EXAMEN_ANGELICA_TORRES.R

acile

2023-03-31

```
                  #ANGELICA TORRES GARCÍA
                     #31/03/2023
                  #MATRICULA 2173388

library(repmis)
suelo<- source_data("https://www.dropbox.com/s/3pi3huovq6qce42/obs.csv?dl=1")
```

```
## Downloading data from: https://www.dropbox.com/s/3pi3huovq6qce42/obs.csv?dl=1
```

```
## SHA-1 hash of the downloaded data file is:
## a88edff139da590ccb918ba2fd00b18d2d839509
```

```
suelo$zone <- factor(suelo$zone)
as.factor(suelo$zone)
```

```
##   [1] 2 2 1 1 2 1 1 2 2 1 2 2 1 1 1 2 2 2 3 3 3 4 4 3 3 3 3 3 3 3 2 2 4 4 4 4 4
##  [38] 4 4 4 4 4 4 3 2 3 3 2 4 3 3 3 3 2 3 3 3 3 3 3 2 2 4 4 4 4 3 3 3 3 3 2 3 3
##  [75] 3 3 3 3 3 3 3 4 2 2 2 3 3 3 2 4 4 3 3 3 3 3 2 3 3 3 3 4 3 4 2 2 3 3 2 3 4
## [112] 4 2 4 3 3 3 3 4 2 2 2 4 4 4 4 4 2 2 3 3 4 4 4 4 2 2 2 2 3 3 2 2 2 2 3 3
## Levels: 1 2 3 4
```

```
suelo$wrb1 <- factor(suelo$wrb1)
as.factor(suelo$wrb1)
```

```
##   [1] 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 3 3 3 3 3 3 1 1 3 3 1 1 3 3 3 3 3 3 1 1 1 1 1
##  [38] 1 1 1 1 1 1 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 1 1 3 3 3 3 3 3 3 3
##  [75] 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 1 1 3 3 3 3 3 3 3 3 3 1 1 1 3 1 3 3 3 3 3 3 1
## [112] 1 3 1 3 3 3 3 1 3 3 3 3 1 1 1 1 1 3 3 3 3 1 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3
## Levels: 1 2 3
```

```
# ACTIVIDAD_1 ------------------------------------------------------------

summary(suelo$Clay1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   21.00   30.00   31.27   39.00   72.00
```

```
summary(suelo$Clay2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.00   27.00   36.00   36.75   47.00   75.00
```

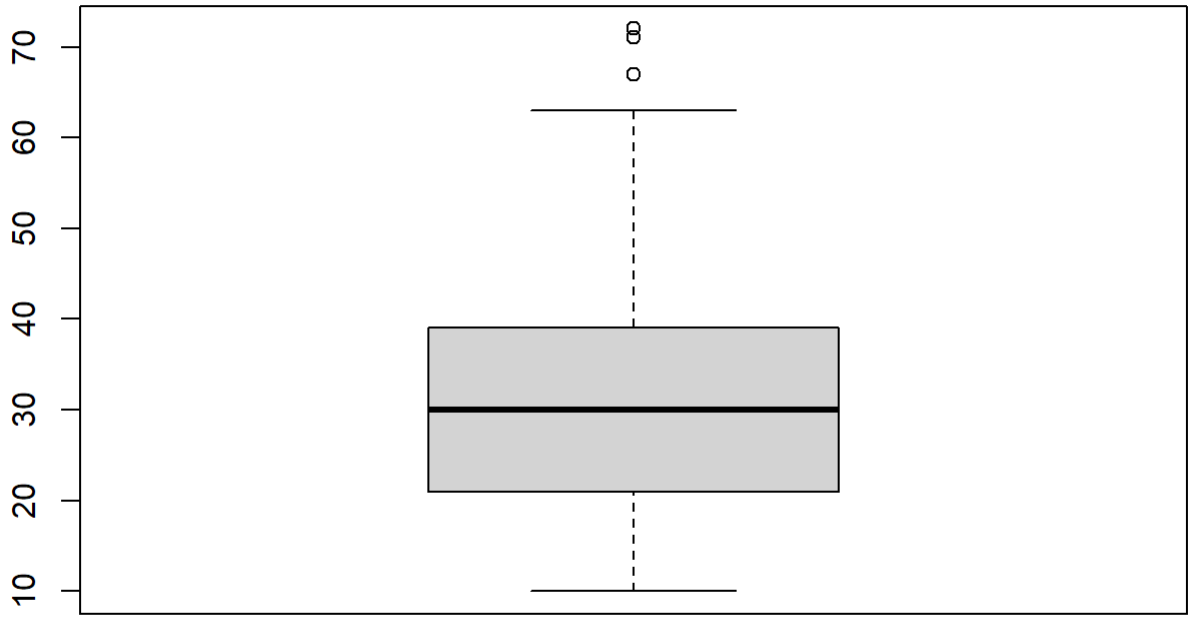
```
summary(suelo$Clay5)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.00   36.50   44.00   44.68   54.00   80.00
```

```
# ACTIVIDAD_2 ------------------------------------------------------------

stem(suelo$Clay1)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   1 | 000222233333444
##   1 | 55555567788889999
##   2 | 000011112222233344444
##   2 | 555555555566788999
##   3 | 00000001122223333333444 4
##   3 | 556666677889999
##   4 | 022233334
##   4 | 55555667899
##   5 | 02334
##   5 | 55689
##   6 | 123
##   6 | 7
##   7 | 12
```

```
#p2 con sesgo

# ACTIVIDAD_3 ------------------------------------------------------------

boxplot(suelo$Clay1)
```




```
#P3 si existen.

#P4

which(suelo$Clay1 > 65)
```

```
## [1]   1   2 106
```

```
# ACTIVIDAD_4 ------------------------------------------------------------

mean(suelo$Clay1)
```

```
## [1] 31.27211
```

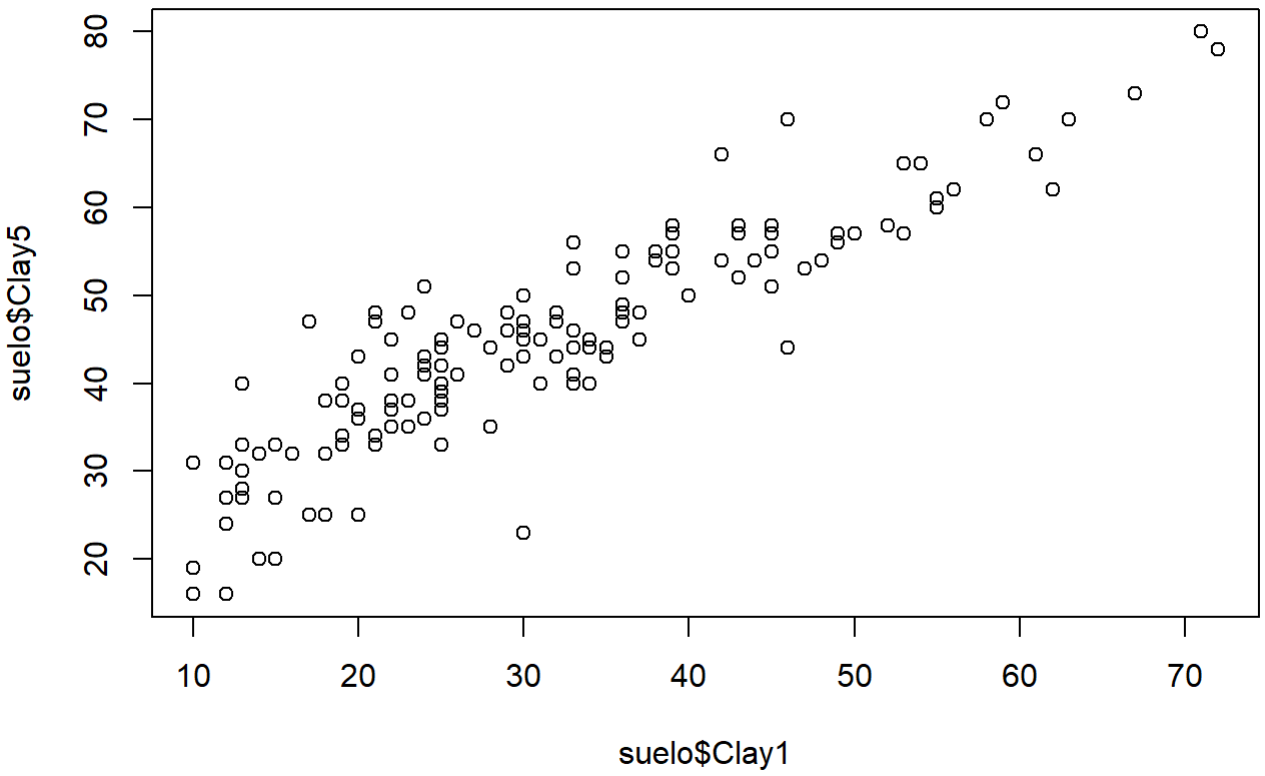
```
#P5

#el valor de p nos indica quen si hay una diferencia significativa. las muestras de clay 1 tiene
n una distribucion normal.

t.test(suelo$Clay1,mu=30)
```

```
##
##  One Sample t-test
##
## data:  suelo$Clay1
## t = 1.1067, df = 146, p-value = 0.2702
## alternative hypothesis: true mean is not equal to 30
## 95 percent confidence interval:
##  29.00045 33.54377
## sample estimates:
## mean of x
##  31.27211
```

```
# ACTIVIDAD_5 ------------------------------------------------------------

#P6

cor.test(suelo$Clay1, suelo$Clay5)
```

```
##
##  Pearson's product-moment correlation
##
## data:  suelo$Clay1 and suelo$Clay5
## t = 24.544, df = 145, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8610227 0.9251946
## sample estimates:
##       cor
## 0.8977721
```

```
#relacion positiva

#P7
# Si es estadisticamente significativa


# ACTIVIDAD_6 ------------------------------------------------------------

#P8

# si es posible

#P9

plot(suelo$Clay5~suelo$Clay1)
```




```
lm(suelo$Clay5~suelo$Clay1)
```

```
##
## Call:
## lm(formula = suelo$Clay5 ~ suelo$Clay1)
##
## Coefficients:
## (Intercept)  suelo$Clay1
##     18.7586       0.8289
```

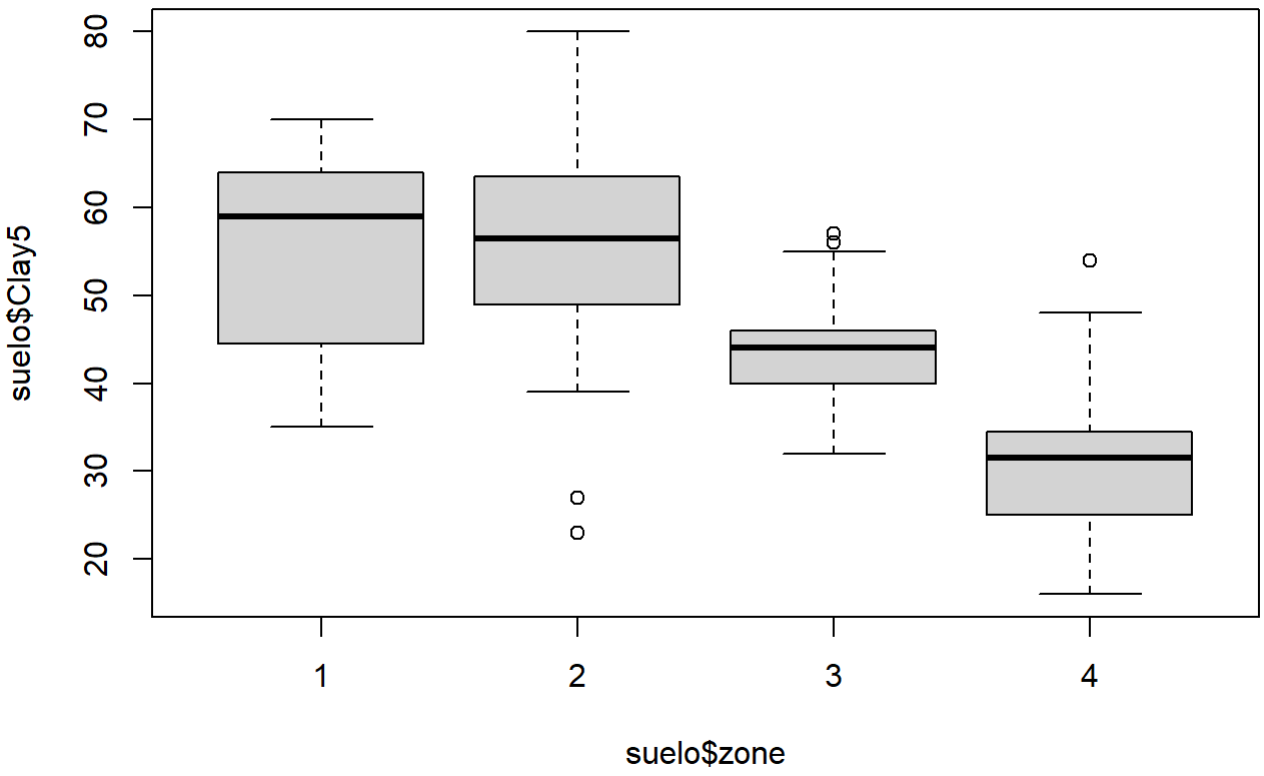
```
regresion <- lm(suelo$Clay5~suelo$Clay1)
regresion
```

```
##
## Call:
## lm(formula = suelo$Clay5 ~ suelo$Clay1)
##
## Coefficients:
## (Intercept)  suelo$Clay1
##     18.7586       0.8289
```

```r
#P10

# si son significativos

summary(regresion)
```

```
##
## Call:
## lm(formula = suelo$Clay5 ~ suelo$Clay1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.6258  -3.1907   0.0055   3.3875  14.1500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.75856    1.15561   16.23   <2e-16 ***
## suelo$Clay1  0.82891    0.03377   24.54   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.687 on 145 degrees of freedom
## Multiple R-squared:  0.806,  Adjusted R-squared:  0.8047
## F-statistic: 602.4 on 1 and 145 DF,  p-value: < 2.2e-16
```

```r
#P11

#cuando R cuadrada ajustada esta mas cercano a 1 es un modelo que predice correctamente.

# ACTIVIDAD_7 ------------------------------------------------------------

#P12

# si existe

#P13

boxplot(suelo$Clay5 ~ suelo$zone)
```




```
#P13
# son diferentes las cuatro zonas

#P14

by(suelo$Clay5, suelo$zone, summary)
```

```
## suelo$zone: 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   35.00   49.25   59.00   55.00   63.00   70.00
## ------------------------------------------------------------
## suelo$zone: 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.00   49.50   56.50   55.95   62.75   80.00
## ------------------------------------------------------------
## suelo$zone: 3
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   32.00   40.00   44.00   43.84   46.00   57.00
## ------------------------------------------------------------
## suelo$zone: 4
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.00   25.00   31.50   31.33   34.25   54.00
```

```
#Las medianas van disminuyendo conforme cambian las zonas de la 1 a la 4

# ACTIVIDAD_8 ------------------------------------------------------------

#P15

varianza <- aov(suelo$Clay5~suelo$zone)
varianza
```

```
## Call:
##    aov(formula = suelo$Clay5 ~ suelo$zone)
##
## Terms:
##                 suelo$zone Residuals
## Sum of Squares    12389.66  11782.31
## Deg. of Freedom          3       143
##
## Residual standard error: 9.077103
## Estimated effects may be unbalanced
```

```
summary(varianza)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## suelo$zone    3  12390    4130   50.12 <2e-16 ***
## Residuals   143  11782      82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# si existen diferencias significativas ya que P nos da menor a 0.05.

#P16

#el valor de P nos indica que existe diferencia significativa en las 4 zonas
#y las zonas 3 y 4 son estadisticamente diferentes entre si.
```