

## BAĞIMSIZ VERİ:

### AİR QUALITY- New York'ta Mayıs-Eylül 1973'te günlük hava kalitesi ölçümleri:

Veri setinde yer alan değerler 1 Mayıs 1973 ile 30 Eylül 1973 zaman diliminde gerçekleşmektedir.

Araştırmanın amacı hava kalitesini arttırarak yaşanılabilir, ferah bir hava sağlamaktır. Böylelikle hava kalitesine zarar veren etkenler ince partiküler madde gibi akciğerlere girmesi , kan dolaşımına girmesi ve kalp hastalığına neden olması engellenebilir. Akciğer kanserine neden olması engellenmektedir. (astım ve akut alt solunum yolu enfeksiyonları) . Hava kalitesini düzelterek gerçekleştirebilecek ölüm oranını azaltabiliriz. Gözlem sayısı 8 olmak üzere 3 değişken ele alınmıştır. Bunlar solar, rüzgar ve sıcaklıktır. Değişkenlerin hava kirliliği üzerinde farklılığının olup olmadığını test edelim.

solarR	7	8	13	20	334	323	322	320
Rüzgar	2.3	2.8	3.4	4.6	18.4	16.6	15.5	15.5
SıcaklıkF	57	58	59	59	97	96	94	94

SolarR: Integer Numeric değişkendir.

Rüzgar: Double Numeric değişkendir.

Sıcaklık\_f: Integer Numeric değişkendir.

### Veri Seti İmport;

```
> veri1 <- read.csv("C:/veri1.txt", sep="")
> view(veri1)
```

Veri setini yukarıda verilen kod ile import ettim.

### Veri Seti Değişkenleri Hakkında İstatistiksel Bilgiler;

```
> summary (veri1)
      solarR      Rüzgar      SıcaklıkF
Min.   : 7.00   Min.   : 2.300   Min.   :57.00
1st Qu.:11.75   1st Qu.: 3.250   1st Qu.:58.75
Median :170.00   Median :10.050   Median :76.50
Mean    :168.38   Mean    : 9.887   Mean    :76.75
3rd Qu.:322.25   3rd Qu.:15.775   3rd Qu.:94.50
Max.    :334.00   Max.    :18.400   Max.    :97.00
```

Summary kodu değişkenlerimizin özetleyici istatistiklerini bulmamızda bize yardımcı olur. Örneğin rüzgar değişkeni için minimum değer 2.30, maksimum değer 18.40'dır. Ortalaması 9.887, medyan değeri ise 10.050'dir. Bununla birlikte 1.çeyreklik değeri 3.25, 3.çeyreklik değeri ise 15.775 olarak bulunmuştur.

## 1) Her Bir Grup İçin Veri Setinin Normal Dağılımını Test Edelim;

```
attach(veri1)

shapiro.test(solarR)

shapiro.test(Rüzgar)

shapiro.test(SıcaklıkF)
```

Attach kodunu kullanarak değişkenlerimizi tanımladık. Shapiro test ile de değişkenlerimizin normalliğini test ettik.

Hipotezlerimiz;

- $H_0$  : Değişkenlerin dağılımı ile normal dağılım arasında fark yoktur.
- $H_s$  : Değişkenlerin dağılımı ile normal dağılım arasında fark vardır.

```
> attach(veri1)
> shapiro.test(solarR)

      shapiro-wilk normality test

data:  solarR
W = 0.69352, p-value = 0.001902
```

SolarR değişkenimizin p-value değeri = 0.001902 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. Yani değişkenimizin normal dağılmadığını %95 güven düzeyinde söyleyebiliriz.

```
> shapiro.test(Rüzgar)

      shapiro-wilk normality test

data:  Rüzgar
W = 0.79141, p-value = 0.02315
```

Rüzgar değişkenimizin p-value değeri = 0.02315 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. Değişkenimizin normal dağılmadığını %5 anlamlılık düzeyinde söyleyebiliriz.

```
> shapiro.test(SıcaklıkF)

      shapiro-wilk normality test

data:  SıcaklıkF
W = 0.71518, p-value = 0.00334
```

SıcaklıkF değişkenimizin p-value değeri = 0.00334 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. Değişkenimizin normal dağılmadığını %95 güven düzeyinde söyleyebiliriz.

## 2) İşaret Testi ve Wilcoxon İşaret Sıra Sayıları Testi;

```
install.packages("DescTools")  
library(DescTools)  
solarR -> C(7, 8, 13, 20, 334, 323, 322, 320 )  
SignTest(x=solarR,mu=15,alternative="greater")  
  
install.packages("stats")  
library(stats)  
Rüzgar = c(2.3,2.8,3.4,4.6,18.4,16.6,15.5,15.5)  
wilcox.test(Rüzgar,mu=3.5,alternative="two.sided")
```

Kodlarını kullanarak önce kütüphaneden “DescTools” ve “stats” paketlerini indirdik. Bu sayede işaret testini ve wilcoxon işaret sıra sayıları testini test edebilir hale geldik.

Hipotezlerimiz;

Ho :  $\theta = 15$

Hs :  $\theta > 15$

## 3) Mann Whitney Testi

```
wilcox.test(veri1$Rüzgar,veri1$SıcaklıkF)  
wilcox.test(veri1$solarR,veri1$Rüzgar)
```

Önce solarR ve sıcaklığı daha sonra solar ve rüzgarı Mann Whitney testini uygulamak için seçtik. Wilcox.test kodu ile de Mann Whitney testini uyguladık.

Hipotezlerimiz;

- Ho : Rüzgar ve sıcaklık değişkenleri arasında konum yönünden farklılık yoktur.  
Hs : Rüzgar ve sıcaklık değişkenleri arasında konum yönünden farklılık vardır.
- Ho : SolarR ve rüzgar değişkenleri arasında konum yönünden farklılık yoktur.  
Hs : SolarR ve rüzgar değişkenleri arasında konum yönünden farklılık vardır.

```
> #MANN-WHITNEY#
> wilcox.test(veri1$Rüzgar, veri1$SıcaklıkF)

      wilcoxon rank sum test with continuity correction

data:  veri1$Rüzgar and veri1$SıcaklıkF
W = 0, p-value = 0.0009148
alternative hypothesis: true location shift is not equal to 0
```

Çıktımızın p – value değeri = 0.0009148 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. Rüzgar ve sıcaklıkF değişkenimizin konum yönünden farklı olduğunu 0.05 anlamlılık düzeyinde söyleyebiliriz.

```
> wilcox.test(veri1$solarR, veri1$Rüzgar)

      wilcoxon rank sum test with continuity correction

data:  veri1$solarR and veri1$Rüzgar
W = 52, p-value = 0.04042
alternative hypothesis: true location shift is not equal to 0
```

Çıktımızın p – value değeri = 0.04042 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. SolarR ve rüzgar değişkenlerimiz arasında konu yönünden farklılık olduğunu %95 güven düzeyinde söyleyebiliriz.

#### 4) Kruskal Wallis Testi

```
attach(yeniveri1)
kruskal.test(veri1)
```

Kodlarını kullanarak testimizi gerçekleştiriyoruz. Burada bir diğer kodla da aynı sonuçların çıktığını göreceğiz.

Hipotezlerimiz;

- $H_0$  : Gruplar arasında konum yönünden farklılık yoktur.  
 $H_s$  : Gruplar arasında konum yönünden farklılık vardır.

```
> kruskal.test(veri1)

      kruskal-wallis rank sum test

data:  veri1
kruskal-wallis chi-squared = 10.333, df = 2, p-value = 0.005703
```

Çıktımda görüldüğü gibi p – value değeri = 0.005 <  $\alpha$  = 0.05 olduğu için Ho hipotezi reddedilir. Yani gruplar arasında konum yönünden farklılığın olduğunu %95 güven düzeyinde söyleyebiliriz. Farklılığı yaratan grubu bulmak için Post-Hoc testi uygulayalım.

	olcum	tur
1	7.0	1
2	8.0	1
3	13.0	1
4	20.0	1
5	334.0	1
6	323.0	1
7	322.0	1
8	320.0	1
9	2.3	2
10	2.8	2
11	3.4	2
12	4.6	2
13	18.4	2
14	16.6	2
15	15.5	2
16	15.5	2
17	57.0	3
18	58.0	3
19	59.0	3
20	59.0	3
21	97.0	3
22	96.0	3
23	94.0	3

Showing 1 to 23 of 24 entries

Veri girişini yanda verildiği gibi yapıyoruz. Böylelikle kodumuz

```
kruskal.test(olcum~tur,data=yeniveri1)
```

şeklinde oluyor.

## BAĞIMLI VERİ:

### BAKTERİ

Bir biyolog bir bakteri türünün belli bir zaman aralığında 3 farklı ısı düzeyinde çoğalma hızını (bakteri sayısı/1000) araştırmak istemektedir. Bunun için 25,45 ve 65 sıcaklıklarında 5'er ölçüm yapılması planlanmıştır. Bu deney ile amaçlanan bakterilerin bu sıcaklıklardaki ortalama üreme sayıları arasında farklılık olup olmadığını araştıralım.

25 derece	45 derece	60 derece
1	1	4
7	2	3
1	5	11
2	3	13
7	16	15

### Veri Seti İmport;

```
> veri2 <- read.delim("C:/veri2.txt")
> view(veri2)
```

Veri setini yukarıda verilen kod ile import ettim.

### Veri Seti Değişkenleri Hakkında İstatistiksel Bilgiler;

```
> attach(veri2)
> summary (veri2)
  x25.derece  x45.derece  x60.derece
Min.   :1.0   Min.   : 1.0   Min.   : 3.0
1st Qu.:1.0   1st Qu.: 2.0   1st Qu.: 4.0
Median :2.0   Median : 3.0   Median :11.0
Mean   :3.6   Mean   : 5.4   Mean   : 9.2
3rd Qu.:7.0   3rd Qu.: 5.0   3rd Qu.:13.0
Max.   :7.0   Max.   :16.0   Max.   :15.0
```

Summary kodu değişkenlerimizin özetleyici istatistiklerini bulmamızda bize yardımcı olur. Örneğin X60.derece değişkeni için minimum değer 3.0, maksimum değer 15.0'dır. Ortalaması 9.2, medyan değeri ise 11.0'dır. Bununla birlikte 1.çeyreklik değeri 4.0, 3.çeyreklik değeri ise 13.0 olarak bulunmuştur.

## 1) Her Bir Grup İçin Veri Setinin Normal Dağılımını Test Edelim;

```
shapiro.test(X25.derece)

shapiro.test(X45.derece)

shapiro.test(X60.derece)
```

Attach kodunu kullanarak değişkenlerimizi tanımladık. Shapiro test ile de değişkenlerimizin normalliğini test ettik.

- $H_0$  : Değişkenlerin dağılımı ile normal dağılım arasında fark yoktur.  
 $H_a$  : Değişkenlerin dağılımı ile normal dağılım arasında fark vardır.

```
> shapiro.test(X25.derece)

      shapiro-wilk normality test

data:  X25.derece
W = 0.75383, p-value = 0.03228

> shapiro.test(X45.derece)

      shapiro-wilk normality test

data:  X45.derece
W = 0.76645, p-value = 0.04197

> shapiro.test(X60.derece)

      shapiro-wilk normality test

data:  X60.derece
W = 0.88168, p-value = 0.317
```

Çıktılarda görüldüğü gibi X25.derece değişkenimizin p – value değeri = 0.03228 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. X25.derece değişkeninin dağılımının normal dağılmadığını 0.05 anlamlılık düzeyinde söyleyebiliriz.

Aynı şekilde X45.derece değişkenimizin p – value değeri = 0.04197 <  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi reddedilir. X45.derece değişkeninin dağılımının normal dağılmadığını 0.05 anlamlılık düzeyinde söyleyebiliriz.

Son olarak X60.derece değişkenimizin p – value değeri = 0.317 >  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi kabul edilir. X60.derece değişkeninin dağılımının normal dağılım gösterdiğini 0.05 anlamlılık düzeyinde söyleyebiliriz.

## 2) İki Örneklem Wilcoxon İşaret Sıra Sayıları Testi;

```
library(stats)

wilcox.test(X25.derece,X45.derece)

wilcox.test(X45.derece,X60.derece)
```

Kodlarını kullanarak iki örneklem wilcoxon işaret sıra sayıları testini uyguladık.

Hipotezlerimiz;

- $H_0$  : Gruplar arasında konum yönünden farklılık yoktur.  
 $H_s$  : Gruplar arasında konum yönünden farklılık vardır.

```
> wilcox.test(X25.derece,X45.derece)

      wilcoxon rank sum test with continuity correction

data:  X25.derece and X45.derece
W = 10.5, p-value = 0.7496
alternative hypothesis: true location shift is not equal to 0
```

Çıktıda görüldüğü gibi  $p - \text{value değeri} = 0.7496 > \alpha = 0.05$  olduğu için  $H_0$  hipotezi kabul edilir. Yani gruplar arasında (X25.derece ve X45.derece) konum yönünden farklılığın olmadığını %95 güven düzeyinde söyleyebiliriz.

```
> wilcox.test(X45.derece,X60.derece)

      wilcoxon rank sum test with continuity correction

data:  X45.derece and X60.derece
W = 7.5, p-value = 0.3457
alternative hypothesis: true location shift is not equal to 0
```

Çıktıda görüldüğü gibi  $p - \text{value değeri} = 0.3457 > \alpha = 0.05$  olduğu için  $H_0$  hipotezi kabul edilir. Yani gruplar arasında (X60.derece ve X45.derece) konum yönünden farklılığın olmadığını %95 güven düzeyinde söyleyebiliriz.



### 3) Friedman Testi ;

```
friedman.test(veri~derece|blok,data=yeniveri)
```

Kodunu kullanarak testimizi uyguluyoruz.

Diğer testlerden farklı olarak burada veri girişimizde farklılıklar yapıyoruz. Yeniveri olarak veri seti oluşturduk.

	derece	veri	blok
1	1	1	1
2	2	1	1
3	3	4	1
4	1	7	2
5	2	2	2
6	3	3	2
7	1	1	3
8	2	5	3
9	3	11	3
10	1	2	4
11	2	3	4
12	3	13	4
13	1	7	5
14	2	16	5
15	3	15	5

Hipotezlerimiz;

- $H_0$  : Yayılım yönünden gruplar arasında farklılık yoktur.  
 $H_a$  : Yayılım yönünden gruplar arasında farklılık vardır.

```
> friedman.test(veri~derece|blok,data=yeniveri)

Friedman rank sum test

data:  veri and derece and blok
Friedman chi-squared = 3.2632, df = 2, p-value = 0.1956
```

Çıktıda görüldüğü gibi p – value değeri = 0.1956 >  $\alpha$  = 0.05 olduğu için  $H_0$  hipotezi kabul edilir. Yani veri kümesinde yayılım yönünden gruplar arasında fark olmadığını 0.05 anlamlılık düzeyinde söyleyebiliriz. Post-Hoc testi uygulamamıza gerek yoktur.