

ANOVA sin Interacción

Sebastián Sánchez Sandí

2025-08-02

Preparación

Los datos para realizar el análisis se encuentran en *tortugas1.csv*.

```
base = read.csv("tortugas1.csv")
str(base)
```

```
## 'data.frame':    16 obs. of  3 variables:
## $ genero   : int  1 1 1 1 1 1 1 1 2 2 ...
## $ cond     : chr  "alimento" "alimento" "alimento" "alimento" ...
## $ proteina: num  42.8 43.1 40.4 46.6 38.9 40.3 37.5 42.9 42.2 38.7 ...
```

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

```
base$genero = factor(base$genero)
base$cond = factor(base$cond)
levels(base$genero) = c("macho", "hembra")
```

Este experimento cuenta con 2 factores cada uno con 2 niveles. Esto indica que el experimento tiene 4 tratamientos. El factor de diseño es la condición de la alimentación de las tortugas.

```
table(list(base$genero, base$cond))
```

```
##           .2
## .1      alimento dieta
## macho         4      4
## hembra         4      4
```

Cada tratamiento tiene 4 repeticiones.

Efectos

Como es un experimento con dos tratamientos, entonces podemos calcular los efectos de los dos tratamientos. El efecto de la condición de la alimentación se representa como $\alpha_i = \mu_{i.} - \mu$ y el genero $\beta_j = \mu_{.j} - \mu$

```
mediaGeneral = mean(base$proteina)
mediaCond = with(data = base, tapply(proteina, cond, mean))
mediaGenero = with(data = base, tapply(proteina, genero, mean))
```

```
alpha = mediaCond - mediaGeneral
beta = mediaGenero - mediaGeneral
```

```
alpha
```

```
## alimento    dieta
##  1.55625 -1.55625
```

```
beta
```

```
## macho hembra
##  1.91875 -1.91875
```

Ahora calculamos la media bajo cada tratamiento de los datos observados.

```
mediaCI = with(data = base, tapply(proteina, list(cond, genero), mean))
mediaCI
```

```
## macho hembra
## alimento 43.225 39.175
## dieta    39.900 36.275
```

Ahora para comparar calculamos la media bajo un modelo sin interacción. Sin interacción $\mu_{ij} = \mu + \alpha_i + \beta_j$, con interacción $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$. Las medias bajo interacción son iguales a las medias observadas.

```
m1 = mediaGeneral + alpha[1] + beta[1]
m2 = mediaGeneral + alpha[2] + beta[1]
m3 = mediaGeneral + alpha[1] + beta[2]
m4 = mediaGeneral + alpha[2] + beta[2]
```

```
mediaSI = c(m1, m2, m3, m4)
mediaSI
```

```
## alimento    dieta alimento    dieta
## 43.11875 40.00625 39.28125 36.16875
```

Construimos una tabla para compararlos entre si.

```
comp = cbind(as.vector(mediaCI), mediaSI)
colnames(comp) = c("mediaCI", "mediaSI")
row.names(comp) = c("alimento-macho", "dieta-macho", "alimento-hembra", "dieta-hembra")
comp
```

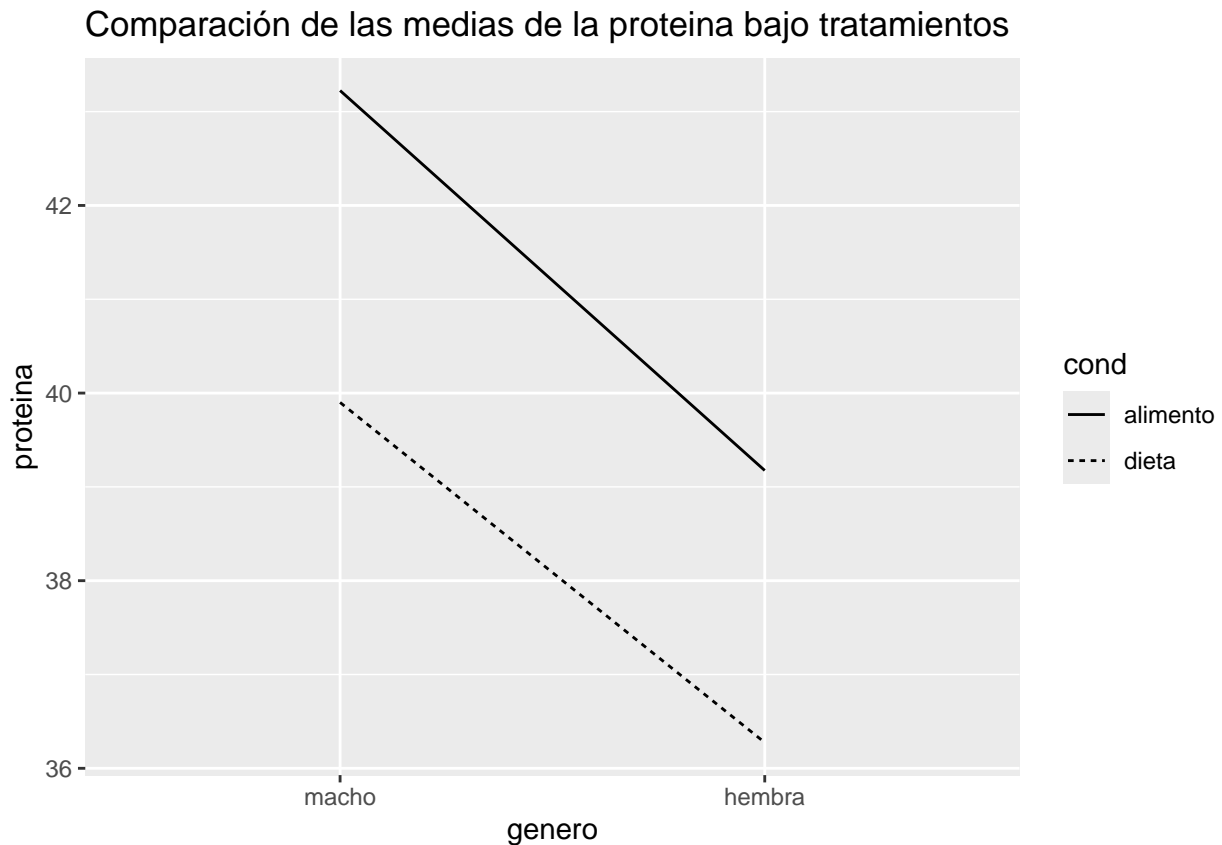
```
## mediaCI mediaSI
## alimento-macho 43.225 43.11875
## dieta-macho    39.900 40.00625
## alimento-hembra 39.175 39.28125
## dieta-hembra    36.275 36.16875
```

Como podemos ver las medias son bastante similares entre cada uno de los modelos. Esto nos da un indicio de que puede que no exista interacción.

Ahora para poder ver de forma gráfica la interacción construimos un gráfico de líneas con la librería ggplot2.

```
library(ggplot2)

ggplot(data = base, aes(x = genero, y = proteina, group = cond)) +
  stat_summary(fun = "mean", geom = "line", aes(linetype = cond)) +
  labs(title = "Comparación de las medias de la proteina bajo tratamientos")
```



En el gráfico se pueden ver que las diferencia entre las medias bajo cada nivel de condición de alimentación es similar para cada nivel de genero. Esto también es una indicación de que no puede haber interacción. Utilizando aov podemos obtener los efectos simples y los efectos de interacción.

```
mod = aov(proteina~cond*genero, data = base)
model.tables(mod)
```

```
## Tables of effects
##
##  cond
##  cond
## alimento  dieta
##  1.5562  -1.5562
##
##  genero
```

```
## genero
## macho hembra
## 1.9187 -1.9187
##
## cond:genero
##          genero
## cond      macho hembra
## alimento  0.10625 -0.10625
## dieta    -0.10625  0.10625
```

Efectos de Interacción en el Modelo

Ahora construimos el modelo con lm.

```
options(contrasts = c("contr.sum", "contr.poly"))
mod = lm(proteina~cond*genero, data = base)
mod$coefficients
```

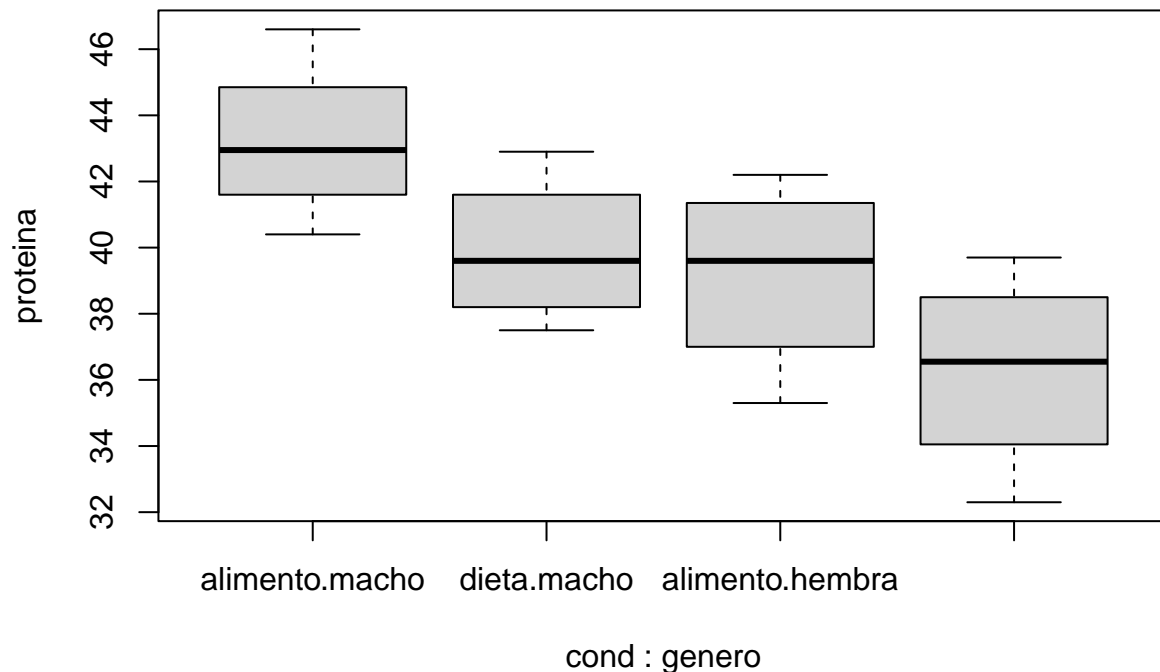
```
## (Intercept)      cond1      genero1 cond1:genero1
##      39.64375      1.55625      1.91875      0.10625
```

Los coeficientes en este caso son los efectos simples y el efecto de interacción.

Varianza del Error

Construimos un gráfico de cajas para visualizar la variabilidad dentro de cada tratamiento.

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



Según el gráfico parece que todos los tratamientos provienen de distribuciones con la misma varianza.

```
v = with(data = base, tapply(proteina, list(cond, genero), var))
v
```

```
##          macho  hembra
## alimento 6.522500 8.715833
## dieta    5.306667 9.602500
```

Para verificar que existe homocedasticidad vamos a utilizar la prueba de bartlett.

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

```
##
## Bartlett test of homogeneity of variances
##
## data:  proteina by interaction(cond, genero)
## Bartlett's K-squared = 0.28408, df = 3, p-value = 0.963
```

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

Prueba de Hipótesis de Interacción

Las visualizaciones y las estimaciones nos dan indicios de que puede que no exista interacción entre los factores, pero vamos a formalizarlo con una prueba ANOVA.

```
mod = lm(proteina~cond*genero, data = base)
anova(mod)
```

```
## Analysis of Variance Table
##
## Response: proteina
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cond           1  38.751   38.751   5.1415 0.04263 *
## genero          1  58.906   58.906   7.8157 0.01617 *
## cond:genero     1   0.181    0.181   0.0240 0.87955
## Residuals     12  90.443    7.537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La hipótesis que estamos poniendo a prueba es la siguiente:

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

En este caso como la probabilidad de comete error tipo 1 es mayor a 0.05 no rechazamos la hipótesis nula. Se espera que no exista interacción entre condición de alimentación y genero, en otras palabras, el efecto de la condición de alimnetación es independiente al genero.

Estimaciones bajo el Modelo sin Interacción

Como no rechazamos la hipótesis de interacción, entonces podemos utilizar el modelo sin interacción que es más simple.

```
mod2 = lm(proteina~cond+genero, data = base)
mod2$coef
```

```
## (Intercept)      cond1      genero1
##    39.64375     1.55625     1.91875
```

Ahora en este caso los coeficientes del modelo son simplemente los efectos simples de los factores. Como no existe interacción se pasa a la siguiente hipótesis sobre el efecto del factor de diseño:

- $H_0 : \alpha_1 = \alpha_2 = 0$ con $H_a : \text{al menos un efecto es distinto a } 0$

```
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: proteina
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cond           1  38.751   38.751   5.5588 0.03472 *
## genero          1  58.906   58.906   8.4501 0.01224 *
## Residuals     13  90.623    6.971
## ---
## 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 en promedio al menos una las condiciones de alimentación sea distinta.

Construimos ahora una cota inferior para esta diferencia para verificar si es relevante.

```
h = c(0, 2, 0)
coef = mod2$coefficients
L = h %*% coef
seL = sqrt(diag(t(h)%*%vcov(mod2)%*%h))
t = qt(1 - 0.05, 13)
lim = L - t*seL
lim
```

```
##           [,1]
## [1,] 0.7746315
```

Entonces ahora sabemos que la cota inferior para la diferencia entre la media de proteína generada bajo alimentación en abundancia y en diatea puede ser tan baja como 0.77 con un 95% de confianza independientemente de si son machos o hembras.