

examen_Jonathan-Mata.R

Jana0

2023-03-31

```
#Examen práctico MFC-202
```

```
#31/03/2023
```

```
#Jonathan Abelardo Mata Hernandez
```

```
# Conjunto de datos -----
```

```
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
```

```
head(suelo)
```

```
##   V1      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
```

```
is.factor(suelo$zone)
```

```
## [1] FALSE
```

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

```
# ACTIVIDADES -----
```

```
#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
```

```
#entre más profundidad la media se incrementa, esto nos dice que tiene una tendencia positiva(+)
```

```
#Actividad 2:
#pregunta 1
stem(suelo$Clay1)
```

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 1 | 000222233333444
## 1 | 5555556778889999
## 2 | 00001111222223334444
## 2 | 555555555566788999
## 3 | 000000011222233333334444
## 3 | 556666677889999
## 4 | 022233334
## 4 | 5555667899
## 5 | 02334
## 5 | 55689
## 6 | 123
## 6 | 7
## 7 | 12
```

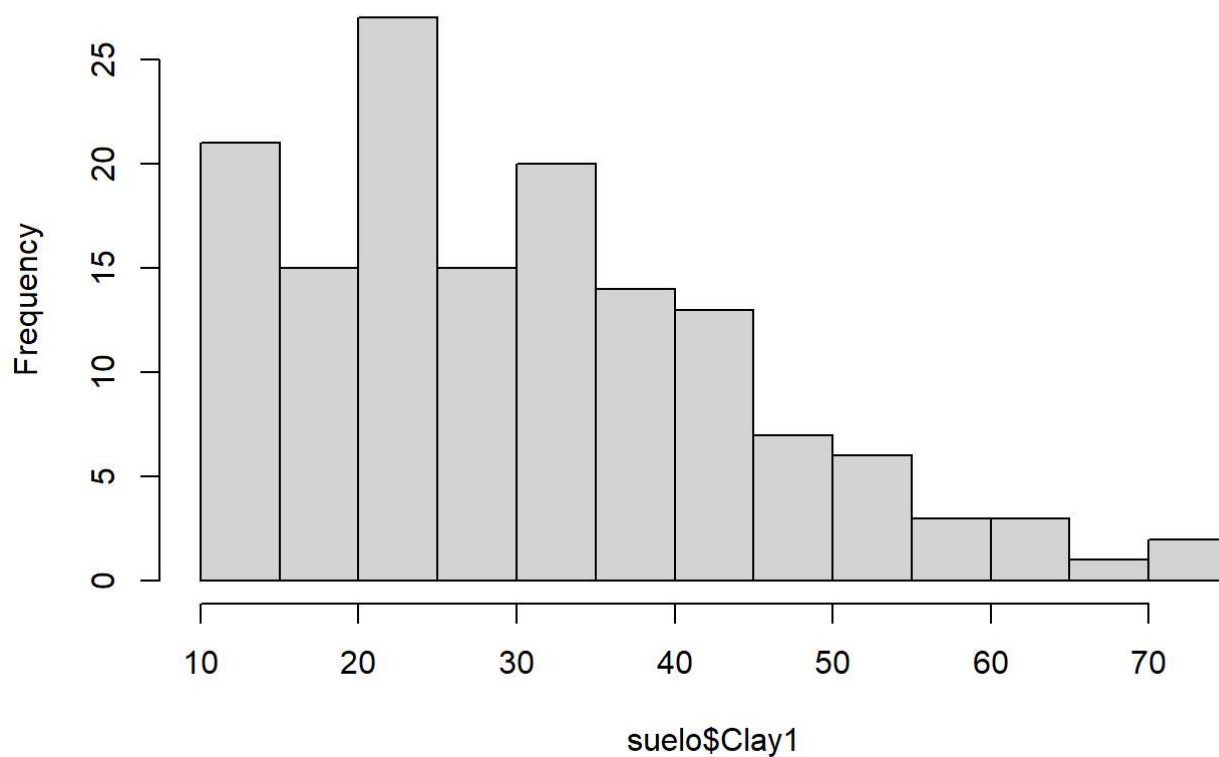
#pregunta 2

#Los datos de contenido de arcilla siguen una distribución simetrica o son sesgo?

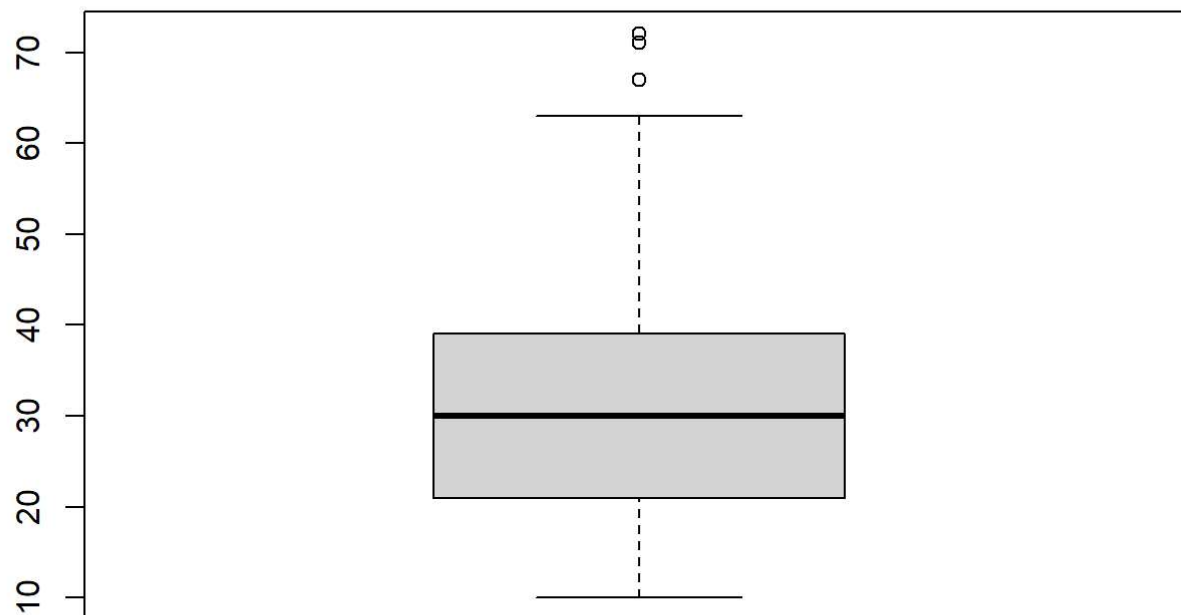
#R= con sesgo, se comprueba con el histograma los datos se van hacia la izquierda.

```
hist(suelo$Clay1)
```

Histogram of suelo\$Clay1



```
#Actividad 3  
boxplot(suelo$Clay1)
```



```
#pregunta 3  
#¿existe evidencia de outliers? si, existen.
```

```
#pregunta 4  
#¿puede identificar cuales observaciones son mediante una simple restricción de datos?  
#R= si mediante la funcion
```

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

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

```
#Actividad 4  
mean(suelo$Clay1)
```

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

#pregunta 5

#¿estime si el contenido de arcilla promedio en los suelos tropicales de 30% es significativo diferente que la media observada en el campo experimental TCP?
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
```

#si es sig

#Actividad 5

#pregunta 6

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
```

```
#pregunta 7
```

```
#¿la correlación es estadísticamente significativa?R= si es estadísticamente sig.
```

```
#Actividad 6
```

```
#pregunta 8
```

```
#es posible determinar una ecuación significativa para predecir el comportamiento del contenido de arcilla en el perfil inferior clay5?
```

```
#R= si es posible
```

```
#pregunta 9
```

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

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

```
regresion
```

```
##
```

```
## Call:
```

```
## lm(formula = suelo$Clay5 ~ suelo$Clay1)
```

```
##
```

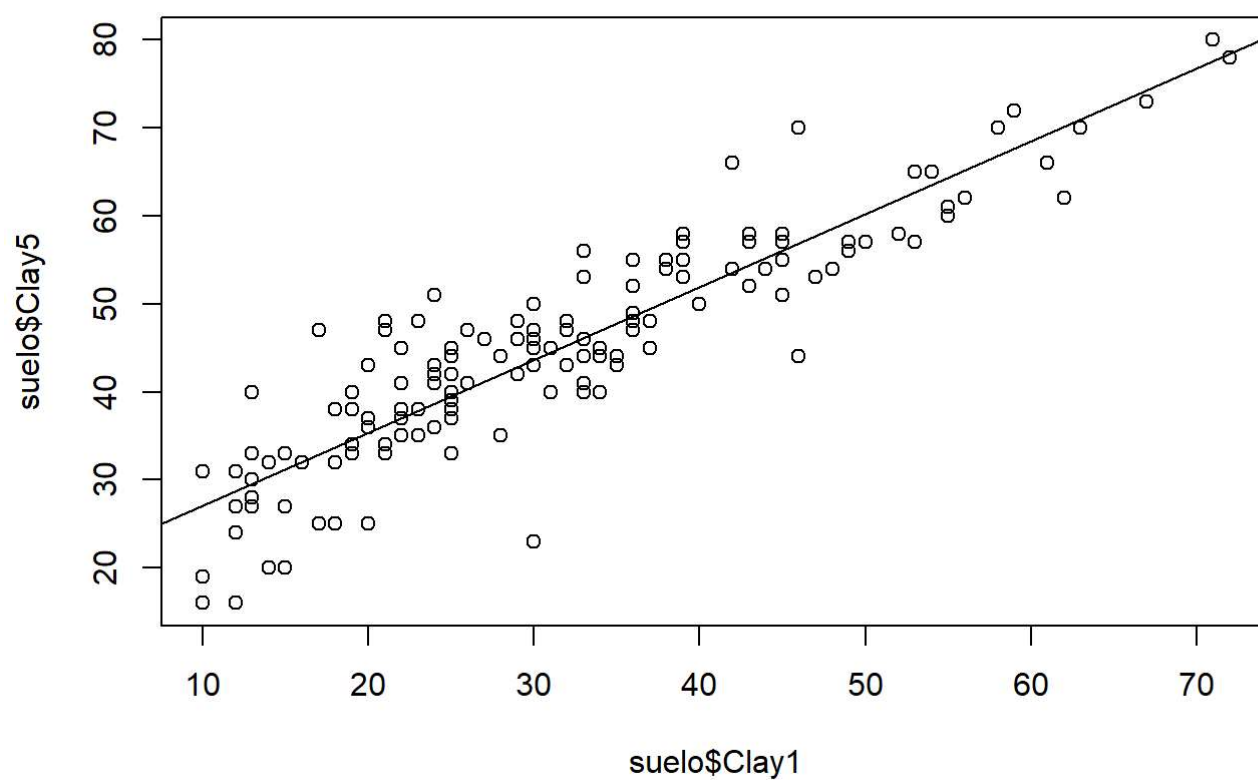
```
## Coefficients:
```

```
## (Intercept)  suelo$Clay1
```

```
##      18.7586      0.8289
```

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

```
abline(regresion)
```



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

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

```
#pregunta 10
#¿son ambos parametros alfa y beta 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= si, amnos son significativos

#pregunta 11

#¿cual es el porcentaje de varianza explicado por el metodo aplicado?
 anova(regresion)

```
## Analysis of Variance Table
##
## Response: suelo$Clay5
##              Df Sum Sq Mean Sq F value    Pr(>F)
## suelo$Clay1   1 19482.5  19482.5   602.4 < 2.2e-16 ***
## Residuals    145  4689.5    32.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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
```

#se revisa la R cuadrada ajustada que es del 80%, así que es confiable porque predice y está cerca de 1 el mresultado (0.80).

#Actividad 7

#pregunta 12

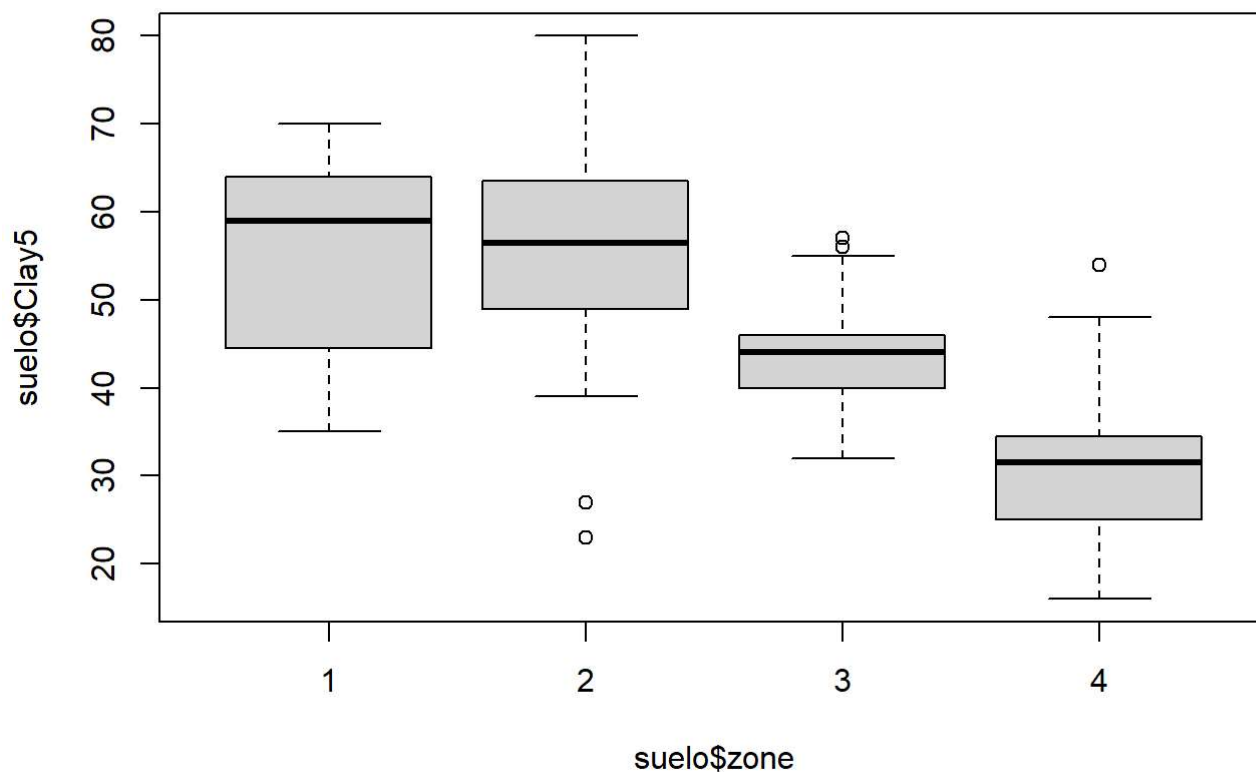
#¿existe una fórmula de identificar la variación entre las cuatro zonas que se encuentran en el estudio?

#si existe, mediante el análisis de varianza

#pregunta 13

#realice una inspección visual del contenido de arcilla en el perfil 30-50cm (clay5) y las cuatro zonas (zone) presentes en el área de estudio. ¿existen indicios de que las cuatro zonas son diferentes en cuanto al contenido de arcilla en el perfil 30 a 50 cm?

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



#R=si es diferente diferencias en los contenidos de arcilla, analizando la grafica se puede observar claramente.

#Pregunta 14

#Observa alguna tendencia en los datos en las diferentes zonas?

#R= el valor maximo disminuye conforme cambia la zona.

```
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

#determine mediante un analisis de varianza con las variable contenido de arcilla en el perfil 3 0-50 y el factor o tratamiento de zonas.

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

```
##              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
```

```
TukeyHSD(anova)
```

```
## 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
```

#Pregunta 15

#existen diferencias significativas entre el contenido de arcilla del perfil 30-50 cm y las zonas de estudio

si existe diferencias significativas, a excepción de zona 1 y 2.

#pregunta 16

#en caso de existir diferencias ¿cuales son diferentes estadísticamente entre si en el contenido de arcilla en el perfil de 30-50?

#la zona 1 y 2 no tienen diferencias significativas pero, la zona 3-1, 4-1, 3-2, 4-2, 4-3, tienen diferencias significativas.