

# Clase 09 Análisis de varianza

OCE 386 - Introducción al análisis de datos con R.

Dr. José A. Gallardo | Pontificia Universidad Católica de  
Valparaíso

03 October 2021

# PLAN DE LA CLASE

## 1.- Introducción

- ▶ ¿Qué es un análisis de varianza?.
- ▶ Modelos lineales en Anova.
- ▶ Hipótesis y supuestos.
- ▶ Interpretar resultados de análisis de varianza con R.

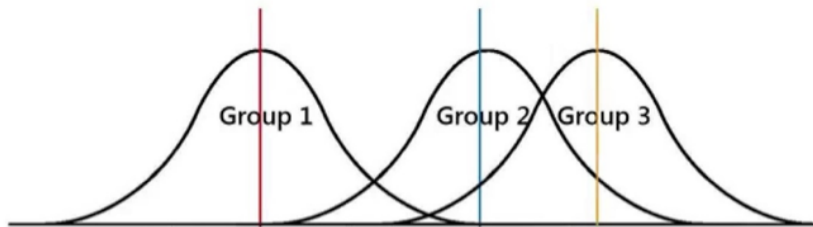
## 2.- Práctica con R y Rstudio cloud

- ▶ Realizar pruebas de hipótesis: Anova y posteriores.
- ▶ Realizar gráficas avanzadas con ggplot2.
- ▶ Elaborar un reporte dinámico en formato pdf.

# ANOVA

## ¿Qué es el análisis de varianza?

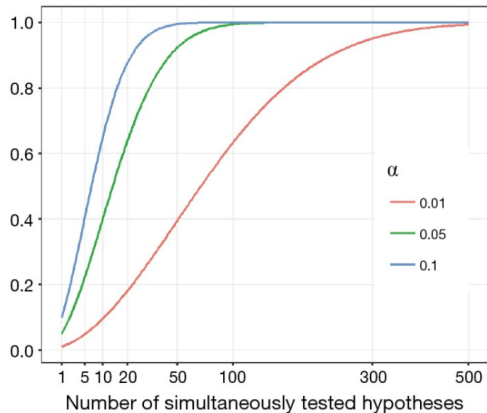
Herramienta básica para analizar el efecto de uno o más factores (cada uno con dos o más niveles) en un experimento.



# PROBLEMA DE LAS COMPARACIONES MÚLTIPLES

## ¿Por qué preferir anova y no múltiples t-test?

Porque con una t-test normal se incrementa la tasa de error al aumentar el número de comparaciones múltiples.



# ANOVA: MODELOS, HIPÓTESIS Y SUPUESTOS

## Modelos lineales

$$\text{Respuesta} \sim \mu + \alpha + \epsilon$$

$$\text{Respuesta} \sim \mu + \alpha + \beta + \epsilon$$

$$\text{Respuesta} \sim \mu + \alpha + \beta + \alpha^*\beta + \epsilon$$

## 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

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

## Supuestos:

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

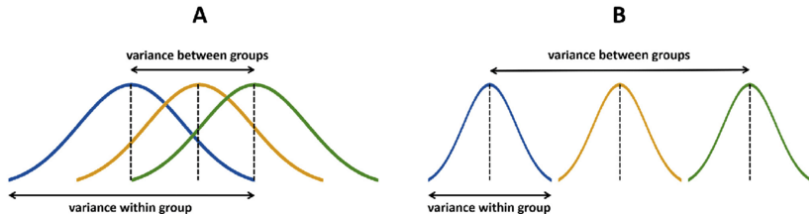
# ANOVA PARA COMPARAR MEDIAS

¿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.



# 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.

## R Documentation anova {stats}

### Anova de una vía

```
res.aov <- aov(respuesta ~ group, data = my_data, ...)  
summary(res.aov)
```

*aov()*, sólo para diseños balanceados

```
##              Df Sum Sq Mean Sq F value Pr(>F)  
## group         2  3.766   1.8832    4.846 0.0159 *  
## Residuals    27 10.492   0.3886  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```



## R Documentation TukeyHSD {stats}

```
TukeyHSD(res.aov)
```

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = weight ~ group, data = my_data)
##
## $group
##           diff           lwr           upr           p adj
## trt1-ctrl -0.371 -1.0622161  0.3202161  0.3908711
## trt2-ctrl  0.494 -0.1972161  1.1852161  0.1979960
## trt2-trt1  0.865  0.1737839  1.5562161  0.0120064
```

## R Documentation anova {stats} 2

### Anova de dos vías con interacción

```
aov1 <- aov(respuesta ~ group1 * group2)
aov2 <- aov(respuesta ~ group1 + group2 + group1 * group2)
aov1 = aov2
summary(aov1)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## supp           1   205.4    205.4   12.317 0.000894 ***
## dose           1 2224.3   2224.3  133.415 < 2e-16 ***
## supp:dose       1    88.9     88.9    5.333 0.024631 *
## Residuals      56   933.6     16.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

# PRÁCTICA ANÁLISIS DE DATOS

- ▶ Guía de trabajo práctico disponible en drive y Rstudio.cloud.  
**Clase\_08**
- ▶ El trabajo práctico se realiza en Rstudio.cloud.  
**Guía 8 Anova y posteriores**

# RESUMEN DE LA CLASE

- ▶ **Elaborar hipótesis**
- ▶ **Realizar análisis de varianza**
  - ▶ 1 factor.
  - ▶ 1 factor con posteriores.
  - ▶ 2 factores con interacción.
- ▶ **Realizar gráficas avanzadas con ggplot2**