Suma de cuadrados de contrastes

Para definiar varios constrates lineales a la vez:

$$H_0 : C\mu = \mathbf{0}$$

donde la matriz de contrastes de definen como los coeficientes de cada contraste en filas (cada fila corresponde a un constraste lineal):

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \dots \\ c_{k1} & c_{k2} & \dots & c_{kp} \end{pmatrix}$$

Se construye el siguiente estadístico

$$F = \frac{C\hat{\mu}\left[\widehat{Var}(C\bar{\mu})\right)^{-1}(C\hat{\mu})^t}{\nu_1}$$

que sigue una F de Fisher con ν_1 grados de libertat en el numerador y ν_2 grados de libertad en el denominador. ν_1 es el rango de la matriz C (o sea, cuantos contrastes hay **linealmente independientes** -no confundir con ortogonales-) mientras que ν_2 son los grados de libertad del error (tabla ANOVA).

$$C\hat{\mu} = C \begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \dots \\ \bar{y}_{p\cdot} \end{pmatrix} = \begin{pmatrix} c_{11}\bar{y}_{1\cdot} & c_{12}\bar{y}_{2\cdot} & \dots & c_{1p}\bar{y}_{p\cdot} \\ c_{21}\bar{y}_{1\cdot} & c_{22}\bar{y}_{2\cdot} & \dots & c_{3p}\bar{y}_{p\cdot} \\ \dots \\ c_{k1}\bar{y}_{1\cdot} & c_{k2}\bar{y}_{2\cdot} & \dots & c_{kp}\bar{y}_{p\cdot} \end{pmatrix} = \begin{pmatrix} \hat{L}_1 \\ \hat{L}_2 \\ \dots \\ \hat{L}_k \end{pmatrix} =$$

$$\widehat{Var}(C\hat{\mu}) = C\widehat{Var}(\hat{\mu})C^t = C\widehat{Var}(\begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \dots \\ \bar{y}_{p\cdot} \end{pmatrix})C^t = C \begin{pmatrix} \hat{\sigma}^2/n & & & 0 \\ & \hat{\sigma}^2/n & & \\ & & \ddots & \\ 0 & & & \hat{\sigma}^2/n \end{pmatrix} C^t = \frac{\hat{\sigma}^2}{n} CC^t$$

Si los contrastes son ortogonales, luego CC^t es una matriz diagonal cuyos elementos de la diagonal son $\sum_{j=1}^p c_{1j}^2, \dots, \sum_{j=1}^p c_{kj}^2$, y ν_1 es el número de contrastes (k):

$$\widehat{Var}(C\hat{\mu}) = \frac{\hat{\sigma}^2}{n} \begin{pmatrix} \sum_{j=1}^p c_{1j}^2 & & & 0 \\ & \sum_{j=1}^p c_{2j}^2 & & \\ & & \ddots & \\ 0 & & & \sum_{j=1}^p c_{kj}^2 \end{pmatrix}$$

Y la inversa de la varianza,

$$\left[\widehat{Var}(C\hat{\mu})\right]^{-1} = \frac{n}{\hat{\sigma}^2} \begin{pmatrix} \frac{1}{\sum_{j=1}^p c_{1j}^2} & & & 0 \\ & \frac{1}{\sum_{j=1}^p c_{2j}^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sum_{j=1}^p c_{kj}^2} \end{pmatrix}$$

Así el estadístico F cuando los contrastes lineales son ortogonales es

$$F = \frac{\frac{n}{\hat{\sigma}^2} \begin{pmatrix} \hat{L}_1 & \hat{L}_2 & \dots & \hat{L}_k \end{pmatrix} \begin{pmatrix} \frac{1}{\sum_{j=1}^p c_{1j}^2} & & & 0 \\ & \frac{1}{\sum_{j=1}^p c_{2j}^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sum_{j=1}^p c_{kj}^2} \end{pmatrix} \begin{pmatrix} \hat{L}_1 \\ \hat{L}_2 \\ \dots \\ \hat{L}_k \end{pmatrix}}{k} =$$

$$= \frac{\frac{n}{\hat{\sigma}^2} \begin{pmatrix} \frac{\hat{L}_1}{\sum_{j=1}^p c_{1j}^2} & \frac{\hat{L}_2}{\sum_{j=1}^p c_{2j}^2} & \cdots & \frac{\hat{L}_k}{\sum_{j=1}^p c_{kj}^2} \end{pmatrix} \begin{pmatrix} \hat{L}_1 \\ \hat{L}_2 \\ \dots \\ \hat{L}_k \end{pmatrix}}{k}$$

$$= \frac{\frac{n}{\hat{\sigma}^2} \left(\frac{\hat{L}_1^2}{\sum_{j=1}^p c_{1j}^2} + \frac{\hat{L}_2^2}{\sum_{j=1}^p c_{2j}^2} + \dots + \frac{\hat{L}_k^2}{\sum_{j=1}^p c_{kj}^2} \right)}{k}$$

$$= \frac{\frac{1}{\hat{\sigma}^2} \left(\frac{n\hat{L}_1^2}{\sum_{j=1}^p c_{1j}^2} + \frac{n\hat{L}_2^2}{\sum_{j=1}^p c_{2j}^2} + \dots + \frac{n\hat{L}_k^2}{\sum_{j=1}^p c_{kj}^2} \right)}{k}$$

$$= \frac{\frac{1}{\hat{\sigma}^2} \left(SS_{L_1} + SS_{L_2} + \dots + SS_{L_k} \right)}{k}$$

$$= \frac{\left(SS_{L_1} + SS_{L_2} + \dots + SS_{L_k} \right)/k}{\hat{\sigma}^2}$$

$$= \frac{\left(SS_{L_1} + SS_{L_2} + \dots + SS_{L_k} \right)/k}{MSE} \sim F_{\nu_2}^k$$

Así, se identifica como:

- $SS_{L_1} + SS_{L_2} + \dots + SS_{L_k}$ es la suma de cuadrados del test conjunto de los k contrastes lineales ortogonales,
- k es la suma de grados de libertad de cada constraste lineal por separado,

Ejemplos

Usando los datos del ejercicio 1 de la lista:

```r
y <- c( 94.09,  90.45,  99.38,  73.56,  74.39,
        98.81, 103.55, 115.23, 129.06, 117.61,
       197.18, 207.31, 177.50, 226.05, 222.74,
       102.93, 117.51, 119.92, 112.01, 101.10,
        83.14,  89.59,  87.76,  96.43,  82.94)

grup <- factor(rep(1:5, each=5))
```

Y usando los contrastes planteados en el ejercicio 2:

```
C1 <- c(1,-1,0,0,0)  # Premuda vs ayuno
C2 <- c(0,1,-1,0,0)  # Ayuno vs 60g
C3 <- c(0,0,1,-1,0)  # 60 vs 80g
C4 <- c(0,0,0,1,-1)  # 80 g vs mezcla
```

definimos la matriz de contrastes colocando los coeficientes por filas para seguir la formulación antes planteada.

```
MatC <- rbind(C1,C2,C3,C4)
```

Ya hemos visto que **no** son contrastes ortogonales. Pero podemos plantear el test de significación conjunto de los 4 contrastes lineales

```
medias <- tapply(y, grup, mean)
model <- aov(y ~ grup)
tablaANOVA <- summary(model)
MSE <- tablaANOVA[[1]][2,3]
dferror <- tablaANOVA[[1]][2,1]
n <- 5
```

```
L <- MatC%*%medias
VarC <- (MSE/n)*MatC%*%t(MatC)
nu1 <- qr(MatC)$rank
nu2 <- dferror
```

El estadístico F se calcula como:

```
Ftest <- t(L)%*%solve(VarC)%*%L/(qr(MatC)$rank)
Ftest
```

```
         [,1]
[1,] 78.08026
```

Y el p-valor correspondiente es

```
pvalor <- 1-pf(Ftest, nu1, nu2)
pvalor
```

```
             [,1]
[1,] 6.481593e-12
```

Recordamos la significación individual de cada contraste

```
library(gmodels)
fit.contrast(model, "grup", coeff=MatC)
```

```
       Estimate Std. Error    t value     Pr(>|t|)
grupC1  -26.478   7.886937  -3.357197 3.135828e-03
grupC2  -93.304   7.886937 -11.830194 1.748178e-10
grupC3   95.462   7.886937  12.103811 1.167421e-10
grupC4   22.722   7.886937   2.880966 9.237985e-03
```

Pero al realizar múltiples contrastes podemos inflar la probabilidad de error de tipo I. Para tener en cuenta la multiplicidad de contrastes podemos usar el método de Scheffé para el test de significación de múltiples contrastes lineales (no hace falta que sean ortogonales pero sí independientes):

Rechazaremos la H0 del contraste lineal j cuando

$$|L_j| > \sqrt{(k-1)F_{\alpha;k-1;\nu_2}}\sqrt{MSE\sum_{i=1}^{p} c_{ki}^2/n_i}$$

,

donde k es el número de contrastes linealmente independientes (no hace falta que sean ortogonales!!), en este ejemplo $k = 4$.

```
alpha <- 0.05
k <- nrow(MatC)

resultados <- data.frame("L"=double(),"Frontera"=double(),"Decisión"=character())
for (i in 1:k){
  frontera <- sqrt(k*qf(1-alpha,k,nu2))*sqrt(MSE*sum(MatC[i,]^2)/n)
  resultados <- rbind(resultados,
                      data.frame("L"=L[i],
                                 "Frontera"=frontera,
                                 "Decisión"=ifelse(abs(L[i])>frontera, "RH0","AH0")))
}

rownames(resultados) <- paste0("C",1:4)
library(knitr)
kable(resultados)
```

	L	Frontera	Decisión
C1	-26.478	26.70439	AH0
C2	-93.304	26.70439	RH0
C3	95.462	26.70439	RH0
C4	22.722	26.70439	AH0

Fíjate que el valor frontera va cambiando ya que los coeficientes (`MatC[i,]`) de los contrastes también son distintos en general.