

PRINCIPIOS DE DISEÑO EXPERIMENTAL: DISEÑOS DE DOS FACTORES

ISADORE NABI

I. ALGUNOS ASPECTOS TEÓRICOS GENERALES	2
I.I. Experimentos Factoriales: Generalidades	2
I.II. Diseños Factoriales $2k$	8
I.II. I. Introducción: Sobre los Diseños nk	8
I.II. II. Diseños $2k$	8
II. CASO DE APLICACIÓN: EFECTO DEL RÉGIMEN ALIMENTICIO EN EL NIVEL DE PROTEÍNAS EN LA SANGRE DE LAS CHELONIA MYDAS	8
II.I. Planteamiento del Problema de Investigación	8
II.II. Verificando el número de repeticiones por tratamiento	10
II.III. Obteniendo la media general de la respuesta.	10
II.IV. Estimando las varianzas de los 4 tratamientos.	14
II.V. Estimación de efectos simples e interacción	15
II.VI. Configuración del Modelo Lineal	16
II.VII. Obteniendo la tabla de ANOVA para probar la hipótesis de no-interacción entre condición y especie.	19
II.VIII. Modelo Sin Interacción	21
III. CASO DE APLICACIÓN: EFECTO DEL RÉGIMEN ALIMENTICIO EN EL NIVEL DE PROTEÍNAS EN LA SANGRE DE LAS CHELONIA MYDAS Y KINOSTERNON SCORPIOIDE	30
III.I. Generalidades	30
III.II. Verificando el número de repeticiones por tratamiento.	32
III.III. Calculando la media general de la respuesta.	32
III.IV. Efectos Simples	32
III.V. Calculando los promedios observados de los 6 tratamientos	33
III.VI. Análisis Descriptivo	34
III.VII. Varianzas de los Tratamientos	36
III.VIII. Estimando los efectos simples y de interacción	38
III.IX. Estimación de los Contrastes	39
III.X. Estimación de los promedios con el modelo con interacción entre efectos	42

III.XI. Estimación manual individual del nivel promedio de proteína en sangre de las chelonia bajo un régimen alimenticio estricto con coeficientes y vector de contrastes _____	43
III.XII. Estimación manual del nivel promedio de proteína en sangre de las chelonia para los demás regímenes alimenticios (tratamientos) _____	43
10. Tabla de ANOVA para probar la hipótesis de NO-interacción entre condición y especie _____	44
III.XIII. Planteamiento de la Hipótesis Nula _____	45
III.XIV. Estimación del Cuadro Medio Residual _____	45
III.XV. Cuadrado Medio de Condición y Especie _____	46
III.XVI. Estimación de los Cuadrados Medios a partir de los Efectos Simples _____	46
III.XVII. Estimación de la Suma de Cuadrados de Interacción A Partir de los Efectos de Interacción _____	47
III.XIX. Conclusión de Prueba de Hipótesis con Valor F _____	47
III.XX. Planteamiento Formal de los Contrastes de Comparación de Interacción Entre Especies _____	48
III.XXI. Construyendo la Matriz de Contraste de Coeficientes _____	48
III.XXI. Estimación de los Contrastes _____	50
III.XXII. Estimación de la Varianza de los Contrastes _____	50
III.XXIII. Verificando la Significancia Estadística de los Contrastes _____	51
III.XXIV. Estimando Cotas Inferiores _____	52
IV. REFERENCIAS _____	53

I. ALGUNOS ASPECTOS TEÓRICOS GENERALES

I.I. Experimentos Factoriales: Generalidades

Un factor es una variable categórica (o categorizada) con varios niveles. Los experimentos factoriales se utilizan cuando a cada observación se le aplican tratamientos que son el resultado de la combinación de dos o más factores. En este tipo de estudios, el interés consiste en estudiar los efectos sobre la variable respuesta de dos o más factores simultáneamente, así como la interacción existente entre estos factores.

Como señalan (Melo, López, & Melo, 2020, págs. 401-404), el término *experimento factorial* o *arreglo factorial* hace referencia a la constitución de los tratamientos o combinaciones de niveles de tratamientos (sea en el sentido de magnitudes cuantitativas o de los tipos en los que se clasifica un determinado tratamiento) que se desea comparar. El término experimento factorial no implica afectación alguna en el significado de lo que se conoce como *diseño de tratamientos*, pues este se refiere a la selección de factores que se desea estudiar, los niveles de los factores a ensayar y la combinación de estos.

En suma, un arreglo factorial es una familia de técnicas cuantitativas posibles para abordar el estudio de algún fenómeno de la realidad, mientras que el diseño de los tratamientos es la metodología científica (natural o social) que guía la investigación científica y que hace uso de técnicas cuantitativas como herramientas de análisis. Es relevante aclarar que el diseño de tratamientos es independiente del diseño experimental, puesto que el último hace referencia a la manera en que los tratamientos se aleatorizan con las diferentes unidades experimentales y a la forma en que se controla la variabilidad natural de las mismas.

Así, el diseño experimental puede ser completamente aleatorizado, de tipo bloques completamente aleatorizados, cuadros latinos, etc., y para cada uno de estos diseños se puede tener un arreglo factorial específico.

En muchos experimentos, el éxito o fracaso del ensayo depende más de la selección de los tratamientos a comparar que de la elección del diseño mismo. Sin embargo, la selección de ambos es importante, luego ninguno de los dos debe descuidarse en la planeación de un experimento controlado. En un experimento factorial se investigan simultáneamente los efectos de cierto número de diferentes factores. La necesidad de estudiar conjuntamente varios factores obedece principalmente a dos razones:

1. Encontrar un modelo que describa el comportamiento general del fenómeno en estudio. Esto se restringe al rango de variación de los niveles de los factores.
2. Optimizar la respuesta o variable independiente, es decir, encontrar la combinación de niveles de los factores que optimizan esa respuesta.

Los tratamientos en el análisis factorial están constituidos por todas las combinaciones que se forman de los distintos niveles de los factores. Por ello, la característica esencial que hace necesario el estudio conjunto de factores es la posibilidad de que el efecto de un factor cambie en presencia de los niveles de otro factor (presencia de interacción), es decir, que los factores interactúen, lo cual conduce al concepto de interacción entre ellos. Si se estudia un factor en forma separada, el resultado puede ser diferente al que resultaría de un estudio conjunto, y es más difícil describir el comportamiento general o encontrar la combinación óptima de niveles cuando se hace de forma separada.

Señalan los autores antes citados que Finney (1955) presenta un experimento de factores por separado que consiste en determinar las condiciones óptimas de almacenaje de los pescados en barcos pesqueros. Los factores estudiados fueron: temperatura, duración y método de empaque (proporción de hielo y pescado). La respuesta de interés es una medida de la calidad del pescado al descargue.

Al investigar únicamente la temperatura, se deben tener varios niveles de temperatura y mantener constante la duración y el empaque a niveles arbitrarios. Una vez obtenida una temperatura óptima (manteniendo los niveles constantes de duración y empaque), se investiga otro factor. Por ejemplo, el empaque con la temperatura óptima y un nivel arbitrario de duración. Si el empaque óptimo encontrado no es el que se seleccionó en la primera etapa, se deberá estudiar de nuevo la temperatura haciendo necesarios ajustes sucesivos. Esto se conoce como experimentación secuencial.

Si el tiempo de obtención de la variable de respuesta es corto y el costo económico es bajo, se puede seguir este procedimiento de forma secuencial. En caso contrario, es más conveniente el uso de experimentos factoriales. Los experimentos agrícolas en general no tienen esta característica, de ahí que estas técnicas de experimentación factorial se desarrollaran en el sector agropecuario con los trabajos pioneros de Fisher y Yates entre 1920 y 1930 en la estación agrícola experimental de Rothamsted en Inglaterra.

Los experimentos factoriales deben ser usados cuando los factores no son independientes, puesto que en general se busca evaluar el efecto de interacción. Algunas de las ventajas de esta clase de experimentos son:

1. Al obtener información sobre varios factores sin aumentar el tamaño del experimento hay economía en el material experimental.
2. Se amplía la base de la inferencia en relación con un factor, ya que se estudia en las diferentes condiciones representadas por los niveles de otros factores.
3. Se puede obtener una estimación de la interacción de los efectos, es decir, se determina el grado y la forma en la cual se modifica el efecto de un factor en presencia de los niveles de los otros factores.
4. El conjunto de los tratamientos en el diseño factorial es óptimo para estudiar efectos principales e interacciones.
5. Se puede trabajar con un subconjunto de tratamientos (fracciones del diseño).

Entre las desventajas más importantes destacan:

1. El gran número de combinaciones de tratamientos cuando se estudian muchos factores o muchos niveles. En forma simultánea esto tiene dos efectos:
 - 1.1. Si se desea usar bloques completos, es difícil encontrar grupos de unidades experimentales homogéneos para asignar todos los

tratamientos dentro del bloque. Esto se puede eliminar usando el principio de confusión.

- 1.2. Se aumenta el costo del experimento al tener muchas unidades experimentales. Este problema se minimiza usando experimentos factoriales fraccionados. En este caso, se lleva a cabo el análisis estadístico considerando solo una parte de los tratamientos posibles.
2. Difícil interpretación, principalmente de las interacciones de orden superior (interacciones de más de tres efectos).

La interacción entre los factores se presenta cuando los niveles de un factor no producen medias poblacionales que guarden las mismas relaciones al considerar cada uno de los niveles de otro factor (hay grandes cambios en la magnitud de estos). Los coeficientes, parámetros o pesos son los niveles medios de los factores, los factores son las variables explicativas y los tratamientos son las combinaciones lineales de los factores y sus niveles.

Es conveniente comenzar la numeración de niveles en cero. En el caso de factores cuantitativos, el nivel cero es el más bajo (generalmente ausencia de tratamiento, conocido también como “placebo”).

Los efectos de un factorial que se estudian se relacionan con los efectos principales y de interacción y se denotan por las letras A, B, C, AB, ABC, etc. Los niveles de un factor se identifican con subíndices a_i, b_j, c_k, \dots . Los tratamientos se denotan de varias formas:

1. Con letras y números, por ejemplo, $a_1 b_2 c_3, a_1 b_0 c_1, \dots$, teniendo entonces que $a_1 b_2 c_3 = T_1, a_1 b_0 c_1 = T_2, \dots$
2. Se indica el factor únicamente con los números y su respectivo orden. En este caso, los tratamientos del numeral anterior se re-expresarían como $123 = T_1, 101 = T_2, \dots$

La interacción entre los factores se presenta cuando los niveles de un factor no producen medias poblacionales que guarden las mismas relaciones al considerar cada uno de los niveles del segundo factor (hay grandes cambios en la magnitud de estos).

knitr::include_graphics("FT1.JPG")

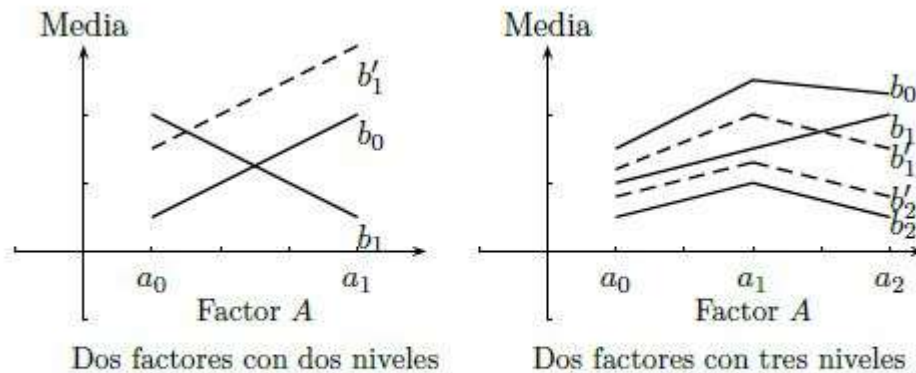


Figura 9.1: Interacción entre dos factores.

#Figura 1: Comparación de Medias con Interacciones para dos y tres factores

#Fuente: (Melo, López, & Melo, 2020, pág. 404).

Si la gráfica anterior estuviese conformada por las líneas b_0 , b'_1 y b_0 , b'_1 y b'_2 , no se tendría efecto de interacción. En ese caso, el efecto de cambiar los niveles de A sería el mismo para todos los niveles de B . Si las líneas son las dadas por b_0 , b_1 y b_0 , b_1 y b_2 , el efecto de cambio de niveles de A sobre la respuesta es diferente dependiendo de los niveles de B (hay interacción) y viceversa.

Para definir la interacción entre tres factores, se toma un patrón de interacción entre dos de ellos, y si este patrón cambia al considerar los niveles del tercer factor, se tiene interacción entre los tres factores. Para estudiar el efecto de una interacción, se debe evaluar primero la hipótesis de no interacción, la cual se "prueba" con una razón de cuadrados medios (prueba F); si la hipótesis no se rechaza, se considera que no hay interacción. Si se rechaza la hipótesis nula,

entonces mediante pruebas de comparación múltiple se debe investigar el patrón de la interacción.

I.II. Diseños Factoriales 2^k

I.II. I. Introducción: Sobre los Diseños n^k

Como señalan (Melo, López, & Melo, 2020, pág. 405), los diseños factoriales se usan ampliamente en experimentos que incluyen varios factores, buscando estudiar el efecto conjunto de los factores sobre la respuesta. Hay varios casos especiales del diseño factorial que son importantes debido a su uso generalizado en investigación. Se destacan los arreglos 2^k porque constituyen las bases de otros diseños de gran valor práctico, que en general pueden describirse como diseños n^k . El más importante de estos arreglos se presenta cuando se ensayan k factores, cada uno con dos niveles. En este caso, se habla de arreglos factoriales 2^k . Los niveles de estos factores pueden ser cuantitativos o cualitativos.

I.II. II. Diseños 2^k

Así, existen 2^k tratamientos que son las combinaciones de los niveles de los k factores. Para poder estimar la varianza del error es necesario repetir el experimento al menos 2 veces. El número de veces que se repita cada tratamiento se conoce como réplicas r . En total se necesitan $r2^k$ observaciones para realizar el experimento completo.

II. CASO DE APLICACIÓN: EFECTO DEL RÉGIMEN ALIMENTICIO EN EL NIVEL DE PROTEÍNAS EN LA SANGRE DE LAS CHELONIA MYDAS

II.I. Planteamiento del Problema de Investigación

Se desea determinar si la falta de alimento afecta el nivel de proteínas en sangre en tortugas de la especie *Chelonia Mydas* (de agua salada) y si la falta de alimento afecta diferente a los machos que a las hembras. Para ello, se escogen 8 machos y 8

hembras. A cada uno se le asigna aleatoriamente una de las siguientes condiciones: 1) dieta regulada y 2) alimento en abundancia. Se registra el nivel de proteína en gramos por decilitro (gr/dl).

Tras cargar el archivo `tortugas1.csv`, es recomendable verificar que las variables 'genero' y 'condicion' sean factores con los niveles adecuados. En el caso de 'genero' los códigos corresponden a macho (1) y hembra (2).

```
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 ...

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

Los factores que se incluyen en el experimento son dos, el que se investiga actualmente es la condición de alimento y adicional a este está el factor género. El *factor de diseño* es la condición de alimento, pues el interés del estudio es analizar el efecto que tiene cada una de las condiciones sobre la proteína promedio que tienen en sangre las tortugas que se alimentan según cada uno de los dos regímenes alimenticios.

A cada tortuga se le puede asignar aleatoriamente una condición, con lo cual las diferencias que se observen en la variable de respuesta (nivel de proteína en sangre) se pueden ligar a la condición estudiada estableciendo una relación de causa y efecto; de manera complementaria, se incluye el género porque

posiblemente el comportamiento no es el mismo en ambos géneros. Es importante incluir este último factor (género) puesto que el estudio también tiene como objetivo analizar si el efecto de la condición es el mismo en cada género.

Este análisis se debe concentrar primero en verificar si el efecto que tiene la condición de alimento es el mismo en ambos géneros. De ser así, debe cuantificarse ese efecto por separado en cada especie. Si el efecto es el mismo en ambos géneros, entonces se puede dar una conclusión general sobre el efecto que tiene cada condición en la proteína promedio sin diferenciar esta conclusión por género. Este experimento cuenta con un total cuatro tratamientos, los cuales se obtienen al combinar las dos condiciones para los dos géneros, lo cual significa que es un experimento factorial de tipo $2^{k=2}$.

II.II. Verificando el número de repeticiones por tratamiento

```
table(cond,genero)
```

```
##      genero
## cond  macho hembra
## alimento  4   4
## dieta    4   4
```

La tabla anterior muestra que de las 16 observaciones en total (tamaño de muestra global), a 4 machos se le dio una dieta regulada, otros 4 alimento en abundancia, mientras que a 4 hembras se les reguló la dieta y a otras cuatro se les dio alimento en abundancia.

II.III. Obteniendo la media general de la respuesta.

```
medgen=mean(proteina)
round(medgen,2)

## [1] 39.64
```

De forma complementaria, pueden obtenerse los efectos simples de cada 'condición' (*i.e.*, de cada régimen alimenticio), es decir, las diferencias entre cada media de condición y la media general. Esto es similar a los τ cuando hay un solo factor, aunque aquí con el fin de diferenciación se denotará como α , que no debe confundirse con el nivel de significancia (ante la posibilidad de ambigüedad se denotará como α^*).

```
alfa=tapply(proteina,cond,mean)-medgen
round(alfa,2)

## alimento  dieta
##  1.56  -1.56
```

Así, pueden obtenerse los efectos simples de cada 'genero', llamando a estos efectos β .

```
beta=tapply(proteina,genero,mean)-medgen
round(beta,2)

## macho hembra
##  1.92  -1.92
```

Adicionalmente, pueden obtenerse los promedios observados de los 4 tratamientos usando la sintaxis 'tapply', pero haciendo una lista con los factores para que se crucen y obteniendo la media para cada combinación o tratamiento: 'tapply(Y,list(X1,X2),mean)'.

```
med=tapply(proteina,list(cond,genero),mean)
med

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

Es recomendable obtener promedios estimados bajo el modelo sin interacción. Puede hacer esto manualmente. Úsese la estimación a partir del modelo sin interacción:

$$\mu_{ij} = \mu + \alpha_i + \beta_j$$

```
mest=medgen+rep(alfa,2)+rep(beta,each=2)
round(mest,2)

## alimento  dieta alimento  dieta
## 43.12  40.01  39.28  36.17
```

Adicionalmente, deben compararse los promedios estimados bajo el modelo sin interacción con los promedios observados.

```
M=cbind(as.vector(med),mest)
colnames(M)=c("Observados","Estimados")
round(M,2)

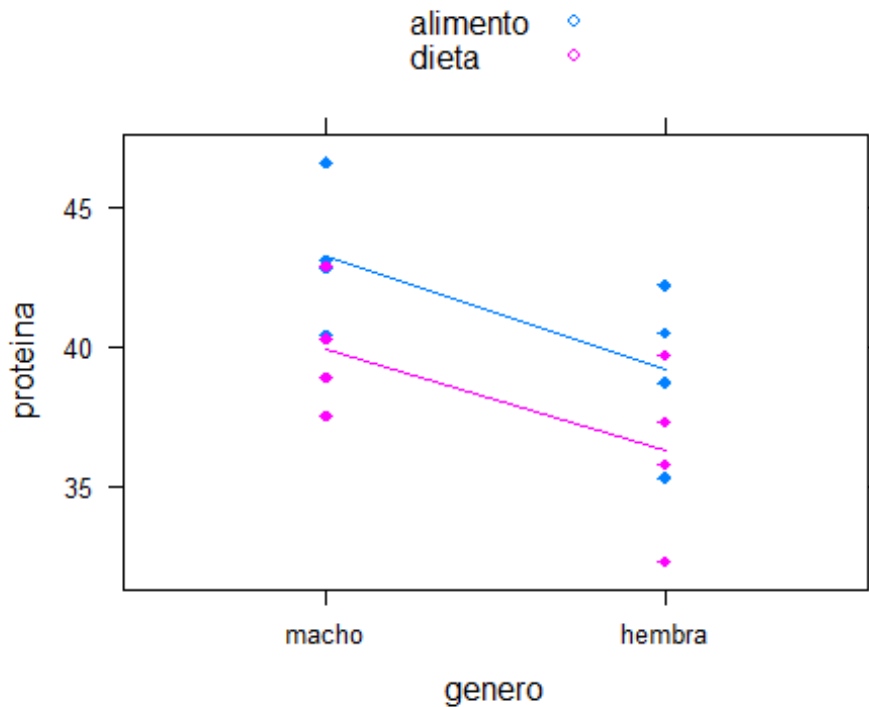
##      Observados Estimados
## alimento  43.23  43.12
## dieta     39.90  40.01
## alimento  39.17  39.28
## dieta     36.27  36.17
```

De lo anterior se puede concluir que los promedios estimados son muy parecidos a los observados, lo cual indica que hay muy poca interacción.

Pueden graficarse los valores de 'proteína' separados por 'especie' con la finalidad de estudiar descriptivamente si existe alguna interacción entre 'condicion' y 'especie'. Es recomendable decidir cuál de las variables a graficar se localizará en el eje X con base en el objetivo del estudio. Para ello, puede utilizarse la sintaxis 'xyplot' de la librería 'lattice', indicando "type=c('p','a')" para que se coloquen

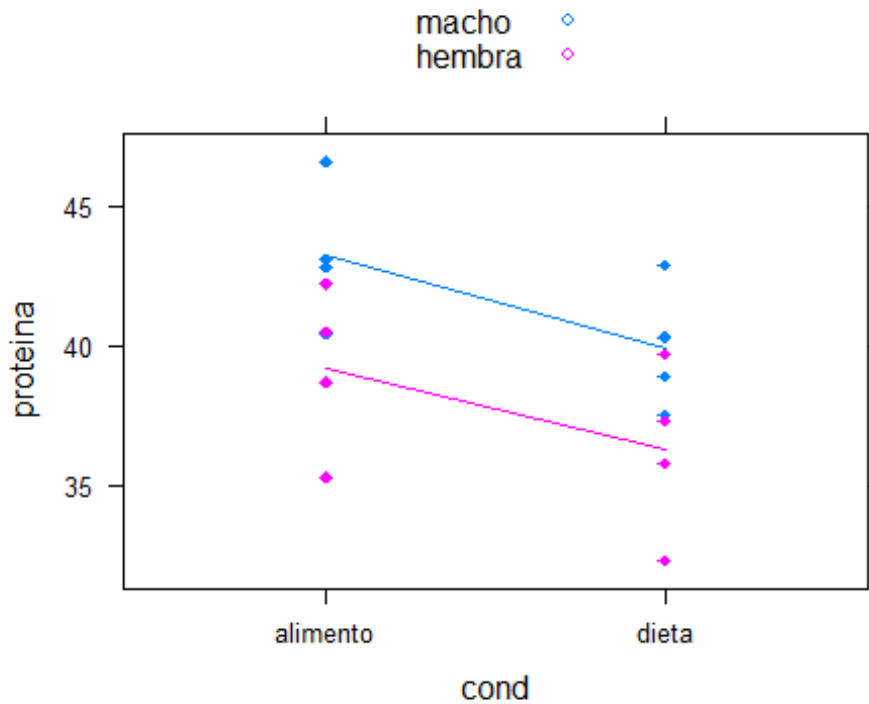
puntos en las observaciones y además, puede agregarse el promedio de cada tratamiento mediante `'xyplot(Y~X1,groups=X2,pch=18=c("p","a"),auto.key=T)'`.

```
library(lattice)
xyplot(proteina~genero,groups=cond,
pch=18,type=c("p","a"),auto.key=T)
```



Complementariamente, debe construirse un gráfico con los valores de 'proteina' separados por 'condicion'.

```
xyplot(proteina~cond,groups=genero,
pch=18,type=c("p","a"),auto.key=T)
```



Obsérvese la distancia que hay entre cada par de medias para cada género, es decir, la media en alimento contra dieta para machos, y lo mismo para hembras, lo que parecería indicar que el sexo no es relevante en el impacto que el régimen alimenticio tiene en el nivel de proteína en sangre de la especie de tortugas estudiada o, en otros términos, el nivel de proteína en sangre según el régimen alimenticio (dieta o alimentación libre) es independiente del sexo del animal. Lo anterior puede realizarse también con un gráfico en que se separen los sexos. Seguramente alguno de los dos será preferido del investigador, pudiendo con ello optar por el uso de la sintaxis pertinente que más se adecúe a sus necesidades concretas.

II.IV. Estimando las varianzas de los 4 tratamientos.

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

```
v
```

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

Es recomendable verificar si se cumple el supuesto de homoscedasticidad. Para ello, puede utilizarse la prueba de Bartlett a través de la sintaxis 'bartlett.test' indicando "Y~interaction(X1,X2)".

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

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

Ante la hipótesis de homoscedasticidad se obtiene una probabilidad asociada de 0.96 al error tipo I ($\alpha = 0.05$), por lo que se decide no rechazar esa hipótesis debido a que no se tiene evidencia de heteroscedasticidad. Así, se continúa suponiendo que los datos provienen de poblaciones cuyas variancias son iguales.

Tras la verificación de homogeneidad de varianza, obténgase una medida de la variabilidad del error asumiendo homoscedasticidad.

```
v1=mean(v)
round(v1,2)

## [1] 7.54
```

II.V. Estimación de efectos simples e interacción

Para estimar los efectos simples y la interacción puede utilizarse la sintaxis `model.tables`. Debe escribirse primero el modelo con interacción entre 'condicion' y

'especie' con la función 'aov'. Además, recuérdese que la interacción se agrega a un modelo de cualquiera de las siguientes formas: "Y~X1+X2+X1:X2" o "Y~X1*X2".

```
mod1=aov(proteina~cond*genero)
model.tables(mod1)

## 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
```

Obtenidos los resultados anteriores, puede observarse que se obtienen los mismos efectos simples de la sección 3.

II.VI. Configuración del Modelo Lineal

Puede construirse el modelo lineal mediante la sintaxis 'lm' y almacenar el resultado en una estructura llamada 'mod2'. Por defecto R usa el modelo de tratamiento referencia. Para cambiar al modelo de suma nula, que es el de interés

en este tipo de investigaciones, debe escribirse

'options(contrasts=c("contr.sum","contr.poly"))' antes de la línea de código de 'mod2'.

```
options(contrasts=c("contr.sum","contr.poly"))
mod2=lm(proteina~cond*genero)
```

Tras la estimación del modelo anterior, debe observarse la matriz de diseño usando 'model.matrix(mod2)'. Verifíquese a qué corresponde cada columna poniendo atención a los códigos y relacionándolos con los niveles de cada factor.

```
contrasts(cond)

##      [,1]
## alimento  1
## dieta    -1

contrasts(genero)

##      [,1]
## macho    1
## hembra  -1

model.matrix(mod2)

## (Intercept) cond1 genero1 cond1:genero1
## 1      1  1  1      1
## 2      1  1  1      1
## 3      1  1  1      1
## 4      1  1  1      1
## 5      1 -1  1     -1
## 6      1 -1  1     -1
## 7      1 -1  1     -1
```

```

## 8      1 -1  1   -1
## 9      1  1 -1   -1
## 10     1  1 -1   -1
## 11     1  1 -1   -1
## 12     1  1 -1   -1
## 13     1 -1 -1    1
## 14     1 -1 -1    1
## 15     1 -1 -1    1
## 16     1 -1 -1    1

## attr(,"assign")
## [1] 0 1 2 3
## attr(,"contrasts")
## attr(,"contrasts")$cond
## [1] "contr.sum"
##
## attr(,"contrasts")$genero
## [1] "contr.sum"

```

Como puede observarse, la matriz de diseño tiene una columna para condición con códigos 1 y -1. Similarmente una columna para género con códigos 1 y -1; el -1 corresponde siempre al tratamiento que no aparece, que es dieta para hembras. La columna de interacción es el producto de las otras dos columnas.

Con relación a lo anterior, es recomendable extraer los coeficientes del modelo. Tras ello, el lector observará el efecto de la interacción entre factores. ¿Qué representa dicha cantidad?

```

round(mod2$coef,2)

## (Intercept)      cond1      genero1 cond1:genero1
##      39.64      1.56      1.92      0.11

```

Cuando dos factores interactúan, los efectos de un factor dependen del nivel en que se encuentre el otro factor. Cuando dos factores interactúan, el efecto de un factor va a depender del nivel en que se esté midiendo el otro factor, por lo que el efecto principal resulta difícil de interpretar.

Los efectos de interacción son todos iguales en valor absoluto a 0.11. Esta cantidad representa cuánto se debe restar o sumar a la media observada en cada tratamiento para obtener promedios en condición de no-interacción.

II.VII. Obteniendo la tabla de ANOVA para probar la hipótesis de no-interacción entre condición y especie.

```
anova(mod2)

## 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 nula que es de interés probar es si existe interacción entre condición y género. No interesa probar los efectos simples si no se ha eliminado la interacción. Formalmente se tiene que

$$H_0: (\alpha * \beta)_{ij} = 0$$

En lenguaje natural, H_0 afirma que el efecto de la condición de alimento sobre la proteína promedio no depende del género, es decir, es el mismo en cualquiera de los dos géneros.

Para verificar la hipótesis nula anterior, es recomendable observar el Cuadrado Medio Residual (CMR) y compararlo con la estimación de la variancia del error obtenida más arriba. ¿Qué significa en este contexto el CMR?

```
round(anova(mod2)[4,3],2)
```

```
## [1] 7.54
```

El valor obtenido coincide con la media de las varianzas obtenida anteriormente y es una medida global de la variabilidad promedio de la predicción respecto a la respuesta dada su variabilidad dentro de cada tratamiento.

Como complemento al CMR, se tiene el Cuadrado Medio de Interacción (CMI) es 0.18 y mide la magnitud general del efecto de interacción. Si esta magnitud tiende a ser pequeña, se tiene un caso con poca interacción, lo que sería una evidencia débil de interacción entre condición y especie. En caso contrario, si esta magnitud tiende a ser grande hay más evidencia de presencia de interacción.

Puede extraerse de la columna de cuadrados medios de tabla ANOVA, los cuadrados medios correspondientes a la interacción entre efectos y los cuadrados medios correspondientes a los residuos y dividirlos (siendo numerador y denominador, respectivamente) para obtener la proporción en que la variabilidad de los efectos de interacción representan de la variabilidad residual.

```
f=0.181/7.537
```

```
round(f,2)
```

```
## [1] 0.02
```

A la luz del resultado, la variabilidad de los efectos de interacción es sumamente baja con respecto a la variabilidad residual. Adicionalmente, se observó en la tabla ANOVA que la fiabilidad asociada a la probabilidad de cometer error tipo I (aquí $\alpha = 0.05$) es de aproximadamente $p = 0.88$, con lo cual se falla en rechazar la hipótesis de no-interacción. Se asume que el efecto de la condición de alimento es el mismo cuando se trata de machos que en el caso de las hembras, a la luz de las pruebas anteriores.

II.VIII. Modelo Sin Interacción

Puesto que se decidió asumir que no hay interacción entre condición y género, ahora se procede a usar el modelo sin interacción. El modelo sin interacción se construye mediante la sintaxis 'lm' y puede ser llamado mediante 'mod3'.

```
mod3=lm(proteina~cond+genero)
```

¿Cuánto es el efecto simple de cada condición y de cada género? ¿Cómo se interpretan los resultados?

```
round(mod3$coef,2)
```

```
## (Intercept)   cond1   genero1
##    39.64     1.56     1.92
```

El efecto de la condición 1 (alimento en abundancia) es 1.56, lo que indica que con alimento el nivel de proteína sube 1.56gr/ml con respecto al promedio general (el intercepto, cuyo valor es de 39.64). Similarmente, el efecto de la condición 2 (dieta regulada) es que baja el nivel promedio de proteína 1.56gr/ml con respecto al promedio general. Los machos (género 1) presentan un nivel de proteína que está

1.92gr/ml sobre la media general, mientras que las hembras tienen una media de proteína 1.92gr/ml por debajo de la media general¹.

Puesto que no hay interacción, se desea simplemente verificar si hay un efecto de la condición sobre la respuesta promedio sin tomar en cuenta el género. Por tanto, la hipótesis nula es:

$$H_0: \alpha_i = 0$$

Lo que H_0 así planteada significa en este contexto es que la condición de alimento no tiene ningún efecto sobre el nivel de proteína promedio, es decir, ambas condiciones producen el mismo promedio de proteína en sangre. Lo anterior sucede en ambos géneros, puesto que se determinó que dicho efecto era independiente del género.

El Cuadrado Medio de Condición es una medida de la distancia entre las medias de las dos condiciones. Similarmente, el Cuadrado Medio de Género es una medida de la distancia entre las medias de los dos géneros. Estos son estimados a continuación

```
round(anova(mod3)[1,3],2)
```

```
## [1] 38.75
```

```
round(anova(mod3)[2,3],2)
```

```
## [1] 58.91
```

Como puede observarse, el Cuadrado Medio de Condición es de 38.75, mientras que el Cuadrado Medio de Género es de 58.91.

¹ Este resultado sobre machos y hembras se obtuvo mediante el uso de las sintaxis 'beta=tapply(proteina,genero,mean)-medgen' y 'round(beta,2)' antes realizado.

Estos cuadrados medios pueden estimarse a partir de los efectos simples.

```
table(cond)

## cond
## alimento  dieta
##      8      8

round(sum(8*alfa^2)/1,2)

## [1] 38.75

table(genero)

## genero
## macho hembra
##      8      8

round(sum(8*beta^2)/1,2)

## [1] 58.91
```

Si se observa el Cuadrado Medio Residual y se compara con la estimación de la variancia obtenida a partir de la media de las varianzas de los 4 tratamientos, se verifica que estas no son iguales, ¿por qué ocurre esto?

```
v2=anova(mod3)[3,3]
round(v2,2)

## [1] 6.97

round(v1,2)

## [1] 7.54
```

La varianza que se obtiene a partir de la media de las varianzas ($v_2 = 7.54$) asume un modelo sin interacción, donde los residuales se calculan con respecto a la media

observada. En contraste con lo anterior, el Cuadrado Medio Residual en este caso ($v_1 = 6.97$) se obtiene con los residuales calculados con respecto a las medias estimadas en el modelo sin interacción. Debido a que la interacción es muy pequeña, las medias se movieron muy poco y las dos cantidades obtenidas son muy similares.

Sabiendo que 'mod2=lm(proteina~ cond*especie)'² y que 'mod3=lm(proteina~cond+genero)'³, es de interés conocer en cuánto aumentó la Suma de Cuadrados Residual en 'mod3' con relación a 'mod2'.

```
round(anova(mod2)[4,2],2)
## [1] 90.44

round(anova(mod3)[3,2],2)
## [1] 90.62

round(anova(mod3)[3,2]-anova(mod2)[4,2],2)
## [1] 0.18

round(anova(mod2)[3,2],2)
## [1] 0.18
```

La magnitud de la diferencia entre la suma de cuadrados residuales de los dos modelos (que es 0.18) es igual que la magnitud la suma de cuadrados de interacción, lo cual es siempre así porque:

² Que implica que 'mod2' es igual a $Proteína = \beta_0 + \beta_{cond1}cond1 + \beta_{genero1}genero1 + \beta_{cond1*genero1}cond1 * genero1 + Error$.

³ Que implica que 'mod3' es igual a $Proteína = \beta_0 + \beta_{cond1}cond1 + \beta_{genero1}genero1 + Error$.

1. En 'mod2' lo que no se consideró fueron los efectos de interacción, mientras que en 'mod3' se consideraron tales efectos.
2. Lo anterior, dado que la suma de cuadrados total es fija⁴ (en este caso igual a 188.2794) y se compone de la suma de cuadrados de regresión (igual a 97.83687 para 'mod2' y a 97.65625 para 'mod3') y de la suma de cuadrados residual (igual a 90.443 para 'mod2' y a 90.623 para 'mod3'), significa que lo que a la suma de cuadrados de regresión de 'mod3' le falta tomar en consideración para ser igual a la suma de cuadrados total (que es de 188.2794) es el valor de los cuadrados medios correspondientes al efecto de interacción de 'mod2' (igual a 0.181), que en el caso de 'mod3' se encuentran en la suma de cuadrados residual (por esta razón la suma de cuadrados residual es superior en 'mod3' que en 'mod2').

knitr::include_graphics("FT1.JPG")

⁴ La suma de cuadrados total SCT se calcula como la suma de las diferencias al cuadrado entre los i -ésimos miembros y_i de la media muestral Y y la media muestral \bar{y} (no debe confundirse a \bar{y} con los i -ésimos valores estimados \hat{y}_i), por lo que para una misma muestra siempre será la misma. Gracias a lo anterior, la suma de cuadrados total puede desglosarse en: 1. Suma de cuadrados de regresión SCR, que son la suma de las distancias \hat{y}_i (la estimación de cada observación) respecto a la media muestral \bar{y}), 2. Suma de cuadrados del error SCE (la distancia de cada observación estimada respecto a al valor real de las observaciones de las cuales se hace la estimación en cuestión).

SUMA DE CUADRADOS		
VARIABILIDAD	FÓRMULA	RELACIÓN
Residuos	$SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$SCE = STC - SCR$
Regresión	$SCR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$SCR = STC - SCE$
Total	$STC = \sum_{i=1}^n (y_i - \bar{y})^2$	$STC = SCR + SCE$

#Figura 2: Descomposición de la Suma de Cuadrados Total

#Fuente: (Marco Sanjuán, 2022).

knitr::include_graphics("FT1.JPG")

```

> anova(mod2)
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
> sum(anova(mod2)[,2])
[1] 188.2794
> anova(mod3)
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
> sum(anova(mod3)[,2])
[1] 188.2794
> round(anova(mod3)[3,2]-anova(mod2)[4,2],2)
[1] 0.18
> round(anova(mod2)[3,2],2)
[1] 0.18
> sum(anova(mod3)[,2])-anova(mod3)[3,2]
[1] 97.65625
> sum(anova(mod3)[,2])-anova(mod2)[4,2]
[1] 97.83687
> 97.83687-97.65625
[1] 0.18062

```

#Figura 3: Obtención de SCT, SCR y SCE de 'mod2' y 'mod3' en RStudio

#Fuente: Elaboración propia.

Es posible realizar conclusiones inferenciales con los resultados de la prueba de hipótesis mediante su valor de probabilidad asociado, localizado en la columna F de la tabla ANOVA.

```
f=38.751/6.971
```

```
round(f,2)
```

```
## [1] 5.56
```

Se tiene que la variabilidad asociada a las medias de condición es mayor que la variabilidad residual. Además, se observa que se tiene una probabilidad asociada baja ($p = 0.035$) con relación al nivel de significancia $\alpha = 0.05$. De esta forma, se rechaza la hipótesis que afirma que no hay efecto de la condición. Se concluye, con un nivel de significancia de 0.05, que las dos condiciones producen medias de proteína diferentes.

Debido a que se concluyó que sí hay diferencia en las medias de proteína según condición, vale la pena establecer una cota inferior (denotada como "LIM") para la magnitud de esa diferencia.

```
round(mod3$coef,2)

## (Intercept)  cond1  genero1
##    39.64    1.56    1.92

contrasts(cond)

##      [,1]
## alimento  1
## dieta    -1

# Estimando el promedio del régimen alimenticio abierto (alimento) y del régimen
alimenticio restringido (dieta)

c(1,1,0)-c(1,-1,0)

## [1] 0 2 0

##En c(1,1,0), se coloca el primer 1 para indicar que se estime el intercepto, el
segundo para indicar que se estime la condición 1 (que es el régimen alimenticio
abierto) y el 0 para indicar que no se toma en consideración el sexo del animal.

##En c(1,-1,0), se coloca el primer 1 para indicar que se estime el intercepto, el
```

segundo para indicar que se estime la condición 2 (que es el régimen alimenticio restringido) y el 0 para indicar que no se toma en consideración el sexo del animal.

```
c1=c(0,2,0)
```

```
##Almacenando el resultado como c1
```

```
# Realizando la estimación del contraste L (primera forma de cálculo)
```

```
L=t(c1)%*%mod3$coef
```

```
ee=sqrt(t(c1)%*%vcov(mod3)%*%c1)
```

```
t=qt(0.95,13) #13 son los grados de libertad de SCE
```

```
LIM=L-t*ee
```

```
round(LIM,2)
```

```
## [1]
```

```
## [1,] 0.77
```

```
(m=tapply(proteina,cond,mean))
```

```
## alimento dieta
```

```
## 41.2000 38.0875
```

```
CME=anova(mod3)[3,3]
```

```
(ee2=sqrt((2/8)*CME))
```

```
## [1] 1.320133
```

```
q2=(m[1]-m[2])/ee2
```

```
p=ptukey(q2*sqrt(2),2,13,lower.tail = F)
```

```
p
```

```
## alimento
## 0.03472289

t2=qt(0.95,13)
LIM2=(m[1]-m[2])-t2*ee2
round(LIM2,2)

## alimento
## 0.77

detach(base)
```

Se concluye con 0.95 de confianza que con alimento en abundancia se obtiene un promedio de proteína en sangre al menos 0.77gr/ml mayor que con dieta. Esto sucede tanto en machos como en hembras.

III. CASO DE APLICACIÓN: EFECTO DEL RÉGIMEN ALIMENTICIO EN EL NIVEL DE PROTEÍNAS EN LA SANGRE DE LAS CHELONIA MYDAS Y KINOSTERNON SCORPIOIDE

III.I. Generalidades

Se desea determinar si la falta de alimento afecta el nivel de proteínas en sangre en tortugas de las especies *Kinosternum scorpioides* (de agua dulce) y *Chelonia midas* (de agua salada) y si el efecto es diferente de una especie a otra. Para ello, se escogen 8 ejemplares de cada especie y se asigna aleatoriamente a cada tortuga una de las siguientes condiciones:

- 1) Dieta estricta.
- 2) Dieta balanceada.
- 3) Alimento en abundancia.

Se registra el nivel de proteína en gramos por decilitro (gr/dl).

Es recomendable cargar el `tortugas2.csv` y verificar que la variable **condición** sea un factor con los niveles adecuados.

```
base=read.csv("tortugas2.csv")
base$cond=as.factor(base$cond)
levels(base$cond)=c("estricta","balanceada","abundancia")
base$especie=as.factor(base$especie)
str(base)

## 'data.frame':  24 obs. of  3 variables:
## $ especie : Factor w/ 2 levels "chelonia","kynosternon": 1 1 1 1 1 1 1 1 1 1 ...
## $ cond   : Factor w/ 3 levels "estricta","balanceada",,..: 1 1 1 1 2 2 2 2 3 3 ...
## $ proteina: int  32 30 34 35 40 43 43 42 38 35 ...

attach(base)
```

Como puede observarse, los factores que se incluyen en el experimento son dos, el que se está investigando es la condición de alimento y se tiene otro factor que es la especie de tortuga.

Con relación a las características de los factores y el alcance del experimento, debe decirse que el factor de diseño es la condición de alimento, pues el interés del estudio es analizar el efecto que tiene cada una de las condiciones sobre la proteína promedio que tienen las tortugas que se alimentan de cada forma. A cada tortuga se le puede asignar aleatoriamente una condición, con lo cual las diferencias que se observen se pueden vincular a la condición (régimen alimenticio) viendo en ello una relación de causa y efecto. Por otra parte, se incluye la especie porque posiblemente el comportamiento no es el mismo en ambas especies. Es importante este factor puesto que el estudio también tiene como objetivo analizar si el efecto de la condición es el mismo en cada especie, es decir, analizar la interacción.

Con relación a los aspectos centrales del análisis, interesa primero ver si el efecto que tiene la condición de alimento es el mismo en ambas especies y, de ser así, debe cuantificarse ese efecto por separado en cada especie. Si el efecto es el mismo en ambas especies entonces se puede dar una conclusión general sobre el efecto que tiene cada condición en la proteína promedio sin diferenciar esta conclusión por especie.

Existe un total de 6 tratamientos en este experimento, que se obtienen de estudiar las tres condiciones para cada una de las dos especies.

III.II. Verificando el número de repeticiones por tratamiento.

```
table(cond,especie)
##      especie
## cond  chelonia kynosternon
## estricta    4      4
## balanceada  4      4
## abundancia  4      4
```

Se observa que es balanceado, es decir, todos los tratamientos son asignados a un número igual de unidades experimentales (Marín Diazaraque, 2022, pág. 14).

III.III. Calculando la media general de la respuesta.

```
medgen=mean(proteina)
medgen
## [1] 36
```

III.IV. Efectos Simples

Pueden obtenerse los efectos simples de cada **condición**, es decir, las diferencias entre cada media de condición y la media general; esto es similar al modelo que usa τ , pero dicho modelo plantea el escenario de un solo factor. Por distinción

respecto a tal modelo, los valores de los efectos simples para **condición** se denotarán como α .

```
alfa=tapply(proteina,cond,mean)-medgen
alfa
## estricta balanceada abundancia
## -3.375 2.375 1.000
```

Deben estimarse también los efectos simples de cada **especie**, que se denotarán como β .

```
beta=tapply(proteina,especie,mean)-medgen
beta
## chelonia kynosternon
## 1 -1
```

III.V. Calculando los promedios observados de los 6 tratamientos

Es recomendable utilizar la sintaxis `tapply`, adicionándole una lista con los factores para que puedan interactuar entre sí y así obtener la media para cada combinación o tratamiento: `tapply(Y,list(X1,X2),mean)`.

```
med=tapply(proteina,list(cond,especie),mean)
med
## chelonia kynosternon
## estricta 32.75 32.50
## balanceada 42.00 34.75
## abundancia 36.25 37.75
```

Además, pueden obtenerse los promedios estimados bajo el modelo sin interacción. Esto puede realizarse manualmente vía el modelo sin interacción:

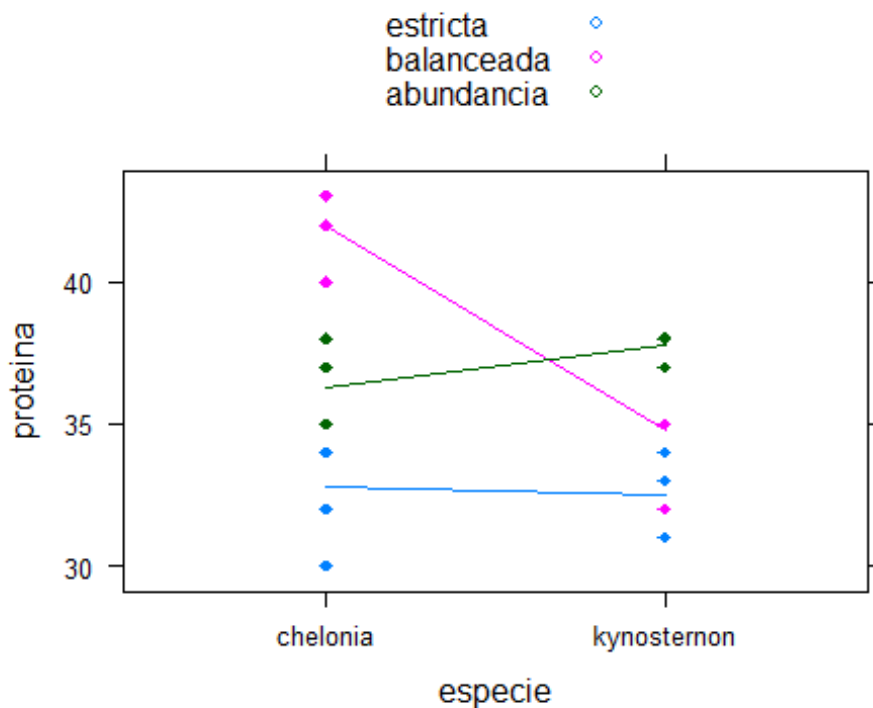
$$\mu_{ij} = \mu + \alpha_i + \beta_j$$

```
mest=medgen+rep(alfa,2)+rep(beta,each=3)
mest
## estricta balanceada abundancia estricta balanceada abundancia
## 33.625 39.375 38.000 31.625 37.375 36.000
```

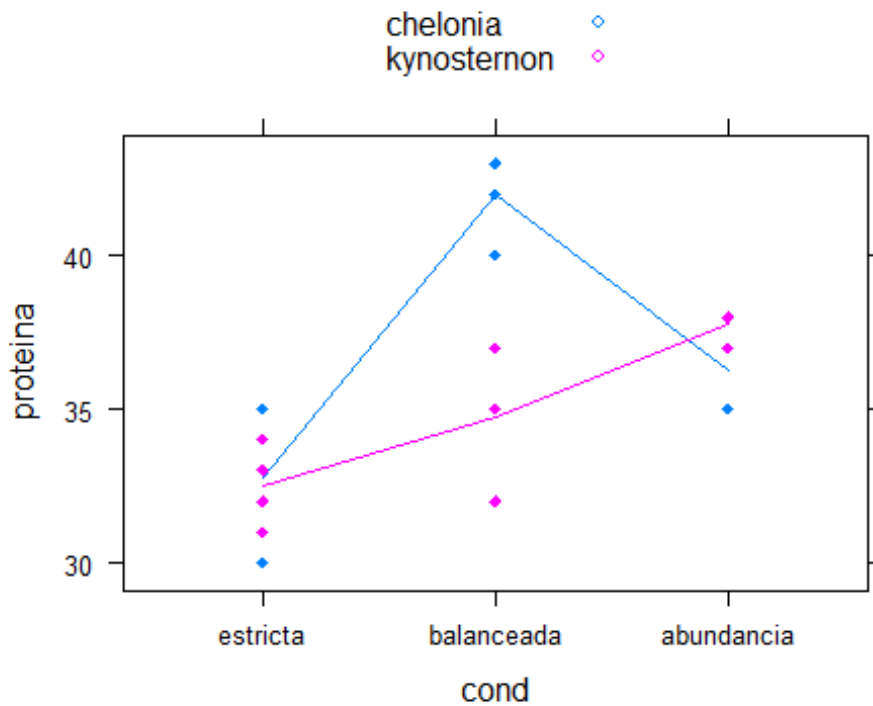
III.VI. Análisis Descriptivo

Puede construirse un gráfico de los valores de 'proteína' para interpretar descriptivamente si existe interacción entre 'condicion' y 'especie'.

```
library(lattice)
xyplot(proteina~especie,groups=cond,
pch=18,type=c("p","a"),auto.key=T)
```



```
xyplot(proteina~cond,groups=especie,
pch=18,type=c("p","a"),auto.key=T)
```



En el primer gráfico se observa que existe interacción entre 'especie' y 'condicion', puesto que, al graficar las medias de proteína en sangre según régimen alimenticio para ambas especies, las rectas que unen dichos promedios para cada especie se interceptan, por lo que parece existir una interacción en los efectos (Frost, 2022). En el segundo gráfico, se observa que la trayectoria del nivel de proteína en sangre para cholonía no depende de la de kynosternon y viceversa, por lo cual se puede concluir que no existe interacción.

A causa de lo anterior, deben compararse los promedios observados con los promedios estimados y, a partir de ello, obtener la cantidad que debe sumarse o restarse a cada promedio observado para obtener una estimación que cumpla con el supuesto de no-interacción (tales cantidades son las *interacciones*)

$efint = med - mest$

efint

```
##      chelonia kynosternon
## estricta  -0.875   0.875
## balanceada  2.625  -2.625
## abundancia -1.750   1.750
```

Estas cantidades deben restarse al promedio observado para obtener la estimación que asegura que no haya interacción. Por ejemplo, en el caso de 'estricta' para chelonia se debe sumar 0.875 (porque tiene ya signo negativo), en el caso de balanceada se debería restar 2.625 (porque tiene signo positivo) y en el caso de abundancia se debería sumar 1.75 (porque tiene ya signo negativo). De esta forma los promedios distarían una cantidad idéntica a la que se obtendría en el caso correspondiente para kynosternon (lo que implica rectas paralelas). Estas cantidades se llaman *efectos de interacción*.

La lógica es simple: si se probó gráficamente que existe interacción, debe usarse el modelo $\mu_{ij} = \mu + \alpha_i + \beta_j$ y, puesto que μ_{ij} es el modelo sin interacción, para obtener las ij medias de nivel de proteína en sangre bajo tal escenario (descartando del análisis los efectos de interacción factorial), según el régimen alimenticio i para la especie de tortuga j , deben restársele a la media general los efectos de interacción existentes entre los α_i factores (régimenes alimenticios) a distintos niveles β_j (para cada especie). Lo anterior explica porque las cantidades referidas antes se adicionan o sustraen en función del signo que tengan, porque el criterio es que deben sustraerse tales efectos de las μ_{ij} estimadas.

III.VII. Varianzas de los Tratamientos

```
v=tapply(proteina,list(cond,especie),var)
```

```
v
```

```
##      chelonia kynosternon
## estricta  4.916667  1.666667
```

```
## balanceada 2.000000 4.250000
## abundancia 2.250000 0.250000
```

Los tratamientos son seis porque las diferentes aplicaciones de regímenes alimenticios en las dos especies son seis, que matemáticamente se expresa como nCr , que significa que se aplican (matemáticamente hablando, se combinan) tres tratamientos sobre dos especies diferentes.

Debe verificarse si se cumple el supuesto de homoscedasticidad. Puede usarse la prueba de Bartlett con mediante `bartlett.test` indicando $Y \sim \text{interaction}(X1, X2)$.

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

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

Ante la hipótesis de homoscedasticidad se obtiene una probabilidad asociada de 0.39, por lo que se decide no rechazar esa hipótesis, con lo cual no se tiene evidencia de heteroscedasticidad.

Cuando se calcularon las varianzas en los diferentes tratamientos se ven grandes diferencias ya que la menor es 0.25 y la mayor 4.91, sin embargo, la prueba no logra demostrar formalmente que las varianzas son diferentes seguramente por la poca cantidad de réplicas en cada tratamiento (de ahí la importancia de las pruebas gráficas para muestras pequeñas, donde las pruebas formales no son fiables - debido a su misma estructura matemática, concebida bajo una visión filosófica de largo plazo-, para comparar sus resultados con los de tales pruebas formales). Para continuar el ejercicio, se continuará suponiendo que los datos provienen de

poblaciones cuyas varianzas son iguales, a pesar de que existe evidencia lógica de que no es así.

Adicionalmente, puede estimarse una medida de la variabilidad del error asumiendo homoscedasticidad.

```
round(mean(v),2)
```

```
## [1] 2.56
```

III.VIII. Estimando los efectos simples y de interacción

Esto puede hacerse usando la sintaxis `model.tables`. Para ello, debe escribirse primero el modelo con interacción entre 'condicion' y 'especie' con la función 'aov'. Recuérdese que la interacción se agrega en un modelo de cualquiera de las siguientes formas (que son equivalentes para RStudio): 1) $Y \sim X_1 + X_2 + X_1:X_2$, 2) $Y \sim X_1 * X_2$.

```
mod1=aov(proteina~cond*especie)
```

```
model.tables(mod1)
```

```
## Tables of effects
```

```
##
```

```
## cond
```

```
## cond
```

```
## estricta balanceada abundancia
```

```
## -3.375 2.375 1.000
```

```
##
```

```
## especie
```

```
## especie
```

```
## chelonia kynosternon
```

```
## 1 -1
```

```
##
```

```
## cond:especie
##      especie
## cond   chelonia kynosternon
## estricta -0.875  0.875
## balanceada 2.625 -2.625
## abundancia -1.750  1.750
```

Comparando estos últimos resultados con los obtenidos antes de ellos (estimados manualmente), se obtienen los mismos efectos simples (alpha y beta), así como los mismos efectos de interacción.

III.IX. Estimación de los Contrastes

Para contrastar los efectos (en ello reside la estimación de contrastes), debe escribirse el modelo con la sintaxis `lm` (llamándolo `mod2`). Es importante tener presente que, por su configuración preestablecida, R usa el modelo de tratamiento referencia. Por ello, debe cambiarse al modelo de suma nula usando `'options(contrasts=c("contr.sum","contr.poly"))'`.

```
options(contrasts=c("contr.sum","contr.poly"))
mod2=lm(proteina~cond*especie)
```

Puede observarse la matriz de diseño usando `'model.matrix(mod2)'` con la finalidad de identificar con qué determinaciones formales corresponde cada columna, así como también vincular los códigos de las mismas con los niveles de cada factor.

```
contrasts(cond)

##      [,1] [,2]
## estricta    1  0
## balanceada  0  1
## abundancia -1 -1
```

```
contrasts(especie)
```

```
##      [1]
```

```
## chelonia    1
```

```
## kynosternon -1
```

"model.matrix()" es la sintaxis que sirve para presentar de forma ordenada cómo se estiman en el modelo de contrastes L cada uno de los efectos y las interacciones entre efectos

```
model.matrix(mod2)
```

```
## (Intercept) cond1 cond2 especie1 cond1:especie1 cond2:especie1
```

```
## 1      1  1  0  1      1      0
```

```
## 2      1  1  0  1      1      0
```

```
## 3      1  1  0  1      1      0
```

```
## 4      1  1  0  1      1      0
```

```
## 5      1  0  1  1      0      1
```

```
## 6      1  0  1  1      0      1
```

```
## 7      1  0  1  1      0      1
```

```
## 8      1  0  1  1      0      1
```

```
## 9      1 -1 -1  1     -1     -1
```

```
## 10     1 -1 -1  1     -1     -1
```

```
## 11     1 -1 -1  1     -1     -1
```

```
## 12     1 -1 -1  1     -1     -1
```

```
## 13     1  1  0 -1     -1      0
```

```
## 14     1  1  0 -1     -1      0
```

```
## 15     1  1  0 -1     -1      0
```

```
## 16     1  1  0 -1     -1      0
```

```
## 17     1  0  1 -1      0     -1
```

```
## 18     1  0  1 -1      0     -1
```

```
## 19     1  0  1 -1      0     -1
```

```
## 20     1  0  1 -1      0     -1
```



```
## 21      1 -1 -1 -1      1      1
## 22      1 -1 -1 -1      1      1
## 23      1 -1 -1 -1      1      1
## 24      1 -1 -1 -1      1      1
## attr(,"assign")
## [1] 0 1 1 2 3 3
## attr(,"contrasts")
## attr(,"contrasts")$cond
## [1] "contr.sum"
##
## attr(,"contrasts")$especie
## [1] "contr.sum"
```

La matriz de diseño tiene dos columnas para condición con códigos 0, 1 y -1 y, de forma similar, una columna para especie con códigos 1 y -1. El -1 corresponde siempre al tratamiento que no se está analizando, el cual es darle comida en abundancia las kynosternon. La columna de interacción es el producto de las otras dos columnas.

Adicionalmente, deben extraerse los coeficientes para obtener los efectos simples y de interacción que no aparecen debido a la restricción de suma nula.

```
mod2$coef
## (Intercept)      cond1      cond2      especie1 cond1:especie1
##      36.000     -3.375      2.375      1.000     -0.875
## cond2:especie1
##      2.625
```

Se multiplica por -1 la suma de los efectos del régimen alimenticio 1 (cond1) y 2 (cond2) para obtener el efecto del régimen alimenticio 3 (cond3)

##Sin interacción entre régimen alimenticio (cond1, cond2, cond3) y especie

$-(-3.375 + 2.375)$

[1] 1

##Con interacción entre régimen alimenticio (cond1, cond2, cond3) y especie

$-(-0.875 + 2.625)$

[1] -1.75

El efecto de la condición 1 (estricta) es -3.375, el de la condición 2 (dieta regulada) es 2.375. Aquí falta el de la condición 3 (abundancia) que se obtiene como $-(-3.375 + 2.375) = 1$. El efecto de la especie 1 (chelonía) es 1 y falta la otra especie cuyo efecto es -1. El efecto de interacción para cond1-especie1 es -0.875 por lo que para cond1-especie2 es 0.875. El efecto de interacción para cond2-especie1 es 2.625 por lo que para cond2-especie2 es -2.625. Se deben obtener los últimos con cond3-especie1: $-(-0.875 + 2.625) = -1.75$ y para cond3-especie2 1.75.

III.X. Estimación de los promedios con el modelo con interacción entre efectos

A continuación, se presenta un ejemplo de estimación de los promedios con el modelo que tiene interacción.

$(\text{Cond1Especie2} = 36.00 - 3.375 - 1 + 0.875)$

[1] 32.5

$(\text{Cond3Especie1} = 36.00 - (-3.375 + 2.375) + 1 - (-0.875 + 2.625))$

[1] 36.25

$(\text{Cond3Especie2} = 36.00 - (-3.375 + 2.375) - 1 + (-0.875 + 2.625))$

[1] 37.75

$(\text{Cond2Especie2} = 36.00 + 2.375 - 1 - 2.625)$

```
## [1] 34.75
```

```
med
```

```
##      chelonia kynosternon
```

```
## estricta  32.75  32.50
```

```
## balanceada 42.00  34.75
```

```
## abundancia 36.25  37.75
```

III.XI. Estimación manual individual del nivel promedio de proteína en sangre de las chelonia bajo un régimen alimenticio estricto con coeficientes y vector de contrastes

Pueden utilizarse los coeficientes antes estimados junto con un vector de contrastes para calcular manualmente el promedio de chelonia con dieta estricta, con el objetivo de compararlo con el promedio observado de ese tratamiento.

```
coef=mod2$coef
```

```
round(coef,2)
```

```
## (Intercept)   cond1   cond2  especie1 cond1:especie1
```

```
##      36.00   -3.38    2.38    1.00   -0.88
```

```
## cond2:especie1
```

```
##      2.63
```

```
c(1,1,0,1,1,0)%*%coef
```

```
##      [1]
```

```
## [1,] 32.75
```

III.XII. Estimación manual del nivel promedio de proteína en sangre de las chelonia para los demás regímenes alimenticios (tratamientos)

Chelonia con dieta balanceada:

```
c(1,0,1,1,0,1)%**%coef
```

```
## [1]
```

```
## [1,] 42
```

Chelonia con dieta en abundancia:

```
c(1,-1,-1,1,-1,-1)%**%coef
```

```
## [1]
```

```
## [1,] 36.25
```

Kynostenron con dieta estricta

```
c(1,1,0,-1,-1,0)%**%coef
```

```
## [1]
```

```
## [1,] 32.5
```

Kynostenron con balanceada

```
c(1,0,1,-1,0,-1)%**%coef
```

```
## [1]
```

```
## [1,] 34.75
```

Kynostenron con abundancia

```
c(1,-1,-1,-1,1,1)%**%coef
```

```
## [1]
```

```
## [1,] 37.75
```

10. Tabla de ANOVA para probar la hipótesis de NO-interacción entre condición y especie

```
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
```

III.XIII. Planteamiento de la Hipótesis Nula

$$H_0 = (\alpha * \beta)_{ij} = 0$$

Así, el efecto de la condición de alimento o régimen alimenticio sobre el nivel promedio de proteína en sangre no depende de la especie, es decir, es el mismo en cualquiera de las dos especies.

III.XIV. Estimación del Cuadro Medio Residual

Puede estimarse el Cuadrado Medio Residual y compararlo con la estimación de la varianza del error obtenida en la sección 4. Interpretése de esta forma qué significa el Cuadrado Medio Residual.

```
round(anova(mod2)[4,3],2)
```

```
## [1] 2.56
```

```
round(mean(v),2)
```

```
## [1] 2.56
```

Este valor coincide con la media de las varianzas obtenida anteriormente y es una medida de la variabilidad de la respuesta dentro de cada tratamiento.

Si se observan los grados de libertad de cada fuente de variación puede verificarse que para el factor 'condición' hay 2 grados de libertad porque hay tres condiciones (tres niveles, que son los tipos de régimen alimenticio), mientras que para el factor 'especie' hay 1 grado de libertad porque son dos especies (que son los dos niveles o tipos de especie de tortuga), de lo que se desprende que los grados de libertad con la que cada fuente de variación es estimada son el número de niveles del factor menos uno. Finalmente, para la interacción entre condición y especie hay 2 grados de libertad que se obtienen del producto de los grados de libertad de cada uno de los factores.

III.XV. Cuadrado Medio de Condición y Especie

```
round(anova(mod2)[1,3],2)
```

```
## [1] 72.13
```

```
round(anova(mod2)[2,3],2)
```

```
## [1] 24
```

El cuadrado medio de 'condición' es 72.13 y es una medida de la distancia entre las medias de las condiciones o regímenes alimenticios. Similarmente, el cuadrado medio de 'especie' es 24 y es una medida de la distancia entre las medias de las dos especies.

III.XVI. Estimación de los Cuadrados Medios a partir de los Efectos Simples

```
table(cond)
```

```
## cond
```

```
## estricta balanceada abundancia
```

```
##      8      8      8
```

```
sum(8*alfa^2)/2
```

```
## [1] 72.125
```

```

table(especie)

## especie
## chelonia kynosternon
##      12      12

sum(12*beta^2)/1

## [1] 24

```

III.XVII. Estimación de la Suma de Cuadrados de Interacción A Partir de los Efectos de Interacción

```

table(cond,especie)

##      especie
## cond  chelonia kynosternon
## estricta      4      4
## balanceada    4      4
## abundancia    4      4

sum(4*efint^2)

## [1] 85.75

```

El Cuadrado Medio de Interacción es 42.88 y mide la magnitud general del efecto de interacción. Si esta magnitud tiende a ser pequeña, se tiene un caso con poca interacción, lo cual en este caso de aplicación sería una evidencia de débil interacción entre condición y especie. En caso contrario, si esta magnitud tiende a ser grande hay más evidencia de presencia de interacción.

III.XIX. Conclusión de Prueba de Hipótesis con Valor F

```

f=42.875/2.556
round(f,2)

## [1] 16.77

```

Se observa que la variabilidad de los efectos de interacción es sumamente alta con respecto a la variabilidad residual. Al ser la primera casi 17 veces la segunda, se tiene una probabilidad asociada sumamente baja ($p < 0.001$), con lo cual se rechaza la hipótesis de no-interacción. Se concluye que el efecto de la condición de alimento no es el mismo cuando se trata de *Chelonia mydas* que en el caso de *Kynosternum scorpioides*.

III.XX. Planteamiento Formal de los Contrastes de Comparación de Interacción Entre Especies

Puesto que se ha determinado que sí existe una interacción estadísticamente importante entre la condición y la especie, ahora es necesario investigar cómo actúan los regímenes alimenticios en cada especie. Escriba el modelo con interacción y luego escriba los contrastes adecuados para comparar las condiciones, pero dentro de cada especie.

$$M_{ij} = M + \alpha_i + \beta_j + (\alpha * \beta)_{ij}$$

El primer subíndice indica la condición y el segundo la especie. Así, debe fijarse el segundo subíndice. Se crean los contrastes de tal forma que el contraste estimado sea positivo, es decir, se pone primero aquella media cuya estimación sea mayor.

Se tienen los contrastes para *Chelonia*:

$$M_{21} - M_{11} \quad M_{21} - M_{31} \quad M_{31} - M_{11}$$

Los contrastes para *Kynosternon* son:

$$M_{22} - M_{12} \quad M_{32} - M_{22} \quad M_{32} - M_{12}$$

III.XXI. Construyendo la Matriz de Contraste de Coeficientes

Para hacer el cálculo del primer contraste se deben restar los dos vectores que producen cada una de las medias del contraste:

$$c(1,0,1,1,0,1) - c(1,1,0,1,1,0) = c(0, -1,1,0, -1,1)$$

Para el segundo contraste se usan: $c(1,0,1,1,0,1) - c(1, -1, -1,1, -1, -1) = c(0,1,2,0,1,2)$

Los demás se obtienen de forma similar:

$$c(1,0,1,1,0,1) - c(1,1,0,1,1,0)$$

$$\#\# [1] 0 -1 1 0 -1 1$$

$$c1 = c(0,-1,1,0,-1,1)$$

$$c(1,0,1,1,0,1) - c(1,-1,-1,1,-1,-1)$$

$$\#\# [1] 0 1 2 0 1 2$$

$$c2 = c(0,1,2,0,1,2)$$

$$c(1,-1,-1,1,-1,-1) - c(1,1,0,1,1,0)$$

$$\#\# [1] 0 -2 -1 0 -2 -1$$

$$c3 = c(0,-2,-1,0,-2,-1)$$

$$c(1,0,1,-1,0,-1) - c(1,1,0,-1,-1,0)$$

$$\#\# [1] 0 -1 1 0 1 -1$$

$$c4 = c(0,-1,1,0,1,-1)$$

$$c(1,-1,-1,-1,1,1) - c(1,0,1,-1,0,-1)$$

$$\#\# [1] 0 -1 -2 0 1 2$$

$$c5 = c(0,-1,-2,0,1,2)$$

$$c(1,-1,-1,-1,1,1) - c(1,1,0,-1,-1,0)$$

```
## [1] 0 -2 -1 0 2 1
```

```
c6=c(0,-2,-1,0,2,1)
```

```
cont=cbind(c1,c2,c3,c4,c5,c6)
```

III.XXI. Estimación de los Contrastes

- Dé una estimación de cada contraste.

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

```
row.names(L)=c("ch:bal-est","ch:bal-abu","ch:abu-est","ky:bal-est","ky:abu-  
bal","ky:abu-est")
```

```
L
```

```
##      [1]
```

```
## ch:bal-est 9.25
```

```
## ch:bal-abu 5.75
```

```
## ch:abu-est 3.50
```

```
## ky:bal-est 2.25
```

```
## ky:abu-bal 3.00
```

```
## ky:abu-est 5.25
```

III.XXII. Estimación de la Varianza de los Contrastes

```
var=diag(t(cont)%*%vcov(mod2)%*%cont)
```

```
round(var,2)
```

```
##  c1  c2  c3  c4  c5  c6
```

```
## 1.28 1.28 1.28 1.28 1.28 1.28
```

Las varianzas de todos los contrastes son iguales porque es un diseño balanceado.

III.XXIII. Verificando la Significancia Estadística de los Contrastes

Deben probarse de forma simultánea las hipótesis que establecen que cada contraste es igual a cero. Puesto que los contrastes no son ortogonales, puede usarse la corrección de Bonferroni, es decir, dividir el nivel de significancia por el número de comparaciones.

```
ee=sqrt(var)
t=L/ee
row.names(t)=row.names(L)
round(t,2)

##      [,1]
## ch:bal-est 8.18
## ch:bal-abu 5.09
## ch:abu-est 3.10
## ky:bal-est 1.99
## ky:abu-bal 2.65
## ky:abu-est 4.64

k=6
p=pt(t,18,lower.tail = F)
row.names(p)=row.names(L)
round(p,4)

##      [,1]
## ch:bal-est 0.0000
## ch:bal-abu 0.0000
## ch:abu-est 0.0031
## ky:bal-est 0.0310
## ky:abu-bal 0.0081
## ky:abu-est 0.0001
```

```
round(0.05/k,4)
```

```
## [1] 0.0083
```

Todas las hipótesis, con excepción de la relativa a balanceada-estricta en kynosternon (que tiene una significancia de 0.0310), por lo que se concluye que en cada especie el nivel promedio de proteína es diferente en todas las condiciones, lo que a su vez significa que todas las condiciones (tipos de régimen alimenticio) son relevantes en el estudio del fenómeno (nivel de proteína en sangre) en casi todas las especies (la excepción es la antes mencionada). Complementariamente, para kynosternon no se ha encontrado que la dieta balanceada produzca un nivel promedio de proteína más alto que la dieta estricta.

III.XXIV. Estimando Cotas Inferiores

Deben estimarse cotas inferiores que permitan cuantificar las diferencias que son significativas.

```
tc=qt(1-0.05/5,18)
```

```
LIM=L[-4]-tc*ee[-4]
```

```
names(lim)=row.names(L)[-4]
```

```
round(lim,2)
```

```
## ch:bal-est ch:bal-abu ch:abu-est ky:abu-bal ky:abu-est
```

```
## 6.36 2.86 0.61 0.11 2.36
```

Se concluye con 95% de confianza que el nivel promedio de proteína en chelonia es al menos 6.36gr/ml mayor con dieta balanceada que con dieta estricta y al menos 2.86gr/ml mayor con dieta balanceada que con abundancia. En kynosternon abundancia aventaja a dieta estricta en al menos 2.36gr/ml. Las otras dos diferencias resultan poco relevantes, dado que para el especialista una diferencia importante es más de 1 gr/ml.

IV. REFERENCIAS

- Frost, J. (26 de Julio de 2022). *Understanding Interaction Effects in Statistics*. Obtenido de Regression: <https://statisticsbyjim.com/regression/interaction-effects/>
- Marco Sanjuán, F. J. (26 de Julio de 2022). *Suma total de cuadrados (STC)*. Obtenido de economipedia: <https://economipedia.com/definiciones/suma-total-de-cuadrados-stc.html>
- Marín Diazaraque, J. M. (26 de Julio de 2022). *Introducción al Diseño de Experimentos*. Obtenido de Universidad Carlos III de Madrid: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Disenno/IntroDE.pdf>
- Melo, Ó. O., López, L. A., & Melo, S. E. (2020). *Diseño de Experimentos: Métodos y Aplicaciones*. Bogotá: Universidad Nacional de Colombia.