

examen.R

toryf

2023-03-31

```
#examen
#Alumno: Ing. Luis Miguel Toribio Ferrer
#Matricula: 2173385
#Fecha: 31/03/2023

# 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
```

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

```
## [1] TRUE
```

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

```
## [1] TRUE
```

```
# Actividad a Realizar -----

# Actividad 1 -----

#¿cual es la tendencia del contenido promedio de arcilla (clay) con respecto a la profundidad?

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
```

#RESPUESTA: conforme va incrementado la profundidad del suelo el contenido de arcilla va incrementado, en conclusion tiene una tendencia positiva.

Actividad 2 -----

#grafica la distribucion del primer perfil Clay1 con la funcion stem_and_leaf.

```
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
```

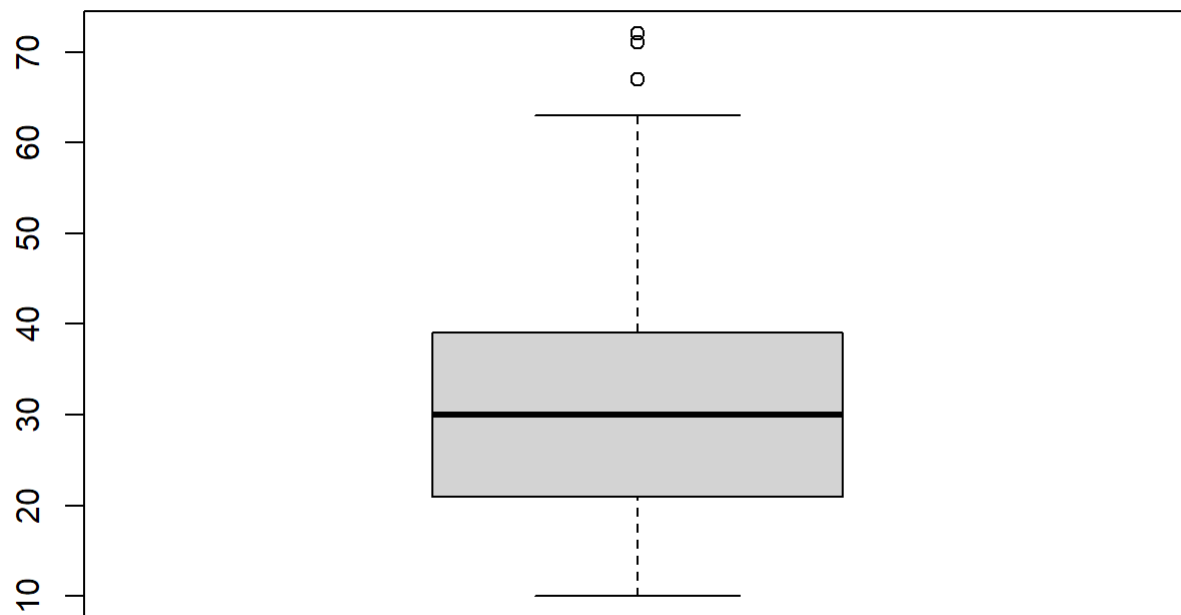
#¿los datos de contenido de arcilla siguen una distribucion simetrica o con sesgo?

#RESPUESTA: Los datos tiene una distribución con sesgo a la izquierda

Actividad 3 -----

#graficar mediante un boxplot para el primer perfil clay1 de arcilla

```
boxplot(suelo$Clay1)
```



#¿existe evidencia de outliers?

#RESPUESTA: si existen evidencias

#puede identificar cuales observaciones son mediante una simple restriccion de datos?

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

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

#RESPUESTA: Las observaciones numero 1, 2, y 106.

Actividad 4 -----

#determine la media de la variable clay1

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

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

#¿estime si el contenido de arcilla promedio en los suelos tropicales de 30 % es significativa diferente a 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
```

#RESPUESTA: no es significativa diferente a la media

Actividad 5 -----

#La primera pregunta del investigador es saber si existe una relación entre el contenido de arcilla en los tres perfiles. De aquí se desprende la siguiente pregunta específica:

#¿Existe una relación positiva, negativa o para nada relacionados, entre los perfiles superior (Clay 1) e inferior (Clay) con el contenido de Arcilla?

```
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
```

```
#RESPUESTA: si existe una relacion positiva
```

```
#¿La correlación es estadísticamente significativa?
```

```
#RESPUESTA: si es estadísticamente significativa
```

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

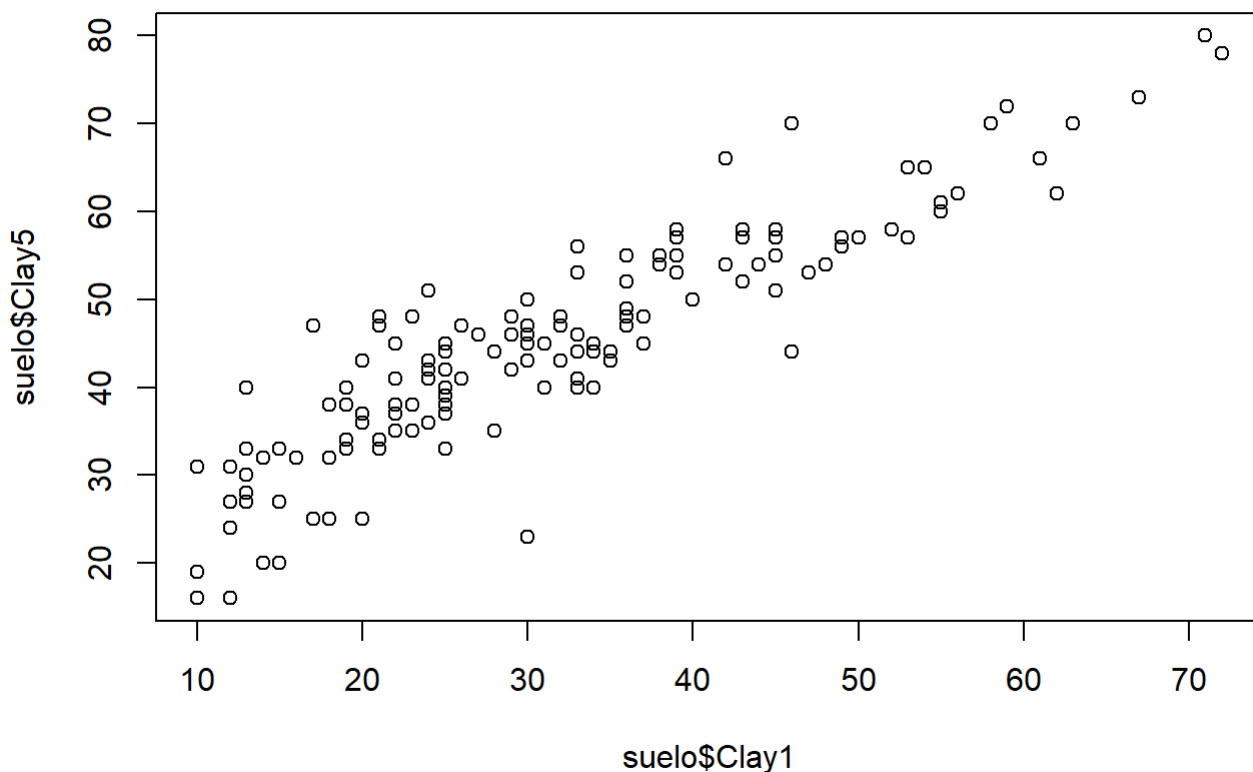
```
#Después de cada análisis de correlación surge la siguiente pregunta:
```

```
#¿Es posible determinar una ecuación significativa para predecir el comportamiento del contenido de arcilla en el perfil inferior Clay5?
```

```
#RESPUESTA: si es posible determinar una ecuación
```

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

```
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
```

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

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

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

#¿Son ambos parámetros 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
```

#RESPUESTA: si ambos parametros son significativos

#¿Cuál es el porcentaje de varianza explicado por el método 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`

#La Adjust R-squared fue de 0.804 obtenido lo que significa que la ecuacion predice de manera correcta. entonces el porcentaje es de 80.4 %.

Actividad 7 -----

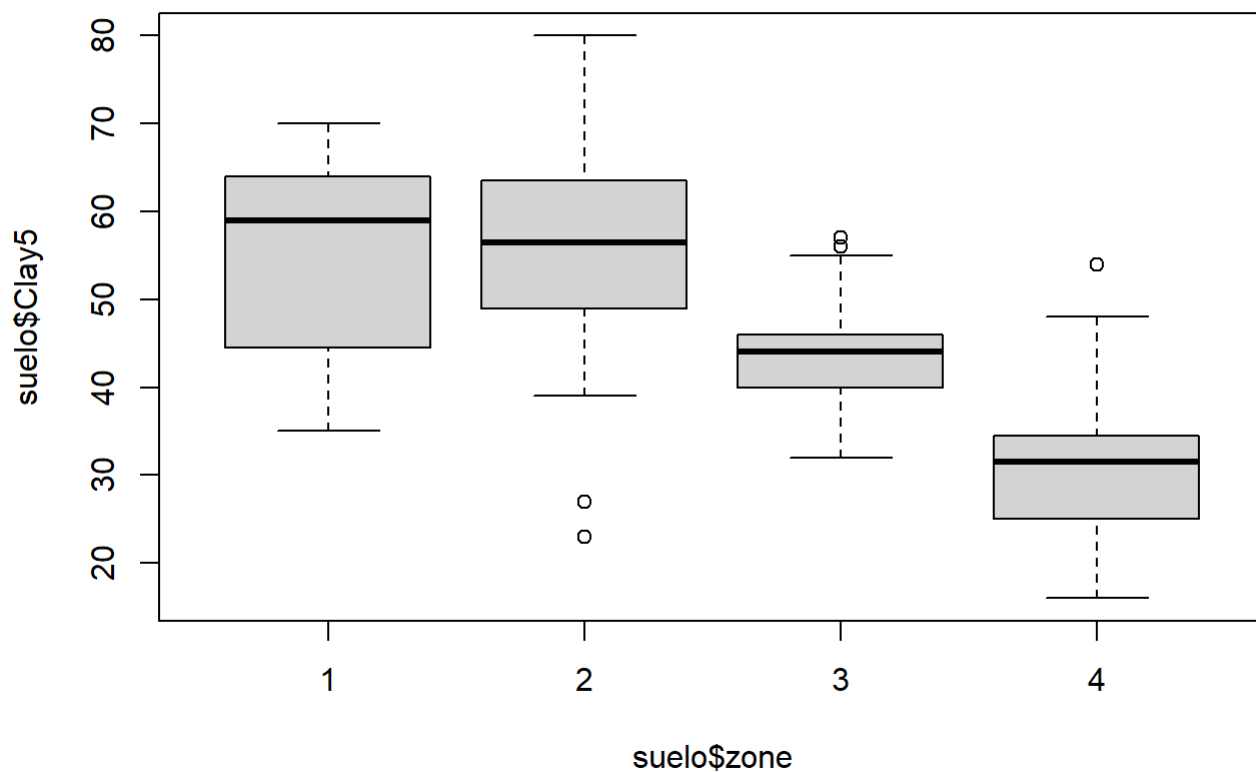
#Con la base de datos se puede predecir que tan variable es el contenido de arcilla en la capa más profunda entre las cuatro zonas identificadas

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

#RESPUESTA: si existe la forma de identificar la variación, mediante el analisis de varianza (anova)

#Realice una inspección visual del contenido de arcilla en el perfil 30-50 cm (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 de 30 a 50 cm.?

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



#RESPUESTA: mediante la inspeccion visual de los boxplot, se encontró que si existe diferencias en los contenidos de arcillas en las cuatro zonas

#¿observa alguna diferencia en los datos en las diferentes zonas?

```
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
```


#RESPUESTA: La tendencia de la media, mediana y 3rd quartil y valores maximo va disminuyendo con forme va cambiando de zona.

Actividad 8 -----

Determine mediante una análisis de varianza con Las variable contenido de arcilla en el perfil 30-50 y el factor o tratamiento 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
```

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

#RESPUESTA: de acuerdo con el analisis de varianza (ANOVA) si existe diferencia significativa en el contenido de arcilla del perfil 30-50 cm y la zona del estudio

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

```
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
```

#RESPUESTA: con base a la prueba de tukey, la zona 1 es diferente estadísticamente con la zona 3 y 4-, la zona 2 es estadísticamente diferente con la zona 3 y 4 y la zona 3 es estadísticamente diferente con la 4.