

Biometría II

TP N° 4

Modelos múltiples

El objetivo de este TP es profundizar el ajuste de modelos lineales generales con mayor grado de complejidad, hasta ahora habíamos estudiado situaciones en donde no se cumplían los supuestos para la validez de pruebas que estiman los parámetros por cuadrados mínimos. En este trabajo práctico sumamos la complejidad de analizar modelos con más de una variable explicativa, pudiendo éstas ser categóricas, continuas o de ambos tipos y pudiendo o no existir interacción entre ellas.

Problema 1. Aclimatación al estrés hídrico en *Eucalyptus globulus*

En los sistemas forestales, luego de la plantación, múltiples factores de estrés abiótico y biótico afectan la supervivencia y el crecimiento de los árboles. El estrés hídrico es el fenómeno más frecuente. Es posible mejorar la eficiencia de una plantación, aplicando prácticas de aclimatación de las plantas en vivero. Una de ellas, es la técnica de restricción de riego que permite aclimatar a las plantas a las condiciones de sequía y se aplica al final del ciclo de producción de las plantas en vivero. Asimismo, se cree que la fertilización con potasio podría favorecer el proceso. El objetivo de este trabajo fue evaluar el comportamiento de plántulas *Eucalyptus globulus* sometidas a tres regímenes de riego y a dos niveles de fertilización potásica al final de ciclo de producción en vivero. Se trabajó con plantas de 8 meses de edad, en envases de 3 l, utilizando como sustrato tierra orgánica. Las plantas fueron asignadas al azar a tres regímenes hídricos: Riego diario (sin estrés hídrico), Riego cada 3 días (estrés moderado) o Riego cada 6 días (estrés alto) y a dos niveles de fertilización potásica: con fertilización potásica (1,8 g/l) y sin fertilización potásica. Al cabo de tres meses de tratamiento en vivero, se evaluó el grado de aclimatación sometiendo a todas las plantas a un período de sequía de 9 días, al cabo de los cuales se midió el contenido relativo de agua (CRA %). El CRA es el contenido porcentual de agua de la planta en relación al contenido de agua a hidratación máxima y constituye una medida del estado hídrico de una planta (cuanto mayor, mejor el estado de la planta). Los resultados, en el archivo *CRA.txt*.

- Identifique la unidad experimental, la variable respuesta, la o las variables explicativas, la cantidad de réplicas del ensayo y el diseño.
U.E: envase de 3l con tierra organica con una planteado.
v.a: contenido relativo de CRA al cabo de 9 dias de exposicion a sequia.
v.e: niveles de estres hidrico (nulo, bajo, alto) y fertilizacion potasica (con/sin).
replicas: N/(3*2).
Disenio: factorial, 2 factores (DCA).
- ¿En función de qué información piensa Ud que se decidió el número de réplicas del experimento?
- Especifique el modelo en términos del problema.

```
Datos <- read.delim("/home/jose/Documents/materias/biome2/2019/tps/tp4/CRA.txt",  
  header = T)  
library(ggplot2) # Para hacer gráficos de interacción  
library(car) # Para prueba de Levene
```

```
## Loading required package: carData  
library(emmeans) #Para comparaciones a posteriori.  
source("/home/jose/Documents/funciones2.R")
```

$$\begin{aligned}CRA_i &= \beta_0 + \beta_1 * [3dias]_i + \beta_2 * [6dias]_i + \beta_3 * [K]_i + \beta_4 * [3dias] * [K]_i + \beta_5 * [6dias] * [K]_i + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \\ i &= 1 : n(30)\end{aligned}$$

- Describa estadísticamente los datos. Efectúe un gráfico de perfiles, ¿qué puede interpretar del mismo?

```

attach(Datos)
str(Datos)

## 'data.frame': 30 obs. of 3 variables:
## $ Riego : Factor w/ 3 levels "cada 3 dias",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Fertilizacion: Factor w/ 2 levels "con K","sin K": 1 1 1 1 1 2 2 2 2 2 ...
## $ CRA : int 77 76 73 75 82 74 73 79 77 62 ...

library(psych)

##
## Attaching package: 'psych'

## The following object is masked from 'package:car':
##
## logit

## The following objects are masked from 'package:ggplot2':
##
## %+%, alpha

describeBy(Datos$CRA, list(Datos$Riego, Datos$Fertilizacion))

##
## Descriptive statistics by group
## : cada 3 dias
## : con K
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 76.6 3.36 76 76.6 1.48 73 82 9 0.56 -1.39 1.5
## -----
## : cada 6 dias
## : con K
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 88.4 5.22 89 88.4 2.97 80 94 14 -0.56 -1.39 2.34
## -----
## : Diario
## : con K
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 69.2 7.6 72 69.2 5.93 58 76 18 -0.43 -1.84 3.4
## -----
## : cada 3 dias
## : sin K
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 73 6.6 74 73 4.45 62 79 17 -0.73 -1.29 2.95
## -----
## : cada 6 dias
## : sin K
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 83.6 5.59 81 83.6 4.45 78 91 13 0.29 -2.07 2.5
## -----
## : Diario
## : sin K
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 67.2 4.66 68 67.2 5.93 61 72 11 -0.21 -2.01 2.08

# medias
tapply(Datos$CRA, list(Datos$Riego, Datos$Fertilizacion), mean)

## con K sin K
## cada 3 dias 76.6 73.0
## cada 6 dias 88.4 83.6
## Diario 69.2 67.2

```

```
# desvío estándar
round(tapply(Datos$CRA, list(Datos$Riego, Datos$Fertilizacion), sd),
      2)
```

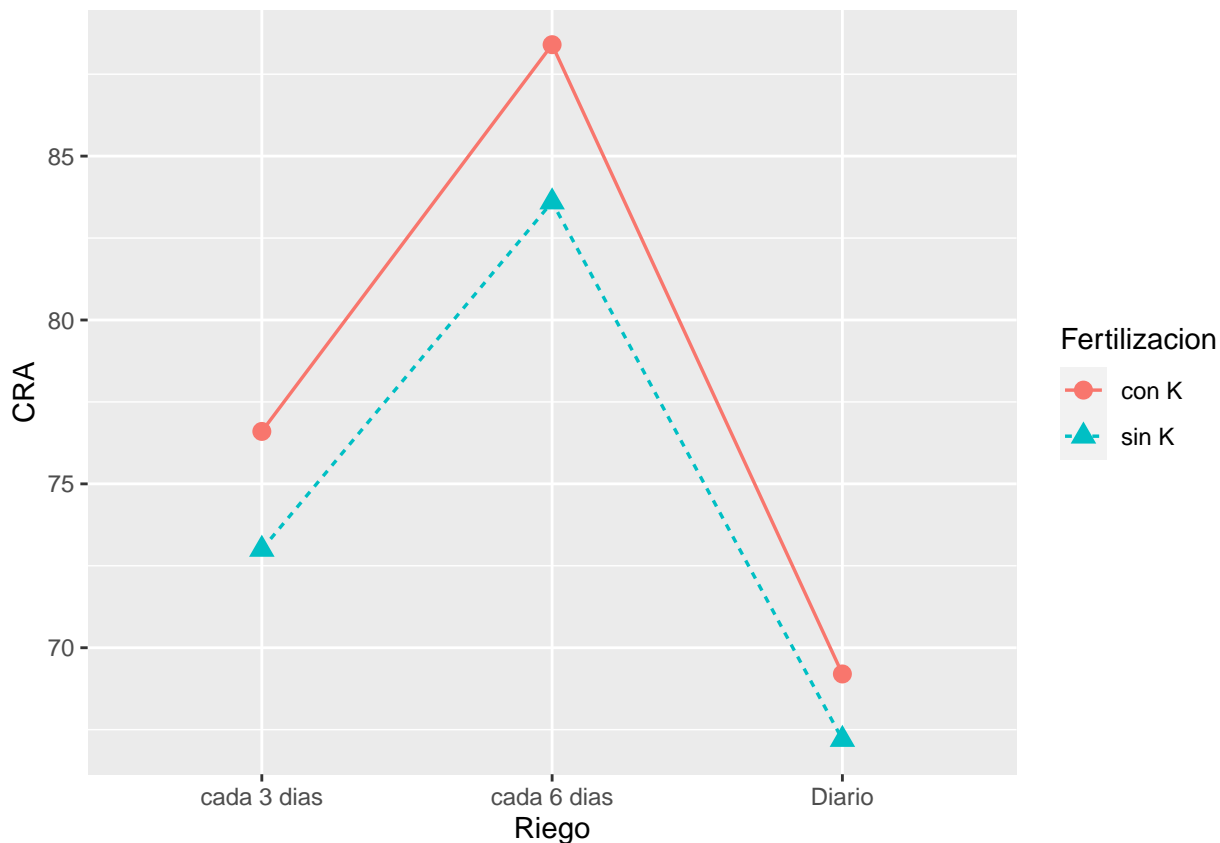
```
##           con K sin K
## cada 3 dias  3.36  6.60
## cada 6 dias  5.22  5.59
## Diario      7.60  4.66
```

Graficos de perfiles

```
medias.Datos <- aggregate(CRA ~ Riego + Fertilizacion, Datos, mean)
summary(medias.Datos)
```

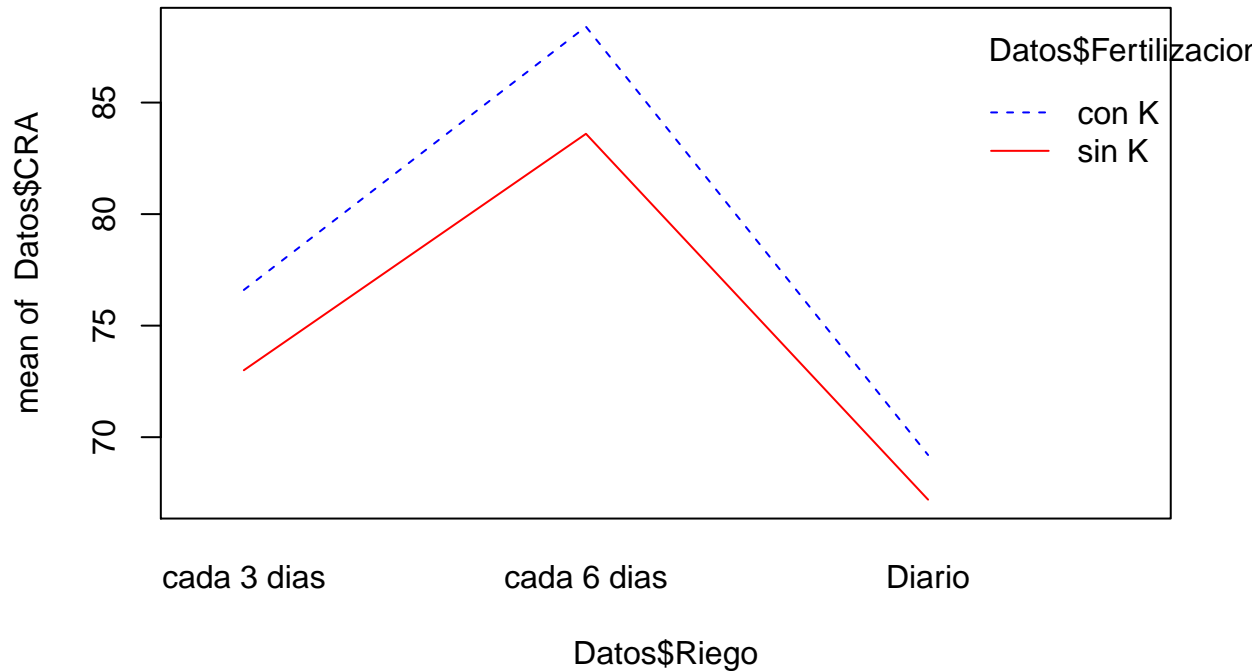
```
##           Riego  Fertilizacion      CRA
## cada 3 dias:2   con K:3      Min.   :67.20
## cada 6 dias:2   sin K:3      1st Qu.:70.15
## Diario      :2                      Median :74.80
##                                           Mean   :76.33
##                                           3rd Qu.:81.85
##                                           Max.   :88.40
```

```
gperfiles <- ggplot(medias.Datos, aes(x = Riego, y = CRA, colour = Fertilizacion,
                                     group = Fertilizacion))
gperfiles + geom_line(aes(linetype = Fertilizacion), size = 0.6) +
  geom_point(aes(shape = Fertilizacion), size = 3)
```



Otra opción gráfica

```
interaction.plot(Datos$Riego, Datos$Fertilizacion, Datos$CRA, col = c(4,
                              2))
```



El grafico de perfiles muestra un incremento en el CRA entre los grupos con diferente regimenes de riego. Sin embargo, no pareciera existir un cambio en la tendencia debido al agregado de fertilizador.

- ¿Se cumplen los supuestos del modelo? Fundamentalmente.

```
modelo1 <- lm(CRA ~ Riego + Fertilizacion + Riego:Fertilizacion, data = Datos)
# anova(modelo1) nos muestra una tabla con las fuentes de
# variación a la que comúnmente se llama "tabla de anova"
anova(modelo1)
```

```
## Analysis of Variance Table
##
## Response: CRA
##
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Riego      2 1619.47   809.73  25.1992 1.27e-06 ***
## Fertilizacion 1   90.13    90.13   2.8050  0.1070
## Riego:Fertilizacion 2    9.87    4.93   0.1535  0.8585
## Residuals  24  771.20    32.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# summary(modelo1) nos muestra los estimadores de los parámetros
# del modelo con sus errores estándar y significación
summary(modelo1)
```

```
##
## Call:
## lm(formula = CRA ~ Riego + Fertilizacion + Riego:Fertilizacion,
##     data = Datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##    -11.2    -3.5     0.5     4.3     7.4
```

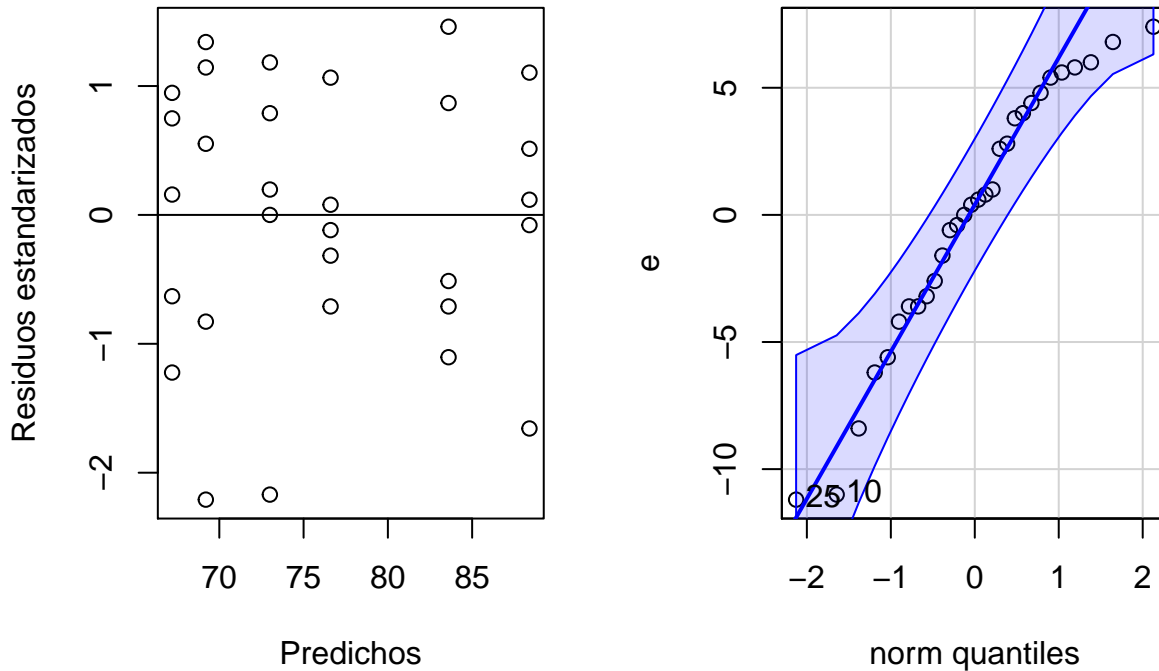
```
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      76.600      2.535  30.216 < 2e-16 ***
## Riegocada 6 dias      11.800      3.585   3.291 0.00308 **
## RiegoDiario       -7.400      3.585  -2.064 0.04998 *
## Fertilizacionsin K   -3.600      3.585  -1.004 0.32533
## Riegocada 6 dias:Fertilizacionsin K -1.200      5.070  -0.237 0.81491
## RiegoDiario:Fertilizacionsin K      1.600      5.070   0.316 0.75506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.669 on 24 degrees of freedom
## Multiple R-squared:  0.6904, Adjusted R-squared:  0.6259
## F-statistic: 10.7 on 5 and 24 DF,  p-value: 1.701e-05
```

```
# Calculamos los residuos y los predichos
e <- resid(modelo1) # residuos
re <- rstandard(modelo1) #residuos estandarizados
re <- round(re, 3)
pre <- predict(modelo1) #predichos
res <- data.frame(Riego, Fertilizacion, CRA, pre, e, re)
head(res)
```

```
##           Riego Fertilizacion CRA  pre    e    re
## 1 cada 3 dias      con K  77 76.6  0.4  0.079
## 2 cada 3 dias      con K  76 76.6 -0.6 -0.118
## 3 cada 3 dias      con K  73 76.6 -3.6 -0.710
## 4 cada 3 dias      con K  75 76.6 -1.6 -0.316
## 5 cada 3 dias      con K  82 76.6  5.4  1.065
## 6 cada 3 dias      sin K  74 73.0  1.0  0.197
```

```
# Supuestos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED")
abline(0, 0)
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PR



```
## [1] 25 10
```

```
leveneTest(CRA ~ Riego * Fertilizacion, Datos, center = mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
```

```
##      Df F value Pr(>F)
```

```
## group 5  0.9349 0.4762
```

```
##      24
```

*Los supuestos de homocedasticidad y de normalidad parecieran cumplirse.**

- Ponga a prueba las hipótesis adecuadas en relación a los objetivos del ensayo. Concluya. Incluya magnitud del efecto. ¿Con algún tratamiento cree que se lograron condiciones que favorezcan la adaptación al estrés hídrico? ¿Recomendaría fertilización con K para mitigar los efectos del estrés hídrico?

```
modelo1 <- lm(CRA ~ Riego * Fertilizacion, data = Datos)
```

```
anova(modelo1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: CRA
```

```
##      Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## Riego      2 1619.47   809.73 25.1992 1.27e-06 ***
```

```
## Fertilizacion  1   90.13    90.13  2.8050  0.1070
```

```
## Riego:Fertilizacion 2    9.87    4.93  0.1535  0.8585
```

```
## Residuals    24   771.20   32.13
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

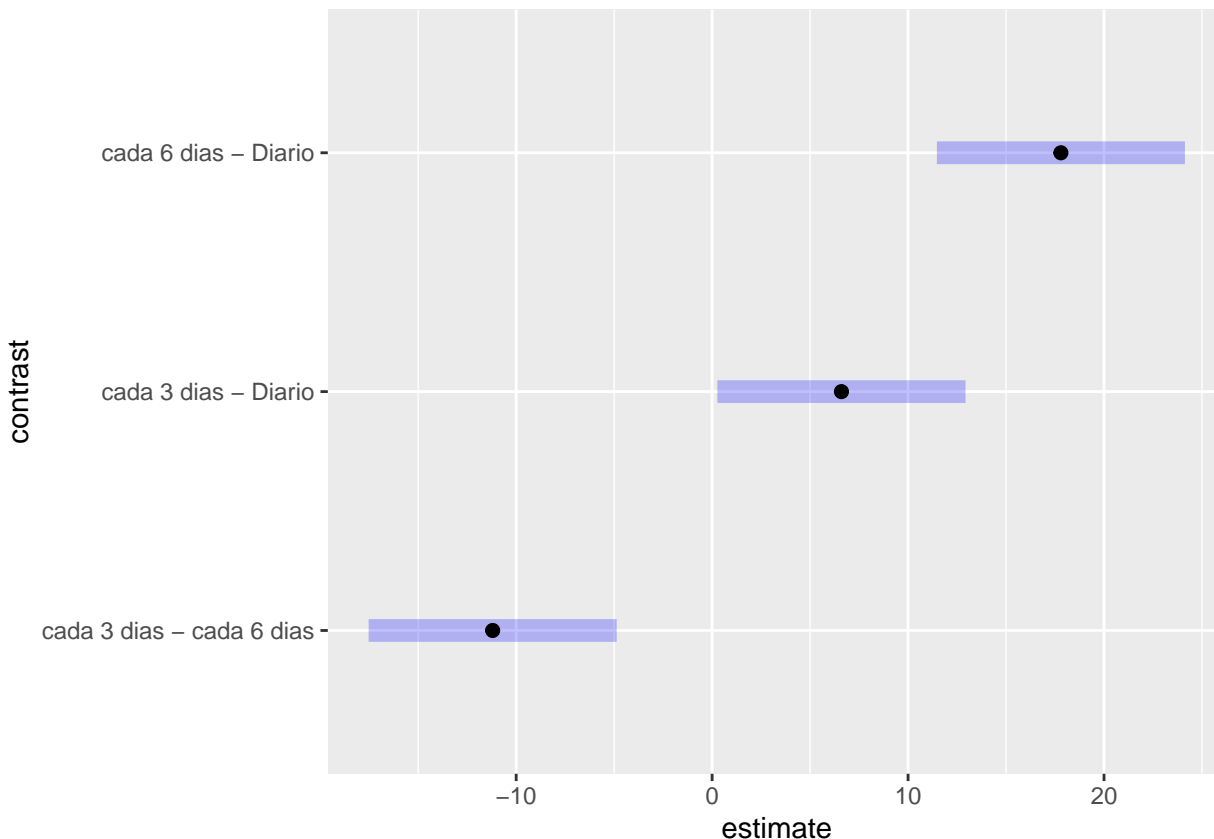
```
(comparacion <- emmeans(modelo1, pairwise ~ Riego))
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
```

```
## Riego      emmean    SE df lower.CL upper.CL t.ratio p.value
```

```
## cada 3 dias 74.8 1.79 24 71.1 78.5 41.728 <.0001
## cada 6 dias 86.0 1.79 24 82.3 89.7 47.976 <.0001
## Diario 68.2 1.79 24 64.5 71.9 38.046 <.0001
##
## Results are averaged over the levels of: Fertilizacion
## Confidence level used: 0.95
##
## $contrasts
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## cada 3 dias - cada 6 dias -11.2 2.54 24 -17.531 -4.87 -4.418 0.0005
## cada 3 dias - Diario 6.6 2.54 24 0.269 12.93 2.603 0.0399
## cada 6 dias - Diario 17.8 2.54 24 11.469 24.13 7.021 <.0001
##
## Results are averaged over the levels of: Fertilizacion
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
plot(comparacion$contrasts)
```



El estrés hídrico genera diferencias en el CRA. En particular, los grupos de riego cada 3 y 6 días mostraron valores de CRA superiores al grupo de riego diario (ver Tabla 1). La fertilización con K no parece haber generado diferencias en cuanto a la adaptación al estrés hídrico por lo que no sería recomendado para mitigar sus efectos.

- Analice los datos considerando a la variable frecuencia de riego como cuantitativa (1, 3 y 6 días). ¿Cómo plantearía el modelo? ¿Esperaría los mismos resultados que en el abordaje analítico anterior? Ajuste este nuevo el modelo y concluya.

```
# pasamos riego a cuantitativa
head(Datos)
```

```
##          Riego Fertilizacion CRA
## 1 cada 3 dias          con K 77
```

```
## 2 cada 3 dias      con K  76
## 3 cada 3 dias      con K  73
## 4 cada 3 dias      con K  75
## 5 cada 3 dias      con K  82
## 6 cada 3 dias      sin K  74

# vemos como se llaman los niveles
levels(Datos$Riego)

## [1] "cada 3 dias" "cada 6 dias" "Diario"

# ahora reemplazamos en la base de datos directamente
levels(Datos$Riego)[levels(Datos$Riego) == levels(Datos$Riego)[1]] <- 3
levels(Datos$Riego)[levels(Datos$Riego) == levels(Datos$Riego)[2]] <- 6
levels(Datos$Riego)[levels(Datos$Riego) == levels(Datos$Riego)[3]] <- 1
# vemos que aun son factores
str(Datos)

## 'data.frame':    30 obs. of  3 variables:
## $ Riego          : Factor w/ 3 levels "3","6","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ Fertilizacion: Factor w/ 2 levels "con K","sin K": 1 1 1 1 1 2 2 2 2 2 ...
## $ CRA           : int  77 76 73 75 82 74 73 79 77 62 ...

# ahora los pasamos a numeric
Datos$Riego = as.numeric(levels(Datos$Riego))[Datos$Riego]
str(Datos)

## 'data.frame':    30 obs. of  3 variables:
## $ Riego          : num  3 3 3 3 3 3 3 3 3 3 ...
## $ Fertilizacion: Factor w/ 2 levels "con K","sin K": 1 1 1 1 1 2 2 2 2 2 ...
## $ CRA           : int  77 76 73 75 82 74 73 79 77 62 ...
```

$$CRA_i = \beta_0 + \beta_1 * dias_i + \beta_2 * [K] + \beta_3 * dias_i * [K] + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

$$i = 1 : n(32)$$

```
modelo2 <- lm(CRA ~ Riego * Fertilizacion, data = Datos)
summary(modelo2)
```

```
##
## Call:
## lm(formula = CRA ~ Riego * Fertilizacion, data = Datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.5000  -3.2750   0.3579   4.4750   7.6000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      65.2421     2.6835  24.312 < 2e-16 ***
## Riego             3.8474     0.6853   5.614 6.7e-06 ***
## Fertilizacionsin K -1.6421     3.7951  -0.433  0.669
## Riego:Fertilizacionsin K -0.5474     0.9692  -0.565  0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.454 on 26 degrees of freedom
## Multiple R-squared:  0.6895, Adjusted R-squared:  0.6537
## F-statistic: 19.25 on 3 and 26 DF, p-value: 8.802e-07
```



```
(riegoLCL <- confint(modelo2)[2])
```

```
## [1] 2.438692
```

```
(riegoHCL <- confint(modelo2)[6])
```

```
## [1] 5.256044
```

Por cada disminución en la frecuencia de riego en un día se espera un efecto de aumento de CRA de entre 2.4386924 y 5.2560444.

Problema 2. Regulación de la producción de aceites esenciales en cedrón *Aloysia citriodora*

El cedrón (*Aloysia citriodora*) ocupa un lugar trascendente en el mercado herbario debido a los atributos sensoriales y medicinales de su aceite esencial. La producción de aceite esencial está regulada por condiciones ambientales que generan estrés en la planta, como la competencia entre plantas de cedrón por recursos, disponibilidad de agua y radiación solar. Se cree que frente al estrés la planta destina mayor cantidad de fotoasimilados a la producción de compuestos secundarios como los aceites esenciales. Una investigadora evaluó el efecto de la competencia (sin competencia, es decir una planta por maceta, y con competencia, es decir 5 plantas por maceta) en tres genotipos correspondientes a tres localidades distintas (Buenos Aires, Mendoza y San Luis) sobre el rendimiento de aceite esencial. Para ello dispuso 30 macetas con sustrato uniforme, a las cuales les asignó al azar los tratamientos. Luego de 6 meses se midió la concentración de aceite en tallo y hojas (ml de aceite/100gr de MS). La investigadora desea responder las siguientes preguntas:

- ¿El efecto de la competencia sobre la producción de aceites esenciales es el mismo en los tres genotipos?
- Si el efecto de la competencia no es el mismo para los tres genotipos, ¿cuál es el genotipo que resulta más afectado y cuál menos?
- Si un productor decide trabajar con alta densidad de plantas ¿qué genotipo le recomendaría para maximizar la producción de aceites esenciales?

Los datos se encuentran en el archivo *Cedron.txt*.

- Explique claramente cuál es la unidad experimental, la variable respuesta y las variables explicatorias. Indique la cantidad de réplicas del ensayo. Describa los datos.

U.E: cada maceta con sustrato con una determinada cantidad de plantas.

v.a: [ml de aceite/100gr masa seca].

v.e: niveles de competencia (con, sin) y genotipo (mendoza, san luis y buenos aires).

replicas: N/(32), osea 5.

Diseño: factorial, 23 (DCA).

```
library(ggplot2) # Para hacer gráficos de interacción
# library(car) # Para prueba de Levene
library(emmeans)
Datos <- read.delim("/home/jose/Documents/materias/biome2/2019/tps/tp4/Cedron.txt",
  header = T)
attach(Datos)
names(Datos)
```

```
## [1] "provincia" "densidad" "aceite"
```

```
str(Datos)
```

```
## 'data.frame': 30 obs. of 3 variables:
## $ provincia: Factor w/ 3 levels "ba","mza","sl": 1 1 1 1 1 1 1 1 1 1 ...
## $ densidad : Factor w/ 2 levels "1 planta por maceta",...: 1 1 1 1 1 2 2 2 2 2 ...
## $ aceite : num 0.99 0.99 1.11 0.82 1.11 1.35 1.41 1.45 1.21 1.18 ...
```

```
library(psych)
describeBy(Datos$aceite, list(Datos$densidad, Datos$provincia))
```

```
##
## Descriptive statistics by group
## : 1 planta por maceta
```

```
## : ba
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 1 0.12 0.99 1 0.18 0.82 1.11 0.29 -0.46 -1.61 0.05
## -----
## : 5 plantas por maceta
## : ba
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 1.32 0.12 1.35 1.32 0.15 1.18 1.45 0.27 -0.13 -2.15 0.05
## -----
## : 1 planta por maceta
## : mza
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 1.43 0.13 1.38 1.43 0.12 1.3 1.59 0.29 0.22 -2.11 0.06
## -----
## : 5 plantas por maceta
## : mza
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 1.84 0.12 1.79 1.84 0.09 1.73 2.04 0.31 0.76 -1.29 0.05
## -----
## : 1 planta por maceta
## : sl
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 1.34 0.13 1.29 1.34 0.07 1.24 1.56 0.32 0.81 -1.27 0.06
## -----
## : 5 plantas por maceta
## : sl
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 5 1.3 0.18 1.29 1.3 0.27 1.07 1.48 0.41 -0.12 -2 0.08
```

```
# medias
```

```
tapply(Datos$aceite, list(Datos$densidad, Datos$provincia), mean)
```

```
## ba mza sl
## 1 planta por maceta 1.004 1.432 1.338
## 5 plantas por maceta 1.320 1.838 1.302
```

```
# desvío estándar
```

```
round(tapply(Datos$aceite, list(Datos$densidad, Datos$provincia),
sd), 2)
```

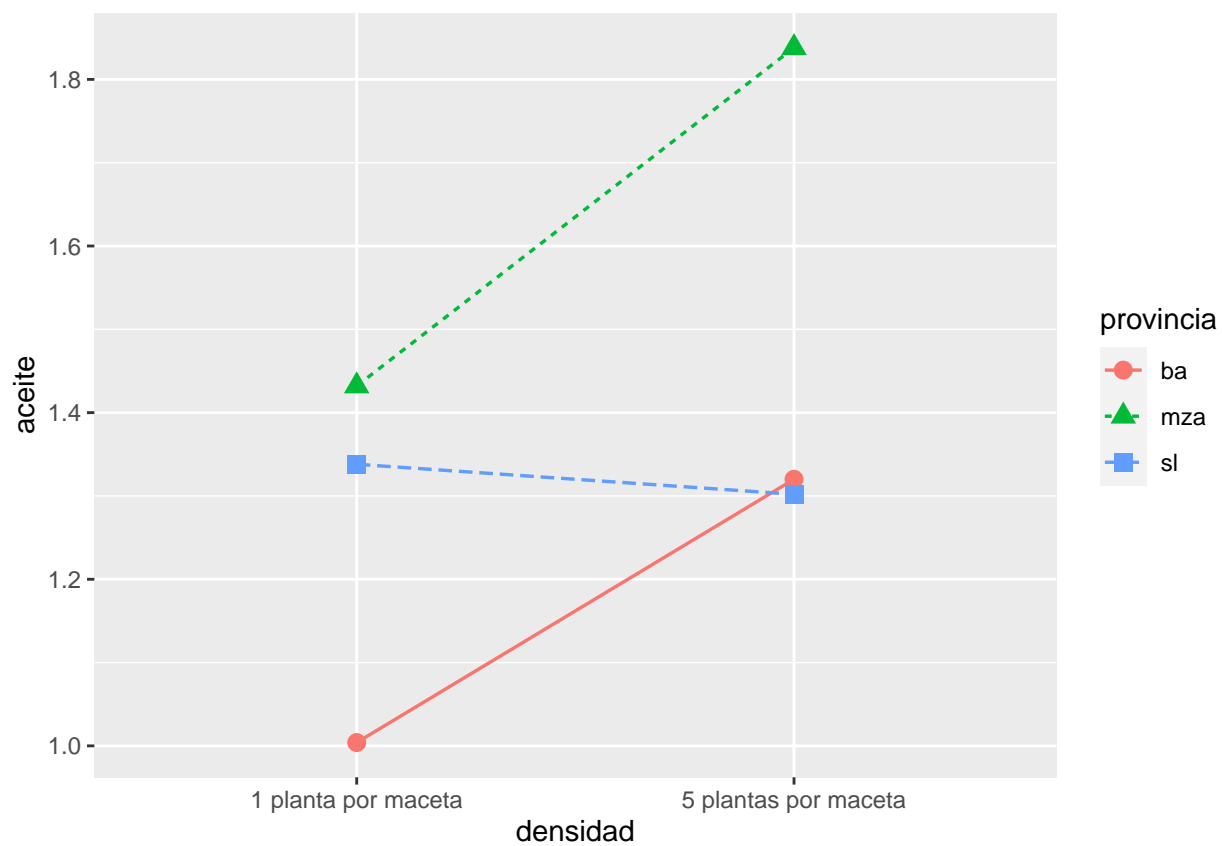
```
## ba mza sl
## 1 planta por maceta 0.12 0.13 0.13
## 5 plantas por maceta 0.12 0.12 0.18
```

```
# Graficos de perfiles
```

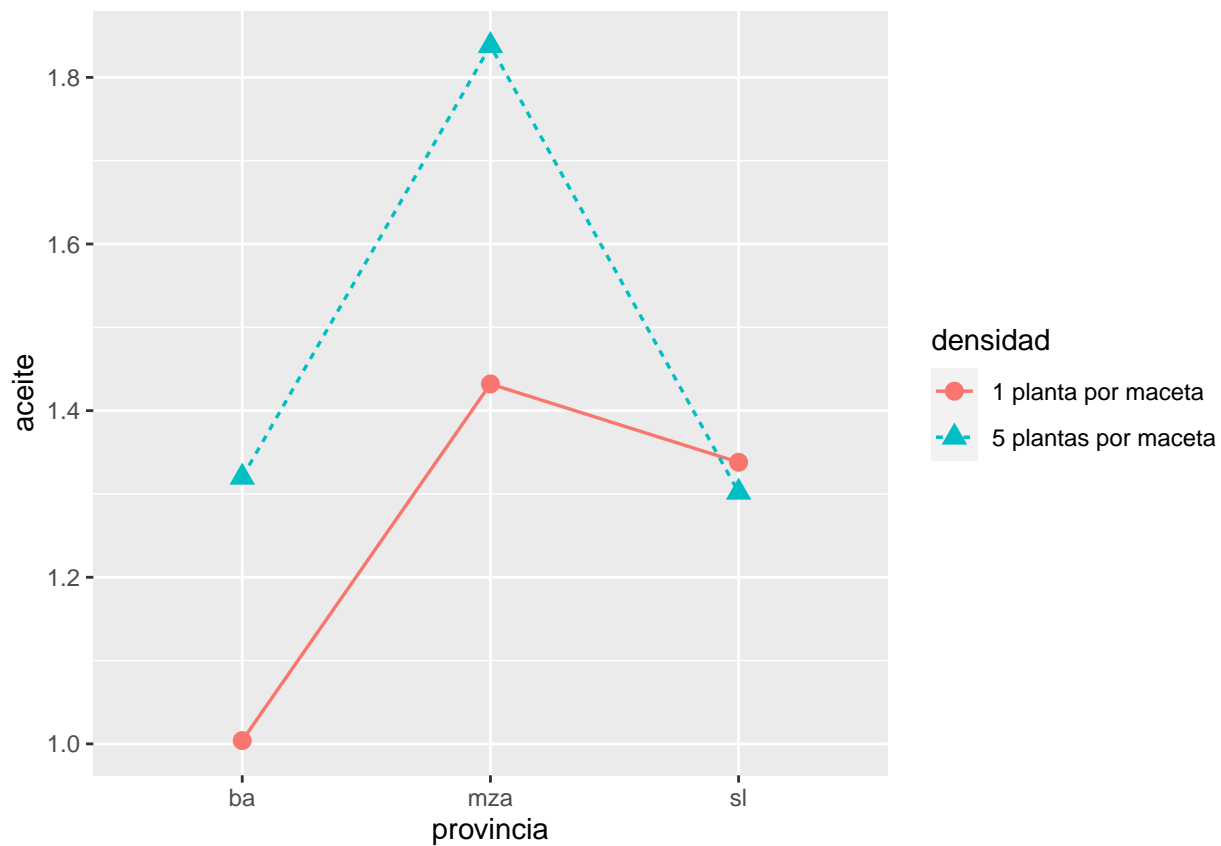
```
(medias.Datos <- aggregate(aceite ~ densidad + provincia, Datos, mean))
```

```
## densidad provincia aceite
## 1 1 planta por maceta ba 1.004
## 2 5 plantas por maceta ba 1.320
## 3 1 planta por maceta mza 1.432
## 4 5 plantas por maceta mza 1.838
## 5 1 planta por maceta sl 1.338
## 6 5 plantas por maceta sl 1.302
```

```
gp <- ggplot(medias.Datos, aes(x = densidad, y = aceite, colour = provincia,
group = provincia))
gp + geom_line(aes(linetype = provincia), size = 0.6) + geom_point(aes(shape = provincia),
size = 3)
```

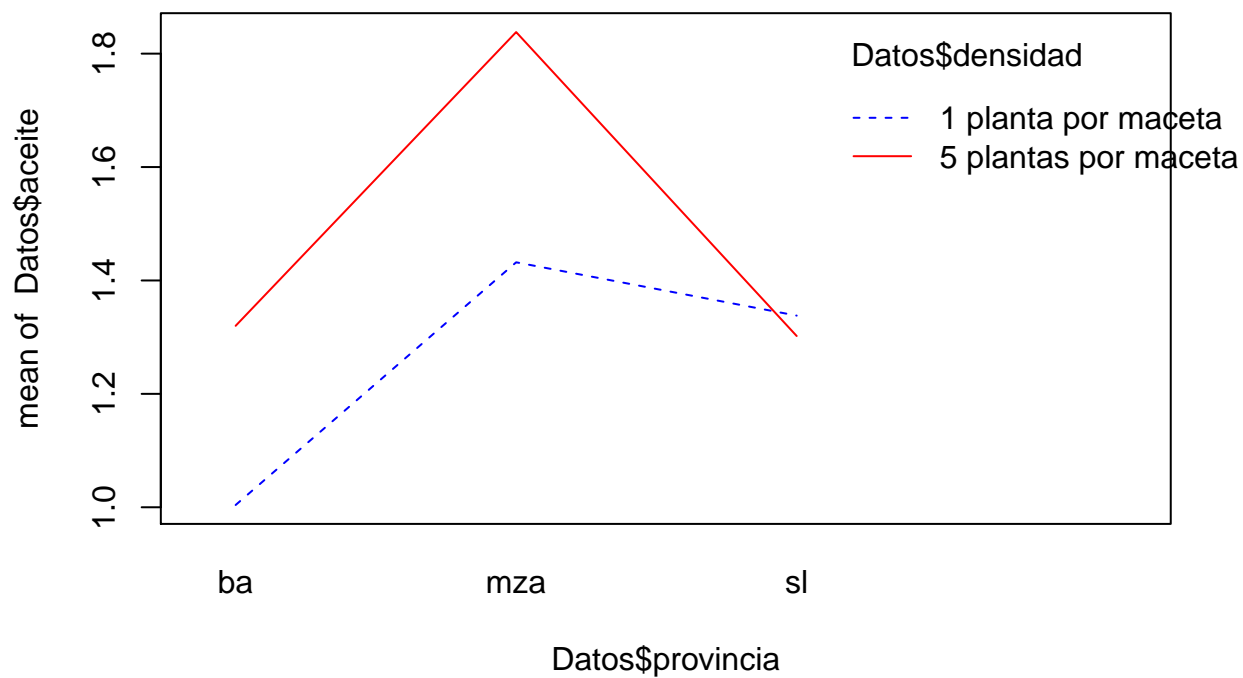


```
gp1 <- ggplot(medias.Datos, aes(x = provincia, y = aceite, colour = densidad,  
  group = densidad))  
gp1 + geom_line(aes(linetype = densidad), size = 0.6) + geom_point(aes(shape = densidad),  
  size = 3)
```



Otra opción gráfica

```
interaction.plot(Datos$provincia, Datos$densidad, Datos$aceite, col = c(4,
2))
```



- Especifique el modelo en términos del problema.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha * \beta_{ij} + \epsilon_{ijk}$$

$$\epsilon_{ijk} \sim NID(0, \sigma^2)$$

y_{ijk} = concentración de aceite en tallos y hojas

α_i = efecto de la competencia, 1 : 2

β_j = efecto del genotipo, 1 : 30

$k = 1 : 5$ (refiere a la replica)

```
modelo1 <- lm(aceite ~ densidad + provincia + densidad:provincia,
  data = Datos)
modelo2 <- lm(aceite ~ densidad * provincia, data = Datos)
# con el signo * se incluyen los factores individuales y la
# interacción Es equivalente la forma de ingreso de ambos
# modelos?
anova(modelo1)
```

```
## Analysis of Variance Table
##
## Response: aceite
##
##          Df Sum Sq Mean Sq F value    Pr(>F)
## densidad    1  0.39216   0.39216   21.9167 9.333e-05 ***
## provincia    2  1.15973   0.57986   32.4067 1.516e-07 ***
## densidad:provincia 2  0.27281   0.13640    7.6231 0.002735 **
## Residuals   24  0.42944   0.01789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(modelo2)
```

```
## Analysis of Variance Table
##
## Response: aceite
##
##          Df Sum Sq Mean Sq F value    Pr(>F)
## densidad    1  0.39216   0.39216   21.9167 9.333e-05 ***
## provincia    2  1.15973   0.57986   32.4067 1.516e-07 ***
## densidad:provincia 2  0.27281   0.13640    7.6231 0.002735 **
## Residuals   24  0.42944   0.01789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- ¿Se verifican los principios de aleatoriedad, replicación y control del error? Fundamente su respuesta.

Aleatoriedad se cumple ya que se asigno al azar cada ue a los tratamientos. La replicacion tambien se verifica ya que hay 5 relicas por tratamiento. El control de error lo suponemos ya que la descripcion del disenio experimental no nos permite establecerlo.

-Verifique los supuestos del modelo.

```
e <- resid(modelo2) # residuos
re <- rstandard(modelo2) #residuos estandarizados
pre <- predict(modelo2) #predichos
res <- data.frame(Datos$provincia, Datos$densidad, Datos$aceite, pre,
  e, round(re, 3))
colnames(res) <- c("provincia", "densidad", "aceite", "Predichos",
  "Residuos", "residuos std")
head(res)
```

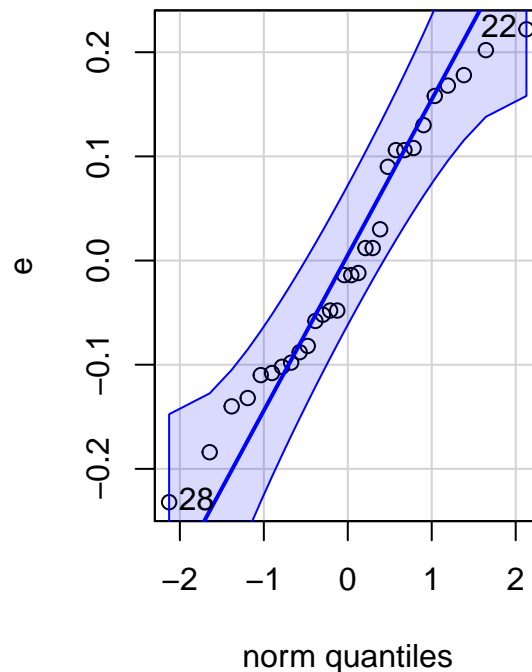
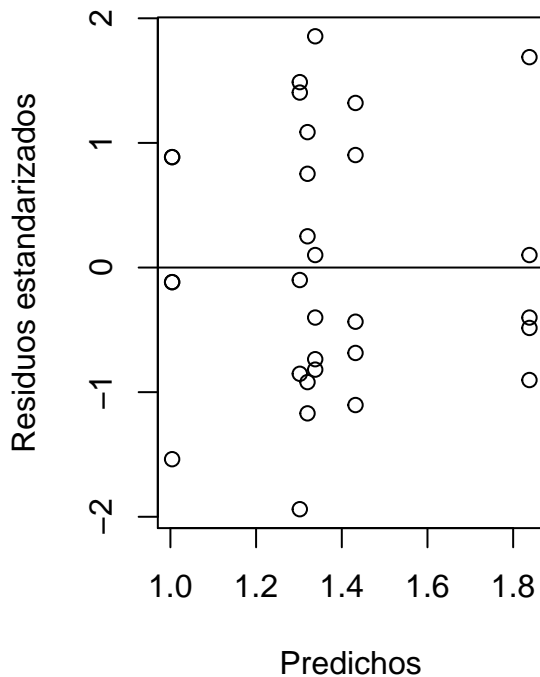
```
## provincia          densidad aceite Predichos Residuos residuos std
## 1      ba 1 planta por maceta 0.99      1.004 -0.014      -0.117
## 2      ba 1 planta por maceta 0.99      1.004 -0.014      -0.117
## 3      ba 1 planta por maceta 1.11      1.004  0.106       0.886
## 4      ba 1 planta por maceta 0.82      1.004 -0.184      -1.538
## 5      ba 1 planta por maceta 1.11      1.004  0.106       0.886
## 6      ba 5 plantas por maceta 1.35      1.320  0.030       0.251
```

```
sum(e)
```

```
## [1] 1.474515e-16
```

```
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED")
abline(0, 0)
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PR



```
## [1] 28 22
```

```
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
##
## data:  e
## W = 0.96125, p-value = 0.3332
```

```
par(mfrow = c(1, 1))
leveneTest(aceite ~ provincia * densidad, Datos, center = mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##      Df F value Pr(>F)
## group 5  0.4298 0.8233
```

24

- ¿Qué ventaja le aporta a la investigadora haber realizado un ensayo con los tratamientos en arreglo factorial?

La ventaja de hacer el ensayo con arreglo factorial es la potencia asociada a estos diseños. Permite poner a prueba hipótesis con una alta confianza.

- Identifique qué pruebas le permitirán responder a cada una de las preguntas de la investigadora.

a) ¿El efecto de la competencia sobre la producción de aceites esenciales es el mismo en los tres genotipos?

```
modelo2
```

```
##
## Call:
## lm(formula = aceite ~ densidad * provincia, data = Datos)
##
## Coefficients:
##                (Intercept)
##                   1.004
##      densidad5 plantas por maceta
##                   0.316
##      provinciamza
##                   0.428
##      provinciasl
##                   0.334
## densidad5 plantas por maceta:provinciamza
##                   0.090
## densidad5 plantas por maceta:provinciasl
##                   -0.352
```

```
anova(modelo2)
```

```
## Analysis of Variance Table
##
## Response: aceite
##              Df Sum Sq Mean Sq F value    Pr(>F)
## densidad      1  0.39216  0.39216  21.9167 9.333e-05 ***
## provincia     2  1.15973  0.57986  32.4067 1.516e-07 ***
## densidad:provincia 2  0.27281  0.13640   7.6231 0.002735 **
## Residuals    24  0.42944  0.01789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

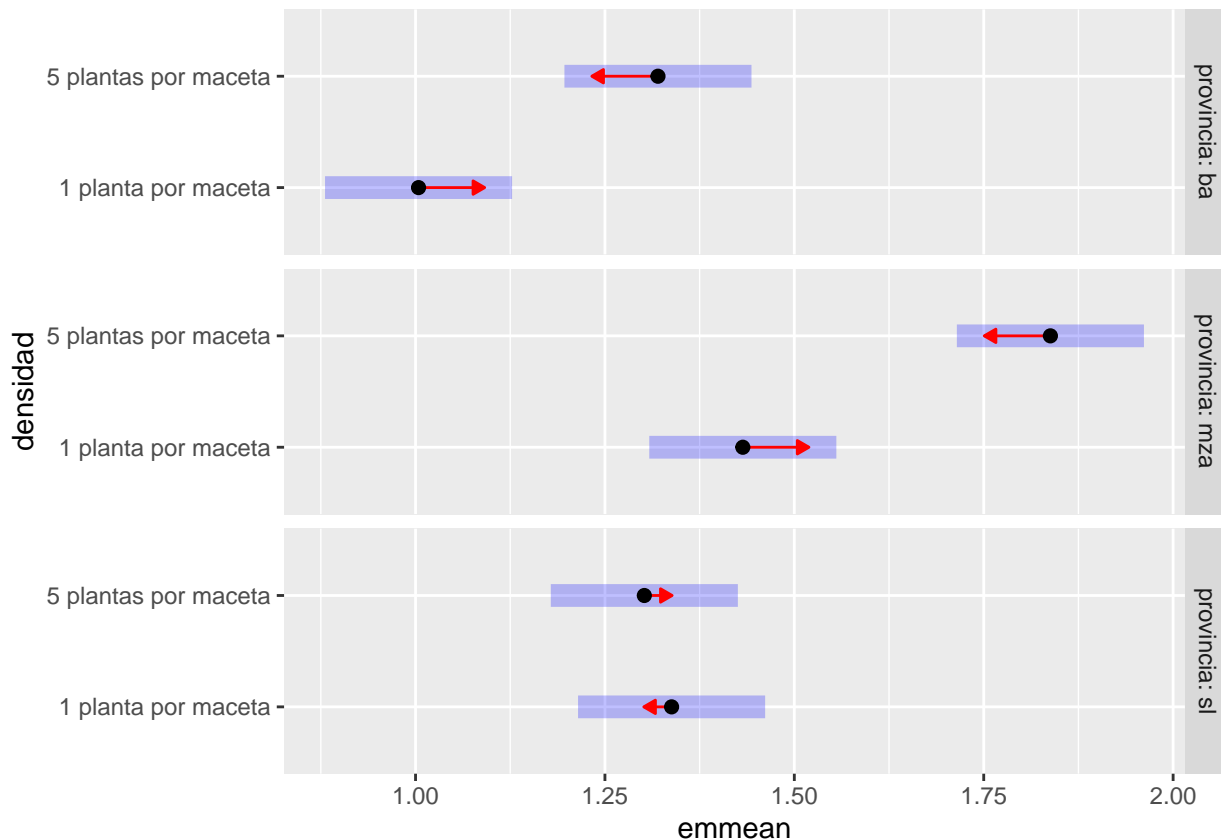
No, dado que la interacción es genotipo y provincia es significativa indicando que la competencia no afecta por igual a todos los genotipos.

b) Si el efecto de la competencia no es el mismo para los tres genotipos, ¿cuál es el genotipo que resulta más afectado y cuál menos?

```
(comparaciones1 <- emmeans(modelo2, pairwise ~ densidad | provincia))
```

```
## $emmeans
## provincia = ba:
##   densidad      emmean      SE df lower.CL upper.CL t.ratio p.value
## 1 planta por maceta    1.00 0.0598 24    0.881    1.13  16.783 <.0001
## 5 plantas por maceta    1.32 0.0598 24    1.197    1.44  22.065 <.0001
##
## provincia = mza:
##   densidad      emmean      SE df lower.CL upper.CL t.ratio p.value
## 1 planta por maceta    1.43 0.0598 24    1.309    1.56  23.938 <.0001
## 5 plantas por maceta    1.84 0.0598 24    1.715    1.96  30.725 <.0001
##
## provincia = sl:
```

```
## densidad      emmean      SE df lower.CL upper.CL t.ratio p.value
## 1 planta por maceta  1.34 0.0598 24   1.215    1.46  22.366 <.0001
## 5 plantas por maceta  1.30 0.0598 24   1.179    1.43  21.765 <.0001
##
## Confidence level used: 0.95
##
## $contrasts
## provincia = ba:
## contrast      estimate      SE df lower.CL
## 1 planta por maceta - 5 plantas por maceta  -0.316 0.0846 24   -0.491
## upper.CL t.ratio p.value
##   -0.141  -3.735  0.0010
##
## provincia = mza:
## contrast      estimate      SE df lower.CL
## 1 planta por maceta - 5 plantas por maceta  -0.406 0.0846 24   -0.581
## upper.CL t.ratio p.value
##   -0.231  -4.799  0.0001
##
## provincia = sl:
## contrast      estimate      SE df lower.CL
## 1 planta por maceta - 5 plantas por maceta   0.036 0.0846 24   -0.139
## upper.CL t.ratio p.value
##    0.211   0.426  0.6742
##
## Confidence level used: 0.95
plot(comparaciones1$emmeans, comparisons = TRUE)
```



Tanto buenos aires como mendoza ven reducida su produccion, mientras que san luis es la menos afectada.

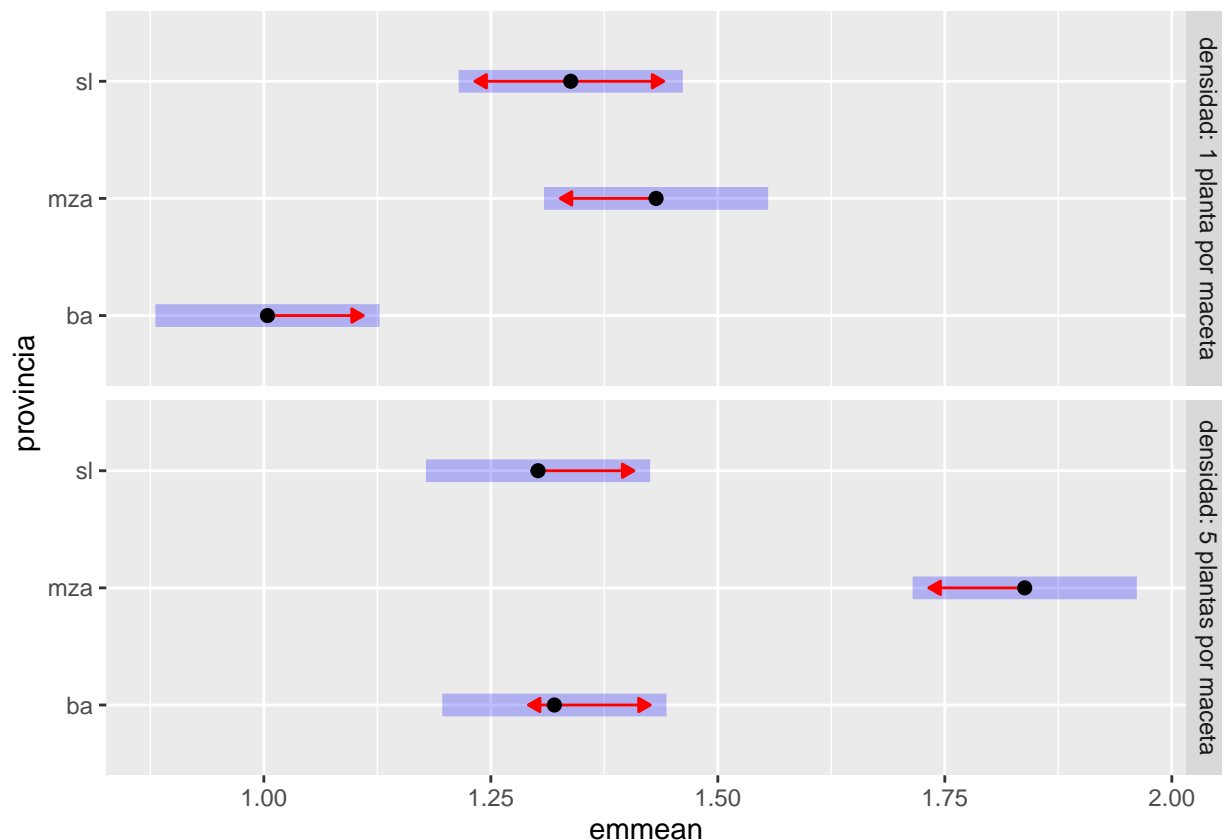
c) Si un productor decide trabajar con alta densidad de plantas ¿qué genotipo le recomendaría para maximizar la

producción de aceites esenciales?

Aca interesa solamente cual es la que tiene mayor produccion en alta densidad.

```
(comparaciones2 <- emmeans(modelo2, pairwise ~ provincia | densidad))
```

```
## $emmeans
## densidad = 1 planta por maceta:
##   provincia emmean      SE df lower.CL upper.CL t.ratio p.value
##   ba         1.00 0.0598 24    0.881    1.13   16.783 <.0001
##   mza         1.43 0.0598 24    1.309    1.56   23.938 <.0001
##   sl          1.34 0.0598 24    1.215    1.46   22.366 <.0001
##
## densidad = 5 plantas por maceta:
##   provincia emmean      SE df lower.CL upper.CL t.ratio p.value
##   ba         1.32 0.0598 24    1.197    1.44   22.065 <.0001
##   mza         1.84 0.0598 24    1.715    1.96   30.725 <.0001
##   sl          1.30 0.0598 24    1.179    1.43   21.765 <.0001
##
## Confidence level used: 0.95
##
## $contrasts
## densidad = 1 planta por maceta:
##   contrast estimate      SE df lower.CL upper.CL t.ratio p.value
##   ba - mza   -0.428 0.0846 24   -0.639   -0.217   -5.059 0.0001
##   ba - sl    -0.334 0.0846 24   -0.545   -0.123   -3.948 0.0017
##   mza - sl    0.094 0.0846 24   -0.117    0.305    1.111 0.5167
##
## densidad = 5 plantas por maceta:
##   contrast estimate      SE df lower.CL upper.CL t.ratio p.value
##   ba - mza   -0.518 0.0846 24   -0.729   -0.307   -6.123 <.0001
##   ba - sl     0.018 0.0846 24   -0.193    0.229    0.213 0.9754
##   mza - sl    0.536 0.0846 24    0.325    0.747    6.336 <.0001
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
plot(comparaciones2$emmeans, comparisons = TRUE)
```



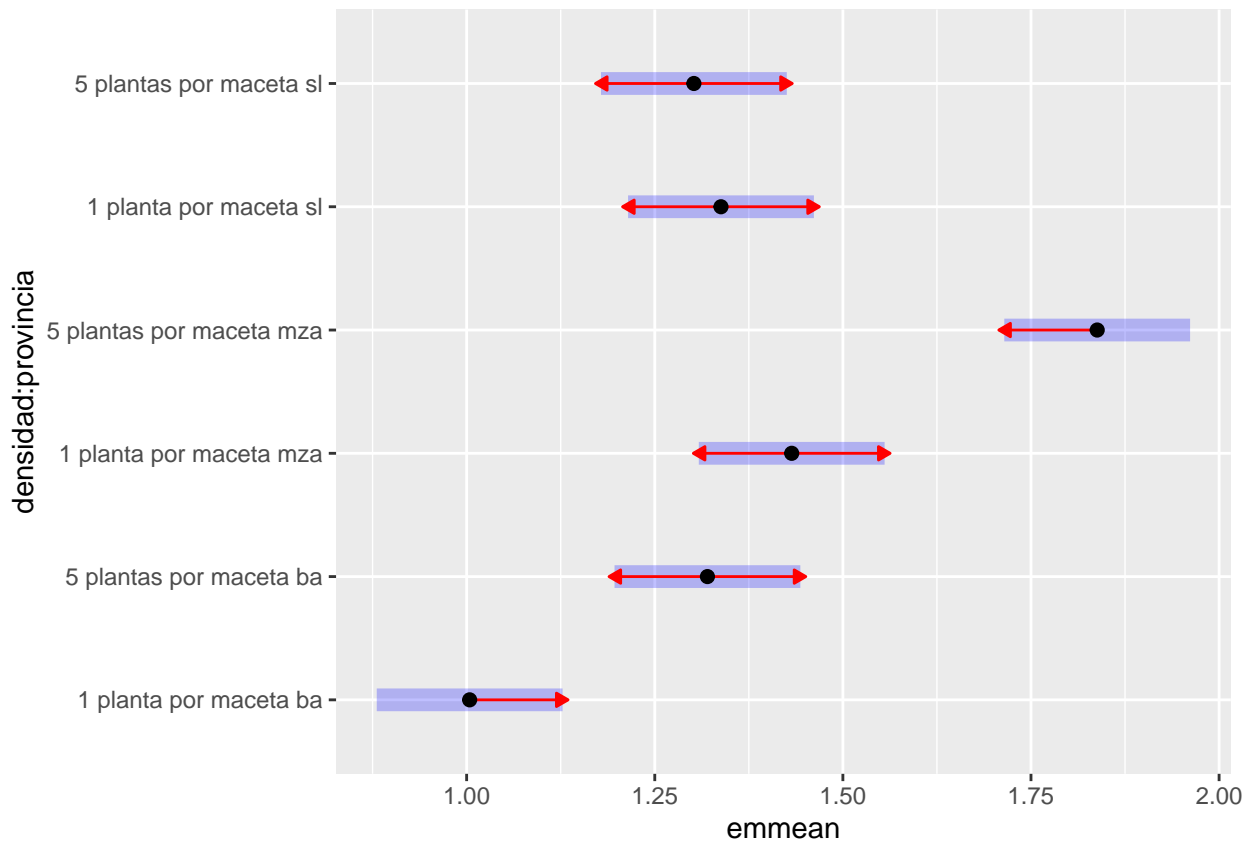
Vemos que mendoza es la que tiene la mayor producción en alta densidad y esta diferencia es mayor que para los otros genotipos. Recuerden que no pueden hacer estos efectos simples en el mismo estudio. Si debieran responder estas preguntas juntas, entonces solo podrían hacer tukey.

- Plantee las hipótesis y concluya en relación a las preguntas de la investigadora. Evalúe e informe la magnitud del efecto en valores absolutos y porcentuales para reforzar sus conclusiones.

```
Efectos_simples <- emmeans(modelo2, pairwise ~ densidad * provincia)
confint(Efectos_simples)
```

```
## $emmeans
## densidad      provincia emmean      SE df lower.CL upper.CL
## 1 planta por maceta ba      1.00 0.0598 24    0.881    1.13
## 5 plantas por maceta ba      1.32 0.0598 24    1.197    1.44
## 1 planta por maceta mza      1.43 0.0598 24    1.309    1.56
## 5 plantas por maceta mza      1.84 0.0598 24    1.715    1.96
## 1 planta por maceta sl      1.34 0.0598 24    1.215    1.46
## 5 plantas por maceta sl      1.30 0.0598 24    1.179    1.43
##
## Confidence level used: 0.95
##
## $contrasts
## contrast                                     estimate      SE df lower.CL
## 1 planta por maceta ba - 5 plantas por maceta ba      -0.316 0.0846 24   -0.578
## 1 planta por maceta ba - 1 planta por maceta mza      -0.428 0.0846 24   -0.690
## 1 planta por maceta ba - 5 plantas por maceta mza      -0.834 0.0846 24  -1.096
## 1 planta por maceta ba - 1 planta por maceta sl       -0.334 0.0846 24   -0.596
## 1 planta por maceta ba - 5 plantas por maceta sl       -0.298 0.0846 24   -0.560
## 5 plantas por maceta ba - 1 planta por maceta mza      -0.112 0.0846 24   -0.374
## 5 plantas por maceta ba - 5 plantas por maceta mza      -0.518 0.0846 24   -0.780
## 5 plantas por maceta ba - 1 planta por maceta sl       -0.018 0.0846 24   -0.280
## 5 plantas por maceta ba - 5 plantas por maceta sl       0.018 0.0846 24   -0.244
```

```
## 1 planta por maceta mza - 5 plantas por maceta mza -0.406 0.0846 24 -0.668
## 1 planta por maceta mza - 1 planta por maceta sl 0.094 0.0846 24 -0.168
## 1 planta por maceta mza - 5 plantas por maceta sl 0.130 0.0846 24 -0.132
## 5 plantas por maceta mza - 1 planta por maceta sl 0.500 0.0846 24 0.238
## 5 plantas por maceta mza - 5 plantas por maceta sl 0.536 0.0846 24 0.274
## 1 planta por maceta sl - 5 plantas por maceta sl 0.036 0.0846 24 -0.226
## upper.CL
## -0.0544
## -0.1664
## -0.5724
## -0.0724
## -0.0364
## 0.1496
## -0.2564
## 0.2436
## 0.2796
## -0.1444
## 0.3556
## 0.3916
## 0.7616
## 0.7976
## 0.2976
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 6 estimates
plot(Efectos_simples, comparisons = TRUE)
```



Problema 3. Efecto de la granivoría sobre *Ipomopsis gossypifera*

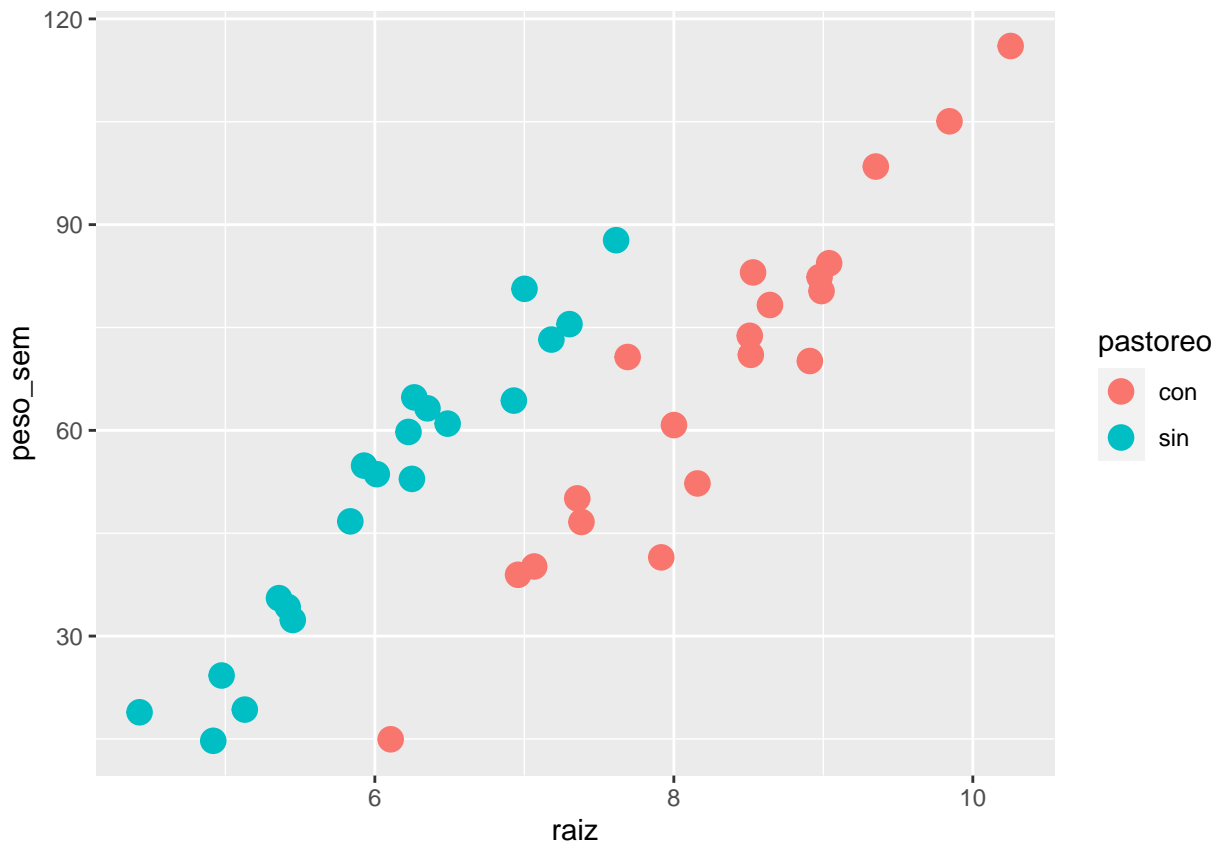
La herbivoría puede producir diferentes efectos sobre las plantas depredadas, desde la muerte de la planta hasta la sobrecompensación, un incremento en la biomasa de las plantas depredadas (respecto a la de las no depredadas),

frecuentemente a expensas de la asignación de recursos a reproducción. Se hizo un experimento de clausura para evaluar el efecto de la granivoría por ratones sobre la producción de semillas de una especie de planta bianual (*Ipomopsis* sp.). Se dispusieron un total de 40 plantas al azar en un área; la mitad de ellas estuvieron en situación de clausura desde el comienzo de la experiencia; mientras que a las otras 20 se las dejó con libre acceso para los ratones granívoros las primeras dos semanas en que estaban elongando su tallo, y después fueron clausuradas para permitir que rebroten. Dado que se sabe que el tamaño de las plantas puede estar relacionado con su producción de semillas, se midió el diámetro de la parte superior de la raíz de cada planta al inicio de la experiencia. Al final de la temporada de crecimiento se midió el peso seco de los frutos producidos por cada planta (en mg). Los datos están disponibles en el archivo *Ipomopsis.txt*.

```
datos <- read.delim("/home/jose/Documents/materias/biome2/2019/tps/tp4/Ipomopsis.txt",
  header = T)
library(ggplot2) # Para hacer gráficos de interacción
```

- Calcule el peso medio de las semillas producidas por planta para cada tratamiento. Haga un gráfico de dispersión de peso de semillas en función del diámetro de la raíz para los dos grupos (con y sin clausura total de ratones).

```
ggplot(datos, aes(x = raiz, y = peso_sem, colour = pastoreo)) + geom_point(size = 4)
```



```
res = data.frame()
for (i in 1:length(levels(datos$pastoreo))) {
  res[levels(datos$pastoreo)[i], 1] = mean(datos[datos$pastoreo ==
    levels(datos$pastoreo)[i], 2])
}
res
```

```
##          V1
## con 67.9405
## sin 50.8805
```

- Indique cual es la variable dependiente o respuesta y la/s variables explicatorias. ¿De que tipo es cada una? ¿Qué predicciones efectuaria en función de la informacion presentada?

U.E: cada parcela.

v.a: peso de semillas (mg).

v.e: cuantitativa continua, diametro de la raiz (unidades?) y cualitativa discreta niveles de pastoreo (con y sin).
replicas: 20.

Disenio: factorial, (DCA).

Observado el grafico de dispersion de las variables se observa un incremento en el peso de las semillas en el area con clausura, sin embargo no hay evidencias de una interaccion entre la condicion de clausura y el diametro de la raiz.

- Escriba el modelo teórico en términos del problema.

$$E(\text{Peso semillas}_i) = \beta_0 + \beta_1 * [\text{diametro raiz}] + \beta_2 * [\text{pastoreo}] + \beta_3 * [\text{diametro raiz}] * [\text{pastoreo}]$$

$$\text{Peso semillas}_i \sim N(0, \sigma^2)$$

$$i = 1 : n(40)$$

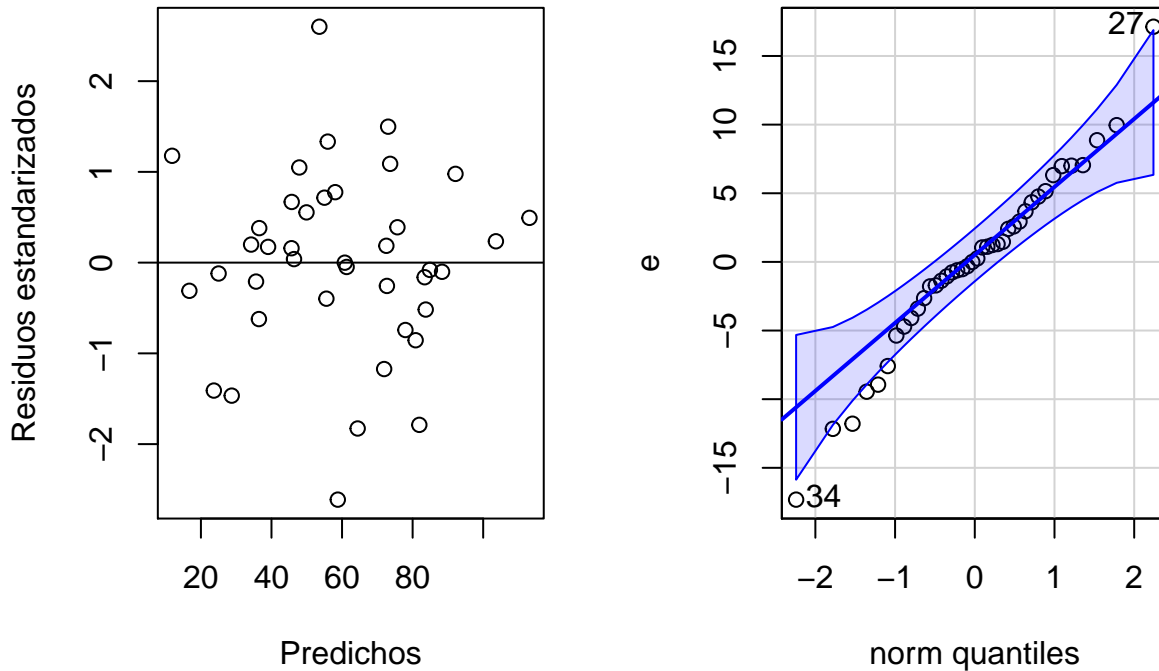
- Evalúe si hay una relación significativa entre el diámetro inicial de la raíz de una planta y el peso de las semillas que produjo al madurar. ¿Esa relación difiere entre el grupo clausurado y el no clausurado? Calcule las ecuaciones de regresión lineal; si la interacción fuera no significativa compare las ecuaciones a partir de los coeficientes del modelo que la incluye y de los de un modelo simplificado que no la incluya. ¿Por que es valido, en este caso evaluar, un modelo sin interaccion?

```
# planteo el modelo con interaccion
modelo1 <- lm(peso_sem ~ raiz * pastoreo, datos)
summary(modelo1)

##
## Call:
## lm(formula = peso_sem ~ raiz * pastoreo, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3177  -2.8320   0.1247   3.8511  17.1313
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -125.173     12.811  -9.771 1.15e-11 ***
## raiz             23.240       1.531  15.182 < 2e-16 ***
## pastoreosin     30.806     16.842   1.829  0.0757 .
## raiz:pastoreosin  0.756       2.354   0.321  0.7500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.831 on 36 degrees of freedom
## Multiple R-squared:  0.9293, Adjusted R-squared:  0.9234
## F-statistic: 157.6 on 3 and 36 DF,  p-value: < 2.2e-16

# Supuestos
e <- resid(modelo1) # residuos
re <- rstandard(modelo1) #residuos estandarizados
pre <- predict(modelo1) #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
      main = "Gráfico de dispersión de RE vs PRED")
abline(0, 0)
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PR



```
## [1] 34 27
```

```
shapiro.test(e)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  e
## W = 0.97642, p-value = 0.559
```

```
round(confint(modelo1), 2)
```

```
##              2.5 % 97.5 %
## (Intercept) -151.16 -99.19
## raiz         20.14  26.34
## pastoreosin  -3.35  64.96
## raiz:pastoreosin -4.02  5.53
```

```
summary(modelo1)$adj.r.squared
```

```
## [1] 0.9233645
```

```
AIC(modelo1)
```

```
## [1] 273.0135
```

```
anova(modelo1) # NS, (p=0.75), puedo no incluir la interacción (raiz no es trat)
```

```
## Analysis of Variance Table
```

```
##
```

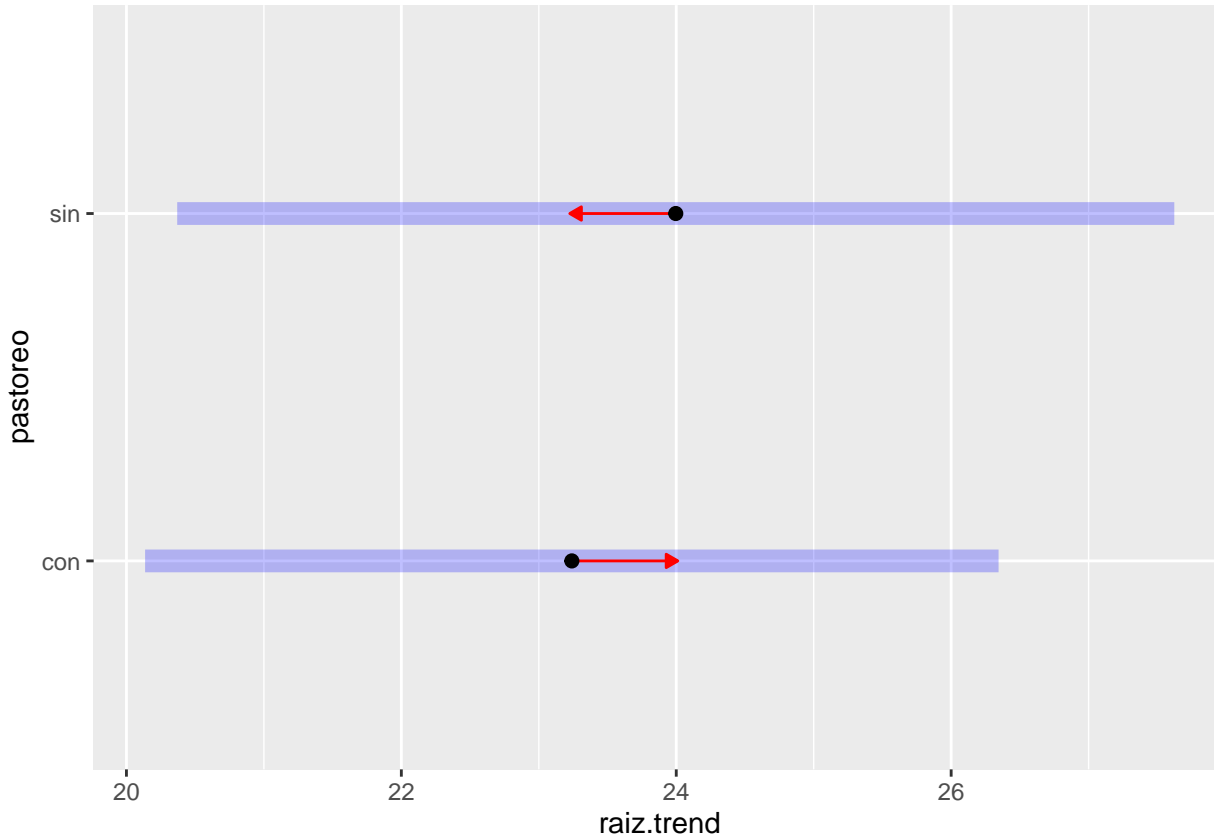
```
## Response: peso_sem
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## raiz           1 16795.0 16795.0 359.9681 < 2.2e-16 ***
## pastoreo       1  5264.4  5264.4 112.8316 1.209e-12 ***
## raiz:pastoreo  1     4.8     4.8   0.1031    0.75
```

```
## Residuals      36 1679.6    46.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Si quisiera utilizar este modelo y comparar pendientes
comp_pendientes <- emtrends(modelo1, ~pastoreo, var = "raiz") #, contr='cld'
pairs(comp_pendientes)

## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## con - sin    -0.756 2.35 36    -5.53     4.02  -0.321  0.7500
##
## Confidence level used: 0.95
plot(comp_pendientes, comparisons = TRUE)
```



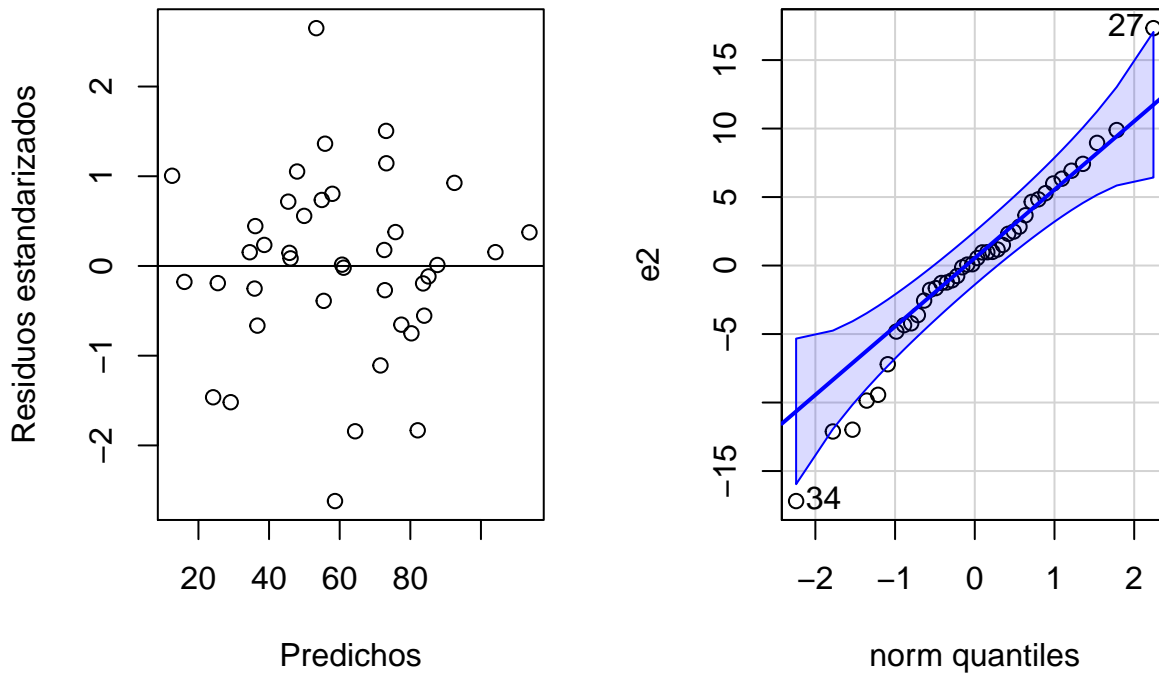
```
# aditivo
modelo2 <- lm(peso_sem ~ raiz + pastoreo, datos)
summary(modelo2)

##
## Call:
## lm(formula = peso_sem ~ raiz + pastoreo, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.1920  -2.8224   0.3223   3.9144  17.3290
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -127.829     9.664  -13.23 1.35e-15 ***
## raiz         23.560     1.149   20.51 < 2e-16 ***
## pastoreosin  36.103     3.357   10.75 6.11e-13 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.747 on 37 degrees of freedom
## Multiple R-squared:  0.9291, Adjusted R-squared:  0.9252
## F-statistic: 242.3 on 2 and 37 DF,  p-value: < 2.2e-16
```

```
# Supuestos
e2 <- resid(modelo2) # residuos
re2 <- rstandard(modelo2) #residuos estandarizados
pre2 <- predict(modelo2) #predichos
par(mfrow = c(1, 2))
plot(pre2, re2, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED")
abline(0, 0)
car::qqPlot(e2)
```

Gráfico de dispersión de RE vs PR



```
## [1] 27 34
```

```
shapiro.test(e2)
```

```
##
## Shapiro-Wilk normality test
##
## data:  e2
## W = 0.97358, p-value = 0.4637
```

```
round(confint(modelo2), 2)
```

```
##           2.5 % 97.5 %
## (Intercept) -147.41 -108.25
## raiz        21.23  25.89
## pastoreosin 29.30  42.91
```



```
summary(modelo2)$adj.r.squared
```

```
## [1] 0.9252221
```

```
AIC(modelo2)
```

```
## [1] 271.1279
```

```
# comparo los dos modelos
```

```
AIC(modelo1, modelo2)
```

```
##           df           AIC
## modelo1    5 273.0135
## modelo2    4 271.1279
```

En este caso es valido estimar un modelo sin interaccion porque el diametro de la raiz no es un tratamiento impuesto por el que realiza el estudio.

- ¿Cuál es el efecto de la herbivoría si comparara a ambos grupos directamente? ¿Cuál es el efecto de la herbivoría una vez descontado el efecto que tiene el tamaño de la planta sobre la producción de frutos?
- Escriba el modelo final seleccionado. ¿Qué porcentaje de la variabilidad en la producción de semillas de *Ipomopsis gossypifera* no explica?

Problema 4. Caracterización de de aguas residuales

```
datos <- read.delim("/home/jose/Documents/materias/biome2/2020/tps/tp4/aguas.txt",
  header = T)
library(ggplot2) # Para hacer gráficos de interacción
```

En el tratamiento de aguas residuales, la demanda bioquímica de oxígeno (DBO) es considerada como un buen parámetro para determinar el grado de contaminación del agua; a mayor DBO menor calidad. Mide la cantidad de O₂ consumido al degradar la materia orgánica de una muestra líquida y se expresa en mg O₂/l.). Como paso previo para el diseño de técnicas de depuración de aguas residuales se llevó a cabo un estudio con el objetivo de relacionar la DBO del agua residual respecto al pH de la misma. Se recolectaron seis muestras de tres tipos de aguas residuales, provenientes de tres fuentes distintas (A, B o C). Los resultados se encuentran en el archivo *aguas.txt*

-Indique cual es la variable respuesta y la/s variables explicatorias. ¿De qué tipo es cada una? Escriba el modelo teorico en terminos del problema.

v.a: DBO en muestra de agua residual.

v.e: fuente (A, B, C, cualitativa discreta) y ph (cuantitativa).

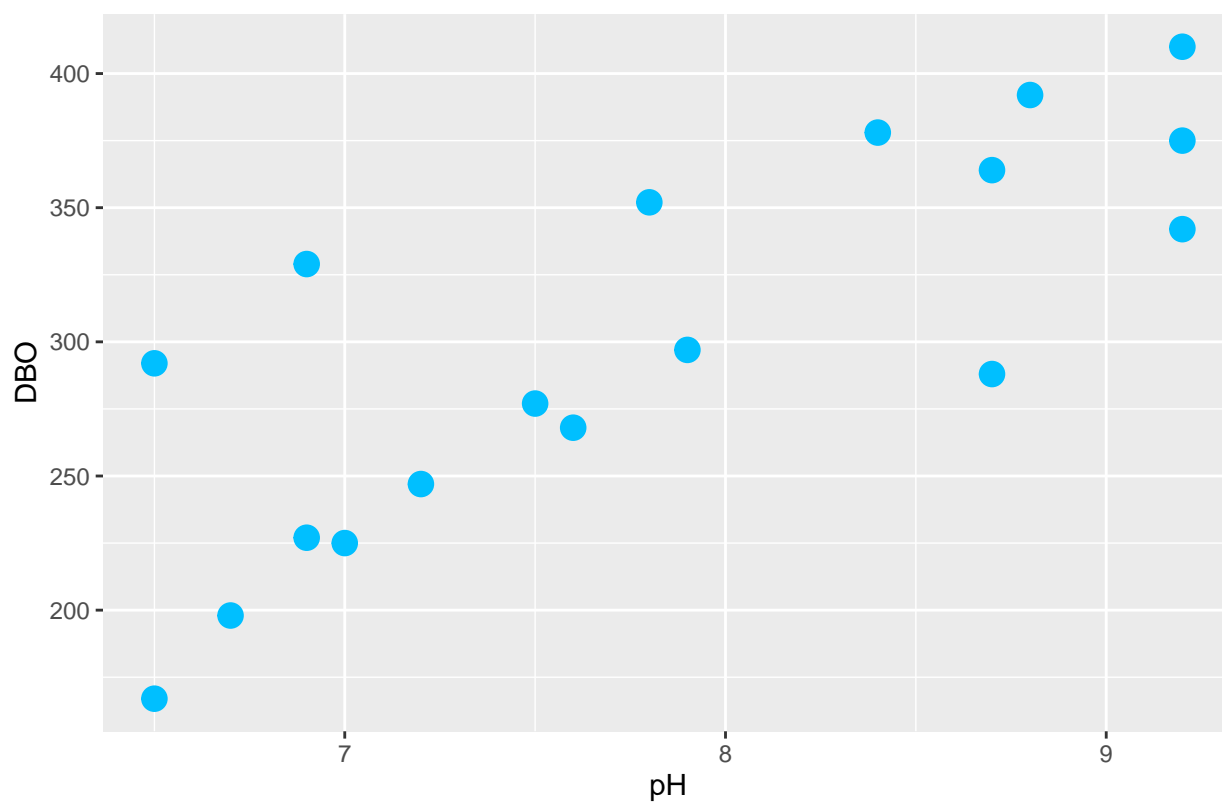
$$DBO_i = \beta_0 + \beta_1 * pH_i + \beta_2 * fteB_i + \beta_3 * fteC_i + \beta_4 * pH * fteB_i + \beta_5 * pH * fteC_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$
$$i = 1 : 18$$

```
attach(datos)
```

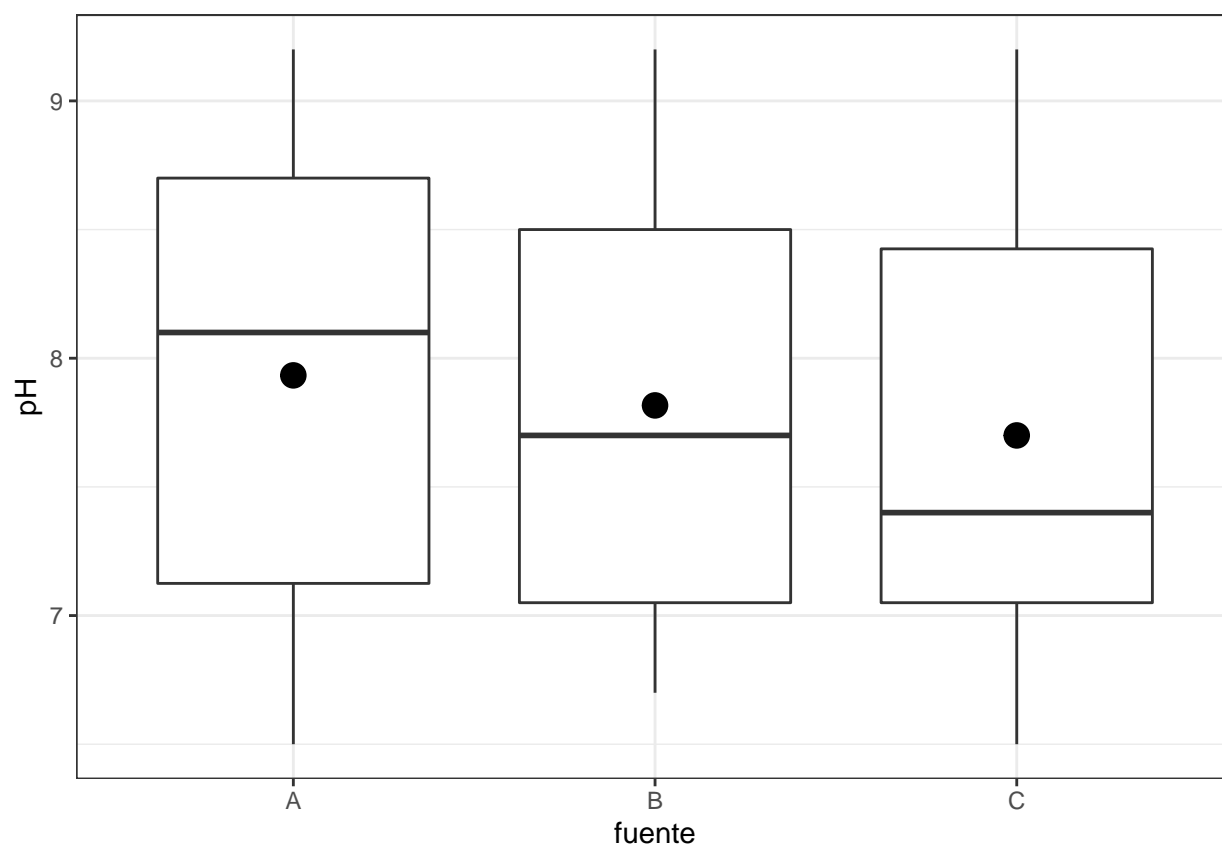
- Evalúe si hay una relación significativa entre la DBO del agua residual respecto al pH. ¿Esa relación difiere entre las fuentes? Analice la validez y el poder explicativo de un modelo que relacione la DBO con el pH del agua, independientemente de la fuente.

```
p <- ggplot(datos, aes(x = pH, y = Y)) + geom_point(aes(), colour = "deepskyblue",
  size = 4)
q <- p + xlab("pH") + ylab("DBO") + ggtitle("DBO vs pH")
q
```

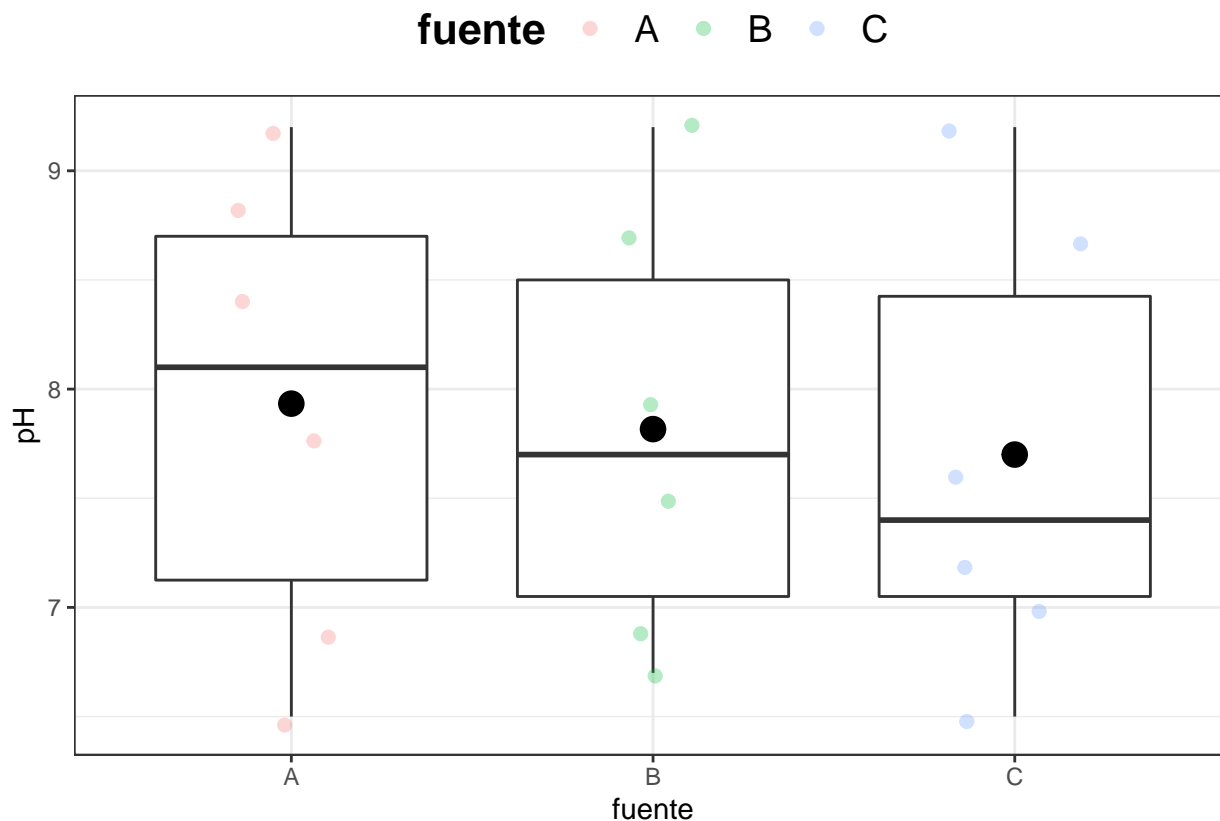
DBO vs pH



```
box <- ggplot(datos, aes(x = fuente, y = pH)) + geom_boxplot() + stat_summary(fun = mean,
  geom = "point", shape = 19, size = 4, color = "black") + theme_bw()
box
```



```
# boxplot
(box <- box + geom_jitter(alpha = 0.3, size = 2, aes(color = fuente),
  position = position_jitter(width = 0.2)) + theme(legend.position = "top",
  legend.text = element_text(size = 14), legend.title = element_text(size = 16,
  face = "bold")))
```



```
modelo1 <- lm(datos$Y ~ fuente * pH, datos)
summary(modelo1)
```

```
##
## Call:
## lm(formula = datos$Y ~ fuente * pH, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.578  -7.275   0.888   9.041  17.745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.417     48.773   0.808  0.43472
## fuenteB      -306.431     71.226  -4.302  0.00103 **
## fuenteC      -197.692     68.795  -2.874  0.01399 *
## pH           40.263      6.102   6.598 2.54e-05 ***
## fuenteB:pH    30.955      8.986   3.445  0.00485 **
## fuenteC:pH    13.561      8.737   1.552  0.14660
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.59 on 12 degrees of freedom
## Multiple R-squared:  0.9701, Adjusted R-squared:  0.9576
## F-statistic: 77.76 on 5 and 12 DF, p-value: 1.016e-08
```

```
anova(modelo1)

## Analysis of Variance Table
##
## Response: datos$Y
##          Df Sum Sq Mean Sq  F value    Pr(>F)
## fuente    2  32893   16447   77.3155 1.395e-07 ***
## pH         1  47288   47288  222.3005 4.164e-09 ***
## fuente:pH  2   2526    1263    5.9374 0.01612 *
## Residuals 12   2553     213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

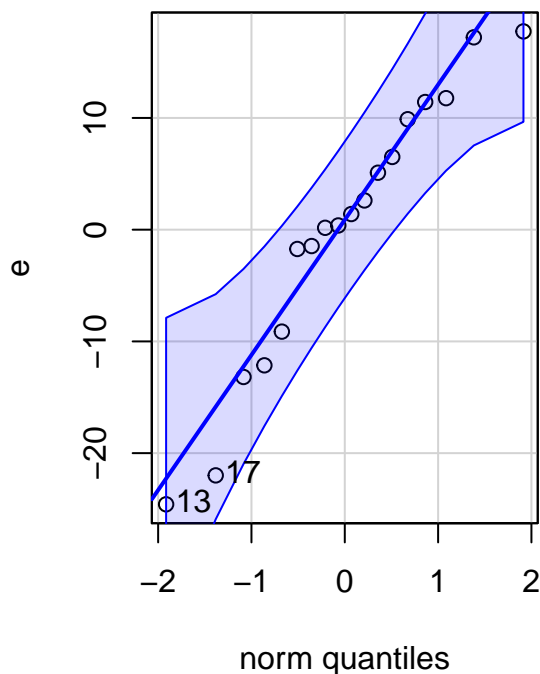
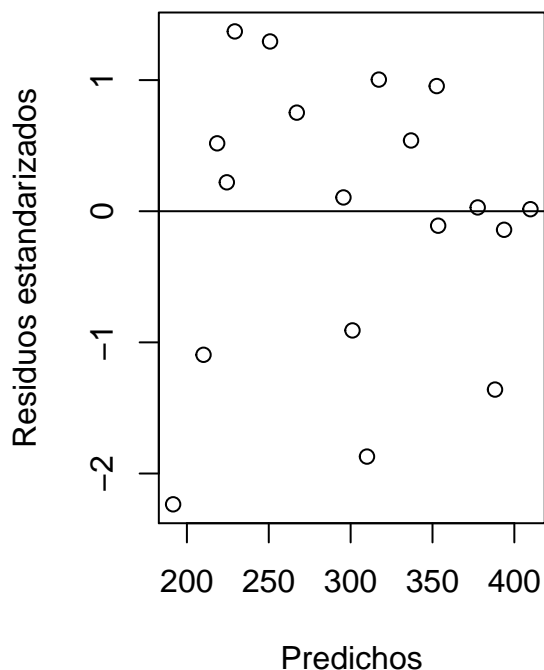
```
# Supuestos
```

```
e <- resid(modelo1) # residuos
sum(e)
```

```
## [1] -2.664535e-15
```

```
re <- rstandard(modelo1) #residuos estandarizados
pre <- predict(modelo1)  #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED")
abline(0, 0)
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PR



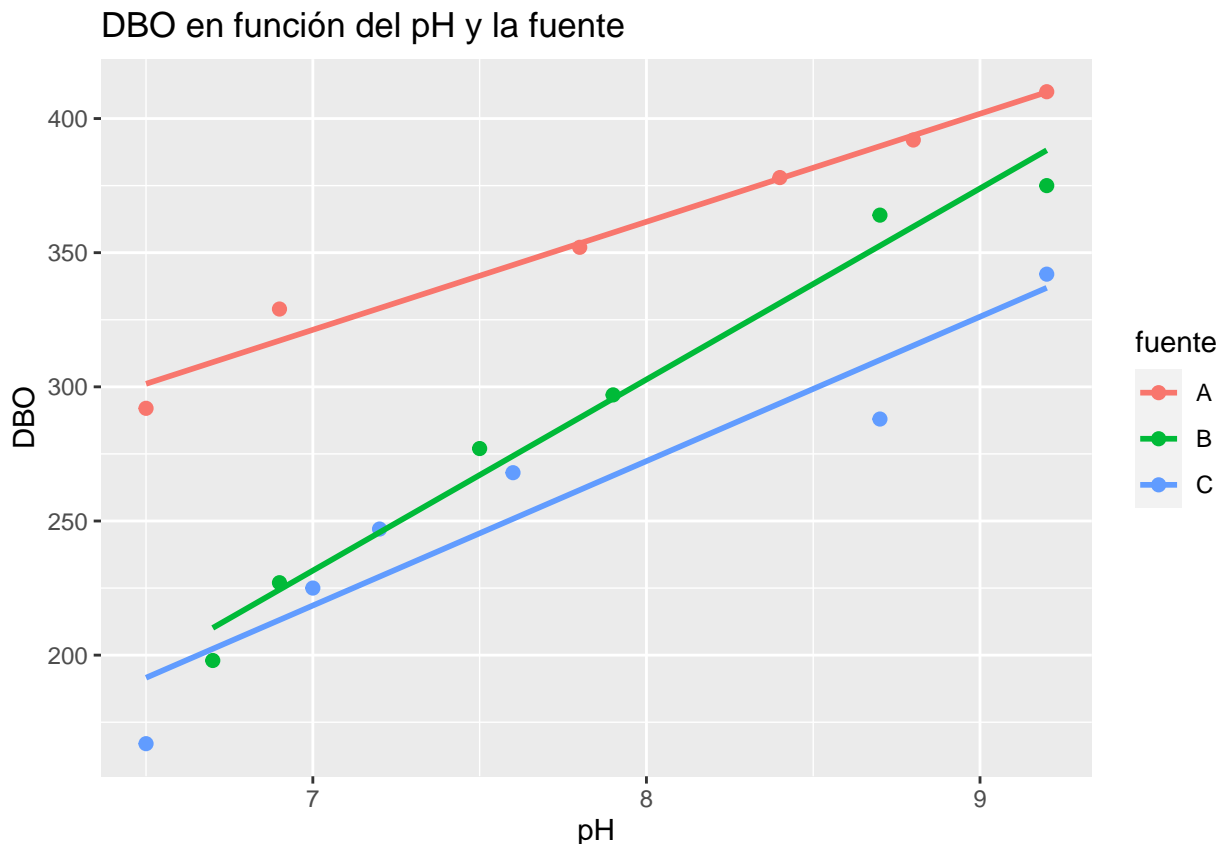
```
## [1] 13 17
```

```
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
```

```
##
## data: e
## W = 0.95103, p-value = 0.4415
# Ecuaciones estimadas para el modelo completo
ggplot(datos, aes(x = pH, y = Y, colour = fuente)) + geom_point(size = 2) +
  xlab("pH") + ylab("DBO") + ggtitle("DBO en función del pH y la fuente") +
  geom_smooth(method = "lm", se = FALSE)

## `geom_smooth()` using formula 'y ~ x'
```



```
# Es un mejor sin el término de interaccion?
modelo2 <- lm(Y ~ pH + fuente, datos)
summary(modelo2)
```

```
##
## Call:
## lm(formula = Y ~ pH + fuente, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.038 -13.640   3.601  10.798  26.374
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -71.899     38.519  -1.867   0.083 .
## pH             54.294      4.755  11.417 1.77e-08 ***
## fuenteB      -62.832     11.010  -5.707 5.42e-05 ***
## fuenteC      -89.998     11.052  -8.143 1.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.05 on 14 degrees of freedom
```

```
## Multiple R-squared:  0.9404, Adjusted R-squared:  0.9277
## F-statistic: 73.68 on 3 and 14 DF,  p-value: 8.14e-09
```

```
anova(modelo2)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: Y
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## pH         1  54856   54856 151.215 6.830e-09 ***
```

```
## fuente      2   25326   12663  34.907 3.628e-06 ***
```

```
## Residuals 14    5079     363
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Supuestos
```

```
e <- resid(modelo2) # residuos
```

```
re <- rstandard(modelo2) #residuos estandarizados
```

```
pre <- predict(modelo2) #predichos
```

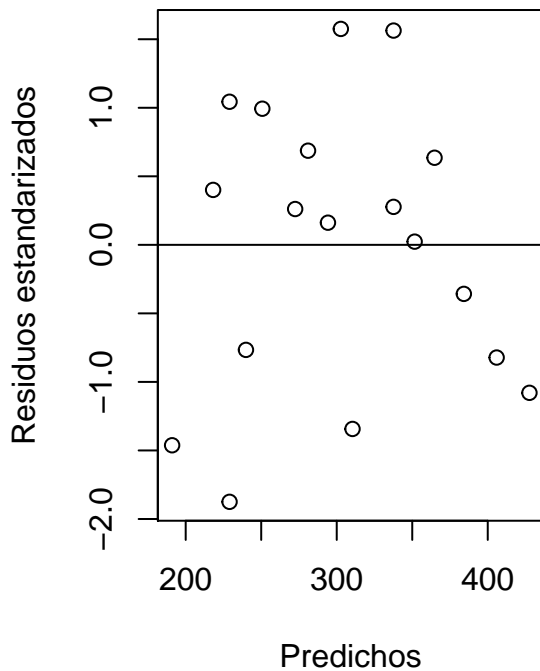
```
par(mfrow = c(1, 2))
```

```
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "RE vs PRED - Modelo 2")
```

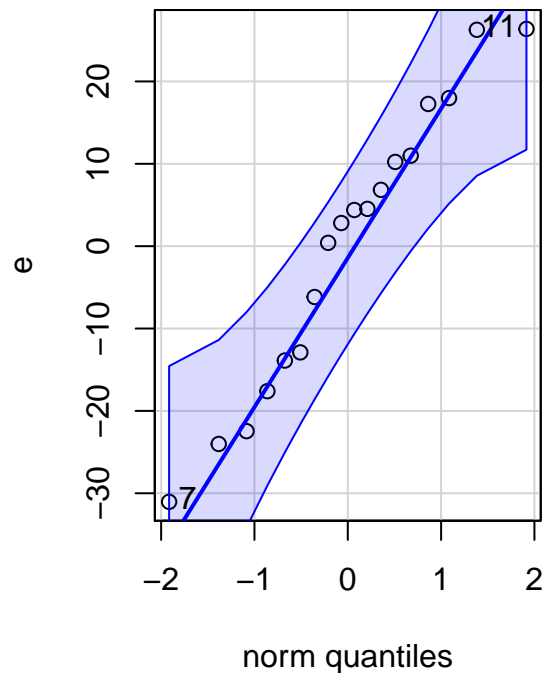
```
abline(0, 0)
```

```
car::qqPlot(e, main = "QQplot - Modelo 2")
```

RE vs PRED – Modelo 2



QQplot – Modelo 2



```
## [1] 7 11
```

```
shapiro.test(e)
```

```
##
```

```
## Shapiro-Wilk normality test
```

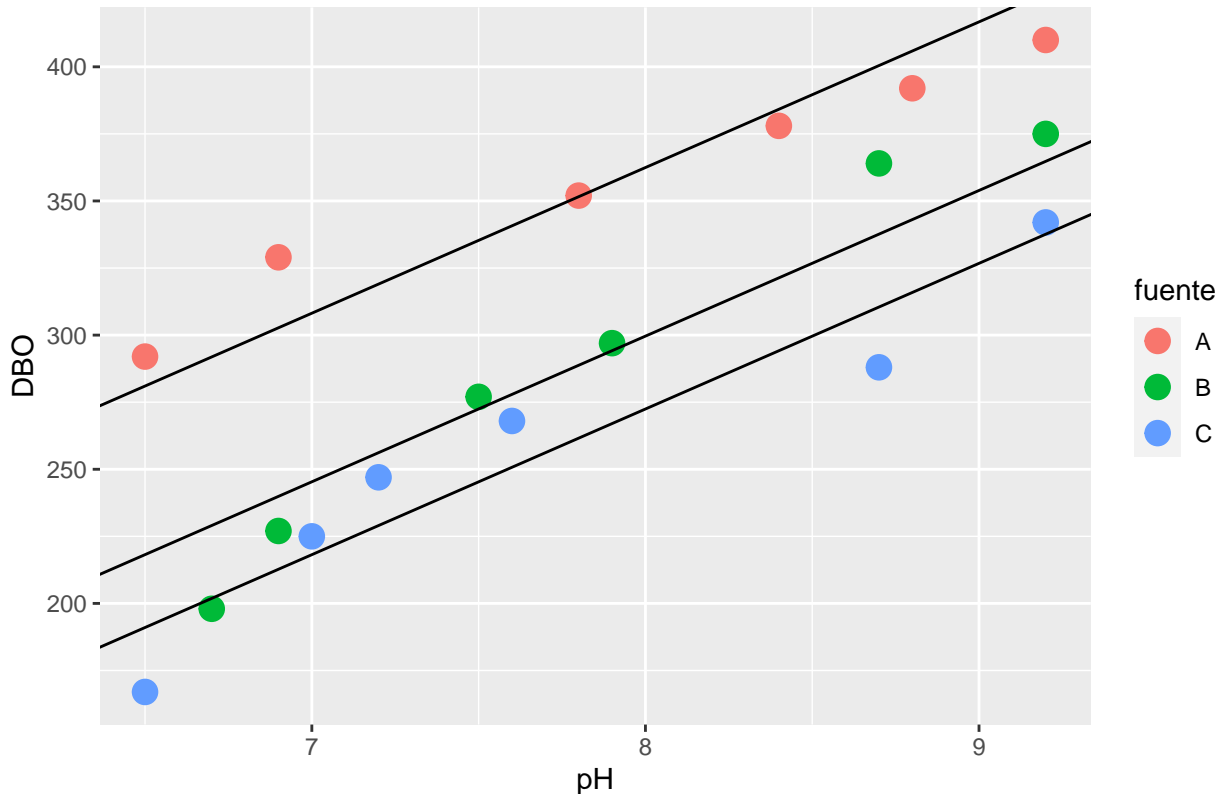
```
##
```

```
## data: e
```

```
## W = 0.96019, p-value = 0.6054
```

```
# plot modelo 2
ggplot(datos, aes(x = pH, y = Y, colour = fuente)) + geom_point(size = 4) +
  xlab("pH") + ylab("DBO") + ggtitle("DBO en función del pH, para cada fuente") +
  geom_abline(intercept = modelo2$coefficients[1], slope = modelo2$coefficients[2]) +
  geom_abline(intercept = modelo2$coefficients[1] + modelo2$coefficients[3],
    slope = modelo2$coefficients[2]) + geom_abline(intercept = modelo2$coefficients[1] +
  modelo2$coefficients[4], slope = modelo2$coefficients[2])
```

DBO en función del pH, para cada fuente



```
# Modelo que no considera las fuentes
modelo3 <- lm(Y ~ pH, datos)
summary(modelo3)
```

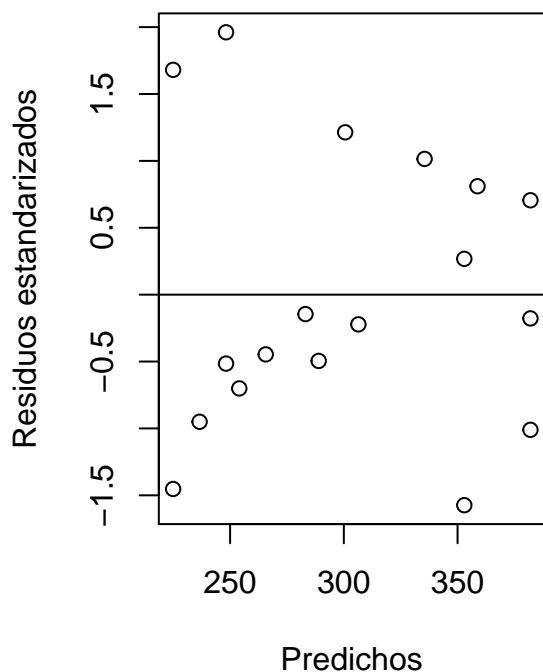
```
##
## Call:
## lm(formula = Y ~ pH, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -64.949 -27.086  -8.222  31.914  80.778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -153.23     85.27  -1.797   0.0912 .
## pH             58.18     10.83   5.373 6.22e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.59 on 16 degrees of freedom
## Multiple R-squared:  0.6434, Adjusted R-squared:  0.6211
## F-statistic: 28.87 on 1 and 16 DF, p-value: 6.221e-05
```

```
anova(modelo3)

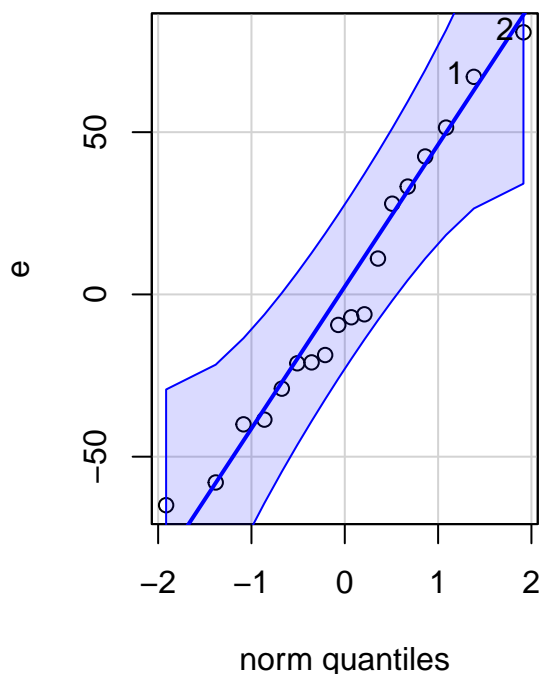
## Analysis of Variance Table
##
## Response: Y
##          Df Sum Sq Mean Sq F value    Pr(>F)
## pH         1  54856    54856  28.867 6.221e-05 ***
## Residuals 16   30405      1900
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Supuestos
e <- resid(modelo3) # residuos
re <- rstandard(modelo3) #residuos estandarizados
pre <- predict(modelo3) #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "RE vs PRED - Modelo 3")
abline(0, 0)
car::qqPlot(e, main = "QQplot -Modelo 3")
```

RE vs PRED – Modelo 3



QQplot –Modelo 3



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

```
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
##
## data:  e
## W = 0.95897, p-value = 0.582
```

- Compare los distintos modelos en base a los gráficos diagnosticos, el R^2 y $R^2_{ajustado}$, y el AIC obtenido. ¿Con cual modelo decide quedarse? Escriba la o las ecuaciones del modelo final.


```
aics <- AIC(modelo1, modelo2, modelo3)
r2adj <- c(summary(modelo1)$adj.r.squared, summary(modelo2)$adj.r.squared,
summary(modelo3)$adj.r.squared)
(comparaciones <- data.frame(aics, r2adj))
```

```
##          df      AIC    r2adj
## modelo1  7 154.2632 0.9575855
## modelo2  5 162.6457 0.9276686
## modelo3  3 190.8575 0.6211002
```

Elegimos entonces el modelo1.

Escribimos las ecuaciones

```
(ordendaA <- modelo1$coefficients[1])
```

```
## (Intercept)
##      39.41715
```

```
(pendienteA <- modelo1$coefficients[4])
```

```
##      pH
## 40.26254
```

```
(ordendaB <- modelo1$coefficients[1] + modelo1$coefficients[2])
```

```
## (Intercept)
##    -267.0143
```

```
(pendienteB <- modelo1$coefficients[4] + modelo1$coefficients[5])
```

```
##      pH
## 71.21718
```

```
(ordendaC <- modelo1$coefficients[1] + modelo1$coefficients[3])
```

```
## (Intercept)
##    -158.2745
```

```
(pendienteC <- modelo1$coefficients[4] + modelo1$coefficients[6])
```

```
##      pH
## 53.82353
```

Para la fuente a:

$40.2625438 \cdot \text{pH} + 39.4171529$

Para la fuente b:

$71.2171838 \cdot \text{pH} + -267.0143198$

Para la fuente c:

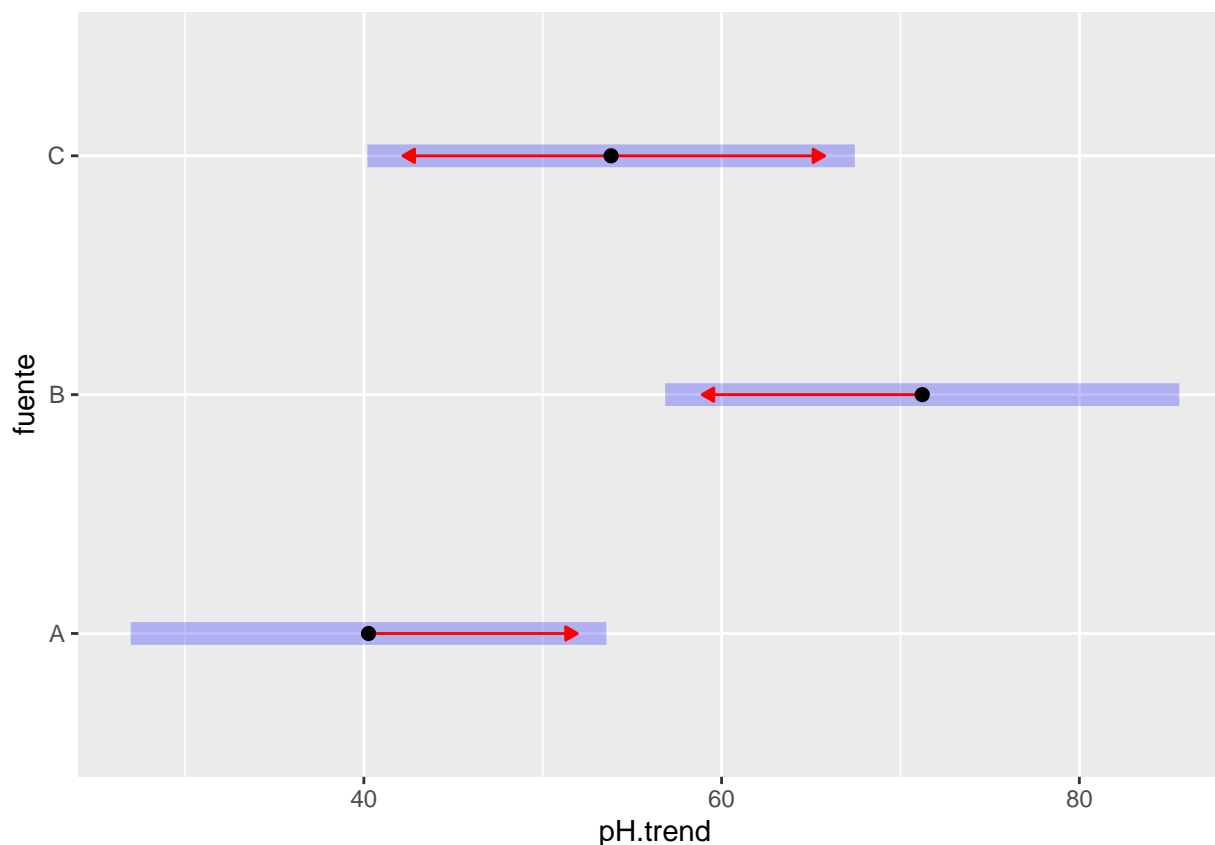
$53.8235294 \cdot \text{pH} + -158.2745098$

Comparacion de pendientes

```
comp_pendientes <- emtrends(modelo1, ~fuente, var = "pH") #, contr='cld'
pairs(comp_pendientes)
```

```
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## A - B          -31.0 8.99 12   -54.93    -6.98  -3.445  0.0125
## A - C          -13.6 8.74 12   -36.87     9.75  -1.552  0.3028
## B - C           17.4 9.09 12    -6.86    41.64   1.914  0.1772
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(comp_pendientes, comparisons = TRUE)
```



```
Efecto_fuente_PH_promedio <- emmeans(modelo1, pairwise ~ fuente | pH)
```

```
Efecto_fuente_PH_promedio #dif entre fuentes para pH promedio
```

```
## $emmeans
```

```
## pH = 7.82:
```

fuente	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
A	354	6.00	12	341	367	59.055	<.0001
B	290	5.95	12	277	303	48.648	<.0001
C	262	6.00	12	249	276	43.750	<.0001

```
##
```

```
## Confidence level used: 0.95
```

```
##
```

```
## $contrasts
```

```
## pH = 7.82:
```

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
A - B	64.5	8.45	12	41.92	87.0	7.629	<.0001
A - C	91.7	8.48	12	69.06	114.3	10.810	<.0001
B - C	27.2	8.45	12	4.67	49.8	3.221	0.0187

```
##
```

```
## Confidence level used: 0.95
```

```
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

```
## P value adjustment: tukey method for comparing a family of 3 estimates
```

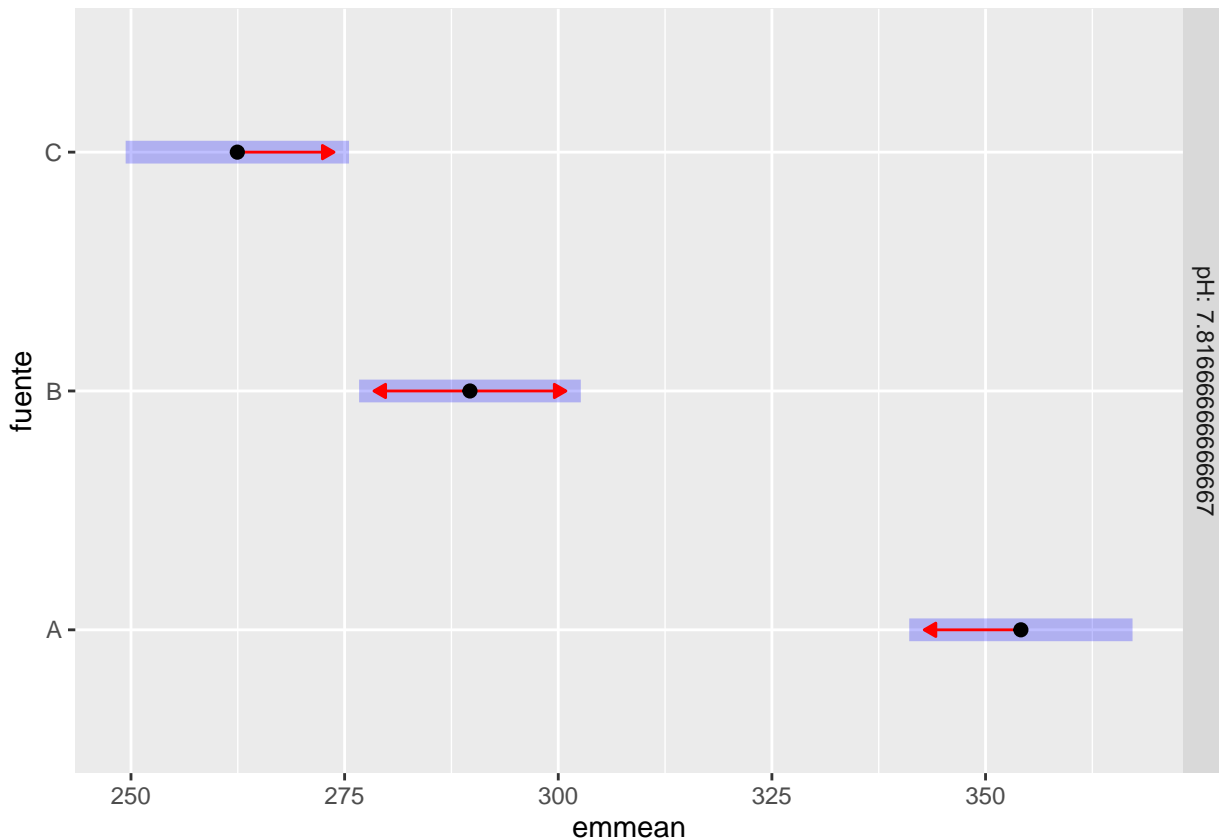
```
confint(Efecto_fuente_PH_promedio)
```

```
## $emmeans
```

```
## pH = 7.82:
```

fuente	emmean	SE	df	lower.CL	upper.CL
A	354	6.00	12	341	367

```
## B      290 5.95 12      277      303
## C      262 6.00 12      249      276
##
## Confidence level used: 0.95
##
## $contrasts
## pH = 7.82:
## contrast estimate SE df lower.CL upper.CL
## A - B      64.5 8.45 12    41.92    87.0
## A - C      91.7 8.48 12    69.06   114.3
## B - C      27.2 8.45 12     4.67    49.8
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
plot(Efecto_fuente_PH_promedio$emmeans, comparisons = TRUE)
```



```
summary(datos$pH)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 6.500 6.925 7.700 7.817 8.700 9.200
```

```
(Efecto_fuente_PH_min_max <- emmeans(modelo1, pairwise ~ fuente:pH,
  cov.reduce = range))
```

```
## $emmeans
## fuente pH emmean SE df lower.CL upper.CL t.ratio p.value
## A 6.5 301 10.58 12 278 324 28.460 <.0001
## B 6.5 196 10.53 12 173 219 18.603 <.0001
## C 6.5 192 9.58 12 171 212 19.999 <.0001
## A 9.2 410 9.76 12 389 431 42.006 <.0001
## B 9.2 388 10.90 12 364 412 35.626 <.0001
## C 9.2 337 11.11 12 313 361 30.324 <.0001
```

```
##
## Confidence level used: 0.95
##
## $contrasts
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## A 6.5 - B 6.5 105.23 14.9 12 55.085 155.4 7.049 0.0002
## A 6.5 - C 6.5 109.55 14.3 12 61.605 157.5 7.675 0.0001
## A 6.5 - A 9.2 -108.71 16.5 12 -164.047 -53.4 -6.598 0.0003
## A 6.5 - B 9.2 -87.06 15.2 12 -138.075 -36.0 -5.732 0.0010
## A 6.5 - C 9.2 -35.78 15.3 12 -87.311 15.8 -2.332 0.2536
## B 6.5 - C 6.5 4.32 14.2 12 -43.498 52.1 0.303 0.9996
## B 6.5 - A 9.2 -213.94 14.4 12 -262.155 -165.7 -14.902 <.0001
## B 6.5 - B 9.2 -192.29 17.8 12 -252.112 -132.5 -10.796 <.0001
## B 6.5 - C 9.2 -141.00 15.3 12 -192.422 -89.6 -9.211 <.0001
## C 6.5 - A 9.2 -218.25 13.7 12 -264.181 -172.3 -15.962 <.0001
## C 6.5 - B 9.2 -196.61 14.5 12 -245.337 -147.9 -13.551 <.0001
## C 6.5 - C 9.2 -145.32 16.9 12 -202.035 -88.6 -8.607 <.0001
## A 9.2 - B 9.2 21.65 14.6 12 -27.479 70.8 1.480 0.6820
## A 9.2 - C 9.2 72.93 14.8 12 23.265 122.6 4.932 0.0036
## B 9.2 - C 9.2 51.28 15.6 12 -0.988 103.6 3.295 0.0556
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 6 estimates
## P value adjustment: tukey method for comparing a family of 6 estimates
```

-Si ahora realizamos el mismo modelo pero centrando el pH a su valor promedio, ¿cual es la ventaja que presenta esta variante? Estime y compare la DBO del agua de las tres fuentes, para un valor de pH promedio.

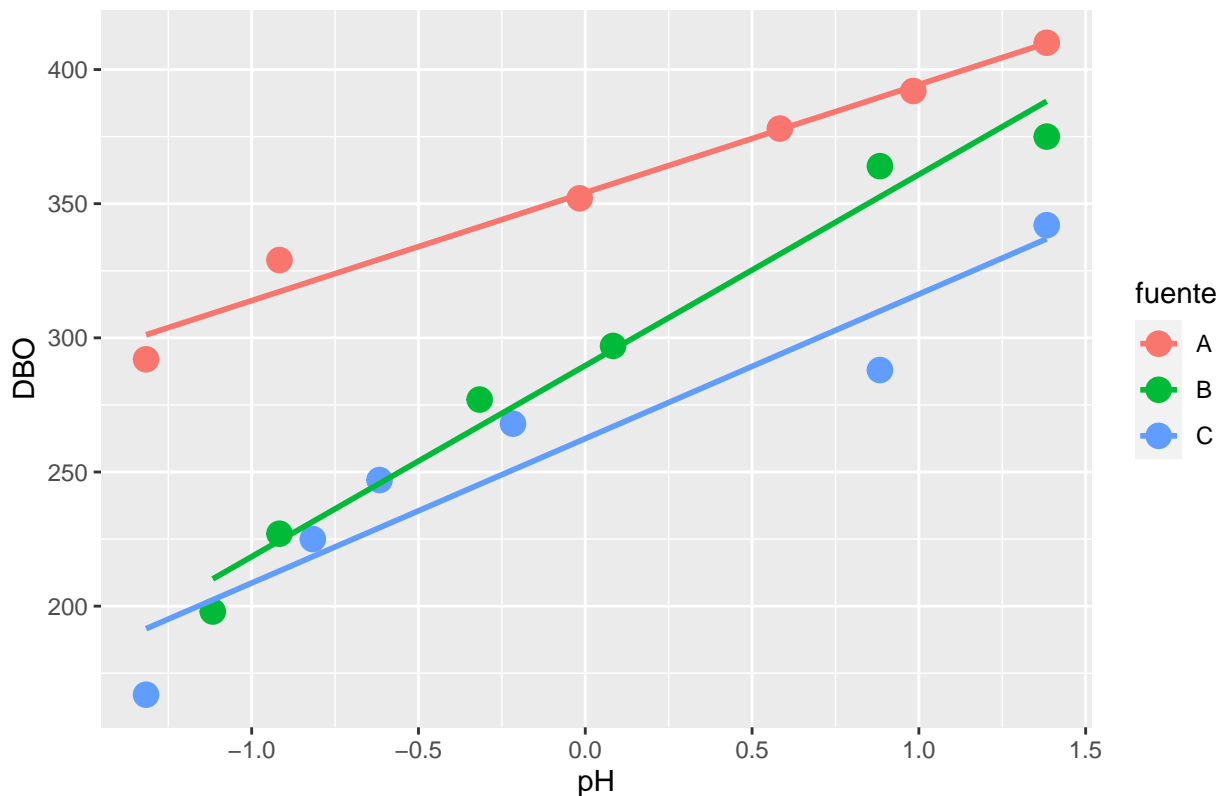
```
mean(pH)
```

```
## [1] 7.816667
```

```
datos$pH_c <- datos$pH - mean(pH)
p <- ggplot(datos, aes(x = pH_c, y = Y, colour = fuente)) + geom_point(size = 4)
q <- p + xlab("pH") + ylab("DBO") + ggtitle("Variación de la DBO en fc del pH y la fuente")
q + geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Variación de la DBO en fc del pH y la fuente



```
modelo4 <- lm(Y ~ pH_c * fuente, datos)
summary(modelo4)
```

```
##
## Call:
## lm(formula = Y ~ pH_c * fuente, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.578  -7.275   0.888   9.041  17.745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    354.136     5.997   59.055 3.67e-16 ***
## pH_c           40.263     6.102    6.598 2.54e-05 ***
## fuenteB       -64.469     8.451   -7.629 6.09e-06 ***
## fuenteC       -91.690     8.482  -10.810 1.53e-07 ***
## pH_c:f fuenteB    30.955     8.986    3.445 0.00485 **
## pH_c:f fuenteC    13.561     8.737    1.552 0.14660
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.59 on 12 degrees of freedom
## Multiple R-squared:  0.9701, Adjusted R-squared:  0.9576
## F-statistic: 77.76 on 5 and 12 DF, p-value: 1.016e-08
```

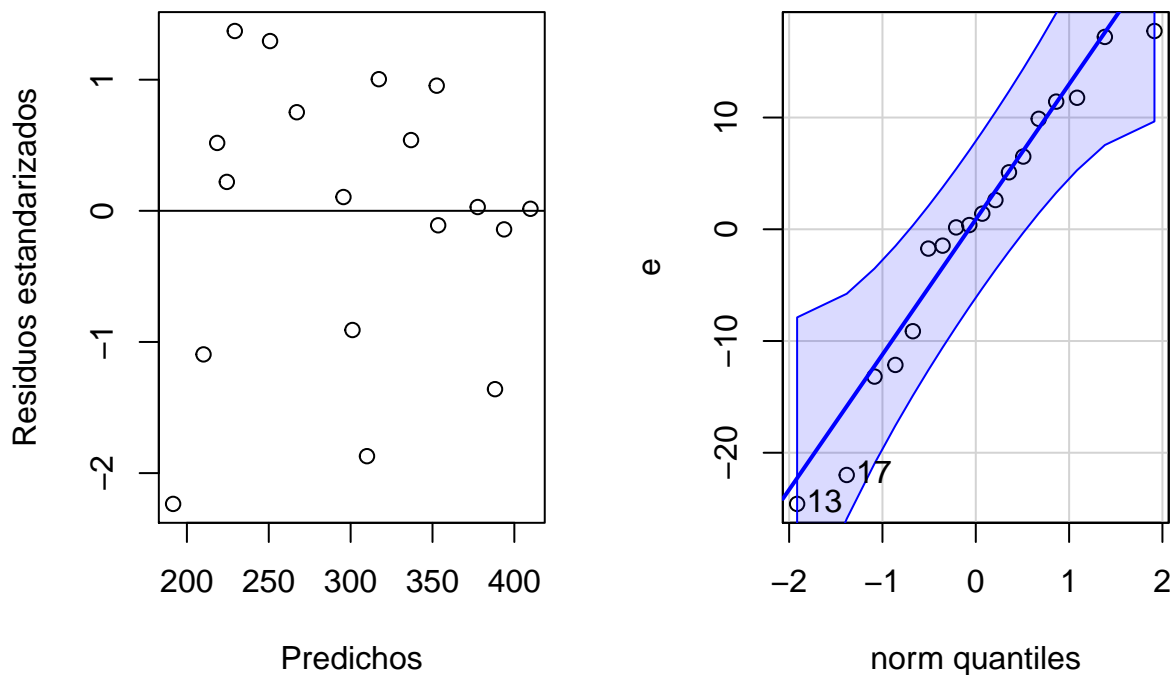
```
# Supuestos
e <- resid(modelo4) # residuos
re <- rstandard(modelo4) #residuos estandarizados
pre <- predict(modelo4) #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
```

```

    main = "Gráfico de dispersión de RE vs PRED")
abline(0, 0)
car::qqPlot(e)

```

Gráfico de dispersión de RE vs PR



```
## [1] 13 17
```

```
shapiro.test(e)
```

```

##
##  Shapiro-Wilk normality test
##
## data:  e
## W = 0.95103, p-value = 0.4415

```

```

# comparaciones
Fuentes_en_PH_PH_c <- emmeans(modelo4, pairwise ~ fuente | pH_c)
Fuentes_en_PH_PH_c

```

```

## $emmeans
## pH_c = 1.48e-16:
##   fuente emmean   SE df lower.CL upper.CL t.ratio p.value
## A         354 6.00 12     341     367  59.055 <.0001
## B         290 5.95 12     277     303  48.648 <.0001
## C         262 6.00 12     249     276  43.750 <.0001
##
## Confidence level used: 0.95
##
## $contrasts
## pH_c = 1.48e-16:
##   contrast estimate   SE df lower.CL upper.CL t.ratio p.value
## A - B         64.5 8.45 12     41.92     87.0   7.629 <.0001
## A - C         91.7 8.48 12     69.06    114.3  10.810 <.0001

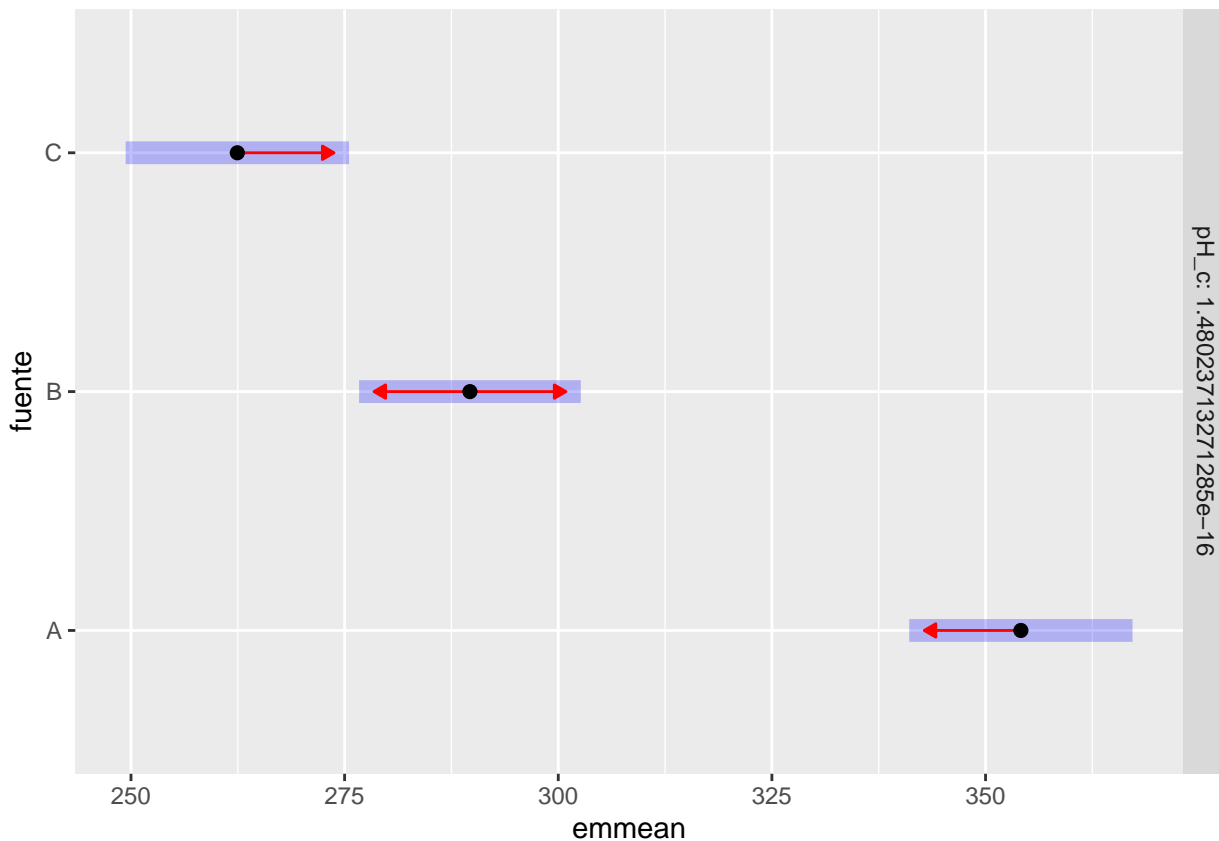
```

```
## B - C      27.2 8.45 12      4.67      49.8      3.221 0.0187
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```
confint(Fuentes_en_PH_PH_c)
```

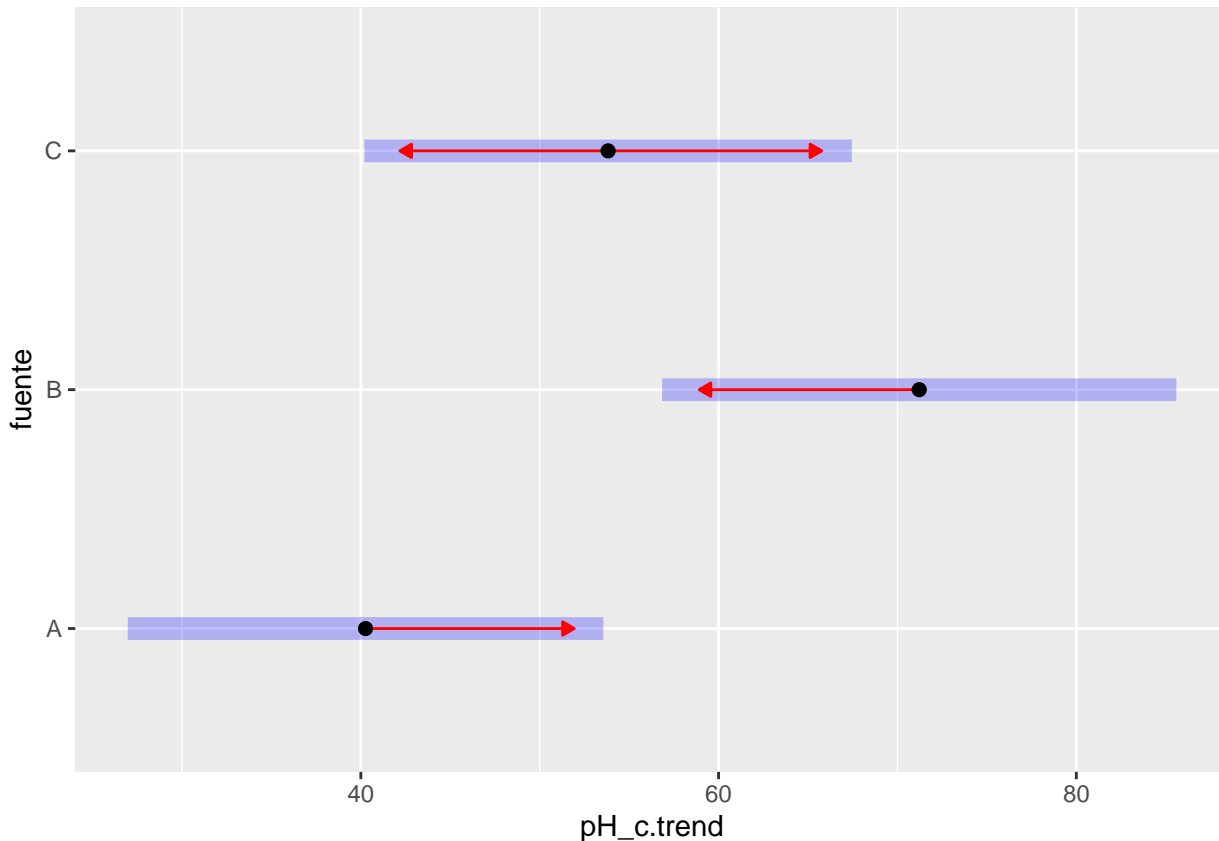
```
## $emmeans
## pH_c = 1.48e-16:
## fuente emmean SE df lower.CL upper.CL
## A      354 6.00 12      341      367
## B      290 5.95 12      277      303
## C      262 6.00 12      249      276
##
## Confidence level used: 0.95
##
## $contrasts
## pH_c = 1.48e-16:
## contrast estimate SE df lower.CL upper.CL
## A - B      64.5 8.45 12      41.92      87.0
## A - C      91.7 8.48 12      69.06     114.3
## B - C      27.2 8.45 12      4.67      49.8
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(Fuentes_en_PH_PH_c$emmeans, comparisons = TRUE)
```



```
# Comparacion de pendientes
comp_pendientes <- emtrends(modelo4, ~fuente, var = "pH_c")
pairs(comp_pendientes)
```

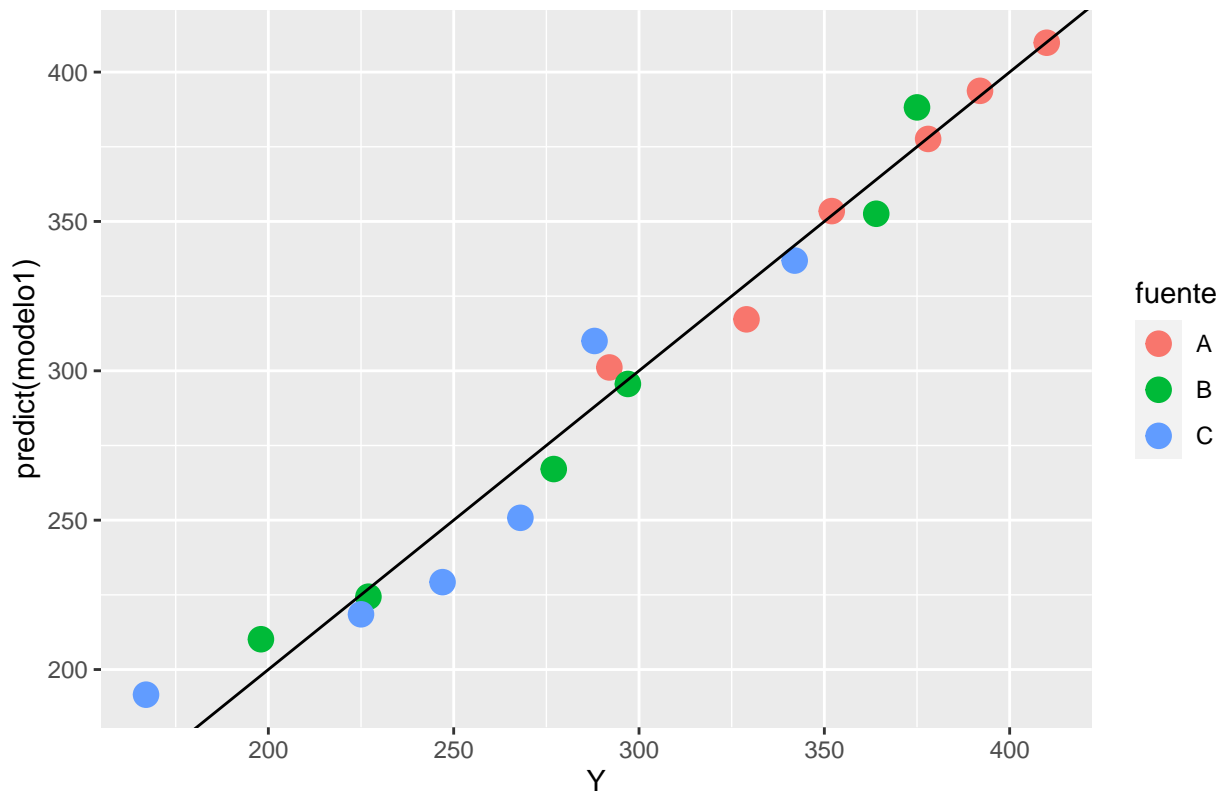
```
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## A - B -31.0 8.99 12 -54.93 -6.98 -3.445 0.0125
## A - C -13.6 8.74 12 -36.87 9.75 -1.552 0.3028
## B - C 17.4 9.09 12 -6.86 41.64 1.914 0.1772
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
plot(comp_pendientes, comparisons = TRUE)
```



- Estime con una confianza del 95% el valor de DBO esperado para aguas residuales provenientes de la fuente A y pH=7.

```
p <- ggplot(datos, aes(x = Y, y = predict(modelo1), colour = fuente)) +
  geom_point(size = 4)
p + geom_abline(intercept = 0, slope = 1) + ggtitle("Predichos vs observados Modelo 3")
```


Predichos vs observados Modelo 3



```
(cor <- cor(predict(modelo1), datos$Y))
```

```
## [1] 0.9849164
```

```
# predicciones Banda de confianza
```

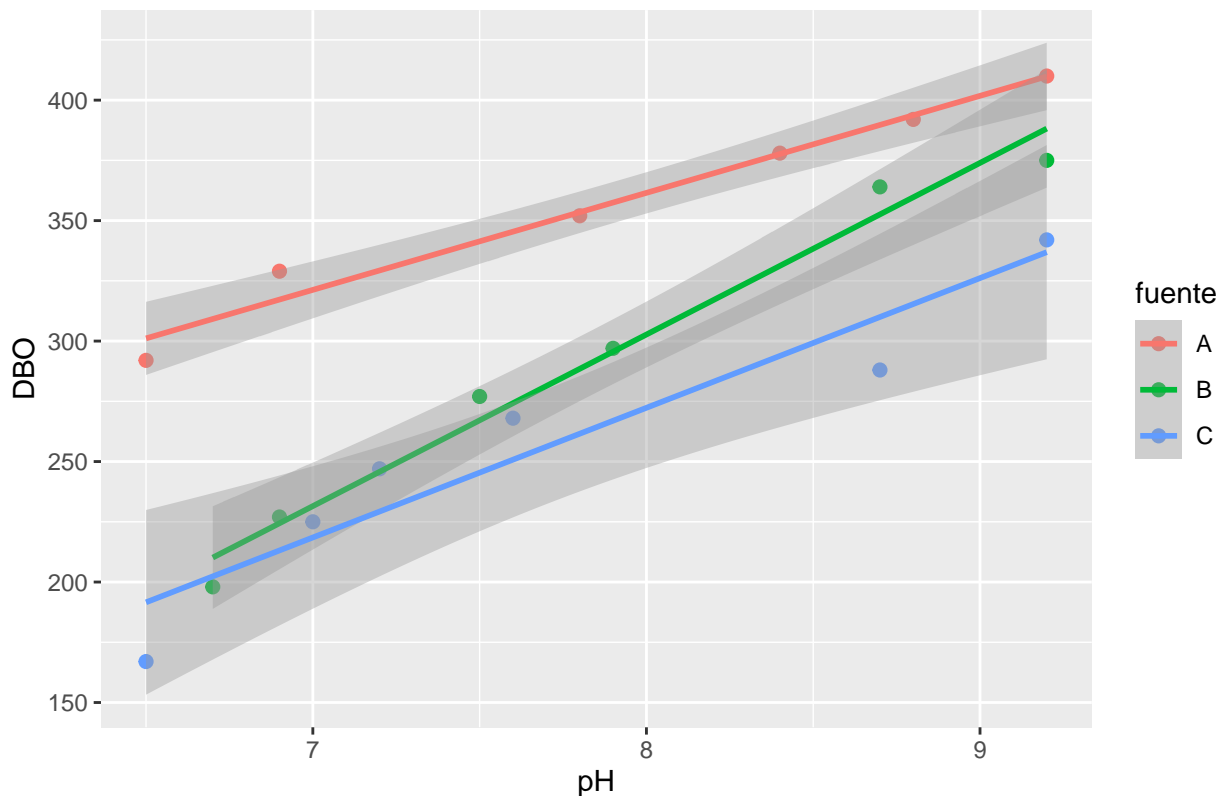
```
p <- ggplot(datos, aes(x = pH, y = Y, colour = fuente)) + geom_point(size = 2)
```

```
q <- p + xlab("pH") + ylab("DBO") + ggtitle("Variación de la DBO en función del pH y la fuente")
```

```
q + geom_smooth(method = "lm", se = TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Variación de la DBO en función del pH y la fuente



```
# predicción de Y para nuevos valores de x
nuevo = data.frame(fuente = "A", pH = 7)
predict(modelo1, nuevo, interval = "predict")
```

```
##      fit      lwr      upr
## 1 321.255 284.7568 357.7531
```

```
# CMe
```

```
CMe <- round(c(summary(modelo1)$sigma^2, summary(modelo2)$sigma^2,
  summary(modelo3)$sigma^2, summary(modelo4)$sigma^2), 2)
```

```
# R2 (no para comparar entre modelos)
```

```
R2 <- c(summary(modelo1)$r.squared, summary(modelo2)$r.squared, summary(modelo3)$r.squared,
  summary(modelo4)$r.squared)
```

```
# R2 ajustado
```

```
R2aj <- c(summary(modelo1)$adj.r.squared, summary(modelo2)$adj.r.squared,
  summary(modelo3)$adj.r.squared, summary(modelo4)$adj.r.squared)
```

```
# AIC
```

```
AIC <- c(AIC(modelo1), AIC(modelo2), AIC(modelo3), AIC(modelo4))
```

```
# Nombre modelo (para generar un data frame bonito)
```

```
modelo <- c(1, 2, 3, 4)
```

```
comp <- cbind(modelo, CMe, round(R2, 2), round(R2aj, 2), AIC)
```

```
colnames(comp) <- c("modelo", "CMe", "R2", "R2 ajust", "AIC")
```

```
comp
```

```
##      modelo      CMe      R2      R2 ajust      AIC
## [1,]      1 212.72 0.97      0.96 154.2632
## [2,]      2 362.77 0.94      0.93 162.6457
## [3,]      3 1900.30 0.64      0.62 190.8575
## [4,]      4 212.72 0.97      0.96 154.2632
```

```
# validación cruzada
```

```
library(caret)
```

```
## Loading required package: lattice
set.seed(123) #para hacer reproducibles los resultados
# Indicamos la función para el entrenamiento
train.control <- trainControl(method = "LOOCV")
# Entrenamos (estimamos) el modelo (n modelos con n-1
# observaciones)
m1loo <- train(Y ~ pH * fuente, data = datos, method = "lm", trControl = train.control)
m2loo <- train(Y ~ pH, data = datos, method = "lm", trControl = train.control)
m3loo <- train(Y ~ pH + fuente, data = datos, method = "lm", trControl = train.control)

# resultados
print(m1loo)
```

```
## Linear Regression
##
## 18 samples
## 2 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 17, 17, 17, 17, 17, 17, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
## 19.25325  0.9228204  14.96035
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
print(m2loo)
```

```
## Linear Regression
##
## 18 samples
## 1 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 17, 17, 17, 17, 17, 17, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
## 46.83017  0.542361  39.60114
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
print(m3loo)
```

```
## Linear Regression
##
## 18 samples
## 2 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 17, 17, 17, 17, 17, 17, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
## 21.87782  0.8993534  18.49416
##
```

```
## Tuning parameter 'intercept' was held constant at a value of TRUE
# dataframe con los resultados
b <- m1loo$results
c <- m2loo$results
d <- m3loo$results
e <- rbind(b, c, d)
e <- e[, 2:4]
# agrego error relativo
e$ER = e$RMSE/mean(datos$Y) * 100
e
```

```
##      RMSE  Rsquared    MAE      ER
## 1 19.25325 0.9228204 14.96035 6.384644
## 2 46.83017 0.5423610 39.60114 15.529535
## 3 21.87782 0.8993534 18.49416 7.254989
```

Problema 5. Relacion entre dieta y colesterol en mujeres

Se estudió la relación entre la edad de las mujeres y el nivel de colesterol en la sangre de mujeres, considerando si la dieta que consumían se basaba principalmente en alimentos de origen vegetal o animal. Para ello se seleccionaron al azar 20 mujeres con distinto tipo de dieta y se les midió el nivel de colesterol en sangre (en mg/ml). Los resultados fueron (archivo disponible en *DietaColesterol.txt*):

- Plantee el modelo. Indique cual es la variable dependiente o respuesta y la/s variables.
- Evalúe estadísticamente si el nivel de colesterol en sangre depende del origen del alimento principalmente consumido y de la edad.
- Centre en el valor de x que considere pertinente según el rango analizado. Interprete la pendiente con y sin centrado.
- Escriba el modelo estimado. Interprete sus coeficientes. ¿Podría estimar con este modelo el nivel de colesterol para un hombre de 51 años de edad? ¿Y para una mujer de 25 años?

Problema 6. Diversidad de aves, anfibios y mamíferos del cono sur

Casi dos siglos después del descubrimiento del “gradiente latitudinal” en la riqueza de especies, los ecólogos y biogeógrafos continúan en la búsqueda de una explicación ampliamente aceptada. La idea de que el control de la energía por parte del clima impulsa el gradiente global de riqueza específica data desde los principios de la biogeografía y ha generado una literatura extensa cuantificando la relación entre riqueza de especies y variables climáticas. La hipótesis de productividad (Hutchinson 1959, Connell y Orias 1964, Brown 1981 y Wright 1983), propone que la energía limita la riqueza de especies a través de cascadas tróficas. Desde este punto de vista, se establecen límites a la riqueza por la energía que fluye a través de redes tróficas, más que la energía total incidente en un área geográfica. Otra hipótesis propuesta (von Humboldt 1808, Turner et al. 1987, y Currie 1991) se basa en las necesidades fisiológicas de los organismos frente a la entrada de energía al ambiente, en lugar de la disponibilidad de alimento. Por supuesto, tanto factores fisiológicos como tróficos pueden subyacer a las asociaciones entre clima y riqueza. Un foco adicional de análisis es evaluar para qué grupos de animales y en qué lugares en el mundo, las variables indicadoras de la energía pura o las variables de agua – energía predicen mejor los gradientes de diversidad. La base de datos diversidad.txt consiste en un gradiente latitudinal de diversidad de aves, mamíferos y anfibios (6 variables de diversidad en total) y variables ambientales relacionadas a las hipótesis planteadas (18 variables totales). Para ello, se grilló América del Sur en celdas de $1^\circ \times 1^\circ$ (lat x long) y se compiló información acerca de las especies presentes (por ejemplo, BirLife Int. data para aves) y de variables climáticas (WorldClim) en cada celda. Se estudia el gradiente latitudinal para la zona central del continente. La base completa está disponible como parte de la descarga del software SAM: Spatial Analysis in Macroecology, <https://www.ecoevol.ufg.br/sam/> Un componente relevante de la diversidad es la composición taxonómica de los ensambles de especies y la cuantificación de sus cambios en relación a los cambios en las condiciones ambientales permite un acercamiento a comprender sus patrones de distribución. En este contexto, les proponemos estudiar la relación entre los cambios en la composición de especies de anfibios (variable a estudiar: Amphibian.Beta... Jaccard) a lo largo del centro del continente sudamericano y los cambios en condiciones de temperatura y precipitación. Los datos se encuentran en el archivo *Diversidad.txt*. Para ello:

- 1- Construya una matriz de gráficos de dispersión para hacer una primera inspección de la relación entre pares de variables. También puede indagar las correlaciones entre pares de variables (dado que son muchas, analice por separado las variables de temperatura y las de precipitación)
- 2- Ajuste un modelo de regresión múltiple con la diversidad beta de anfibios como variable respuesta y las variables explicativas lineales (por separado para temperatura y precipitación). Evalúe los

supuestos necesarios para la validez del modelo. 3- Simplifique el modelo al menor número de parámetros conveniente para cada conjunto de variables y luego establezca un modelo que incorpore a las principales de ambos grupos (i.e., temp y prec), incorporando términos de interacción. Pruebe de volver a trabajar con la estacionalidad de la precipitación "Precipitation.Seasonality". 4- Calcule el VIF, el R² ajustado y el AIC para cada uno de los modelos planteados. Concluya sobre la importancia de los factores de control de la composición taxonómica de los anfibios.

Problema 7. Diversidad de mamíferos del Cono Sur

Bajo el mismo contexto del Problema 8, responda las mismas preguntas pero para mamíferos (variable Mammal.Beta... Jaccard) y compare los resultados de ambos grupos (anfibios vs mamíferos).

Problema 8. Biomasa de comunidades de ratones granívoros

En una comunidad, la biomasa de un nivel trófico intermedio (consumidores) puede estar determinada más bien por la abundancia de sus recursos (un control "desde abajo" o bottom-up) o más bien por la de sus predadores (un control "desde arriba" o top-down). A su vez, todos ellos pueden estar afectados por condicionantes y modificadores de la productividad y la estructura del sistema (como la energía y recursos recibidos por los autótrofos, o la heterogeneidad espacial). Con el objetivo de estudiar algunas variables que podrían determinar la biomasa de las comunidades de ratones granívoros en 16 localidades ubicadas en ambientes desérticos, se realizó un estudio donde se registró: 1) la producción de semillas en esos ambientes, 2) la abundancia de predadores de la comunidad de ratones granívoros, 3) las precipitaciones medias anuales (un limitante de la productividad primaria en estos ambientes), y 4) la cobertura de la vegetación perenne (un factor estructural importante para los ratones). La base de datos se encuentra en el archivo Ratones.txt [a partir de un ejemplo en Crawley 2002]. En base al enunciado:

1- Construya una matriz de gráficos de dispersión para hacer una primera inspección de la relación entre pares de variables. Calcule el coeficiente de correlación entre pares de variables y el factor de inflación de la varianza. ¿Recomendaría incluir las 4 variables en un mismo modelo?

2- Ajuste distintos modelos de regresión múltiple con la biomasa de ratones como variable respuesta y las variables explicativas que considere apropiadas. Ranquee los modelos en función de su AIC y seleccione modelos candidatos.

3- Compare los modelos candidatos. Explore los gráficos de residuos parciales. Seleccione el mejor modelo a su juicio fundamentando su decisión y concluya sobre la importancia de los factores de control de la comunidad de ratones granívoros de desierto.

```
Ratones <- read.delim("/home/jose/Documents/materias/biome2/2019/tps/tp4/martes/ratones.txt")
colnames(Ratones) = c("ratones", "lluvia", "predadores", "cobertura",
  "semillas")
```

- Discuta el tipo de estudio, la unidad experimental o de análisis, el tamaño de la muestra y la población objetivo. Explique cuál es la variable respuesta y tipo de variables explicativas.

Tipo de estudio: Observacional.

UE: Localidad.

N=16.

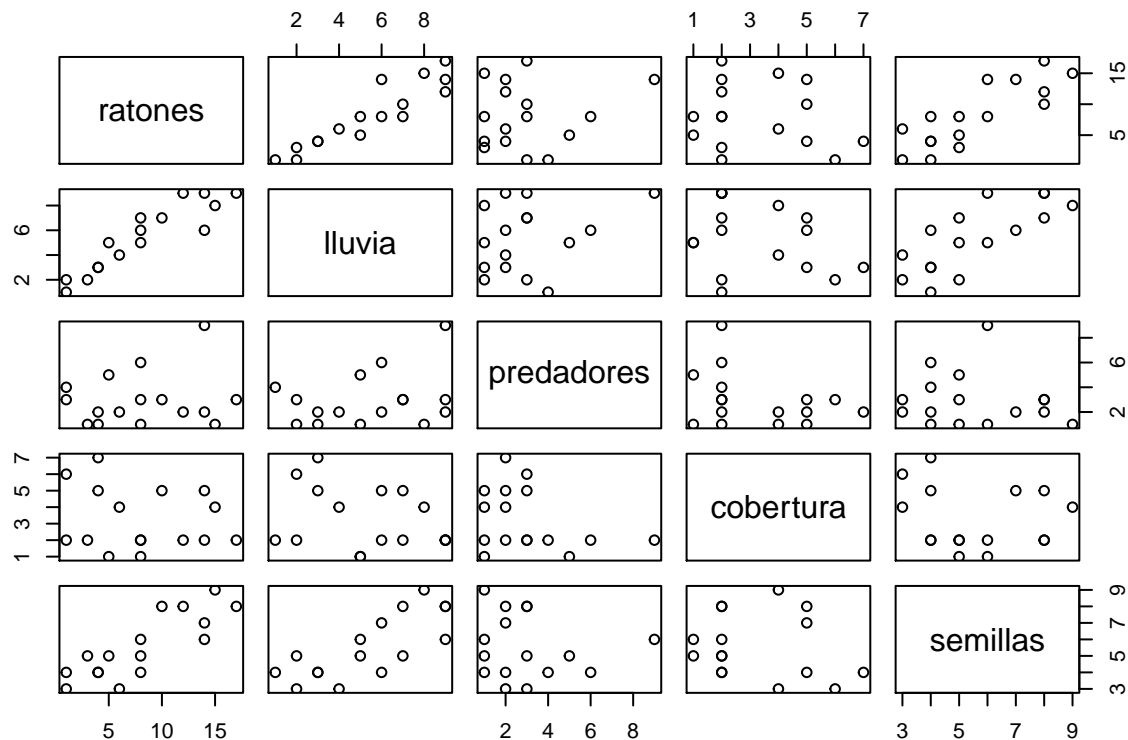
población objetivo: ratones granívoros de ambientes desérticos.

VR: número de ratones.

VE: lluvia, predadores, cobertura, semillas (cuantitativas).

- Construya una matriz de gráficos de dispersión para hacer una primera inspección de la relación entre pares de variables. Calcule el coeficiente de correlación entre pares de variables y el factor de inflación de la varianza. ¿Recomendaría incluir las 4 variables en un mismo modelo?

```
plot(Ratones)
```



```
round(cor(Ratones), 2)
```

```
##          ratones lluvia predadores cobertura semillas
## ratones      1.00  0.92      0.14    -0.18    0.83
## lluvia       0.92  1.00      0.29    -0.30    0.77
## predadores   0.14  0.29      1.00    -0.33   -0.13
## cobertura   -0.18 -0.30     -0.33     1.00   -0.17
## semillas     0.83  0.77     -0.13    -0.17    1.00
```

```
# Correlation matrix
```

```
names(Ratones)
```

```
## [1] "ratones" "lluvia" "predadores" "cobertura" "semillas"
```

```
corr_datos <- round(cor(Ratones), 2)
```

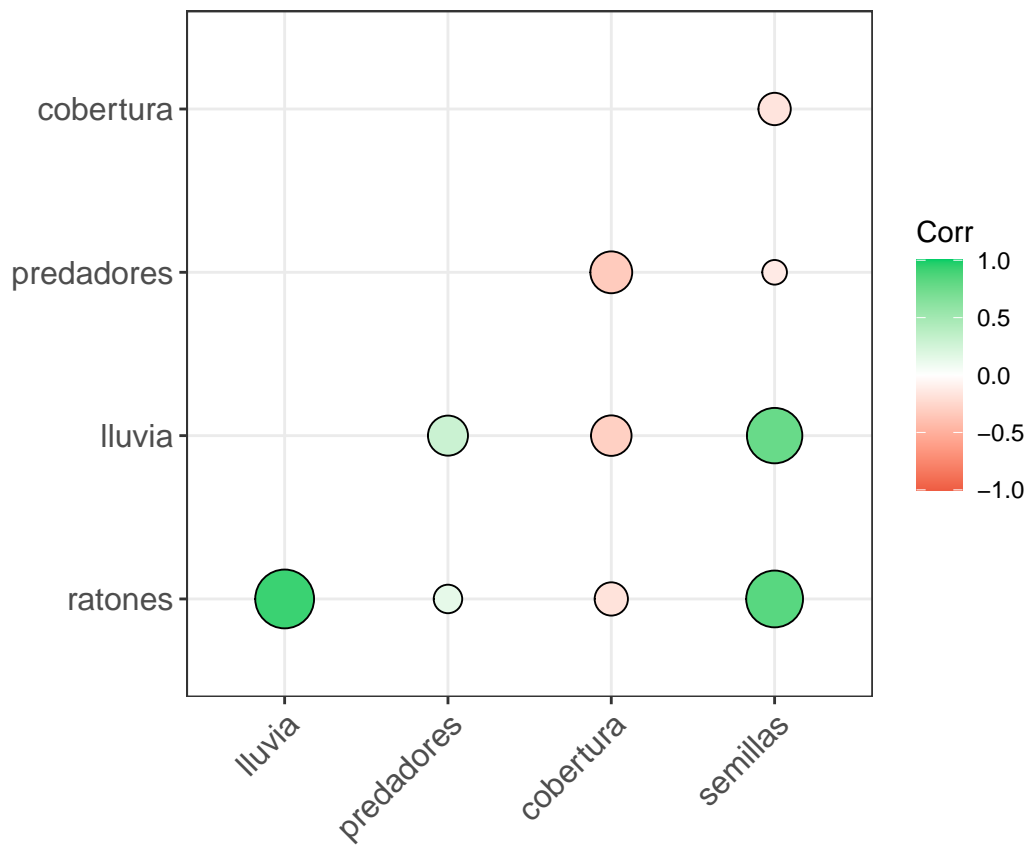
```
print(corr_datos)
```

```
##          ratones lluvia predadores cobertura semillas
## ratones      1.00  0.92      0.14    -0.18    0.83
## lluvia       0.92  1.00      0.29    -0.30    0.77
## predadores   0.14  0.29      1.00    -0.33   -0.13
## cobertura   -0.18 -0.30     -0.33     1.00   -0.17
## semillas     0.83  0.77     -0.13    -0.17    1.00
```

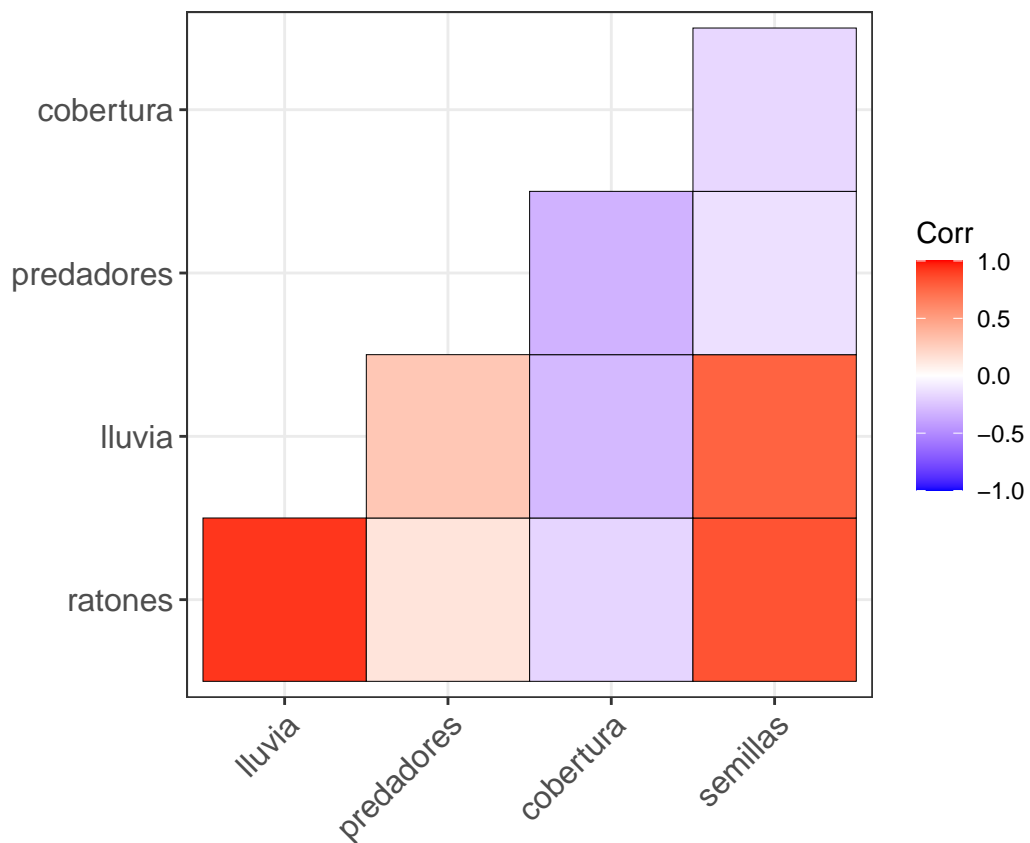
```
# correlaciones
```

```
ggcorrplot(corr_datos, type = "lower", method = "circle", colors = c("tomato2",
  "white", "springgreen3"), outline.color = "black", ggtheme = theme_bw)
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```



```
ggcorrplot(corr_datos, type = "lower", method = "square", outline.color = "black",
  ggtheme = theme_bw)
```

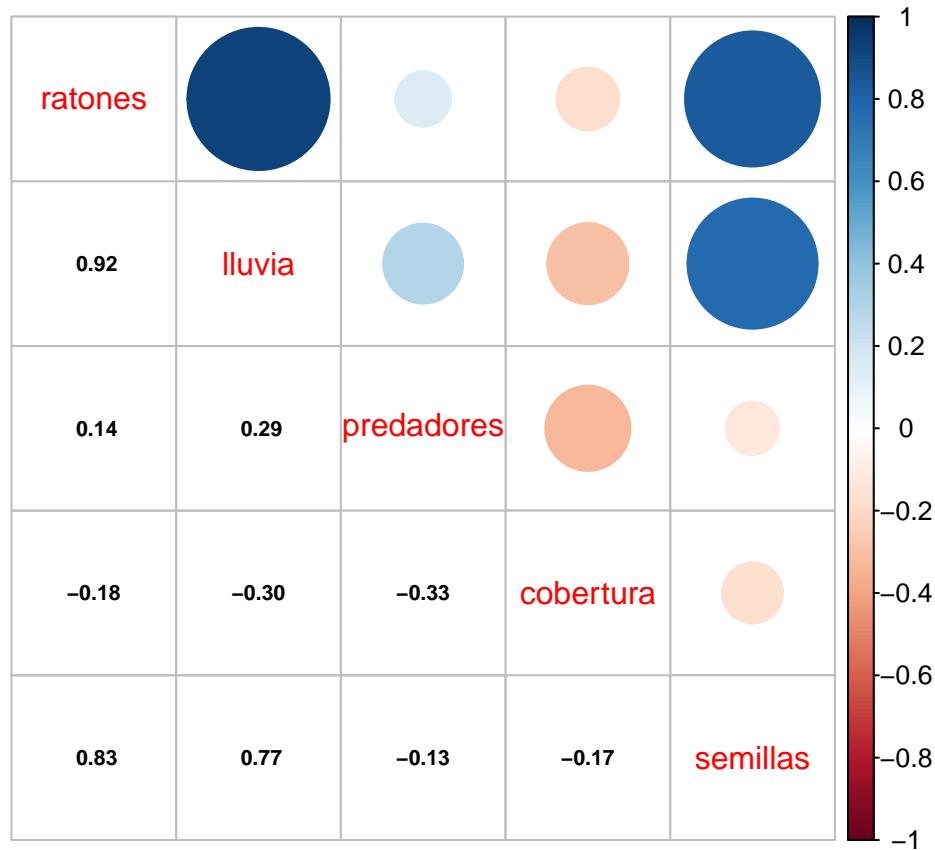


```
library(corrplot)
```

```
## corrplot 0.90 loaded
```

```
# mixto
```

```
corrplot.mixed(corr_datos, lower.col = "black", number.cex = 0.7)
```



```
# ¿Qué variables se encuentran asociadas?
```

- Ajuste distintos modelos de regresión múltiple con la biomasa de ratones como variable respuesta y las variables explicativas que considere apropiadas. Ranquee los modelos en función de su AIC y seleccione modelos candidatos.

```
# modelo aditivo completo (veo el VIF)
```

```
m1 <- lm(ratones ~ lluvia + predadores + cobertura + semillas, Ratones)
```

```
car::vif(m1)
```

```
##      lluvia predadores cobertura semillas
## 3.882096 1.726523 1.186098 3.622148
```

```
# Evaluo posibles datos atípicos en el m1 que podrían afectar la
```

```
# selección
```

```
e <- resid(m1) # residuos
```

```
re <- rstandard(m1) #residuos estandarizados
```

```
pre <- predict(m1) #predichos
```

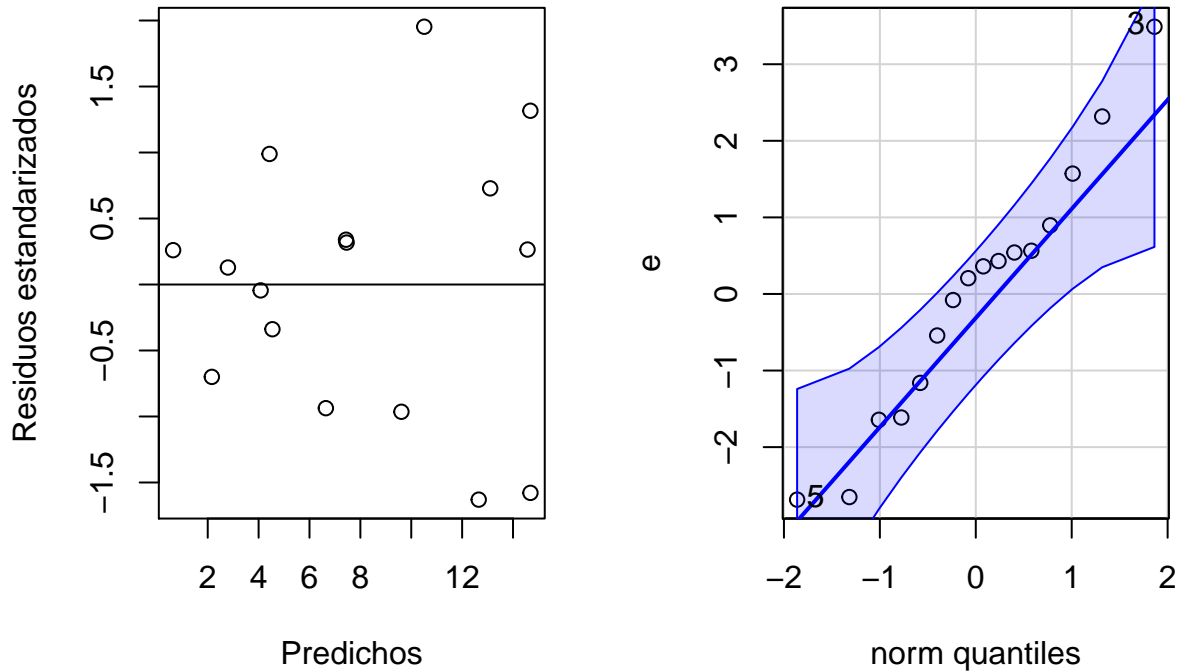
```
par(mfrow = c(1, 2))
```

```
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED")
```

```
abline(0, 0)
```

```
car::qqPlot(e)
```


Gráfico de dispersión de RE vs PR



```
## [1] 3 5
```

```
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
##
## data: e
## W = 0.96674, p-value = 0.7835
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = ratones ~ lluvia + predadores + cobertura + semillas,
##     data = Ratones)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6856 -1.2743  0.2832  0.6472  3.4896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.297654   2.520517  -1.705   0.1162
## lluvia       1.365145   0.375235   3.638   0.0039 **
## predadores  -0.003237   0.310467  -0.010   0.9919
## cobertura    0.233255   0.295734   0.789   0.4469
## semillas     0.779619   0.503049   1.550   0.1495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.977 on 11 degrees of freedom
## Multiple R-squared:  0.8897, Adjusted R-squared:  0.8496
```

```
## F-statistic: 22.18 on 4 and 11 DF, p-value: 3.195e-05
```

- Realice una selección de modelos “hacia atrás”. Evalúe supuestos y utilice algún criterio apropiado para ir reduciendo las variables ingresadas en el modelo. Seleccione el mejor modelo a su juicio, fundamentando su decisión.

```
drop1(m1, test = "F") # Interprete lo que hace la funcion (pueden ayudarse con el help)
```

```
## Single term deletions
```

```
##
```

```
## Model:
```

```
## ratones ~ lluvia + predadores + cobertura + semillas
```

```
##           Df Sum of Sq   RSS   AIC F value    Pr(>F)
```

```
## <none>                42.988 25.813
```

```
## lluvia      1      51.726 94.714 36.452 13.2358 0.003901 **
```

```
## predadores  1       0.000 42.989 23.814  0.0001 0.991869
```

```
## cobertura   1       2.431 45.419 24.694  0.6221 0.446934
```

```
## semillas    1       9.386 52.375 26.973  2.4018 0.149467
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# que variable eliminarían?
```

```
# ...continuamos
```

```
m2 <- update(m1, . ~ . - predadores)
```

```
e <- resid(m2) # residuos
```

```
re <- rstandard(m2) #residuos estandarizados
```

```
pre <- predict(m2) #predichos
```

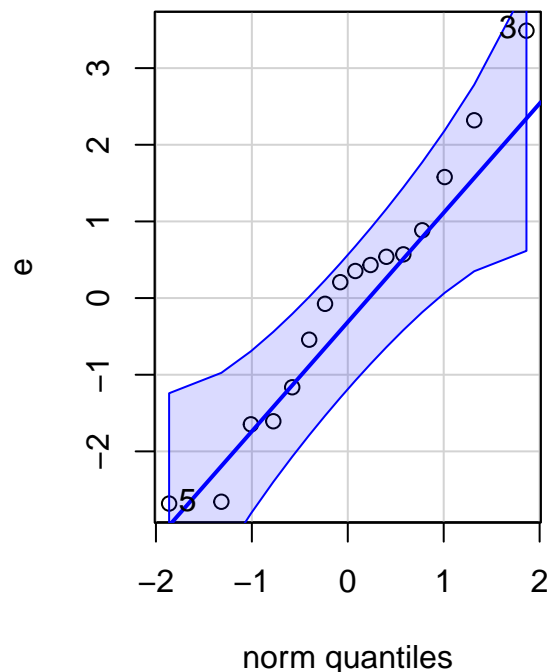
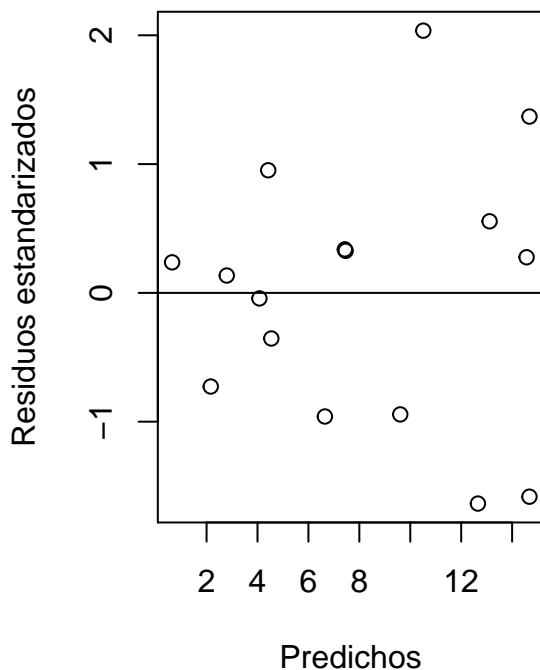
```
par(mfrow = c(1, 2))
```

```
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",  
     main = "Gráfico de dispersión de RE vs PRED m2")
```

```
abline(0, 0)
```

```
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PRE



```
## [1] 3 5
shapiro.test(e)

##
## Shapiro-Wilk normality test
##
## data: e
## W = 0.96646, p-value = 0.7785
# para confirmar que eliminar la variable no afecta la capacidad
# explicativa del modelo
anova(m1, m2)

## Analysis of Variance Table
##
## Model 1: ratones ~ lluvia + predadores + cobertura + semillas
## Model 2: ratones ~ lluvia + cobertura + semillas
## Res.Df RSS Df Sum of Sq F Pr(>F)
## 1 11 42.988
## 2 12 42.989 -1 -0.00042474 1e-04 0.9919
# modelo
summary(m2)

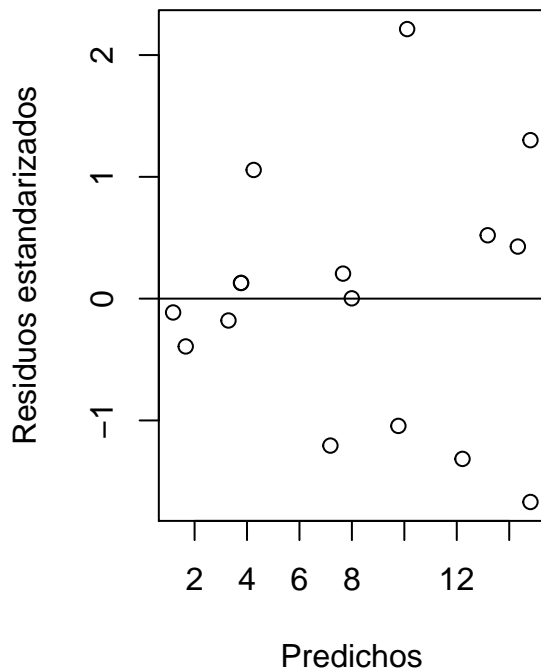
##
## Call:
## lm(formula = ratones ~ lluvia + cobertura + semillas, data = Ratones)
##
## Residuals:
## Min 1Q Median 3Q Max
## -2.6806 -1.2743 0.2804 0.6489 3.4886
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.3145 1.8533 -2.328 0.038213 *
## lluvia 1.3629 0.2953 4.616 0.000594 ***
## cobertura 0.2340 0.2736 0.855 0.409070
## semillas 0.7826 0.3970 1.971 0.072191 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.893 on 12 degrees of freedom
## Multiple R-squared: 0.8897, Adjusted R-squared: 0.8621
## F-statistic: 32.27 on 3 and 12 DF, p-value: 5.024e-06
# corresponde eliminar alguna nueva variable del m2? SI? Por qué?
drop1(m2, test = "F")

## Single term deletions
##
## Model:
## ratones ~ lluvia + cobertura + semillas
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 42.989 23.814
## lluvia 1 76.327 119.315 38.147 21.3061 0.0005944 ***
## cobertura 1 2.621 45.610 22.761 0.7318 0.4090695
## semillas 1 13.922 56.910 26.302 3.8861 0.0721909 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

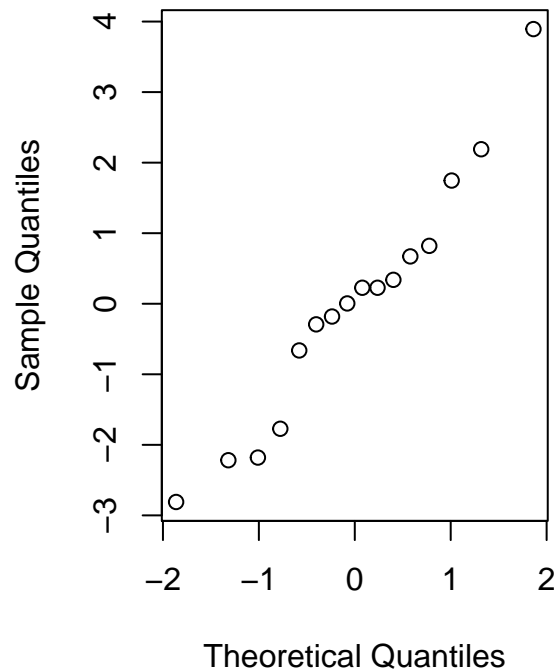
```
# elimino cobertura
```

```
m3 <- update(m2, . ~ . - cobertura)
e <- resid(m3) # residuos
re <- rstandard(m3) #residuos estandarizados
pre <- predict(m3) #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
      main = "Gráfico de dispersión de RE vs PRED m3")
abline(0, 0)
qqnorm(e)
```

Gráfico de dispersión de RE vs PRE



Normal Q-Q Plot

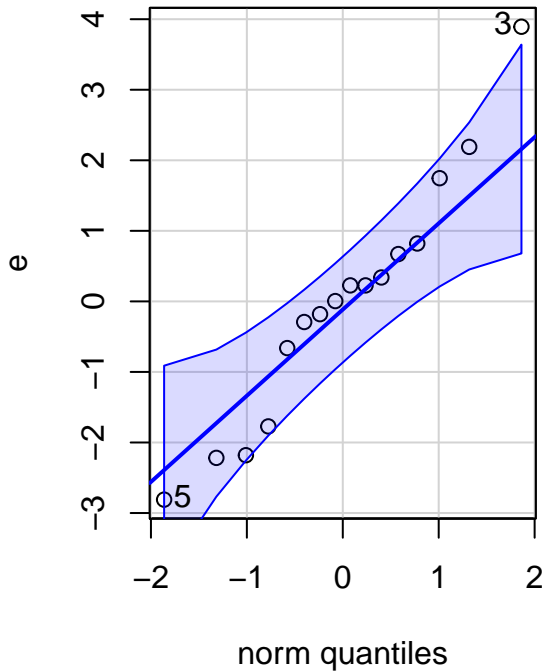


```
car::qqPlot(e)
```

```
## [1] 3 5
```

```
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
##
## data: e
## W = 0.95982, p-value = 0.6585
```



```
anova(m1, m2, m3) # m1 vs m2 y m2 vs m3
```

```
## Analysis of Variance Table
##
## Model 1: ratones ~ lluvia + predadores + cobertura + semillas
## Model 2: ratones ~ lluvia + cobertura + semillas
## Model 3: ratones ~ lluvia + semillas
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      11 42.988
## 2      12 42.989 -1  -0.00042 0.0001 0.9919
## 3      13 45.610 -1  -2.62144 0.6708 0.4302
```

```
summary(m3)
```

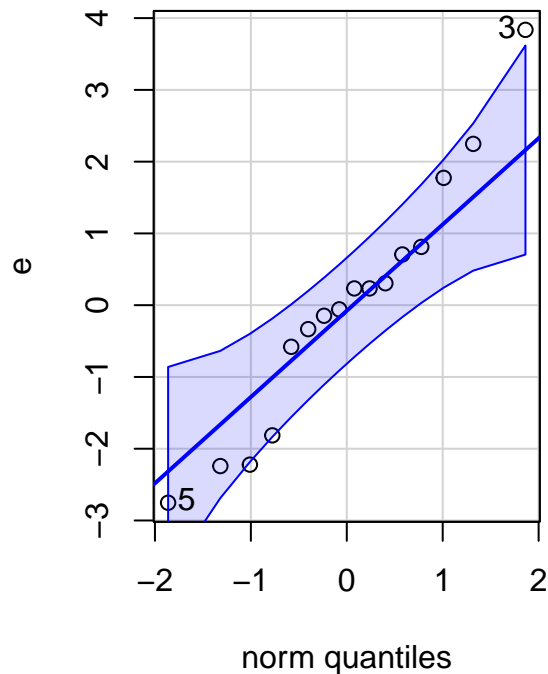
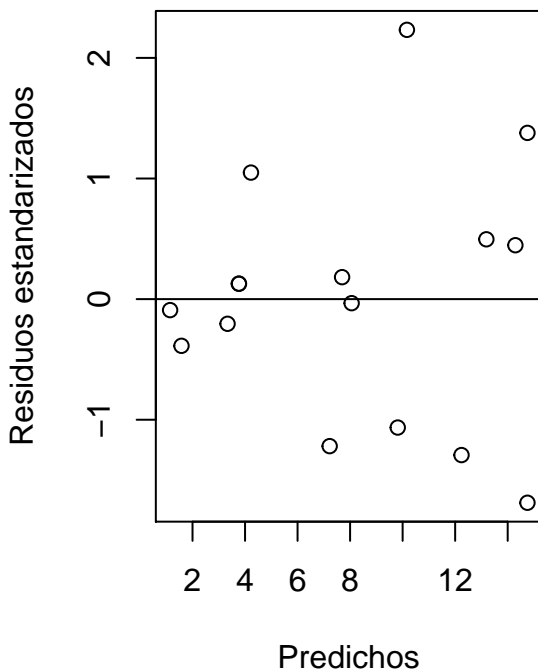
```
##
## Call:
## lm(formula = ratones ~ lluvia + semillas, data = Ratones)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8101 -0.9396  0.1155  0.7081  3.8931
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.3756     1.4780  -2.284 0.039823 *
## lluvia         1.2960     0.2818   4.600 0.000498 ***
## semillas      0.8152     0.3910   2.085 0.057378 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.873 on 13 degrees of freedom
## Multiple R-squared:  0.883, Adjusted R-squared:  0.865
```

```
## F-statistic: 49.04 on 2 and 13 DF, p-value: 8.786e-07
# ¿semillas queda en el modelo o sale?
drop1(m3, test = "F") # cuanto cambia el AIC si elimino semillas? La dejamos o la sacamos?

## Single term deletions
##
## Model:
## ratones ~ lluvia + semillas
##          Df Sum of Sq    RSS   AIC F value    Pr(>F)
## <none>                 45.610 22.761
## lluvia     1    74.227 119.837 36.217 21.1565 0.000498 ***
## semillas   1    15.248  60.858 25.375  4.3462 0.057378 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# ¿Existe interaccion entre lluvia y semillas?
m4 <- update(m3, . ~ . + lluvia:semillas)
e <- resid(m4) # residuos
re <- rstandard(m4) #residuos estandarizados
pre <- predict(m4) #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED m4")
abline(0, 0)
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PRE



```
## [1] 3 5
shapiro.test(e)

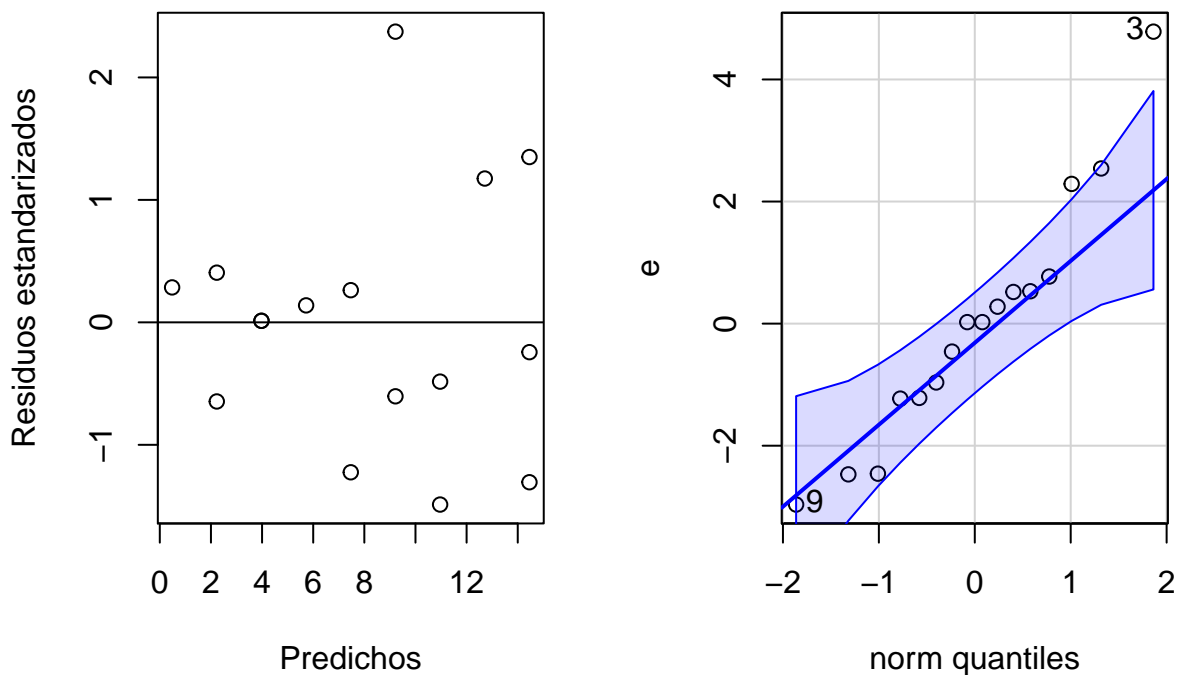
##
## Shapiro-Wilk normality test
```

```
##
## data: e
## W = 0.95662, p-value = 0.6012
anova(m3, m4)

## Analysis of Variance Table
##
## Model 1: ratones ~ lluvia + semillas
## Model 2: ratones ~ lluvia + semillas + lluvia:semillas
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      13 45.610
## 2      12 45.579   1   0.030726 0.0081 0.9298
car::vif(m4)

##          lluvia          semillas lluvia:semillas
##      18.49054         16.79689          54.73122
# ¿Y si excluimos semillas?
m5 <- update(m3, . ~ . - semillas)
e <- resid(m5) # residuos
re <- rstandard(m5) #residuos estandarizados
pre <- predict(m5) #predichos
par(mfrow = c(1, 2))
plot(pre, re, xlab = "Predichos", ylab = "Residuos estandarizados",
     main = "Gráfico de dispersión de RE vs PRED - m5")
abline(0, 0)
car::qqPlot(e)
```

ráfico de dispersión de RE vs PREC



```
## [1] 3 9
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
##
## data: e
## W = 0.94792, p-value = 0.4575
# Y si no quisiera incluir en el mismo modelo lluvia y semillas?
# comparar modelos alternativos
m_lluvia <- lm(ratones ~ lluvia + predadores + cobertura, Ratones)
m_semillas <- lm(ratones ~ semillas + predadores + cobertura, Ratones)
AIC(m_lluvia, m_semillas)

##           df      AIC
## m_lluvia   5 74.37939
## m_semillas  5 83.85843
```

Mas sobre comparacion de modelos

```
# CMe
CMe <- round(c(summary(m1)$sigma^2, summary(m2)$sigma^2, summary(m3)$sigma^2,
  summary(m4)$sigma^2, summary(m5)$sigma^2), 2)
# R2 (no para comparar entre modelos)
R2 <- c(summary(m1)$r.squared, summary(m2)$r.squared, summary(m3)$r.squared,
  summary(m4)$r.squared, summary(m5)$r.squared)
# R2 ajustado
R2aj <- c(summary(m1)$adj.r.squared, summary(m2)$adj.r.squared, summary(m3)$adj.r.squared,
  summary(m4)$adj.r.squared, summary(m5)$adj.r.squared)
# AIC
AIC <- c(AIC(m1), AIC(m2), AIC(m3), AIC(m4), AIC(m5))
# Nombre modelo (para generar un data frame bonito)
modelo <- c(1, 2, 3, 4, 5)
comp <- cbind(modelo, CMe, round(R2, 2), round(R2aj, 2), AIC)
colnames(comp) <- c("modelo", "CMe", "R2", "R2 ajust", "AIC")
comp
```

```
##      modelo  CMe  R2 R2 ajust      AIC
## [1,]      1 3.91 0.89      0.85 73.21943
## [2,]      2 3.58 0.89      0.86 71.21959
## [3,]      3 3.51 0.88      0.86 70.16667
## [4,]      4 3.80 0.88      0.85 72.15589
## [5,]      5 4.35 0.84      0.83 72.78143
```

Validacion cruzada

```
# LOO CV
library(caret)
# comparo mis cuatro modelos candidatos agrego un m5, muy simple
# Indicamos la función para el entrenamiento
train.control <- trainControl(method = "LOOCV")
# train.control<-trainControl(method = 'repeatedcv', repeats = 5)
# Entrenamos (estimamos) el modelo (n modelos con n-1
# observaciones)
m1loo <- train(ratones ~ lluvia + predadores + cobertura + semillas,
  data = Ratones, method = "lm", trControl = train.control)
m2loo <- train(ratones ~ lluvia + cobertura + semillas, data = Ratones,
  method = "lm", trControl = train.control)
m3loo <- train(ratones ~ lluvia + semillas, data = Ratones, method = "lm",
  trControl = train.control)
```



```
m4loo <- train(ratones ~ lluvia * semillas, data = Ratones, method = "lm",
  trControl = train.control)
m5loo <- train(ratones ~ lluvia, data = Ratones, method = "lm", trControl = train.control)
```

```
summary(m1loo)
```

```
# resultados
print(m1loo)
```

```
## Linear Regression
##
## 16 samples
## 4 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 15, 15, 15, 15, 15, 15, ...
## Resampling results:
##
## RMSE      Rsquared    MAE
## 2.257995  0.7948441  1.850323
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
print(m2loo)
```

```
## Linear Regression
##
## 16 samples
## 3 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 15, 15, 15, 15, 15, 15, ...
## Resampling results:
##
## RMSE      Rsquared    MAE
## 2.086466  0.8230007  1.680651
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
print(m3loo)
```

```
## Linear Regression
##
## 16 samples
## 2 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 15, 15, 15, 15, 15, 15, ...
## Resampling results:
##
## RMSE      Rsquared    MAE
## 2.01718  0.8337319  1.538411
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
print(m4loo)
```

```
## Linear Regression
##
```

```
## 16 samples
## 2 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 15, 15, 15, 15, 15, 15, ...
## Resampling results:
##
## RMSE      Rsquared   MAE
## 2.233208  0.7973781  1.693042
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
print(m5loo)
```

```
## Linear Regression
##
## 16 samples
## 1 predictor
##
## No pre-processing
## Resampling: Leave-One-Out Cross-Validation
## Summary of sample sizes: 15, 15, 15, 15, 15, 15, ...
## Resampling results:
##
## RMSE      Rsquared   MAE
## 2.185828  0.8044699  1.667999
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
# dataframe con los resultados
b
```

```
## intercept    RMSE Rsquared    MAE
## 1          TRUE 19.25325 0.9228204 14.96035
```

-Compare los modelos candidatos. Explore los graficos de residuos parciales. Seleccione el mejor modelo a su juicio fundamentando su decision y concluya sobre la importancia de los factores de control de la comunidad de ratones granivoros de desierto.

Interpretacion de los resultados

Analicemos m3

```
summary(m3)

##
## Call:
## lm(formula = ratones ~ lluvia + semillas, data = Ratones)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8101 -0.9396  0.1155  0.7081  3.8931
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.3756     1.4780  -2.284 0.039823 *
## lluvia        1.2960     0.2818   4.600 0.000498 ***
## semillas      0.8152     0.3910   2.085 0.057378 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.873 on 13 degrees of freedom
## Multiple R-squared:  0.883, Adjusted R-squared:  0.865
## F-statistic: 49.04 on 2 and 13 DF, p-value: 8.786e-07

vif(m3)

## lluvia semillas
## 2.438091 2.438091

# escriba la ecuacion final del modelo concluya, control
# bottom-up o top-down?

# grafico 3d modelo final
library(rgl)

## This build of rgl does not include OpenGL functions. Use
## rglwidget() to display results, e.g. via options(rgl.printRglwidget = TRUE).

summary(Datos) #lluvia de 1 a 9, semillas de 3 a 9

## provincia          densidad      aceite
## ba :10      1 planta por maceta :15  Min.   :0.820
## mza:10      5 plantas por maceta:15  1st Qu.:1.202
## sl :10                               Median :1.350
##                               Mean   :1.372
##                               3rd Qu.:1.525
##                               Max.   :2.040

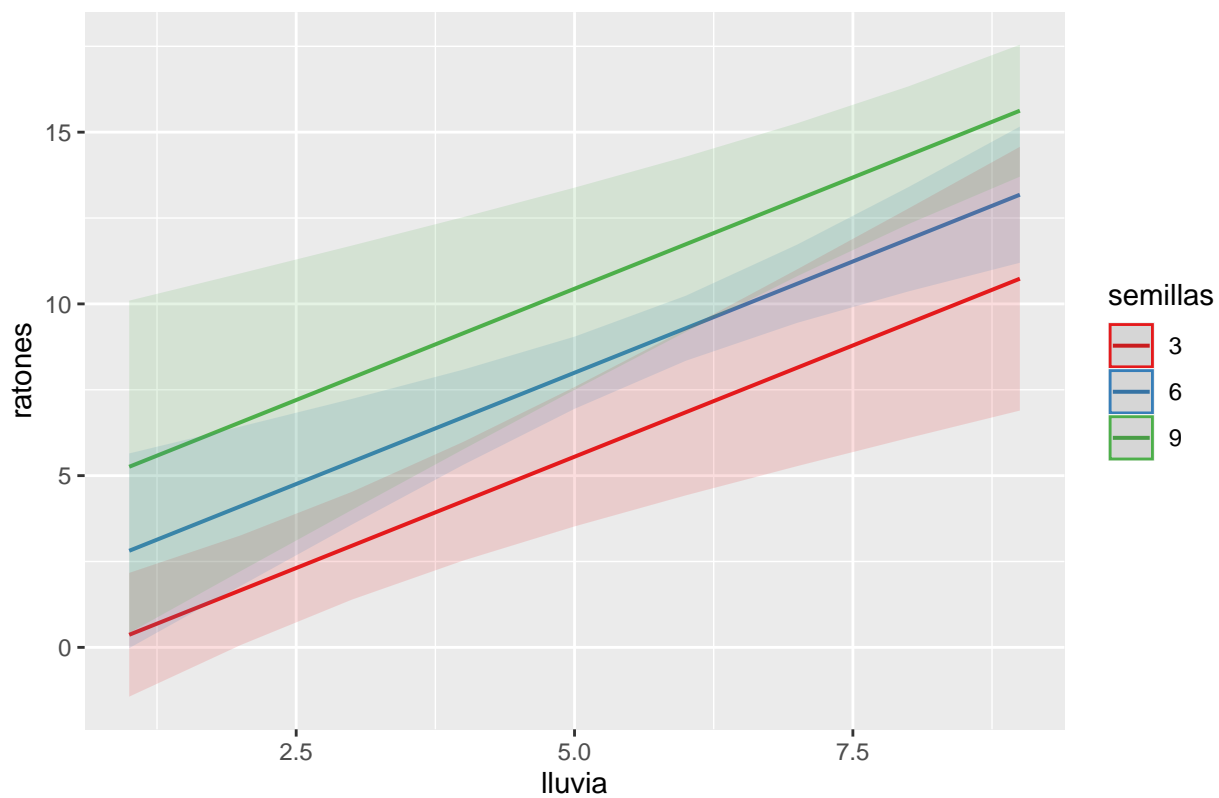
scatter3d(ratones ~ lluvia + semillas, fill = FALSE, data = Ratones)
library(sjPlot)

## Registered S3 methods overwritten by 'parameters':
## method                from
## as.double.parameters_kurtosis  datawizard
## as.double.parameters_skewness  datawizard
## as.double.parameters_smoothness datawizard
## as.numeric.parameters_kurtosis  datawizard
## as.numeric.parameters_skewness  datawizard
## as.numeric.parameters_smoothness datawizard
## print.parameters_distribution  datawizard
## print.parameters_kurtosis      datawizard
## print.parameters_skewness      datawizard
## summary.parameters_kurtosis    datawizard
## summary.parameters_skewness    datawizard

## #refugeeswelcome

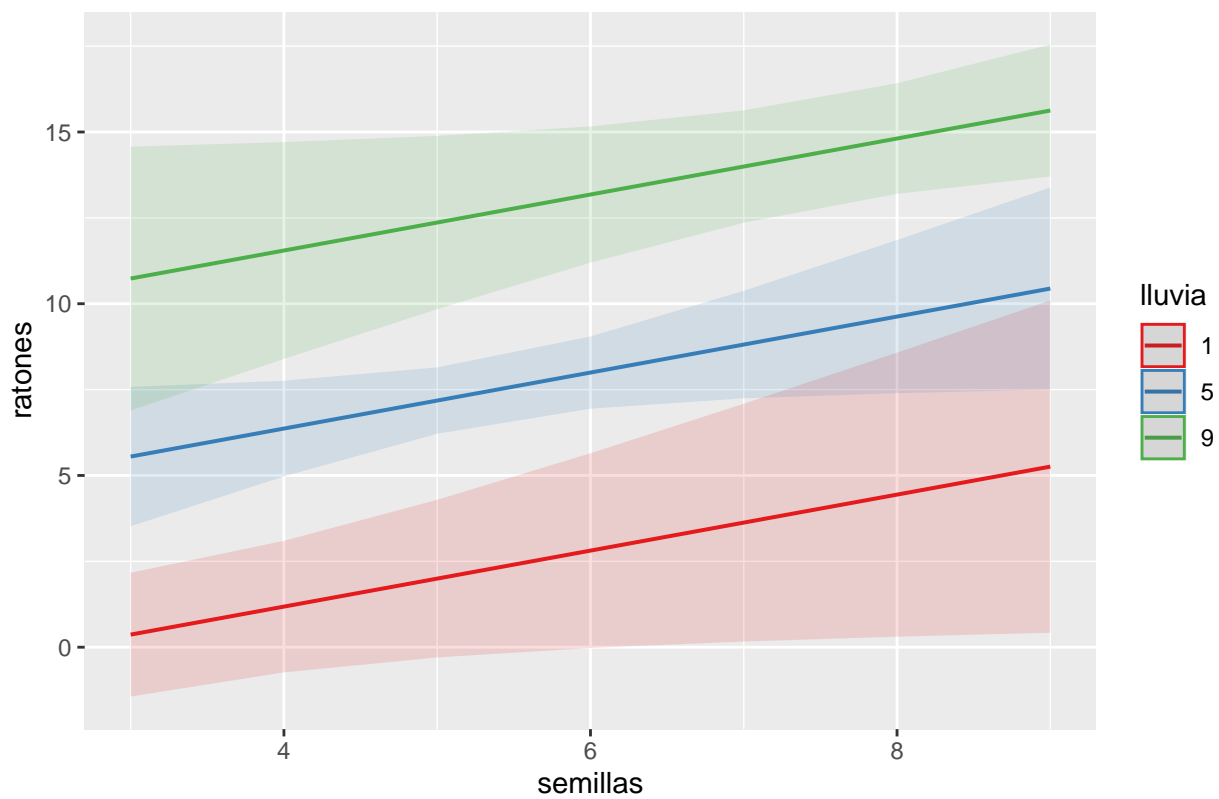
# select only levels 3, 6 y 9 from continuous semillas
plot_model(m3, type = "pred", terms = c("lluvia", "semillas [3, 6, 9]"))
```

Predicted values of ratones



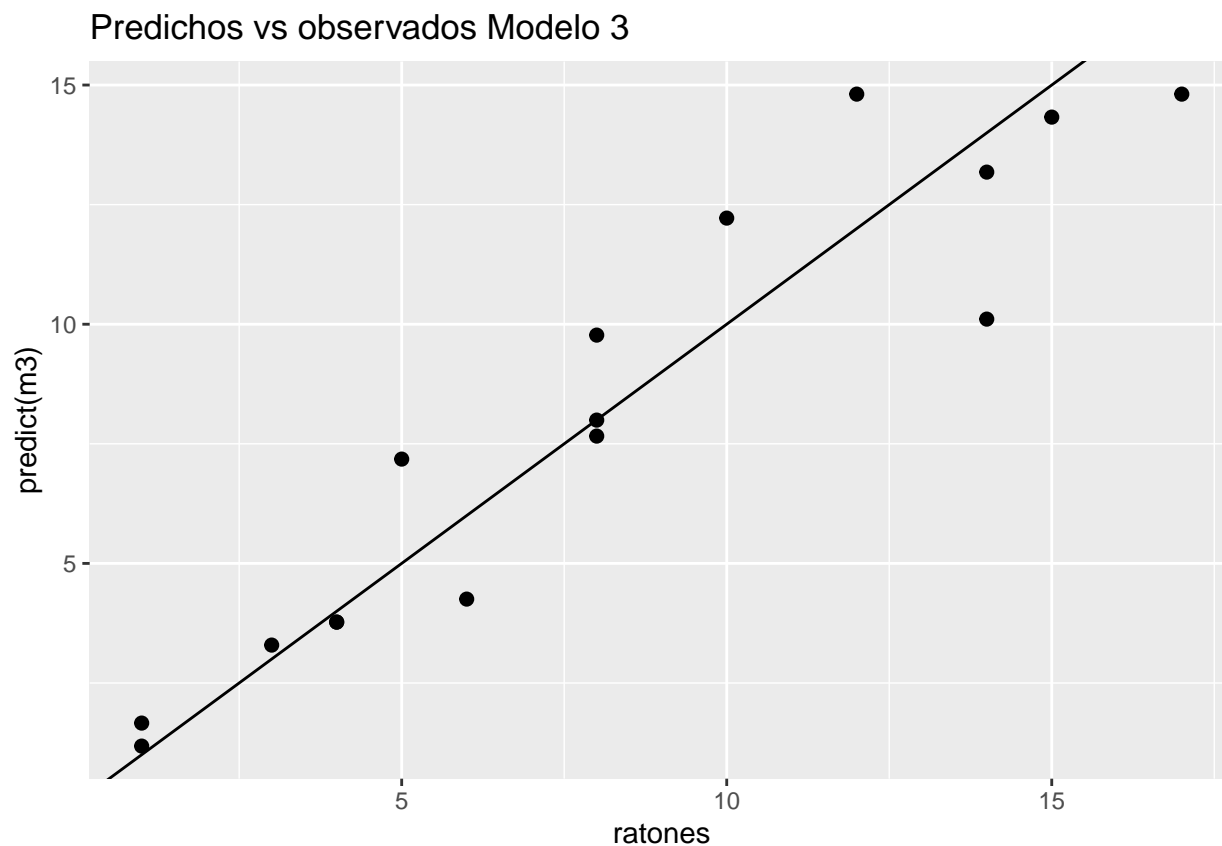
```
plot_model(m3, type = "pred", terms = c("semillas", "lluvia [1, 5, 9]"))
```

Predicted values of ratones



```
# validación Valores observados vs valores predichos por el
# modelo
```

```
p <- ggplot(Ratones, aes(x = ratones, y = predict(m3))) + geom_point(size = 2)
p + geom_abline(intercept = 0, slope = 1) + ggtitle("Predichos vs observados Modelo 3")
```



```
cor(predict(m3), Ratones$ratones)
```

```
## [1] 0.9396681
```

```
(cor(predict(m3), Ratones$ratones))^2
```

```
## [1] 0.8829761
```

```
# Residuos parciales
```

```
library(faraway)
```

```
##
```

```
## Attaching package: 'faraway'
```

```
## The following object is masked from 'package:lattice':
```

```
##
```

```
##     melanoma
```

```
## The following object is masked from 'package:psych':
```

```
##
```

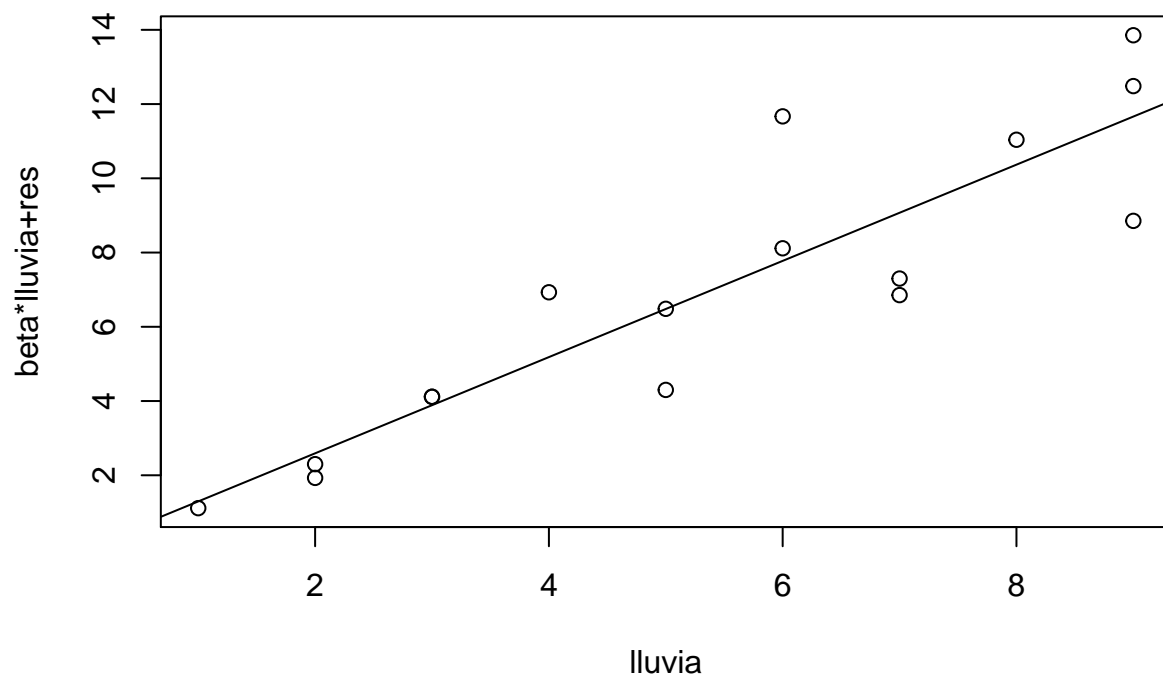
```
##     logit
```

```
## The following objects are masked from 'package:car':
```

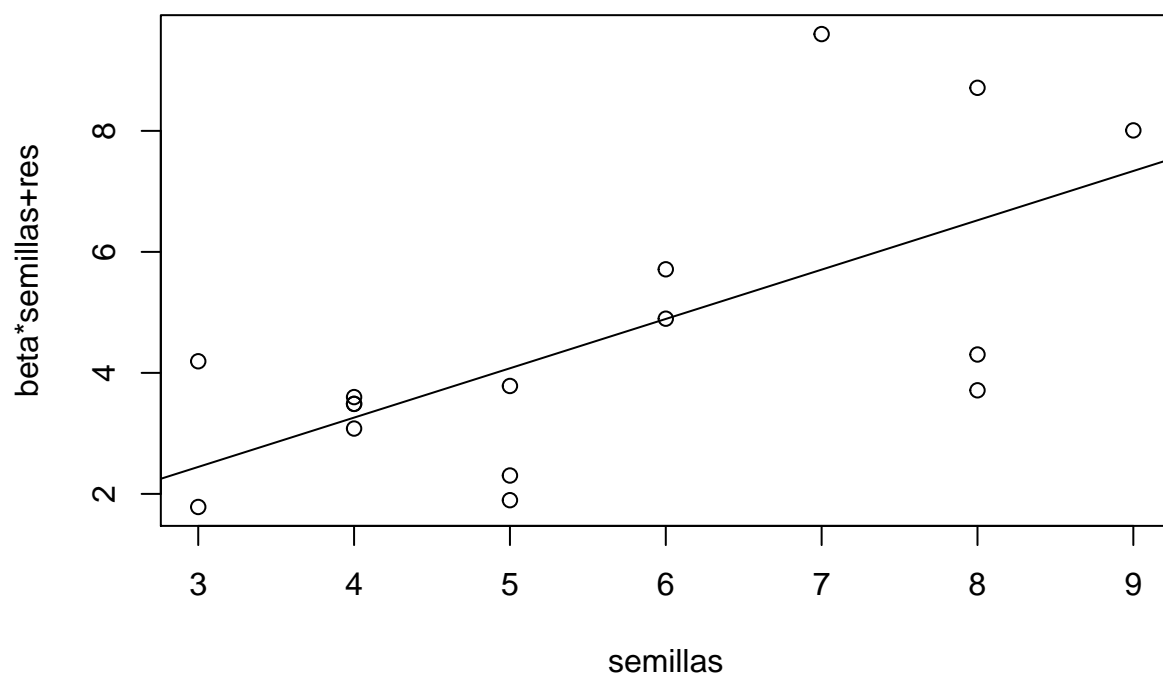
```
##
```

```
##     logit, vif
```

```
prplot(m3, 1) #Datos vs lluvia controlando semillas
```



```
prplot(m3, 2) #Datos vs semillas controlando por lluvia
```



```
# estimación y comparación de la magnitud del efecto
lluviaLCL <- round(confint(m3)[2], 3)
lluviaHCL <- round(confint(m3)[5], 3)
semillaLCL <- round(confint(m3)[3], 3)
semillaHCL <- round(confint(m3)[6], 3)
library(sjPlot)
# tab_model(m3) #sjPlot
```

interpretamos coeficientes parciales

Para el caso de la lluvia, un incremento unitario en la lluvia estara asociado con un incremento entre 0.687 y 1.905 en la abundancia de ratones ajustando por las semillas.

Para el caso de las semillas, un incremento unitario en las semillas estara asociado con un incremento entre -0.03 y 1.66 en la abundancia de ratones ajustando por las condiciones de lluvia.

Cabe destacar que el efecto de las semillas es marginal.

```
library(lm.beta)
lm.beta(m3) #coeficientes estandarizados
```

```
##
## Call:
## lm(formula = ratones ~ lluvia + semillas, data = Ratones)
##
## Standardized Coefficients::
## (Intercept)      lluvia      semillas
##  0.0000000    0.6814159    0.3088474
```

Estudiar los coeficientes estandarizados nos permite evaluar la magnitud de efecto de ambos factores. Vemos que la lluvia es 2 veces mas importante sobre la abundancia de ratones que las semillas.

```
# centrado Modelo 3_c: idem 3 pero con centrado en la media de
# las X
Ratones$lluvia_c <- Ratones$lluvia - mean(Ratones$lluvia)
Ratones$semillas_c <- Ratones$semillas - mean(Ratones$semillas)
m3_c <- lm(ratones ~ lluvia_c + semillas_c, Ratones)
summary(m3_c)
```

```
##
## Call:
## lm(formula = ratones ~ lluvia_c + semillas_c, data = Ratones)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8101 -0.9396  0.1155  0.7081  3.8931
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.1250     0.4683  17.351 2.26e-10 ***
## lluvia_c       1.2960     0.2818   4.600 0.000498 ***
## semillas_c     0.8152     0.3910   2.085 0.057378 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.873 on 13 degrees of freedom
## Multiple R-squared:  0.883, Adjusted R-squared:  0.865
## F-statistic: 49.04 on 2 and 13 DF, p-value: 8.786e-07
# tab_model(m3_c) #sjPlot
```

```

# Seleccion automatica
library(MuMIn)
# modelo aditivo
dredge(lm(ratones ~ lluvia + predadores + cobertura + semillas, data = Ratones,
na.action = "na.fail"))

## Fixed term is "(Intercept)"

## Global model call: lm(formula = ratones ~ lluvia + predadores + cobertura + semillas,
## data = Ratones, na.action = "na.fail")
## ---
## Model selection table
##      (Intrc)   cbrtr lluvi      prddr  smlls df  logLik  AICc delta weight
## 11 -3.3760      1.296      0.8152  4 -31.083  73.8  0.00  0.419
## 3  -1.2660      1.747      3 -33.391  74.8  0.98  0.257
## 7  -0.7011      1.822 -0.322100      4 -32.464  76.6  2.76  0.105
## 12 -4.3140  0.2340  1.363      0.7826  5 -30.610  77.2  3.42  0.076
## 4  -2.5160  0.2859  1.807      4 -32.854  77.3  3.54  0.071
## 15 -3.0960      1.346 -0.066270  0.7522  5 -31.050  78.1  4.30  0.049
## 8  -1.6640  0.2016  1.853 -0.275700      5 -32.190  80.4  6.58  0.016
## 16 -4.2980  0.2333  1.365 -0.003237  0.7796  6 -30.610  82.6  8.75  0.005
## 9  -4.0940      2.1970  3 -38.811  85.6 11.82  0.001
## 13 -6.3300      0.589200  2.2810  4 -37.031  85.7 11.90  0.001
## 10 -3.6740 -0.1007      2.1800  4 -38.776  89.2 15.39  0.000
## 14 -7.2100  0.1646      0.640200  2.3150  5 -36.929  89.9 16.06  0.000
## 1   8.1250      2 -48.246 101.4 27.61  0.000
## 2   9.6890 -0.4811      3 -47.990 104.0 30.18  0.000
## 5   7.1390      0.328600      3 -48.090 104.2 30.38  0.000
## 6   8.7830 -0.4003      0.214200      4 -47.930 107.5 33.69  0.000
## Models ranked by AICc(x)

# modelo con interaccion
dredge(lm(ratones ~ lluvia * predadores * semillas, data = Ratones,
na.action = "na.fail"))

## Fixed term is "(Intercept)"

## Global model call: lm(formula = ratones ~ lluvia * predadores * semillas, data = Ratones,
## na.action = "na.fail")
## ---
## Model selection table
##      (Int)   llv      prd      sml  llv:prd  llv:sml  prd:sml  llv:prd:sml  df
## 6  -3.3760  1.2960      0.8152      4
## 2  -1.2660  1.7470      3
## 4  -0.7011  1.8220 -0.32210      4
## 8  -3.0960  1.3460 -0.06627  0.7522      5
## 22 -3.7780  1.3640      0.9040      -0.01353      5
## 12  0.4447  1.6490 -0.74780      0.06009      5
## 16 -1.6790  1.0620 -0.64580  0.8149  0.08481      6
## 40 -2.3100  1.3350 -0.39120  0.6241      0.05948      6
## 24 -3.5900  1.4340 -0.06961  0.8610      -0.01706      6
## 5  -4.0940      2.1970      3
## 7  -6.3300      0.58920  2.2810      4
## 48 -2.6980  0.9416 -0.22130  1.1320  0.12840      -0.13230      7
## 32 -1.2410  0.9832 -0.65810  0.7282  0.08702  0.01385      7
## 39 -4.3180      -0.22800  1.9320      0.14710      5
## 56 -2.7590  1.4120 -0.38830  0.7223      -0.01503  0.05842      7
## 64 -1.9380  0.7850 -0.21690  0.9888  0.13540  0.02617 -0.14090      8
## 1   8.1250      2
## 3   7.1390      0.32860      3
## 128 -15.7200  3.0430  5.14200  3.7180 -0.66150 -0.38270 -1.27300      0.1579  9

```



```
##      logLik  AICc delta weight
## 6    -31.083  73.8  0.00  0.462
## 2    -33.391  74.8  0.98  0.283
## 4    -32.464  76.6  2.76  0.116
## 8    -31.050  78.1  4.30  0.054
## 22   -31.078  78.2  4.35  0.052
## 12   -32.259  80.5  6.72  0.016
## 16   -30.567  82.5  8.66  0.006
## 40   -30.999  83.3  9.53  0.004
## 24   -31.041  83.4  9.61  0.004
## 5    -38.811  85.6 11.82  0.001
## 7    -37.031  85.7 11.90  0.001
## 48   -30.430  88.9 15.06  0.000
## 32   -30.561  89.1 15.32  0.000
## 39   -36.881  89.8 15.96  0.000
## 56   -30.992  90.0 16.18  0.000
## 64   -30.410  97.4 23.59  0.000
## 1    -48.246 101.4 27.61  0.000
## 3    -48.090 104.2 30.38  0.000
## 128  -29.807 107.6 33.81  0.000
## Models ranked by AICc(x)
```

*# Modelos candidatos? dónde se ubica nuestro 'm3'? que podemos
concluir en base a todos estos criterios acerca de la
capacidad explicativa de cada una de las variables? se le
ocurren otras cosas para seguir probando?*

Problema 9. Asociación entre rasgos paternos y tamaño de la bola de cría en el escarabajo estercolero *Sulcophanaeus* sp

Los rasgos de los padres suelen afectar el desarrollo de rasgos de su cría. Un mecanismo por el cual los padres pueden influir en el fenotipo de la descendencia es a través del nivel de cuidado que proporcionan. *Sulcophanaeus* sp (Coleoptera, Scarabaeidae) es un escarabajo estercolero con cuidados biparentales. El macho construye una bola de cría formada por estiércol, en la que la hembra deposita un huevo, que al eclosionar se alimenta de esa masa. Se sabe que cuanto mayor es el tamaño de la bola de cría, mayor es la disponibilidad de alimento y mayor el desarrollo de la cría, pero se desconocen los factores que regulan el tamaño de las bolas de cría. Los machos presentan además variaciones en el tamaño de sus cuernos, que utilizan durante las peleas por las hembras permitiéndoles incrementar su éxito reproductivo.

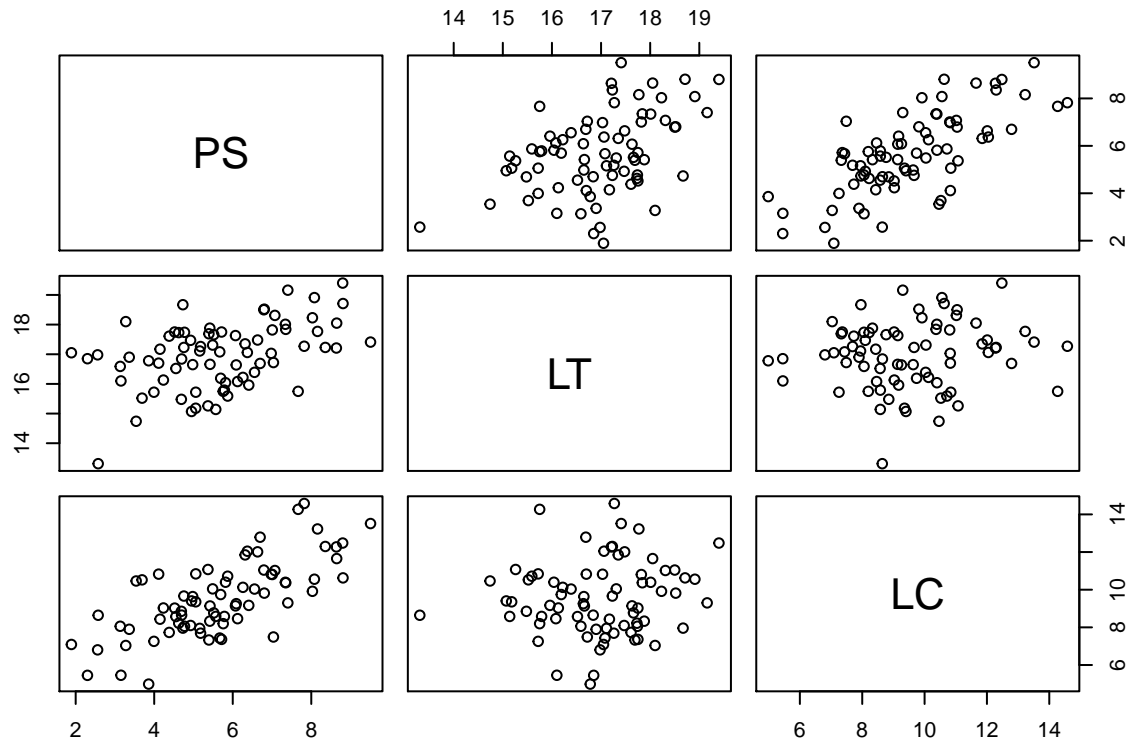
Se desea probar la hipótesis de que ambas características morfológicas de los machos (tamaño corporal y la longitud de los cuernos) afectan en forma sinérgica el tamaño de las bolas de cría. Para demostrarlo se capturaron ejemplares adultos en la provincia de Buenos Aires y se les midió el largo corporal total (LT) y del cuerno (LC), ambos en mm. Se seleccionaron machos de manera de cubrir el rango de combinaciones de tamaños y se los cruzó con una hembra de tamaño promedio. Cada pareja fue mantenida en una cámara de cría individual con estiércol de vaca. Se obtuvieron 75 bolas de cría, a las que se les determinó el peso seco (PS), en gramos. Los resultados se encuentran en el archivo *escarab.csv*

- Explore los datos.

```
escarab <- read.csv("/home/jose/Documents/materias/biome2/2019/tps/tp4/escarab.csv",
  header = T)
summary(escarab)
```

```
##      PS          LT          LC
##  Min.   :1.891   Min.   :13.31   Min.   : 4.985
## 1st Qu.:4.658   1st Qu.:16.16   1st Qu.: 8.147
## Median :5.567   Median :17.06   Median : 9.300
## Mean   :5.669   Mean   :16.96   Mean   : 9.529
## 3rd Qu.:6.798   3rd Qu.:17.73   3rd Qu.:10.761
## Max.   :9.509   Max.   :19.40   Max.   :14.580
```

```
plot(escarab)
```



- Plantee el

modelo que permita poner a prueba la hipótesis de investigación.

```
m1 = lm(PS ~ LT * LC, data = escarab)
summary(m1)
```

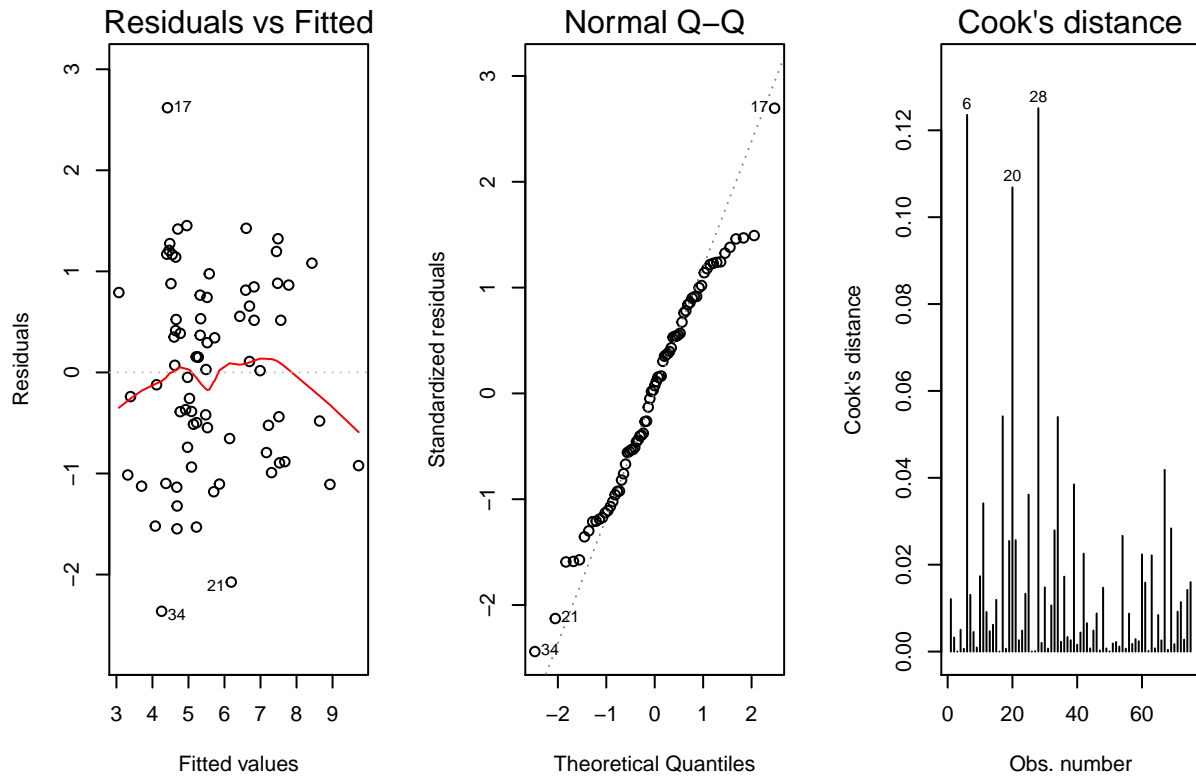
```
##
## Call:
## lm(formula = PS ~ LT * LC, data = escarab)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36338 -0.76682  0.07065  0.77764  2.61904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.1662    11.4714   1.409   0.1631
## LT           -0.9422     0.6743  -1.397   0.1667
## LC           -2.0460     1.1664  -1.754   0.0837 .
## LT:LC         0.1543     0.0684   2.256   0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9861 on 71 degrees of freedom
## Multiple R-squared:  0.6745, Adjusted R-squared:  0.6607
## F-statistic: 49.04 on 3 and 71 DF,  p-value: < 2.2e-16
```

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: PS
##           Df Sum Sq Mean Sq F value    Pr(>F)
## LT         1  41.382   41.382  42.5551 8.676e-09 ***
```

```
## LC          1 96.734  96.734 99.4769 3.830e-15 ***
## LT:LC       1  4.951   4.951  5.0916  0.02712 *
## Residuals 71 69.042   0.972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow = c(1, 3))
plot(m1, which = c(1, 2, 4))
```



```
par(mfrow = c(1, 1))
```

- Estudie si existen violaciones a los supuestos del modelo propuesto.

```
shapiro.test(residuals(m1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(m1)
## W = 0.98353, p-value = 0.4396
```

```
# Hay un outlier. Cambia mucho si lo sacamos?
m2 = lm(PS ~ LT * LC, data = escarab[-17, ])
summary(m2) #no demasiado, lo dejamos
```

```
##
## Call:
## lm(formula = PS ~ LT * LC, data = escarab[-17, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2834 -0.7771  0.1015  0.7646  1.5011
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.9065    10.9551   1.361   0.1780
## LT          -0.8808     0.6438  -1.368   0.1756
## LC          -1.9258     1.1138  -1.729   0.0882 .
## LT:LC        0.1484     0.0653   2.272   0.0261 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.941 on 70 degrees of freedom
## Multiple R-squared:  0.7052, Adjusted R-squared:  0.6925
## F-statistic: 55.81 on 3 and 70 DF,  p-value: < 2.2e-16

library(car)
vif(m1) #da horrible
```

```
##           LT           LC           LT:LC
##  43.19103 402.98992 472.46431
```

- ¿Qué cambia si centra los valores de LT y LC con respecto a sus respectivas medias?

```
# y si centramos?
escarab$LTc <- escarab$LT - mean(escarab$LT)
escarab$LCc <- escarab$LC - mean(escarab$LC)
m1c = lm(PS ~ LTc * LCc, data = escarab)
summary(m1c)

##
## Call:
## lm(formula = PS ~ LTc * LCc, data = escarab)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36338 -0.76682  0.07065  0.77764  2.61904
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.63363    0.11493  49.018 < 2e-16 ***
## LTc         0.52857    0.10419   5.073 3.01e-06 ***
## LCc         0.57130    0.05865   9.741 1.02e-14 ***
## LTc:LCc     0.15434    0.06840   2.256  0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9861 on 71 degrees of freedom
## Multiple R-squared:  0.6745, Adjusted R-squared:  0.6607
## F-statistic: 49.04 on 3 and 71 DF,  p-value: < 2.2e-16

# la interaccion no cambia, los efectos parciales sí
vif(m1c)
```

```
##           LTc           LCc           LTc:LCc
##  1.031232 1.018724 1.030319
```

```
# centramos? como no miramos los coef parciales porq la int es
# signif, podemos elegir no centrar
```

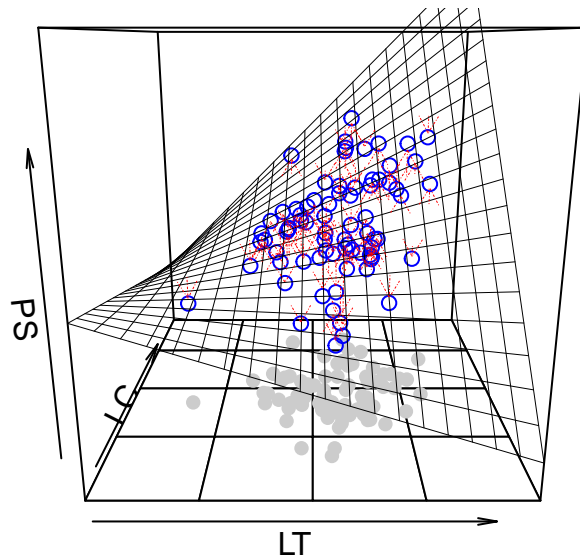
- Ajuste el modelo y concluya.

```
# graficos
library(rgl)
data(iris)
sep.l <- iris$Sepal.Length
sep.w <- iris$Sepal.Width
```

```

pet.l <- iris$Petal.Length
# scatter3d(x = sep.l, y = pet.l, z = sep.w, groups =
# iris$Species, surface=FALSE, grid = FALSE, ellipsoid = TRUE,
# surface.col = c('red', 'blue', 'green')) interaccion
library(rockchalk)
mcGraph3(escarab$LT, escarab$LC, escarab$PS, interaction = TRUE, theta = 0)

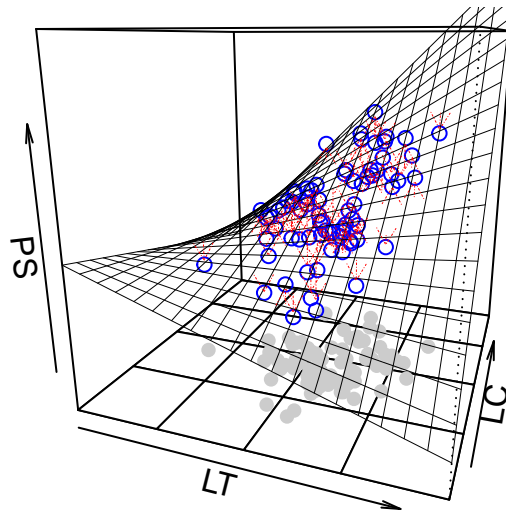
```



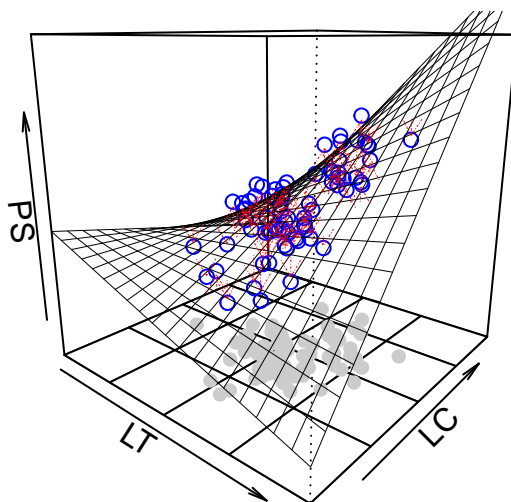
```

mcGraph3(escarab$LT, escarab$LC, escarab$PS, interaction = TRUE, theta = 20)

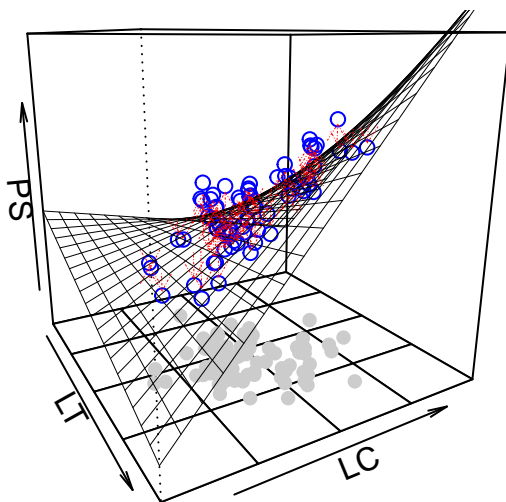
```



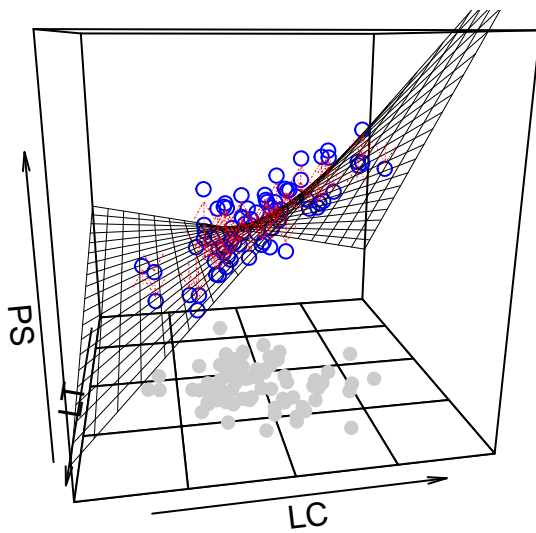
```
mcGraph3(escarab$LT, escarab$LC, escarab$PS, interaction = TRUE, theta = 40)
```



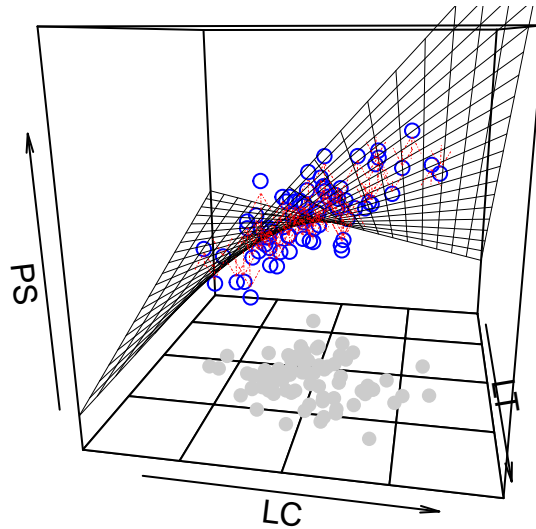
```
mcGraph3(escarab$LT, escarab$LC, escarab$PS, interaction = TRUE, theta = 60)
```



```
mcGraph3(escarab$LT, escarab$LC, escarab$PS, interaction = TRUE, theta = 80)
```



```
mcGraph3(escarab$LT, escarab$LC, escarab$PS, interaction = TRUE, theta = 100)
```



- Interprete el R2 y estudie la relación entre los valores observados de PS en bolas de cría y los predichos por el modelo.

```
summary(m1)
```

```
##
## Call:
## lm(formula = PS ~ LT * LC, data = escarab)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36338 -0.76682  0.07065  0.77764  2.61904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.1662    11.4714   1.409   0.1631
## LT           -0.9422     0.6743  -1.397   0.1667
## LC           -2.0460     1.1664  -1.754   0.0837 .
## LT:LC         0.1543     0.0684   2.256   0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9861 on 71 degrees of freedom
## Multiple R-squared:  0.6745, Adjusted R-squared:  0.6607
## F-statistic: 49.04 on 3 and 71 DF, p-value: < 2.2e-16
```

```
summary(m2)
```

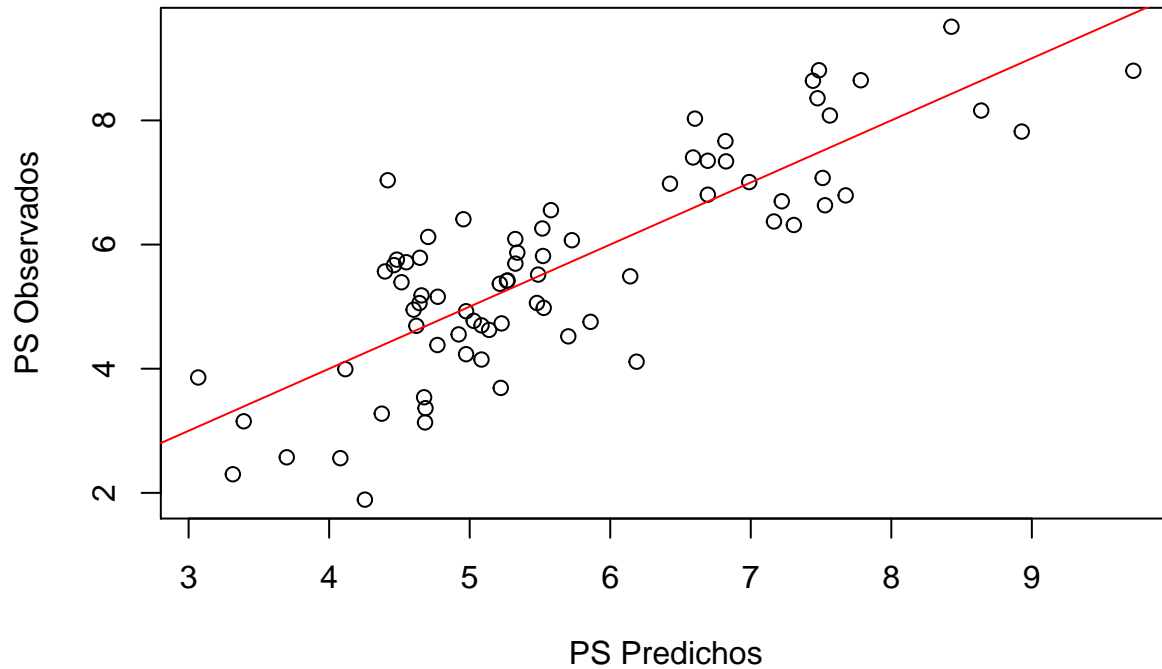
```
##
## Call:
## lm(formula = PS ~ LT * LC, data = escarab[-17, ])
```



```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2834 -0.7771  0.1015  0.7646  1.5011
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.9065     10.9551   1.361   0.1780
## LT           -0.8808      0.6438  -1.368   0.1756
## LC           -1.9258      1.1138  -1.729   0.0882 .
## LT:LC         0.1484      0.0653   2.272   0.0261 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.941 on 70 degrees of freedom
## Multiple R-squared:  0.7052, Adjusted R-squared:  0.6925
## F-statistic: 55.81 on 3 and 70 DF, p-value: < 2.2e-16
```

```
summary(m1c)
```

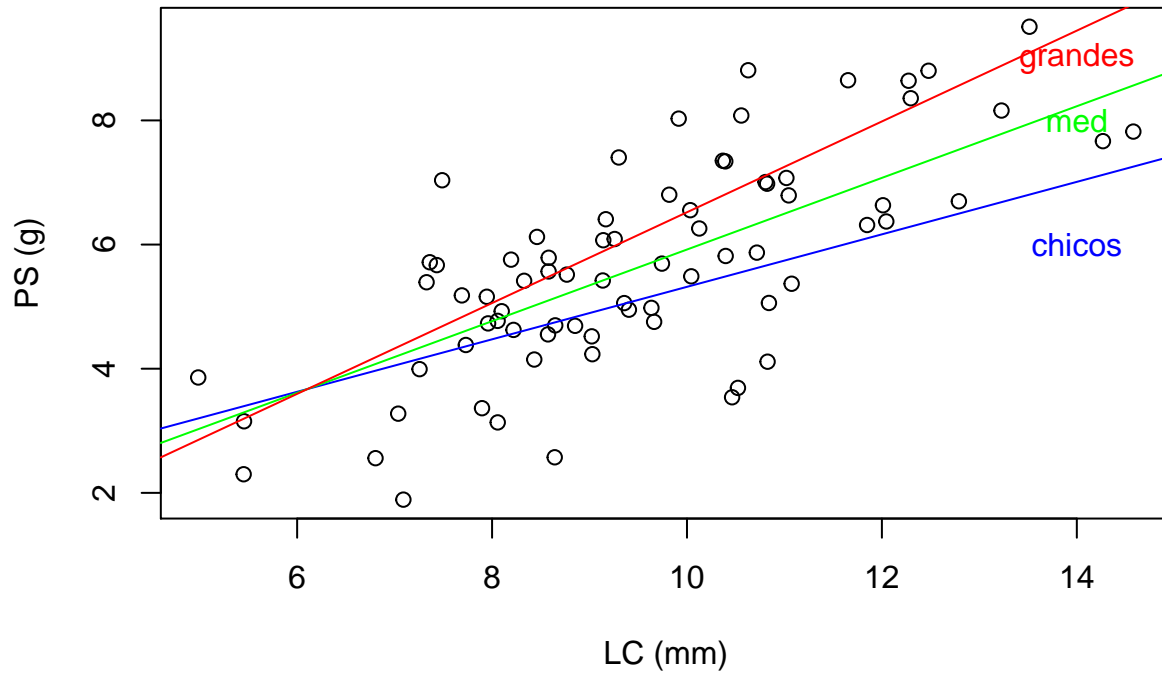
```
##
## Call:
## lm(formula = PS ~ LTc * LCc, data = escarab)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36338 -0.76682  0.07065  0.77764  2.61904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.63363     0.11493  49.018 < 2e-16 ***
## LTc          0.52857     0.10419   5.073 3.01e-06 ***
## LCc          0.57130     0.05865   9.741 1.02e-14 ***
## LTc:LCc      0.15434     0.06840   2.256  0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9861 on 71 degrees of freedom
## Multiple R-squared:  0.6745, Adjusted R-squared:  0.6607
## F-statistic: 49.04 on 3 and 71 DF, p-value: < 2.2e-16
plot(predict(m1), escarab$PS, ylab = "PS Observados", xlab = "PS Predichos")
abline(0, 1, col = "red")
```



- Grafique el

PS esperado de las bolas de cría según el LC del macho para tres tamaños de LT (media \pm DE).

```
# Como estudiamos la interaccion? Graficamente Calculamos la
# ecuacion de PS en función de LC para 3 valores de LT media LT
# = 17, DE=1, elijo 16, 17 y 18 mm y calculo PS vs LC (ver diapo
# 22 teo8)
plot(escarab$LC, escarab$PS, xlab = "LC (mm)", ylab = "PS (g)")
abline(1.091, 0.4228, col = "blue")
abline(0.1488, 0.5771, col = "green")
abline(-0.7934, 0.7314, col = "red")
text(14, 6, "chicos", srt = 0.2, col = "blue")
text(14, 8, "med", srt = 0.2, col = "green")
text(14, 9, "grandes", srt = 0.2, col = "red")
```



- Otro investigador desea realizar el mismo ensayo, pero quiere diferenciar entre dos especies de *Sulcophanaeus*. Escriba la nueva ecuación del modelo en parámetros indicando qué significan en el contexto de esta experiencia.
- ¿Cómo se modificaría el modelo si el PS de cada bola se hubiese determinado tres veces. ¿Y si la experiencia se realizaba en 5 “tandas” distintas?