PAC6

Maria Lucas Gascón

2024-01-17

# Problema 1

Cargamos los datos del problema.

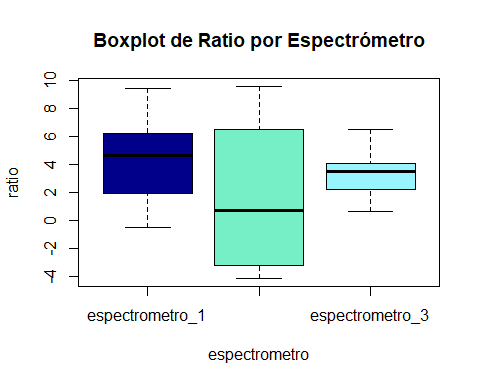
# Seteamos el directorio de trabajo  
setwd("D:/Antiguos estudios/MASTER2/Sem3/Diseño/PAC6/PAC6")  
  
# Importamos los datos  
data <- read.csv("dades\_problema1\_pec6.csv", dec = ",", header = TRUE, sep = ";")  
  
# Marcamos los factores  
data$espectrometro <- as.factor(data$espectrometro)  
data$suelo <- as.factor(data$suelo)  
data$dia <- factor(data$dia, levels = c(1, 2, 3), labels = c("dia1", "dia2", "dia3"))

## 1.1 Análisis descriptivo de los datos

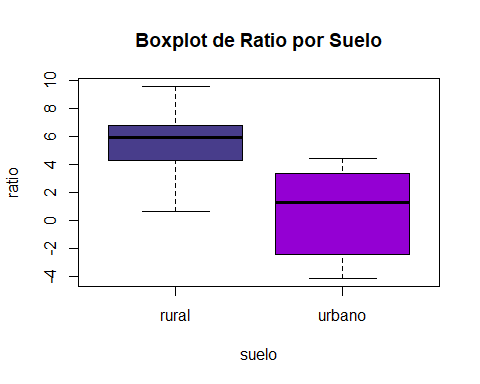
# Primero examinamos la variable ratio  
summary(data$ratio)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -4.1356 0.9882 3.5102 3.1112 5.7919 9.5738

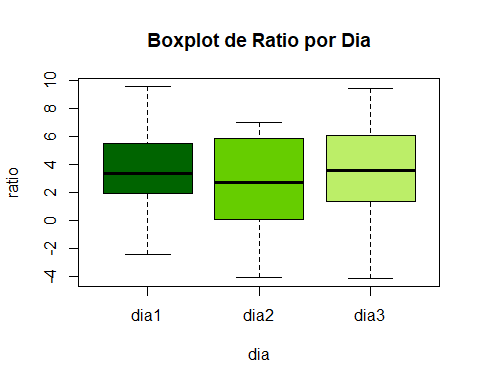
# Box plot de ratio por espectrometro  
boxplot(ratio ~ espectrometro, data = data, col = c("blue4", "aquamarine2", "cadetblue1"), main = "Boxplot de Ratio por Espectrómetro")



# Box plot de ratio por suelo  
boxplot(ratio ~ suelo, data = data, col = c("darkslateblue", "darkviolet"), main = "Boxplot de Ratio por Suelo")



# Box plot de ratio por dia  
boxplot(ratio ~ dia, data = data, col = c("darkgreen", "chartreuse3", "darkolivegreen2"), main = "Boxplot de Ratio por Dia")



## 1.2 Modelo lineal

En este problema, se busca modelar la proporción de 14N a 15N en dos tipos de suelos (urbano, rural) utilizando tres espectrómetros distintos. Se realizaron mediciones en 18 muestras de cada tipo de suelo. Cada muestra se asignó a uno de los tres espectrómetros al azar, y para cada combinación de suelo y espectrómetro, se eligieron 3 días al azar para realizar las mediciones. El objetivo es analizar la influencia de los factores suelo, espectrómetro y día, así como sus interacciones.

El modelo lineal puede expresarse de la siguiente manera:

* es la medida de la proporción de 14N a 15N para la k-ésima observación en el suelo i, espectrómetro j y día k.
* es la media global.
* es el efecto del i-ésimo nivel del factor suelo. La suma de todos los efectos del factor suelo debe ser cero:
* es el efecto del j-ésimo nivel del factor espectrómetro. La suma de todos los efectos del factor espectrómetro debe ser cero:
* es la interacción entre el suelo i y el espectrómetro j. La suma de las interacciones debe ser cero:
* es el efecto del k-ésimo nivel del factor día anidado en la combinación suelo i y espectrómetro j. Como es un efecto aleatorio, no hay restricciones específicas, pero se asume que sigue una distribución normal con media cero:
* es el error aleatorio asociado con la k-ésima observación en la combinación suelo i, espectrómetro j y día k. Se asume que sigue una distribución normal con media cero:

## 1.3 Hipótesis de interés

Para el factor suelo:

* . No hay diferencia significativa en la proporción de 14N a 15N entre los suelos urbanos y rurales.
* Almenos un . Hay al menos una diferencia significativa en la proporción de 14N a 15N entre los suelos urbanos y rurales.

Para el factor espectrómetro:

* . No hay diferencia significativa en la proporción de 14N a 15N entre los espectrómetros.
* Almenos un . Hay al menos una diferencia significativa en la proporción de 14N a 15N entre los tres espectrómetros distintos.

Para la interacción suelo-espectrómetro:

* para todos los . No hay interacción significativa entre el tipo de suelo y el espectrómetro.
* Existe al menos una . Hay al menos una interacción significativa entre el tipo de suelo y el espectrómetro.

Para el efecto aleatorio del día: + . No hay variabilidad significativa asociada con los días anidados en la combinación de suelo y espectrómetro. + . Existe variabilidad significativa asociada con los días anidados en la combinación de suelo y espectrómetro.

Para el error aleatorio: + . No hay variabilidad significativa no explicada por los factores considerados en el modelo. + . Existe variabilidad significativa no explicada por los factores considerados en el modelo.

## 1.4 Estudiar las hipótesis

# Definimos el modelo  
library(lme4)

## Loading required package: Matrix

library(car)

## Loading required package: carData

modelo <- lmer(ratio ~ suelo + espectrometro + suelo:espectrometro + (1 | dia:suelo:espectrometro), data = data)  
  
# Testamos las hipótesis  
  
anova\_modelo <- Anova(modelo, type = "III", test = "Chisq")  
print(anova\_modelo)

## Analysis of Deviance Table (Type III Wald chisquare tests)  
##   
## Response: ratio  
## Chisq Df Pr(>Chisq)   
## (Intercept) 47.6697 1 5.044e-12 \*\*\*  
## suelo 12.5961 1 0.0003866 \*\*\*  
## espectrometro 5.9247 2 0.0516970 .   
## suelo:espectrometro 18.4324 2 9.942e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Especificar que queremos la ANOVA de “tipo III” implica que se atribuye la variabilidad a cada término después de tener en cuenta todos los demás términos, incluidas las interacciones que involucran ese término. Es útil cuando se tienen interacciones y se busca evaluar la contribución de cada variable a la variabilidad explicada, considerando las otras variables y sus interacciones.

Además, como no se han comprobado las suposiciones (normalidad y homocedasticidad), es preferible usar la prueba Chi-cuadrado en lugar de F.

Como podemos ver en los resultados, el suelo y la interacción suelo:espectrofotometro son significativas (pv<0.05), aunque el espectrofotómetro por sí mismo no lo es (pv>0.05). Veamos en detalle lo que esto significa:

* Suelo Significativo: Hay diferencias estadísticamente significativas en la proporción de 14N a 15N entre los suelos urbanos y rurales (pv<0.05, se rechaza H0), después de tener en cuenta los otros efectos en el modelo. En otras palabras, hay evidencia estadística para afirmar que el tipo de suelo tiene un impacto significativo en la proporción de nitrógeno.
* Espectrómetro No Significativo: No hay evidencia estadística suficiente para afirmar que hay diferencias en la proporción de 14N a 15N entre los distintos espectrómetros (pv>0.05, se acepta H0), después de considerar los otros efectos. En resumen, las variaciones observadas en la proporción de nitrógeno no pueden atribuirse de manera significativa a las diferencias entre los espectrómetros utilizados.
* Interacción Suelo-Espectrómetro Significativa: La influencia del espectrómetro en la proporción de nitródeno depende del tipo de suelo. En otras palabras, la diferencia en la proporción de 14N a 15N entre los espectrómetros puede variar según si el suelo es urbano o rural. (pv<0.05, se rechaza H0). Esta interacción indica que el efecto del espectrómetro no es constante en todos los suelos y viceversa.

En resumen, los resultados sugieren que el tipo de suelo y la interacción entre el tipo de suelo y el espectrómetro son factores significativos en la variabilidad de la proporción de 14N a 15N, mientras que el espectrómetro por sí mismo no tiene un efecto significativo en esta proporción.

## 1.5 Comparaciones múltiples

Cómo el factor espectrómetro no mostraba diferencias significativas y el factor suelo sólo tiene 2 niveles; la única comparación múltiple que tiene sentido realizar es la de la interación.

# install.packages("emmeans")  
library(emmeans)

## Warning: package 'emmeans' was built under R version 4.3.2

library(knitr)  
  
# Realizamos las comparaciones múltiples  
emmeans\_result <- emmeans(modelo, pairwise ~ suelo:espectrometro)  
  
# Ajustamos p-valores usando la corrección de Bonferroni  
adjusted\_emmeans <- summary(emmeans\_result, infer = c(TRUE, TRUE), adjust = "bonferroni")  
  
# Imprimimos los valores ajustados  
print(adjusted\_emmeans)

## $emmeans  
## suelo espectrometro emmean SE df lower.CL upper.CL t.ratio p.value  
## rural espectrometro\_1 6.89 0.999 12 3.74620 10.042 6.904 0.0001  
## urbano espectrometro\_1 1.88 0.999 12 -1.26569 5.031 1.885 0.5031  
## rural espectrometro\_2 6.09 0.999 12 2.94315 9.239 6.100 0.0003  
## urbano espectrometro\_2 -2.94 0.999 12 -6.08548 0.211 -2.942 0.0740  
## rural espectrometro\_3 3.60 0.999 12 0.45026 6.746 3.604 0.0217  
## urbano espectrometro\_3 3.14 0.999 12 -0.00981 6.286 3.143 0.0509  
##   
## Degrees-of-freedom method: kenward-roger   
## Confidence level used: 0.95   
## Conf-level adjustment: bonferroni method for 6 estimates   
## P value adjustment: bonferroni method for 6 tests   
##   
## $contrasts  
## contrast estimate SE df lower.CL  
## rural espectrometro\_1 - urbano espectrometro\_1 5.012 1.41 12 -0.141  
## rural espectrometro\_1 - rural espectrometro\_2 0.803 1.41 12 -4.350  
## rural espectrometro\_1 - urbano espectrometro\_2 9.832 1.41 12 4.679  
## rural espectrometro\_1 - rural espectrometro\_3 3.296 1.41 12 -1.857  
## rural espectrometro\_1 - urbano espectrometro\_3 3.756 1.41 12 -1.397  
## urbano espectrometro\_1 - rural espectrometro\_2 -4.209 1.41 12 -9.362  
## urbano espectrometro\_1 - urbano espectrometro\_2 4.820 1.41 12 -0.333  
## urbano espectrometro\_1 - rural espectrometro\_3 -1.716 1.41 12 -6.869  
## urbano espectrometro\_1 - urbano espectrometro\_3 -1.256 1.41 12 -6.409  
## rural espectrometro\_2 - urbano espectrometro\_2 9.029 1.41 12 3.876  
## rural espectrometro\_2 - rural espectrometro\_3 2.493 1.41 12 -2.660  
## rural espectrometro\_2 - urbano espectrometro\_3 2.953 1.41 12 -2.200  
## urbano espectrometro\_2 - rural espectrometro\_3 -6.536 1.41 12 -11.689  
## urbano espectrometro\_2 - urbano espectrometro\_3 -6.076 1.41 12 -11.228  
## rural espectrometro\_3 - urbano espectrometro\_3 0.460 1.41 12 -4.693  
## upper.CL t.ratio p.value  
## 10.165 3.549 0.0600  
## 5.956 0.569 1.0000  
## 14.985 6.962 0.0002  
## 8.449 2.334 0.5670  
## 8.909 2.660 0.3120  
## 0.944 -2.980 0.1721  
## 9.973 3.413 0.0772  
## 3.437 -1.215 1.0000  
## 3.897 -0.889 1.0000  
## 14.181 6.393 0.0005  
## 7.646 1.765 1.0000  
## 8.106 2.091 0.8769  
## -1.383 -4.628 0.0087  
## -0.923 -4.302 0.0154  
## 5.613 0.326 1.0000  
##   
## Degrees-of-freedom method: kenward-roger   
## Confidence level used: 0.95   
## Conf-level adjustment: bonferroni method for 15 estimates   
## P value adjustment: bonferroni method for 15 tests

Como podemos ver, las siguientes combinaciones resultan significativas:

# Extraemos los contrastes y p-valores  
contrast\_df <- as.data.frame(adjusted\_emmeans$contrasts)  
  
# Filtramos por p-valores < 0.05  
significant\_contrasts <- contrast\_df[contrast\_df$p.value < 0.05, c("contrast", "p.value")]  
  
  
# Formateamos la tabla para imprimir  
formatted\_table <- data.frame(  
 Contrast = significant\_contrasts$contrast,  
 P\_Value = sprintf("%.4f", significant\_contrasts$p.value)  
)  
  
# Imprimimos la tabla  
kable(formatted\_table, col.names = c("Contrast", "P-Value"))

| Contrast | P-Value |
| --- | --- |
| rural espectrometro\_1 - urbano espectrometro\_2 | 0.0002 |
| rural espectrometro\_2 - urbano espectrometro\_2 | 0.0005 |
| urbano espectrometro\_2 - rural espectrometro\_3 | 0.0087 |
| urbano espectrometro\_2 - urbano espectrometro\_3 | 0.0154 |

Eso significa que:

* Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo rural es diferente cuando se utiliza el espectrómetro\_1 en comparación con el suelo urbano cuando se utiliza el espectrómetro\_2.
* Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo rural es diferente cuando se utiliza el espectrómetro\_2 en comparación con el suelo urbano cuando también se utiliza el espectrómetro\_2.
* Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo urbano es diferente cuando se utiliza el espectrómetro\_2 en comparación con el suelo rural cuando se utiliza el espectrómetro\_3.
* Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo urbano es diferente cuando se utiliza el espectrómetro\_2 en comparación con el mismo suelo urbano cuando se utiliza el espectrómetro\_3.

Estos p-valores bajos indican que, después de ajustar para múltiples comparaciones, las diferencias observadas en estas combinaciones específicas son estadísticamente significativas. Para el resto de combinaciones no se observan diferencias significativas.

Cabe destacar la importancia de realizar una corrección para minimizar el error de tipo 1, ya que se estan realizando un gran número de comparaciones. En este caso se ha usado la corrección de Bonferroni.