

Examen-Analisis-Estadisticos.R

FCF

2022-10-04

```
#Diego Axayacatl Gonzalez Cuellar
#03/10/22
#Examen Practico MCF

#Carga de datos

TCP <- "https://www.dropbox.com/s/3pi3huovq6qce42/obs.csv?dl=1"

suelo <- read.csv(TCP, header= TRUE, fileEncoding = "latin1")
head(suelo)

##      X          e          n elev zone wrb1 LC Clay1 Clay2 Clay5 CEC1 CEC2 CEC5  OC1 OC2
## 1 1 702638 326959 657      2      3 FF      72      74      78 13.6 10.1 7.1 5.50 3.1
## 2 2 701659 326772 628      2      3 FF      71      75      80 12.6 8.2 7.4 3.20 1.7
## 3 3 703488 322133 840      1      3 FV      61      59      66 21.7 10.2 6.6 6.98 2.4
## 4 4 703421 322508 707      1      3 FV      55      62      61 11.6 8.4 8.0 3.19 1.5
## 5 5 703358 322846 670      2      3 FV      47      56      53 14.9 9.2 8.5 4.40 1.2
## 6 6 702334 324551 780      1      3 FV      49      53      57 18.2 11.6 6.2 5.31 3.2
##      OC5
## 1 1.50
## 2 1.00
## 3 1.30
## 4 1.26
## 5 0.80
## 6 1.08

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

# 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
```

```
#P1: ¿Cual es la tendencia del contenido de Arcilla (Clay) con respecto a la  
#profundidad? La tendencia del contenido de arcilla es aumentar en dependencia  
#del aumento en la profundidad de donde se tomaron las muestras.
```

```
# Actividad 2 -----
```

```
stem(suelo$Clay1)
```

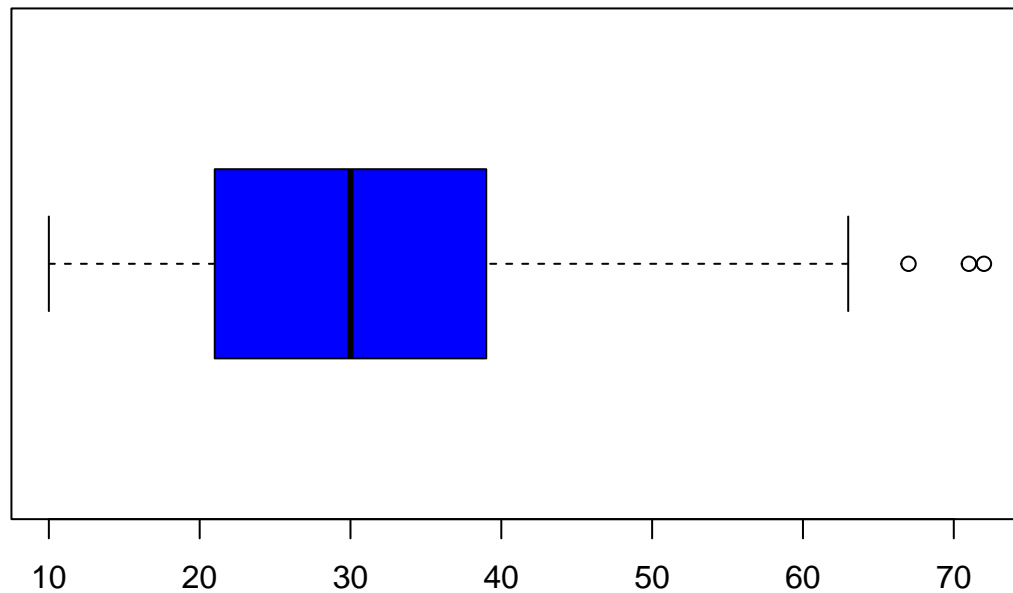
```
##  
## The decimal point is 1 digit(s) to the right of the |  
##  
## 1 | 000222233333444  
## 1 | 55555567788889999  
## 2 | 00001111222223334444  
## 2 | 555555555566788999  
## 3 | 000000011222233333334444  
## 3 | 556666677889999  
## 4 | 022233334  
## 4 | 55555667899  
## 5 | 02334  
## 5 | 55689  
## 6 | 123  
## 6 | 7  
## 7 | 12
```

```
#P2: ¿Los datos de contenido de arcilla siguen una distribucion simetrica  
#o con sesgo? Los datos cuentan con un sesgo a la derecha positivo agrupandose  
#la mayoria en el primer cuantil
```

```
# Actividad 3 -----
```

```
boxplot(suelo$Clay1, horizontal = T, col = "blue",  
        main = "Contenido de Arcilla de 0 a 10 cm")
```

Contenido de Arcilla de 0 a 10 cm



```
which(suelo$Clay1 > 65)
```

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

```
clay1out <- subset(suelo, Clay1 > 65)
```

#P3: ¿Existe evidencia de outliers? Si existe evidencia, por lo menos tres datos

```
clay1out
```

```
##      X      e      n elev zone wrb1 LC Clay1 Clay2 Clay5 CEC1 CEC2 CEC5 OC1
## 1    1 702638 326959 657    2    3 FF   72   74   78 13.6 10.1  7.1 5.5
## 2    2 701659 326772 628    2    3 FF   71   75   80 12.6  8.2  7.4 3.2
## 106 106 696707 327780 623    2    3 FV   67   70   73 22.0 13.0 11.0 4.8
##      OC2 OC5
## 1    3.1 1.5
## 2    1.7 1.0
## 106 2.1 1.2
```

#P4: ¿Puede identificar cuales observaciones son mediante una simple restriccion de datos? Si, las observaciones son 72,71 y 67.

Actividad 4 -----

```
mean(suelo$Clay1)
```

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

```
#P5: ¿Estime si el contenido de Arcilla promedio en los suelos tropicales de 30%
#es significativamente diferente que la media observada en el campo experimental
#Tropenbos Cameroon Programme (TCP)?
```

```
# Actividad 5 -----
```

```
CorC1C5 <- cor.test(suelo$Clay1, suelo$Clay5)
CorC1C5
```

```
##
## 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
```

```
#P6: ¿Existe una relación positiva, negativa o para nada relacionados,
#entre los perfiles superior (Clay1 ) e inferior (Clay5 )
#con el contenido de Arcilla?
#Existe una relacion positiva entre los perfiles del suelo
#P7: ¿La correlación es estadísticamente significativa? Si es una correlacion
#significativa segun el coeficiente
```

```
# Actividad 6 -----
```

```
#P8: ¿Es posible determinar una ecuación significativa para predecir el
#comportamiento del contenido de arcilla en el perfil inferior Clay5 ?
#Si es posible con una regresion y la ecuacion de y=alfa-beta*x
```

```
clay.lm <- lm(suelo$Clay5 ~ suelo$Clay1)
clay.lm
```

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

```
#P9: ¿Cuál es la ecuación final para predecir el comportamiento del contenido
#de arcilla en el perfil más profundo (30-50 cm)?
```

```
# "esti.clay5 <- c(18.7586-(0.8289*x))"
```

```
summary(clay.lm)
```

```
##
## 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
```

#P10: ¿Son ambos parámetros alfa y beta significativos? Si son significativos

```
suelo$residual <- clay.lm$residuals

sum(suelo$residual**2)/145
```

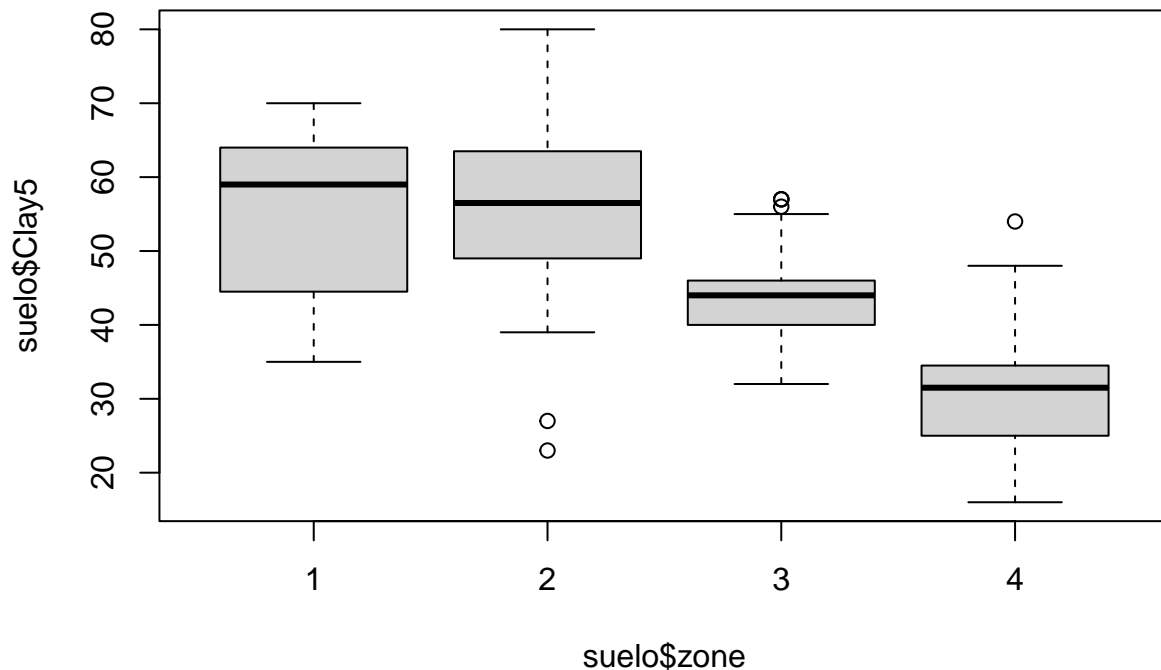
```
## [1] 32.3413
```

#P11: ¿Cuál es el porcentaje de varianza explicado por el método aplicado? Es del 32.34%

Actividad 7 -----

#P12: ¿Existe una forma de identificar la variación entre las cuatro zonas que se encuentran en el estudio? Si mediante un analisis de varianza ANOVA

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



#P13: ¿Existen indicios de que las cuatro zonas son diferentes en cuanto al #contenido de arcilla en el perfil de 30 a 50 cm.? La zona uno varia en mayor #medida con la zona 4, la zona 2 también varia entre la zona 3 y 4 y la zona 4 #es la que cuenta con mayor variacion respecto al resto de zonas y el contenido #de arcilla en el perfil inferior

#P14 ¿Observa alguna tendencia en los datos en las diferentes zonas?
#Si, la concentracion de arcilla en el perfil inferior decrece en dependencia #de la zona, con mayores concentraciones en la zona 1 y menores en las zonas #subsecuentes

```
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
```

```
# Actividad 8 -----
```

```
suelo.aov <- aov(suelo$Clay5 ~ suelo$zone)
summary(suelo.aov)
```

```
##              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
```

```
suelo.aov
```

```
## 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
```

```
#P15 ¿Existen diferencias significativas entre el contenido de arcilla del
#perfil 30-50 cm y las zonas del estudio? Si existe diferencia significativa
#dado por el valor de p que es menor al valor de 0.05
```

```
#P16: En caso de existir diferencias ¿Cuáles zonas son diferentes
#estadísticamente entre si en el contenido de arcilla en el perfil de 30-50 cm?
#Entre la zona 1 y 2 no existen diferencias estadísticas, pero entre las demás
#zonas sí.
```

```
TukeyHSD(suelo.aov, conf.level = 0.95)
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = suelo$Clay5 ~ suelo$zone)
##
## $`suelo$zone`
##              diff              lwr              upr              p adj
## 2-1    0.95000   -8.188656   10.088656  0.9930711
## 3-1  -11.15873  -20.014995   -2.302465  0.0071699
## 4-1  -23.66667  -32.889552  -14.443781  0.0000000
## 3-2  -12.10873  -16.879133   -7.338327  0.0000000
## 4-2  -24.61667  -30.037453  -19.195881  0.0000000
## 4-3  -12.50794  -17.437776   -7.578097  0.0000000
plot(TukeyHSD(suelo.aov))
```

95% family-wise confidence level

