

5 Analisis de varianza

- Diseño factorial: experimentos cubriendo todas las combinaciones de los niveles que pueden tomar los factores experimentales
- Diseño balanceado: mismo numero de observaciones por celda
- Diseño no balanceado: distinto numero de observaciones

Modelo ANOVA

two-way layout with interactions

$$y_{ijk} = \mu + \gamma_i + \beta_j + (\gamma\beta)_{ij} + \varepsilon_{ijk}$$

Termino de error ddp
Normal $(0, \sigma^2)$

Interaccion entre los dos factores

Efecto principal del segundo factor

Efecto principal del primer factor

Media global

K-esima media en la celda (i,j)

- Especificación en R

two-way layout with interactions

$$y \sim a + b + a:b$$

$$y \sim a * b$$

Caso especial $\mu = 0$ $\rightarrow y \sim a + b + a:b - 1$

Test de hipotesis

- F-test sobre los efectos.
- Hipotesis nula: efectos identicos
- Asunciones:
 - Observaciones independientes
 - Celdas normalmente distribuidas
 - Misma varianza

MANOVA

one-way MANOVA

$$y_{ijh} = \mu_h + \gamma_{jh} + \varepsilon_{ijh}$$

Diagram illustrating the components of the MANOVA equation $y_{ijh} = \mu_h + \gamma_{jh} + \varepsilon_{ijh}$:

- y_{ijh} : Medida de la h-esima variable con el nivel i del factor
- μ_h : Media de la h-esima variable
- γ_{jh} : Efecto del nivel i en la variable h
- ε_{ijh} : Terminio de error con ddp Normal multivariada

$$\varepsilon_{ij}^T = (\varepsilon_{ij1}, \varepsilon_{ij2}, \dots, \varepsilon_{ijq}) \quad \text{ddp Normal Multivariada}$$

Hipotesis nula: los efectos son los mismos para todos los niveles

Caso 1 ANOVA

The data in Table 5.1 (from Hand et al., 1994) arise from an experiment to study the gain in weight of rats fed on four different diets, distinguished by amount of protein (low and high) and by source of protein (beef and cereal). Ten rats are randomised to each of the four treatments and the weight gain in grams recorded. The question of interest is how diet affects weight gain.

➤ data(weightgain)

> View(weightgain)

- **Estadísticas sumarias**

```
tapply(weightgain$weightgain, list(weightgain$source, weightgain$type), mean)
```

```
tapply(weightgain$weightgain, list(weightgain$source, weightgain$type), sd)
```

- **Visualización**

```
plot.design(weightgain)
```

- **AOV**

```
wg_aov <- aov(weightgain ~ source * type, data = weightgain)
```

```
summary(wg_aov)
```

- **Visualizacion**

```
interaction.plot(weightgain$type, weightgain$source, weightgain$weightgain)
```

- **coeficientes**

```
coef(wg_aov)
```

```
coef(aov(weightgain ~ source + type + source:type, data = weightgain,  
contrasts = list(source = contr.sum)))
```

Caso 2 ANOVA

The data in Table 5.2 are from a foster feeding experiment with rat mothers and litters of four different genotypes: A, B, I and J (Hand et al., 1994). The measurement is the litter weight (in grams) after a trial feeding period. Here the investigator's interest lies in uncovering the effect of genotype of mother and litter on litter weight.

Table 5.2: `foster` data. Foster feeding experiment for rats with different genotypes of the litter (`litgen`) and mother (`motgen`).

<code>litgen</code>	<code>motgen</code>	<code>weight</code>	<code>litgen</code>	<code>motgen</code>	<code>weight</code>
A	A	61.5	B	J	40.5
A	A	68.2	I	A	37.0
A	A	64.0	I	A	36.3

```
>data(foster)
```

```
> View(foster)
```

- Visualization

```
plot.design(foster)
```

- Analysis (unbalanced design)

```
summary(aov(weight ~ litgen * motgen, data = foster))
```

```
summary(aov(weight ~ motgen * litgen, data = foster))
```

- Posthoc analysis *multiple comparison procedures*
Tukey honest significant differences

```
foster_aov <- aov(weight ~ litgen * motgen, data = foster)
```

```
foster_hsd <- TukeyHSD(foster_aov, "motgen")
```

```
plot(foster_hsd)
```

Caso 3 MANOVA

The data in Table 5.3 (from Hand et al., 1994) give four measurements made on Egyptian skulls from five epochs. The data has been collected with a view to deciding if there are any differences between the skulls from the five epochs. The measurements are:

mb: maximum breadths of the skull,

bh: basibregmatic heights of the skull,

bl: basialveolar length of the skull, and

nh: nasal heights of the skull.

➤ `data(skulls)`

➤ `View(skulls)`

Non-constant measurements of the skulls over time would indicate interbreeding with immigrant populations.

Table 5.3: `skulls` data. Measurements of four variables taken from Egyptian skulls of five periods.

epoch	mb	bh	bl	nh
c4000BC	131	138	89	49
c4000BC	125	121	82	48

- Exploration

```
means <- aggregate(skulls[,c("mb", "bh", "bl", "nh")], list(epoch = skulls$epoch), mean)
```

- Visualization

```
pairs(means[,-1],  
+ panel = function(x, y) {  
+ text(x, y, abbreviate(levels(skulls$epoch)))  
+ })
```

pairs(means)

- **Analysis**

```
skulls_manova <- manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls)
```

```
summary(skulls_manova, test = "Pillai")
```

```
summary(skulls_manova, test = "Wilks")
```

```
summary(skulls_manova, test = "Hotelling-Lawley")
```

```
summary(skulls_manova, test = "Roy")
```

- Posthoc analysis

```
summary.aov(skulls_manova)
```

- Comparando epocas

```
summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,  
+ subset = epoch %in% c("c4000BC", "c3300BC")))
```

```
summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,  
+ subset = epoch %in% c("c4000BC", "c1850BC")))
```

```
summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,  
+ subset = epoch %in% c("c4000BC", "c200BC")))
```

```
summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,  
+ subset = epoch %in% c("c4000BC", "cAD150")))
```