

# **Clase 06 - Análisis de varianza**

## **Curso Análisis de datos con R para Biociencias.**

Dra. Angélica Rueda Calderón

Pontificia Universidad Católica de Valparaíso

15 January 2023

# PLAN DE LA CLASE

## 1.- Introducción

- ▶ ¿Qué es un análisis de varianza?.
- ▶ Hipótesis y supuestos.
- ▶ Interpretar resultados de análisis de varianza con R.
- ▶ Pruebas a posteriori.

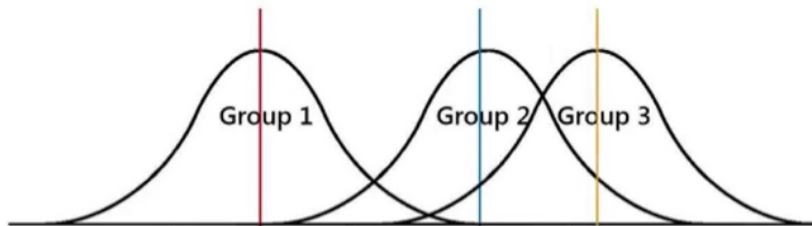
## 2.- Práctica con R y Rstudio cloud

- ▶ Realizar pruebas de hipótesis: Anova y posteriores.
- ▶ Realizar gráficas avanzadas con ggplot2.

# ANOVA

## ¿Qué es un análisis de varianza?

Herramienta básica para analizar el efecto de uno o más factores (cada uno con dos o más niveles) en un experimento.



# HIPÓTESIS EN UNA ANOVA

## Hipótesis factor 1

$$H_0 : \alpha_{1.1} = \alpha_{1.2} = \alpha_{1.3}$$

## Hipótesis factor 2

$$H_0 : \beta_{2.1} = \beta_{2.2} = \beta_{2.3}$$

## Hipótesis interacción

$$H_0 : \alpha^*\beta = 0$$

## Hipótesis Alternativa

$H_A$  : No todas las medias son iguales

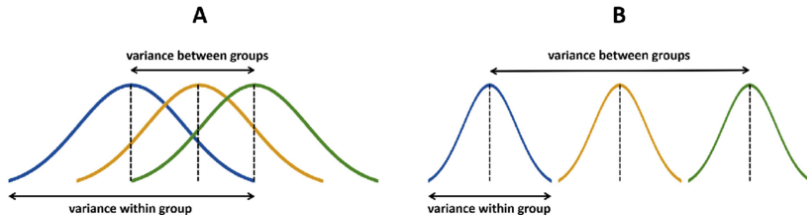
# ANOVA PARA COMPARAR MEDIAS

¿Por qué se llama ANOVA si se comparan medias?

Por que el estadístico **F** es un cociente de varianzas.

$$F = \frac{\sigma_{\text{entregrupos}}^2}{\sigma_{\text{dentrogrupos}}^2}$$

Mientras mayor es el estadístico **F**, más es la diferencia de medias entre grupos.



# SUPUESTOS DE UNA ANOVA

- 1) Independencia de las observaciones.
- 2) Normalidad.
- 3) Homocedasticidad: homogeneidad de las varianzas.

# TEST POSTERIORES (PRUEBAS A POSTERIORI)

## ¿Para qué sirven?

Para identificar que pares de niveles de uno o más factores son significativamente distintos entre sí.

## ¿Cuándo usarlos?

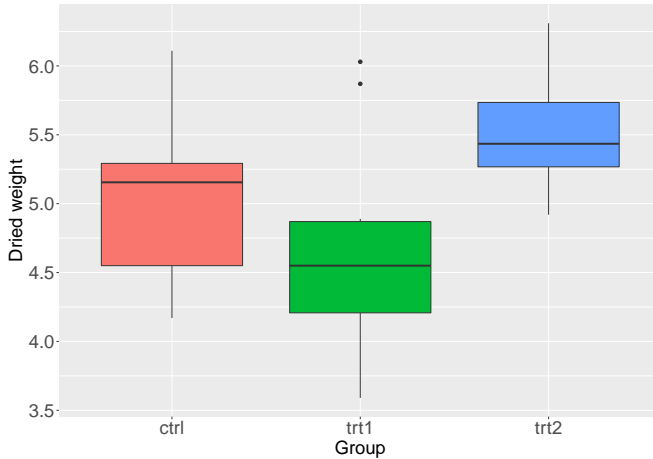
Sólo cuando se rechaza  $H_0$  del ANOVA.

## Tukey test

Es uno de los más usados, similar al *t-test*, pero corrige la tasa de error por el número de comparaciones.

# ESTUDIO DE CASO: CRECIMIENTO DE PLANTAS

## ► PlantGrowth {datasets}





# ANOVA A UNA VÍA (UN CRÍTERIO DE CLASIFICACIÓN)

```
res.aov <- lm(weight ~ group, data = my_data)
anova(res.aov)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: weight
```

```
##           Df  Sum Sq Mean Sq F value  Pr(>F)
```

```
## group      2   3.7663   1.8832   4.8461 0.01591 *
```

```
## Residuals 27 10.4921   0.3886
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

# COMPARACIONES MÚLTIPLES

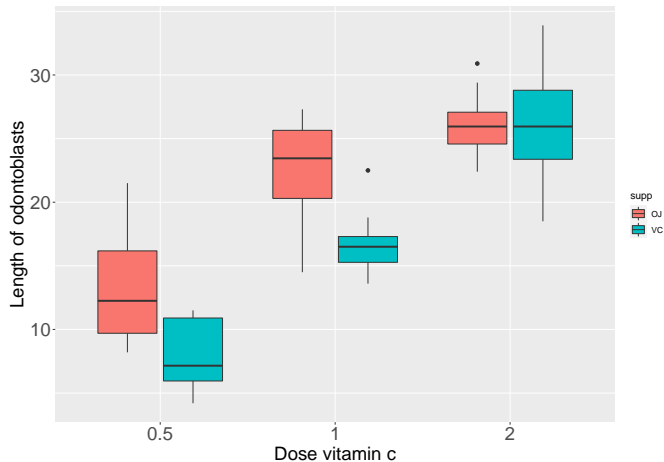
```
fit_anova <- aov(res.aov)
tk <- TukeyHSD(fit_anova)
```

**Table 1:** Prueba de Tukey.

Trat.	Contraste	H0	Diferencia	IC-bajo	IC-alto	p-ajustado
group	trt1-ctrl	0	-0.37	-1.06	0.32	0.39
group	trt2-ctrl	0	0.49	-0.20	1.19	0.20
group	trt2-trt1	0	0.86	0.17	1.56	0.01

# ESTUDIO DE CASO: GUINEA PIGS

## ► ToothGrowth {datasets}



OJ: Orange juice - VC: Vitamin C

# ANOVA DOS VIAS CON INTERACCIÓN

```
res.aov2 <- lm(len ~ dose * supp, data = my_data1)
anova(res.aov2)
```

Analysis of Variance Table

Response: len

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
dose	2	2426.43	1213.22	92.000	< 2.2e-16	***
supp	1	205.35	205.35	15.572	0.0002312	***
dose:supp	2	108.32	54.16	4.107	0.0218603	*
Residuals	54	712.11	13.19			

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# PRÁCTICA ANÁLISIS DE DATOS

- ▶ El trabajo práctico se realiza en Posit.cloud.

## **Guía 06 Anova y posteriores**

# RESUMEN DE LA CLASE

- ▶ **Elaborar hipótesis de anova**
- ▶ **Realizar análisis de varianza**
  - ▶ 1 factor.
  - ▶ 2 factores y con interacción.
  - ▶ pruebas *a posteriori*
- ▶ **Realizar gráficas avanzadas con ggplot2**