

ANOVA con Interacción

Sebastián Sánchez Sandí

2025-08-04

Preparación

Los datos para realizar el análisis se encuentran en *tortugas2.csv*.

```
base = read.csv("tortugas2.csv")
str(base)
```

```
## 'data.frame': 24 obs. of 3 variables:
## $ especie : chr "chelonia" "chelonia" "chelonia" "chelonia" ...
## $ cond : int 1 1 1 1 2 2 2 2 3 3 ...
## $ proteina: int 32 30 34 35 40 43 43 42 38 35 ...
```

Cambiamos el tipo de genero y cond a factor para poder utilizar las herramientas del análisis.

```
base$genero = factor(base$especie)
base$cond = factor(base$cond)
levels(base$cond) = c("estricta", "balanceada", "abundancia")
```

Efectos

Estimamos los efectos simples y de interacción para buscar indicios de interacción.

```
mediaGeneral = mean(base$proteina)
mediaCond = with(data = base, tapply(proteina, cond, mean))
mediaEspecie = with(data = base, tapply(proteina, especie, mean))
```

```
alpha = mediaCond - mediaGeneral; alpha
```

```
##   estricta balanceada abundancia
##   -3.375      2.375      1.000
```

```
beta = mediaEspecie - mediaGeneral; beta
```

```
##   chelonia kynosternon
##           1          -1
```

Una vez tenemos los efectos simples calculamos las medias bajo el modelo con interacción y modelo sin interacción para obtener los efectos de interacción.

```

mediaCI = with(data = base, tapply(proteina, list(cond, especie), mean))
cont = 1
mediaSI = c()
for(i in 1:2) {
  for(j in 1:3) {
    mediaSI[cont] = mediaGeneral + alpha[j] + beta[i]
    cont = cont + 1
  }
}

comp = cbind(as.vector(mediaCI), mediaSI)
colnames(comp) = c("MediaCI", "MediaSI")
row.names(comp) = c("e-ch", "b-ch", "a-ch", "e-ky", "b-ky", "a-ky")
comp

```

```

##      MediaCI MediaSI
## e-ch    32.75  33.625
## b-ch    42.00  39.375
## a-ch    36.25  38.000
## e-ky    32.50  31.625
## b-ky    34.75  37.375
## a-ky    37.75  36.000

```

Como podemos ver en este caso las medias estimadas con un modelo y las medias del otro modelo son relativamente distintas, lo que puede indicar que existe interacción. Construimos un gráfico para visualizar estas diferencias.

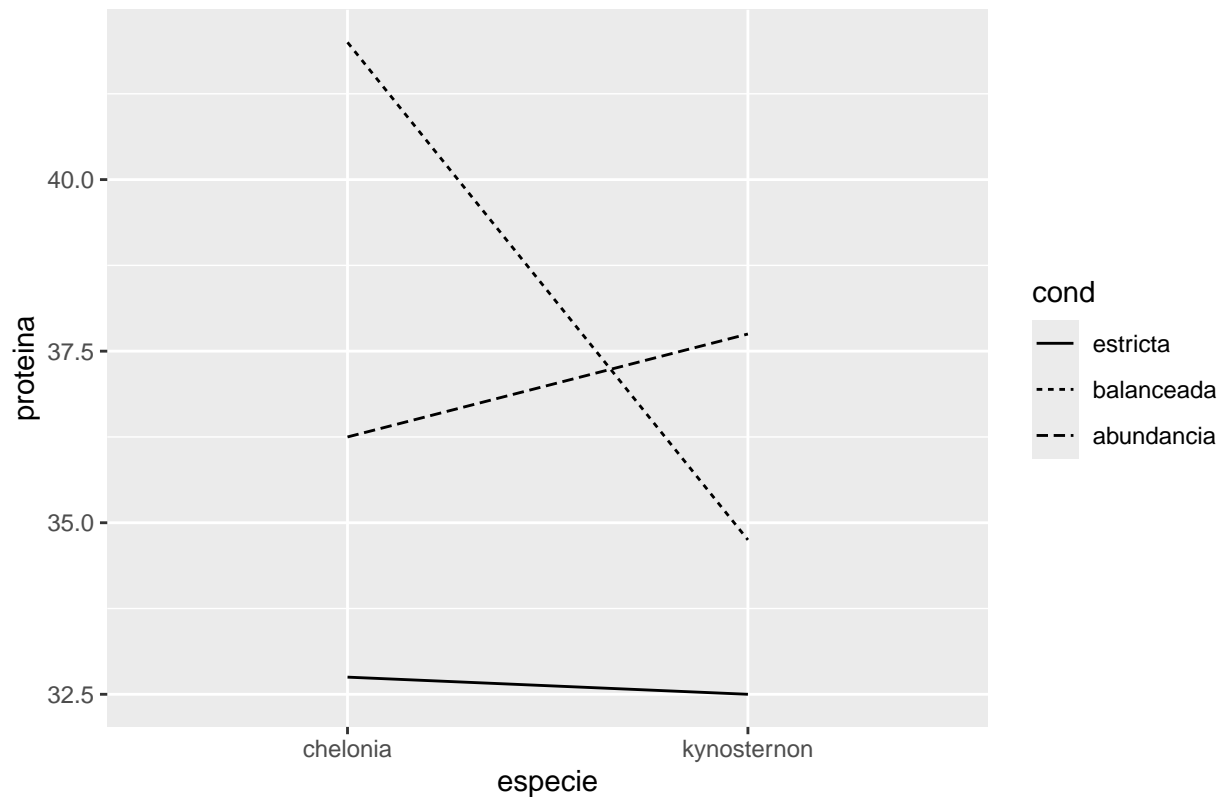
```

library(ggplot2)

ggplot(data = base, aes(x = especie, y = proteina, group = cond)) +
  stat_summary(fun = "mean", geom = "line", aes(linetype = cond)) +
  labs(title = "Comparación medias de proteina para cada especie")

```

Comparación medias de proteína para cada especie



Gráficamente parece evidente que existe interacción ya que las diferencias entre las medias de proteína para cada condición varía dependiendo de la especie. Podemos obtener directamente los efectos de interacción con el siguiente comando:

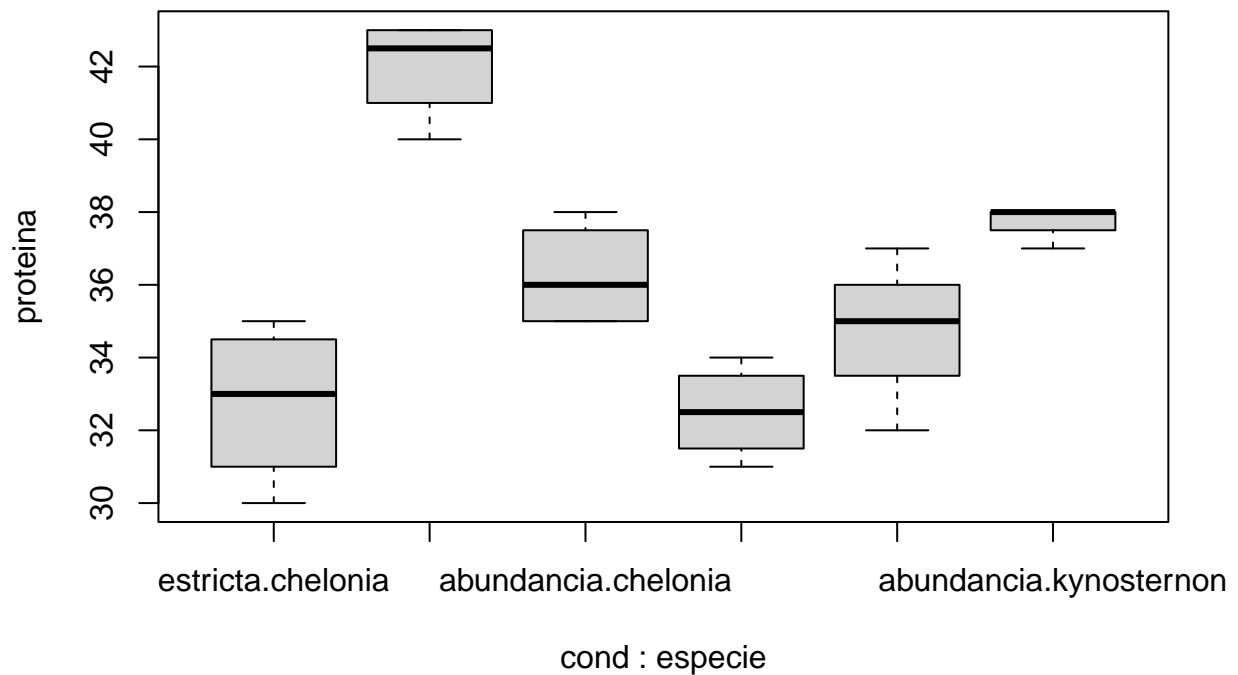
```
mod = aov(proteína~cond*especie, data = base)
model.tables(mod)
```

```
## Tables of effects
##
##  cond
##  cond
##    estricta balanceada abundancia
##    -3.375      2.375      1.000
##
##  especie
##  especie
##    chelonía kynosternon
##           1          -1
##
##  cond:especie
##           especie
##  cond    chelonía kynosternon
##  estricta  -0.875    0.875
##  balanceada 2.625   -2.625
##  abundancia -1.750    1.750
```

Varianza del Error

Pasamos a realizar el análisis de la varianza del error para verificar que se cumple el supuesto de homocedasticidad. Construimos un boxplot para visualizar gráficamente la estimación de la varianza de cada tratamiento.

```
boxplot(proteina~cond*especie, data = base)
```



Según el gráfico parece que no se cumple el supuesto de homocedasticidad, pero vamos a realizar la prueba formal.

```
bartlett.test(proteina~interaction(cond, especie), data = base)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data:  proteina by interaction(cond, especie)  
## Bartlett's K-squared = 5.2091, df = 5, p-value = 0.3909
```

Como la probabilidad de cometer error tipo 1 es mayor a 0.05, entonces no rechazamos la hipótesis nula. Se espera que si se cumpla el supuesto de homocedasticidad.

Hipótesis de Interacción

Ahora pasamos a poner a prueba la hipótesis de interacción:

- $H_0 : (\alpha\beta)_{ij} = 0$ con $H_a : (\alpha\beta)_{ij} \neq 0$

```
mod2 = lm(proteina~cond*especie, data = base)
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: proteina
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cond         2  144.25   72.125  28.2228 2.824e-06 ***
## especie       1   24.00   24.000   9.3913 0.006677 **
## cond:especie  2   85.75   42.875  16.7772 7.710e-05 ***
## Residuals    18   46.00    2.556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como la probabilidad de cometer error tipo 1 es menor a 0.05, entonces rechazamos la hipótesis nula. Se espera que si exista interacción entre los factores, en otras palabras, el promedio de la proteína generada por la tortuga bajo cada dieta varia dependiendo de la especie.

Comparaciones bajo el Modelo con Interacción

Como nos quedamos con el modelo con interacción, entonces las comparaciones se realizan entre los niveles del factor de diseño para cada nivel del otro factor. Estas hipótesis serían:

Especie ch:

- $H_0 : \mu_{11} = \mu_{21}$
- $H_0 : \mu_{11} = \mu_{31}$
- $H_0 : \mu_{21} = \mu_{31}$

Especie ky:

- $H_0 : \mu_{12} = \mu_{22}$
- $H_0 : \mu_{12} = \mu_{32}$
- $H_0 : \mu_{22} = \mu_{32}$

Con esto podemos obtener los vectores de los contrastes y de los coeficientes.

```
c1 = model.matrix(mod2)[1,]
c2 = model.matrix(mod2)[5,]
c3 = model.matrix(mod2)[9,]
c4 = model.matrix(mod2)[13,]
c5 = model.matrix(mod2)[17,]
c6 = model.matrix(mod2)[21,]

h1 = c2 - c1
h2 = c3 - c1
h3 = c2 - c3
h4 = c5 - c4
h5 = c6 - c4
```

```
h6 = c6 - c5

coef = mod2$coefficients
h = cbind(h1, h2, h3, h4, h5, h6)

L = t(h) %*% coef
L
```

```
##      [,1]
## h1 9.25
## h2 3.50
## h3 5.75
## h4 2.25
## h5 5.25
## h6 3.00
```

Ahora una vez tenemos los contrastes calculamos la varianza de los estimadores.

```
seL = sqrt(diag(t(h) %*% vcov(mod2) %*% h))
seL
```

```
##      h1      h2      h3      h4      h5      h6
## 1.130388 1.130388 1.130388 1.130388 1.130388 1.130388
```

Tiene sentido que sea el mismo para todos ya que se cumple el supuesto de homocedasticidad. Con todo esto estandarizamos el estadístico para probar que sean significantes.

```
q = L / seL
round(pt(q, 18, lower.tail = F), 4)
```

```
##      [,1]
## h1 0.0000
## h2 0.0031
## h3 0.0000
## h4 0.0310
## h5 0.0001
## h6 0.0081
```

Comparamos todas estas probabilidades contra $\frac{\alpha}{cd}$ donde c es el número de límites y d el número de comparaciones ya que en cada caso se están realizando 3 comparaciones, si fueran menos el denominador cambia. Este número es 0.017. Rechazamos todas las hipótesis menos una. Calculamos el límite inferior para todos los cuales se encontró diferencia. El estadístico t se busca con una probabilidad de $1 - \frac{\alpha}{cd}$.

```
tc1 = qt(1 - 0.05/3, 18)
tc2 = qt(1 - 0.05/2, 18)

lim1 = round(L[1:3] - tc1 * seL[1:3], 4); lim1
```

```
##      h1      h2      h3
## 6.6452 0.8952 3.1452
```

```
lim2 = round(L[5:6] - tc1 * seL[5:6], 4); lim2
```

```
##      h5      h6  
## 2.6452 0.3952
```

En conclusión, con un 95% de confianza se puede afirmar que para el caso de la especie *chelonina* con una dieta balanceada produce en promedio al menos 6.64 más proteína que la dieta estricta y al menos 3.14 más proteína en promedio que alimento en abundancia. Para la especie *kynosternon* las tortugas con una dieta estricta y una dieta balanceada no presentan una diferencia significativa, pero las que están bajo alimentación en abundancia producen en promedio al menos 2.64 más proteína que bajo alimentación estricta y al menos 0.39 más proteína que dieta balanceada. Es importante destacar que se debe discutir con el investigador si estas diferencias son relevantes para el estudio.