedman)

##   
## Friedman rank sum test  
##   
## data: forFriedman  
## Friedman chi-squared = 5.2, df = 2, p-value = 0.07427

Sonuçta görüldüğü üzere P-Value değerimiz 0.07 > alfa=0.05 yani H0 Hipotezimiz kabul edilir. Farelerin hareket düzeylerinde zamanın bir etkisinin olmadığını söyleyebiliriz. Bu durumda farklılık yaratan grubu bulmamıza gerek yoktur.