clase_4.R

hugop

2023-11-24

```
# Hugo Vazquez
# 2176696
# 19/09/2023

# Importar datos de archivo excel a la consola de R
# Funcion "read.csv"
setwd("C:/UANL_FCF/REPOSITORIOS/Exp_Met_Est_AD2023")


# MAGT ----------------------------------------------------------------------

#Correlacion

library(repmis)

erupciones <- source_data("https://www.dropbox.com/s/liir6sil7hkqlxs/erupciones.csv?dl=0")
```

```
## Downloading data from: https://www.dropbox.com/s/liir6sil7hkqlxs/erupciones.csv?dl=0
```

```
## SHA-1 hash of the downloaded data file is:
## 40a756eef34863f4a2fb08351702e13c6764e3a6
```

```
## Warning in fread(data, sep = sep, header = header, data.table = F,
## stringsAsFactors = stringsAsFactors, : Detected 6 column names but the data has
## 4 columns. Filling rows automatically. Set fill=TRUE explicitly to avoid this
## warning.
```

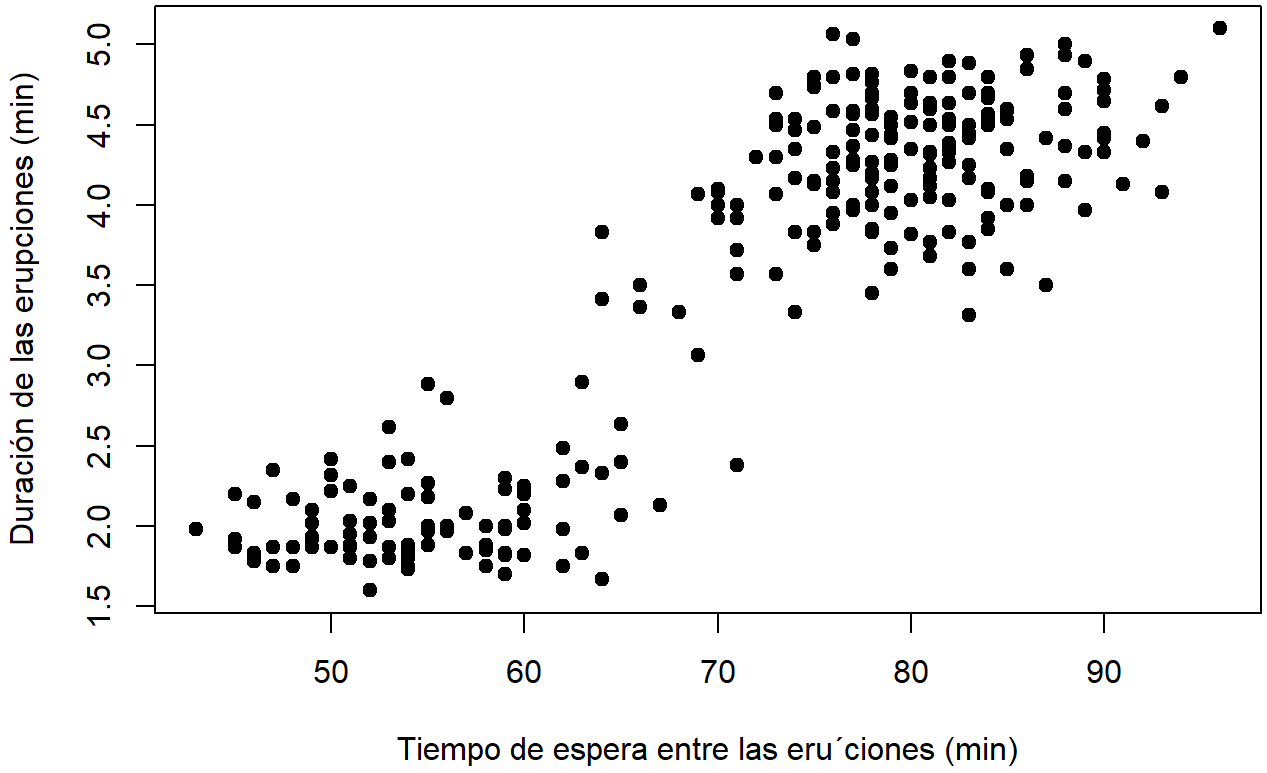
```
## Warning in fread(data, sep = sep, header = header, data.table = F,
## stringsAsFactors = stringsAsFactors, : Stopped early on line 34. Expected 6
## fields but found 7. Consider fill=TRUE and comment.char=. First discarded
## non-empty line: <<<</script><title>Dropbox - erupciones.csv - Simplify your
## life</title>>>
```

```
erupciones <- source_data("https://www.dropbox.com/s/liir6sil7hkqlxs/erupciones.csv?dl=1")
```

```
## Downloading data from: https://www.dropbox.com/s/liir6sil7hkqlxs/erupciones.csv?dl=1
```

```
## SHA-1 hash of the downloaded data file is:
## b07708389ddf62ee20d19c759c88d7dc2d0da3ac
```

```
plot(erupciones$waiting, erupciones$eruptions, xlab = "Tiempo de espera entre las eru´ciones (mi
n)", ylab = "Duración de las erupciones (min)", pch = 19)
```




```
# Estadística descriptiva Erupciones

median(erupciones$eruptions)
```

```
## [1] 4
```

```
sd(erupciones$eruptions)
```

```
## [1] 1.141371
```

```
var(erupciones$eruptions)
```

```
## [1] 1.302728
```

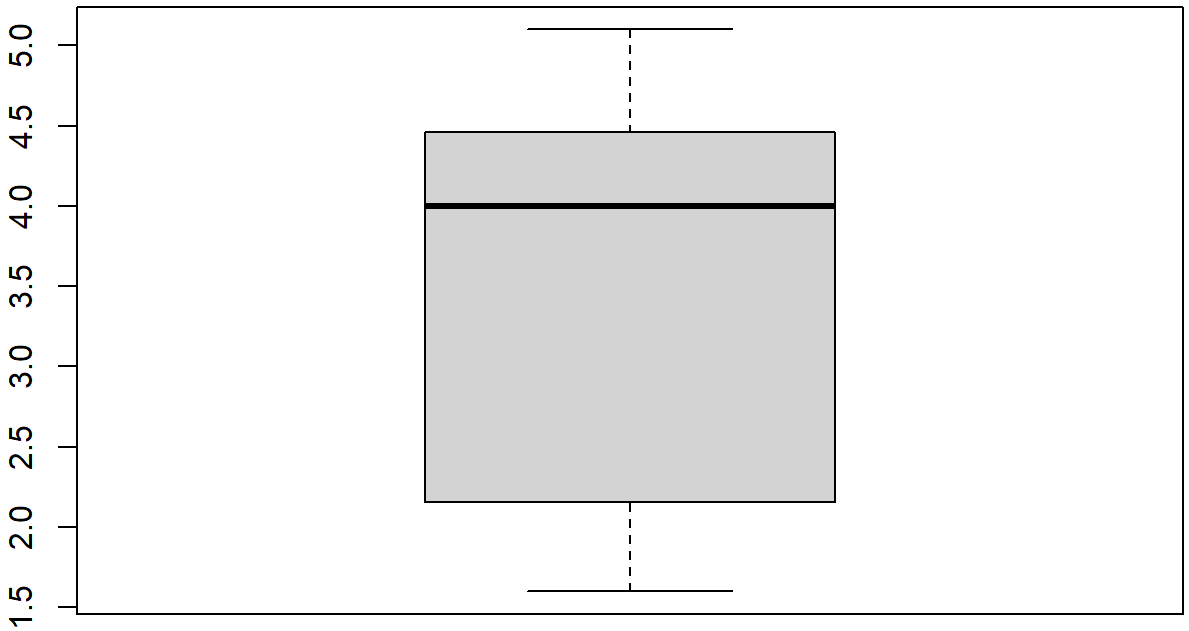
```
mean(erupciones$eruptions)
```

```
## [1] 3.487783
```

```
range(erupciones$eruptions)
```

```
## [1] 1.6 5.1
```

```
boxplot(erupciones$eruptions)
```




```
# Estadística descriptiva de waiting
median(erupciones$waiting)
```

```
## [1] 76
```

```
sd(erupciones$waiting)
```

```
## [1] 13.59497
```

```
var(erupciones$waiting)
```

```
## [1] 184.8233
```

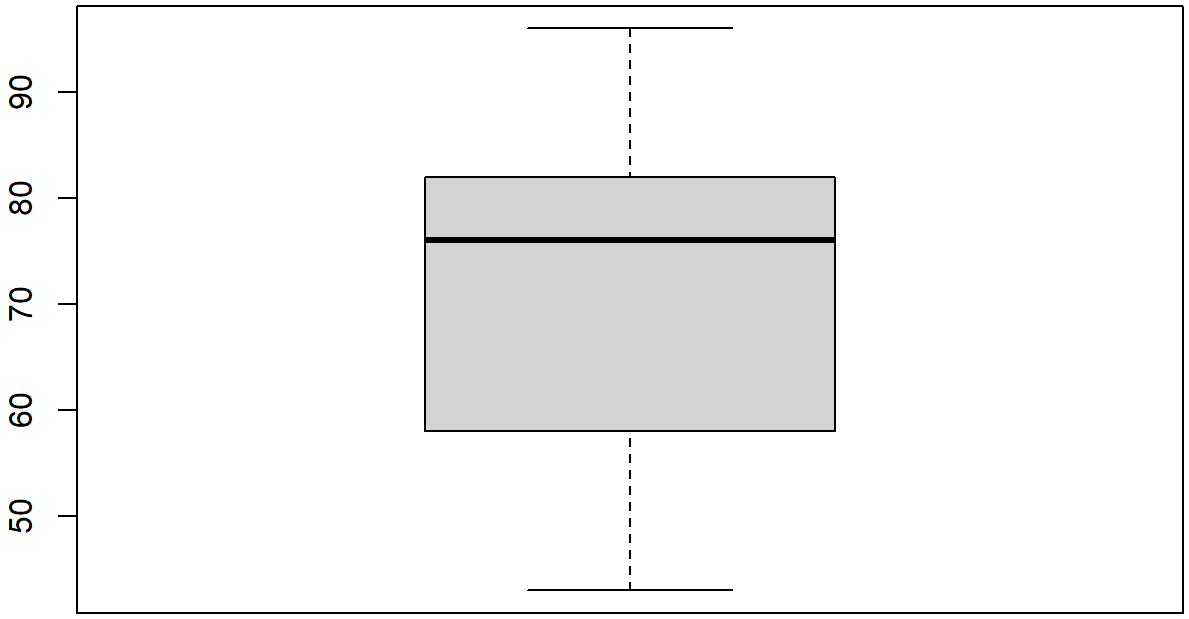
```
mean(erupciones$waiting)
```

```
## [1] 70.89706
```

```
range(erupciones$waiting)
```

```
## [1] 43 96
```

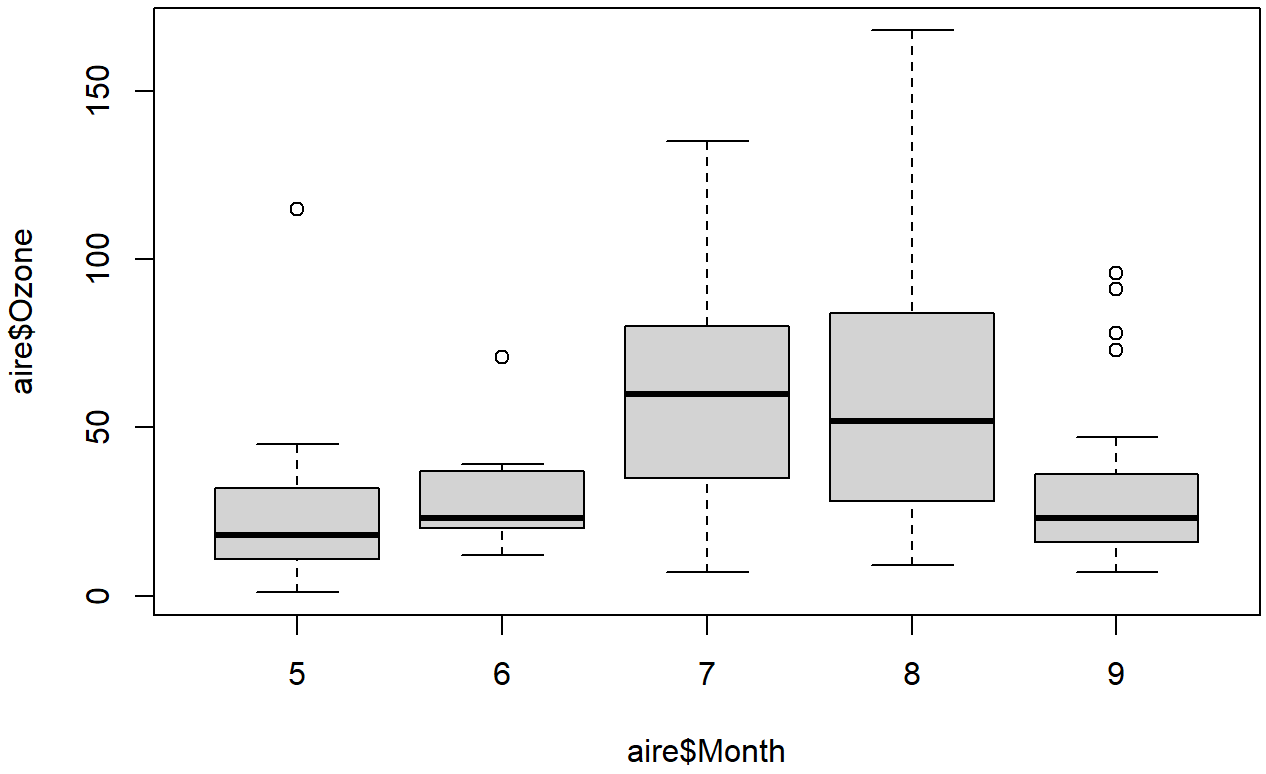
```
boxplot(erupciones$waiting)
```



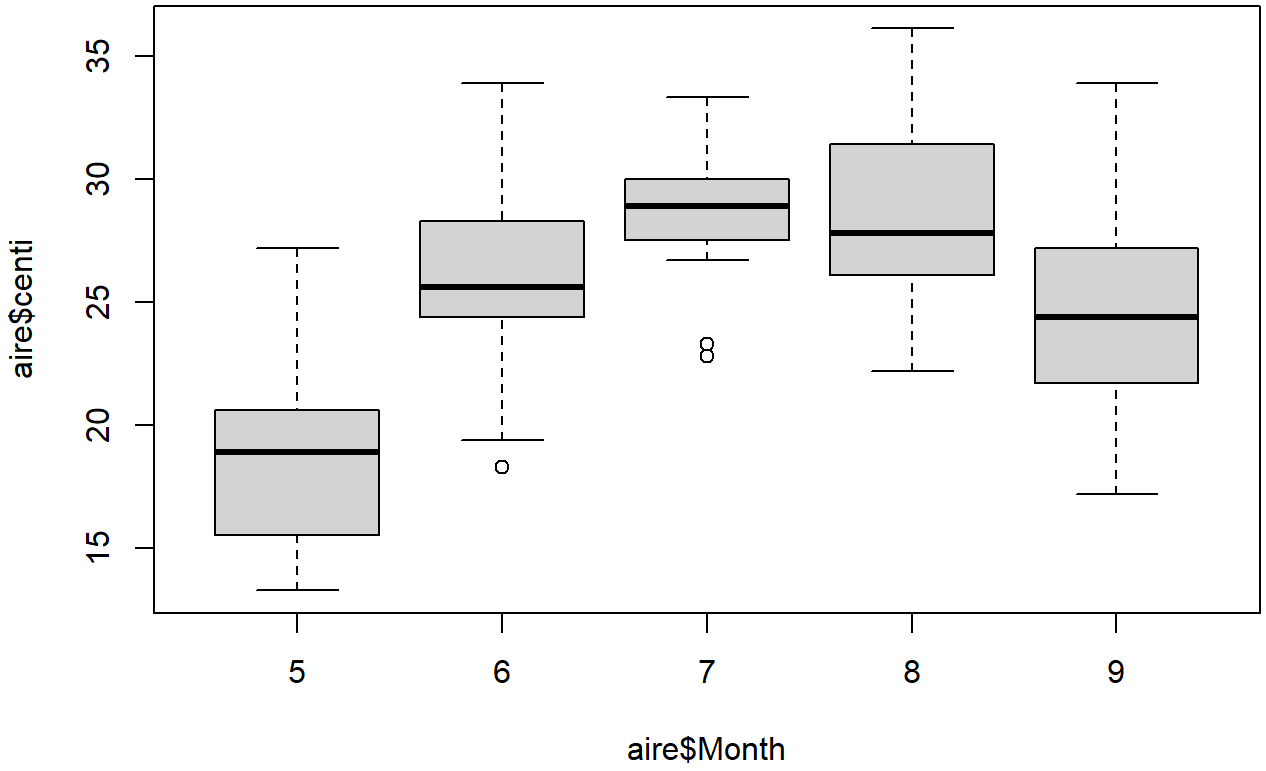

```
# Establecer H0 y H1
# Función de correlación cor.test

cor.test (erupciones$eruptions, erupciones$waiting)
```

```
##
##  Pearson's product-moment correlation
##
## data:  erupciones$eruptions and erupciones$waiting
## t = 34.089, df = 270, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8756964 0.9210652
## sample estimates:
##       cor
## 0.9008112
```

```
#  Prueba de t dependients ------------------------------------------------


# Datos de airquality de la ciudad de NY, USA
# Comparar las variables en dos periodos de verano (junio)
# Otoño (Septiembre)

aire<- airquality
boxplot(aire$Ozone ~ aire$Month)
```



```
aire$centi <- (aire$Temp - 32) / 1.8
aire$centi <- round((aire$Temp - 32)/1.8,1)
boxplot(aire$centi ~ aire$Month)
```




```
# Crear un subconjunto solo con los meses de Junio y Sept

aire.junio <- subset(aire, Month == "6")
aire.sep <- subset(aire, Month == "9")


t.test(aire.junio$Wind, aire.sep$Wind, paired = T)
```

```
##
##  Paired t-test
##
## data:  aire.junio$Wind and aire.sep$Wind
## t = 0.094506, df = 29, p-value = 0.9254
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -1.788913  1.962246
## sample estimates:
## mean difference
##      0.08666667
```

```
boxplot(aire$Wind ~ aire$Month)
```