

## Clase 10 Análisis de varianza

Diplomado en análisis de datos con R para la acuicultura.

Dr. José A. Gallardo y Dra. María Angélica Rueda.  
jose.gallardo@pucv.cl | Pontificia Universidad Católica de  
Valparaíso

22 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.

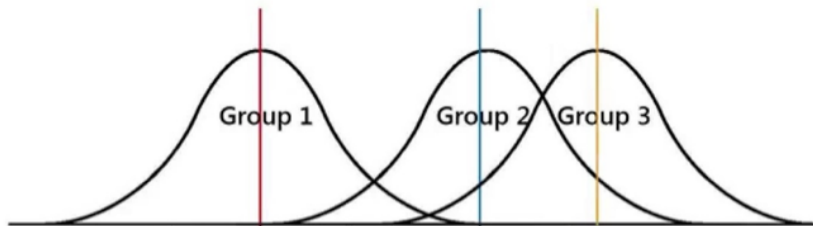
## 2.- Práctica con R y Rstudio cloud

- ▶ Realizar guía 9 + 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?

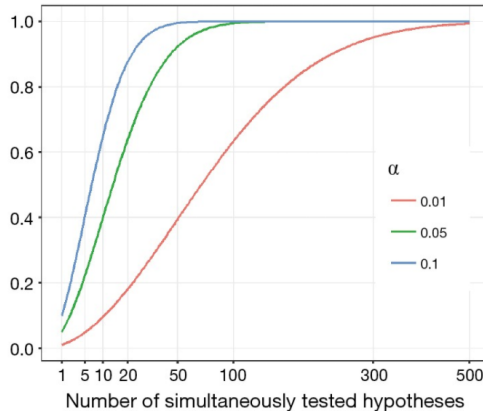
Herramienta básica para analizar el efecto de uno o más factores (cada uno con dos o más niveles) en un experimento.



# 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.



Fuente[1]: [1]:doi:10.21037/jtd.2017.05.34

# ANOVA: MODELOS LINEALES

Una forma muy conveniente de representar una ANOVA es mediante un modelo lineal.

## **Modelo lineal para ANOVA de una vía**

$$y \sim \mu + \alpha + \epsilon$$

## **Modelo lineal para ANOVA de dos vías**

$$y \sim \mu + \alpha + \beta + \epsilon$$

## **Modelo lineal para ANOVA de dos vías con interacción**

$$y \sim \mu + \alpha + \beta + \alpha*\beta + \epsilon$$

# ANOVA: HIPÓTESIS Y SUPUESTOS

## **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<sub>A</sub>** : No todas las medias son iguales

## **Supuestos:**

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

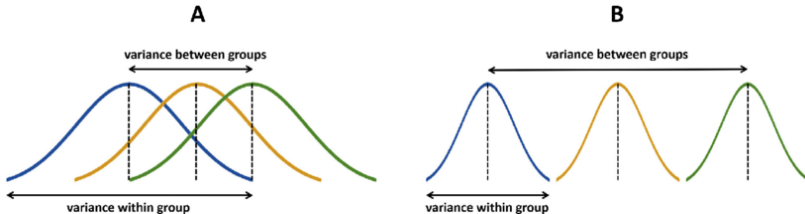
# ANOVA PARA COMPARAR MEDIAS

¿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?

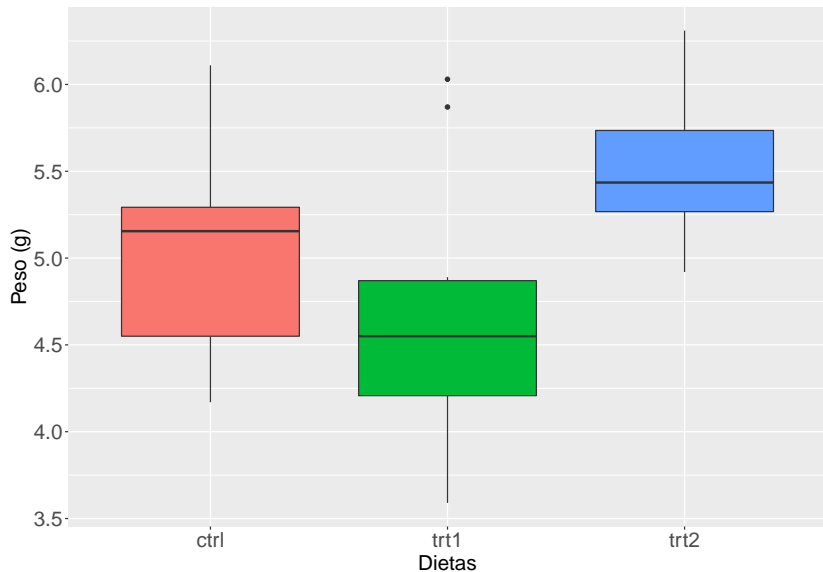
Sólo cuando se rechaza  $H_0$  del ANOVA.

## Tukey test

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



# ESTUDIO DE CASO: EVALUACIÓN DE DIETAS TRUCHA ARCOIRIS



## Anova de una vía.

```
# Crea objeto anova  
res.aov <- aov(Peso ~ Dietas, data = my_data)  
# Imprime resultado en formato tabla.  
anova(res.aov)%>% kable(caption = "Anova de una vía.",  
                        digits=2)
```

Table 1: Anova de una vía.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Dietas	2	3.77	1.88	4.85	0.02
Residuals	27	10.49	0.39	NA	NA

***aov()** sólo para diseños balanceados*

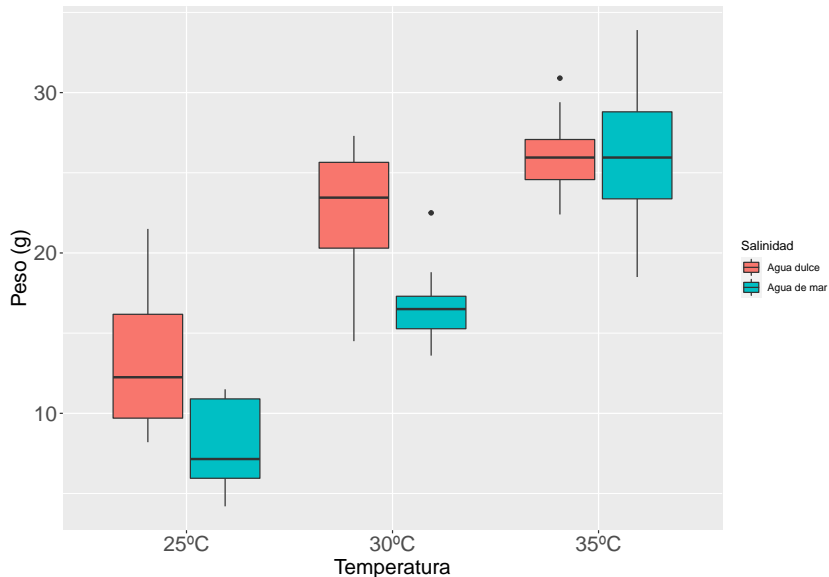
## Prueba de Tukey

```
tk <- TukeyHSD(res.aov)
tidy(tk) %>% kable(caption = "Prueba de Tukey.", digits=2,
col.names=c("Trat.", "Contraste", "H0",
             "Diferencia", "IC-bajo", "IC-alto",
             "p-ajustado"))
```

Table 2: Prueba de Tukey.

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

# ESTUDIO DE CASO: EVALUACIÓN CRECIMIENTO TILAPIA (TEMPERATURA Y SALINIDAD)



## Anova de dos vías con interacción

```
res.aov2 <- aov(Peso ~ Temperatura * Salinidad,  
               data = my_data1)  
anova(res.aov2)%>% kable(caption = "Anova de dos vías.",  
                          digits=3)
```

Table 3: Anova de dos vías.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperatura	2	2426.434	1213.217	92.000	0.000
Salinidad	1	205.350	205.350	15.572	0.000
Temperatura:Salinidad	2	108.319	54.160	4.107	0.022
Residuals	54	712.106	13.187	NA	NA

# PRÁCTICA ANÁLISIS DE DATOS

- ▶ Guía de trabajo práctico disponible en drive y Rstudio.cloud.

## **Clase\_10**

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

## **Guía 10 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**