

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**ENES MÉRIDA**  
**LICENCIATURA EN ECOLOGÍA**

**ESTADÍSTICA APLICADA**  
**Tema 4. Análisis de Varianza**

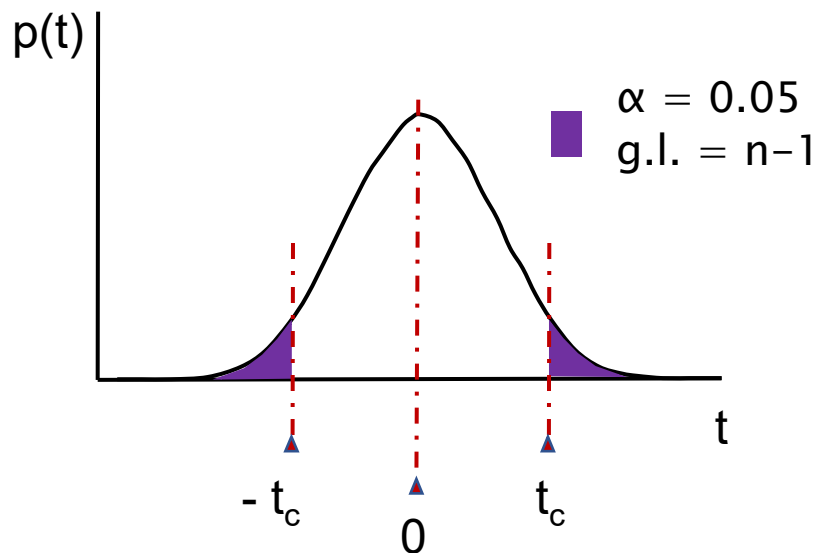
Prof. Edlin J. Guerra Castro

## Prueba $t$ : Un muestreo / grupo

$$t = \frac{\bar{X} - \mu_0}{S_{\bar{X}}}$$

Diferencia entre media de la muestra y un valor de referencia (fijo)

Error estándar: medida de la precisión para estimar la media a partir de la muestra



La distribución de probabilidades de valores de  $t$  Student bajo la hipótesis nula de que  $t = \mu_0$

**Bajo la  $H_0$**

$t \approx 0$

Tienen probabilidades altas de ocurrir

$t \gg 0$

Tienen probabilidades bajas de ocurrir

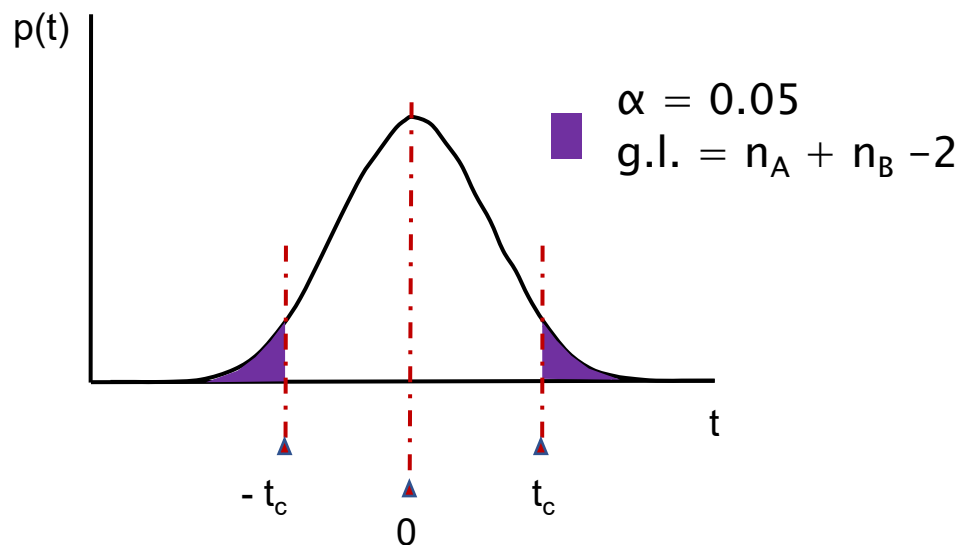
$t \ll 0$

## Prueba $t$ : dos muestreos / grupos

Diferencia entre las  
dos medias

$$t = \frac{\bar{X}_A - \bar{X}_B}{S_p}$$

Error estándar “pooled”: medida  
de la precisión para estimar la  
diferencia de las medias.



La distribución de probabilidades de  
valores de  $t$  *Student* bajo la hipótesis  
nula de que  $t = \mu_0$

**Bajo la  $H_0$**

$\left\{ \begin{array}{l} t \approx 0 \\ t \gg 0 \\ t \ll 0 \end{array} \right. \begin{array}{l} \text{Tienen probabilidades altas de ocurrir} \\ \text{Tienen probabilidades bajas de ocurrir} \end{array}$

# Modelos lineales

**Casos simples:** modelo de la media: una cola, dos colas, tipo de varianzas, tipo de distribución.

## **Casos más complejos:**

1. Análisis de varianza: un factor
2. Análisis de varianza: dos o más factores
  - a) Diseños ortogonales
  - b) Diseños anidados
  - c) Diseños mixtos
3. Regresión lineal simple
4. Regresión lineal múltiple
5. Análisis de covarianza

## **Evaluación de supuestos**

Normalidad

Homocedasticidad

Independencia

Aditividad

## Conflicto entre Argentina y Uruguay por plantas de celulosa (2005-2010)



Las plantas de celulosa se dedican al procesamiento de la madera para la obtención de la principal materia prima para la producción de papel: la pulpa, o pasta.

Residuos tóxicos: varios, principalmente Materia Orgánica.

# Desarrollemos el MHD

- Problema
- Modelos
- Hipótesis científica y/o empírica
- Hipótesis estadística
- Hipótesis nula
- Evaluación de la Hipótesis nula... ¿cómo lo hacemos?
- ¿Conclusión?

# Evaluar la DBO en 4 puntos a lo largo del Río luego de la instalación de la planta



Métodos

10 muestras por localidad

# ANALISIS DE VARIANZA (ANOVA)

Ideal para estudios experimentales o mensurativos

*es una prueba que compara **medias**, pero también puede comparar **varianzas***

SE APLICA A **VARIABLES INDEPENDIENTES CATEGÓRICAS** QUE PUEDEN TENER UN EFECTO EN LA VARIABLE RESPUESTA

$$H_0 : \mu_1 = \mu_2 = \mu_3 \dots = \mu_i$$



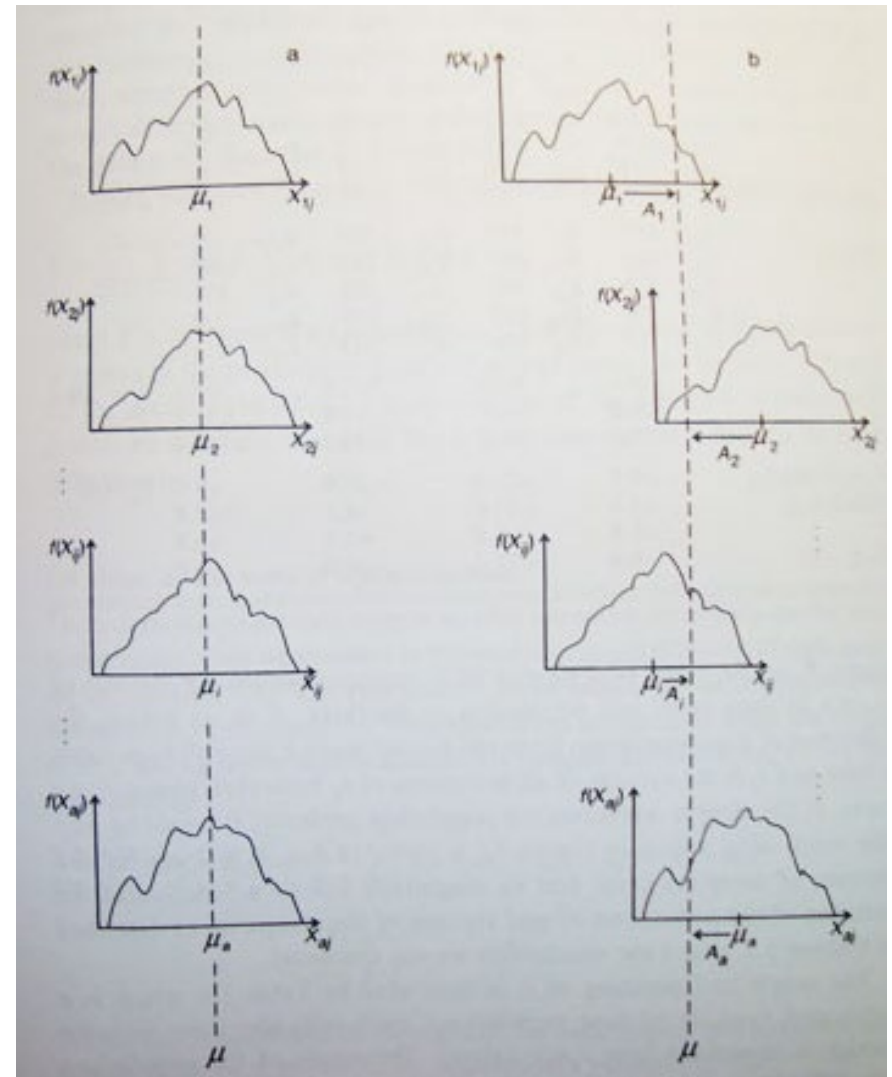
# Modelo lineal

$$H_0: y_{ij} = \mu + e_{ij}$$

$$H_i: y_{ij} = \mu + \bar{A} + e_{ij}$$



Hipótesis a evaluar



# El truco del ANOVA: descomponer la variación total

¿cuánta variación proviene de las diferencias reales? (*si es que existen*)

¿cuánta variación es producto del azar?

Variación Total = Variación entre muestras + Variación dentro de las muestras

The diagram illustrates the decomposition of total variation. Three green arrows point from the terms in the equation above to the corresponding terms in the formula below. The first arrow points from 'Variación Total' to 'SCT'. The second arrow points from 'Variación entre muestras' to 'SCE'. The third arrow points from 'Variación dentro de las muestras' to 'SCD'.

$$SCT = SCE + SCD$$

*En estadística, el azar es un comodín que incluye todo lo desconocido y no incluido en el modelo*

# SUMATORIA CUADRÁTICA TOTAL

	Grupo 1	Grupo 2	Grupo 3	Grupo a
1	$x_{11}$	$x_{21}$	$x_{31}$	$x_{a1}$
2	$x_{12}$	$x_{22}$	$x_{32}$	$x_{a2}$
3	$x_{13}$	$x_{23}$	$x_{33}$	$x_{a3}$
4	$x_{14}$	$x_{24}$	$x_{34}$	$x_{a4}$
5	$x_{15}$	$x_{25}$	$x_{35}$	$x_{a5}$
n	$x_{1n}$	$x_{2n}$	$x_{3n}$	$x_{an}$
Media	media1	media2	media3	Media a
Desv. Est	$s_1$	$s_2$	$s_3$	$s_a$

$$\overline{X} = \frac{\sum_{i=1}^a \overline{X}_i}{a} = \frac{\sum_{i=1}^a \sum_{j=1}^n x_{ji}}{an}$$

# SUMATORIA CUADRÁTICA TOTAL

	Grupo 1	Grupo 2	Grupo 3	Grupo a
1	$x_{11}$	$x_{21}$	$x_{31}$	$x_{a1}$
2	$x_{12}$	$x_{22}$	$x_{32}$	$x_{a2}$
3	$x_{13}$	$x_{23}$	$x_{33}$	$x_{a3}$
4	$x_{14}$	$x_{24}$	$x_{34}$	$x_{a4}$
5	$x_{15}$	$x_{25}$	$x_{35}$	$x_{a5}$
n	$x_{1n}$	$x_{2n}$	$x_{3n}$	$x_{an}$
Media	media1	media2	media3	Media a
Desv. Est	$s_1$	$s_2$	$s_3$	$s_a$

$$SCT = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \overline{X})^2$$

# SUMATORIA CUADRÁTICA TOTAL

	Grupo 1	Grupo 2	Grupo 3	Grupo a
1	$x_{11}$	$x_{21}$	$x_{31}$	$x_{a1}$
2	$x_{12}$	$x_{22}$	$x_{32}$	$x_{a2}$
3	$x_{13}$	$x_{23}$	$x_{33}$	$x_{a3}$
4	$x_{14}$	$x_{24}$	$x_{34}$	$x_{a4}$
5	$x_{15}$	$x_{25}$	$x_{35}$	$x_{a5}$
n	$x_{1n}$	$x_{2n}$	$x_{3n}$	$x_{an}$
Media	media1	media2	media3	Media a
Desv. Est	$s_1$	$s_2$	$s_3$	$s_a$

$$SCT = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^a (\bar{x}_i - \bar{x})^2$$

Variabilidad dentro + Variabilidad entre

Esto es cero (0) si  $H_0$  es cierta

# TABLA ANOVA

VE

$$\frac{\sum_{i=1}^a \sum_{j=1}^n (A_i - \bar{A})^2}{a - 1} + \sigma_e^2$$

VD

$$\sigma_e^2$$

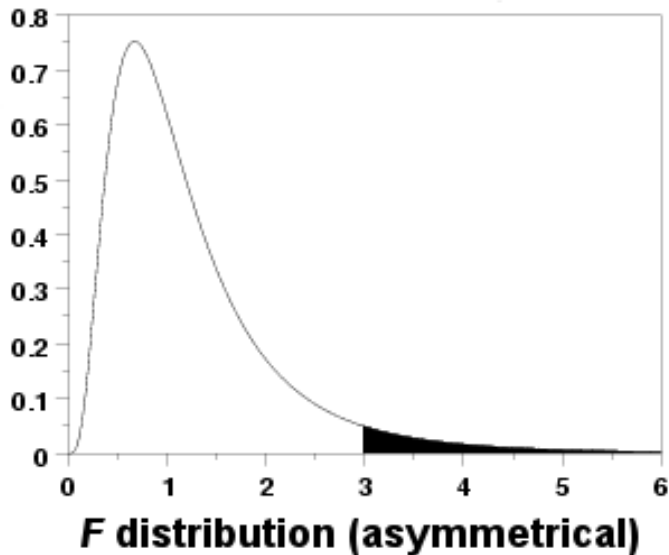
Fuente de variación	Sumatoria cuadrática	Grados de libertad	Cuadrados medios	Valor $F$	Probabilidad
Entre (modelo)	$n \sum_{i=1}^a (\bar{x}_i - \bar{\bar{x}})^2$	$a - 1$	$\frac{SC_{entre}}{gl_{entre}}$	$\frac{CM_{entre}}{CM_{dentro}}$	$\text{¿?}$
Dentro (residual)	$\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$	$a(n - 1)$	$\frac{SC_{dentro}}{gl_{dentro}}$		
Total	$\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{\bar{x}})^2$	$an - 1$			

# CRITERIO Y TABLA FISHER

Criterios:

**Fisher:** si  $p < 0.05$  se rechaza  $H_0$  (si repetimos el experimento 20 veces, el resultado se obtendría una sola vez si  $H_0$  es verdadera)

**Neyman Y Pearson:** Si el estadístico es mayor que el tabulado se rechaza  $H_0$



$$F_{\alpha}(gl_{entre}/gl_{dentro})$$

# SUPUESTOS DEL ANOVA

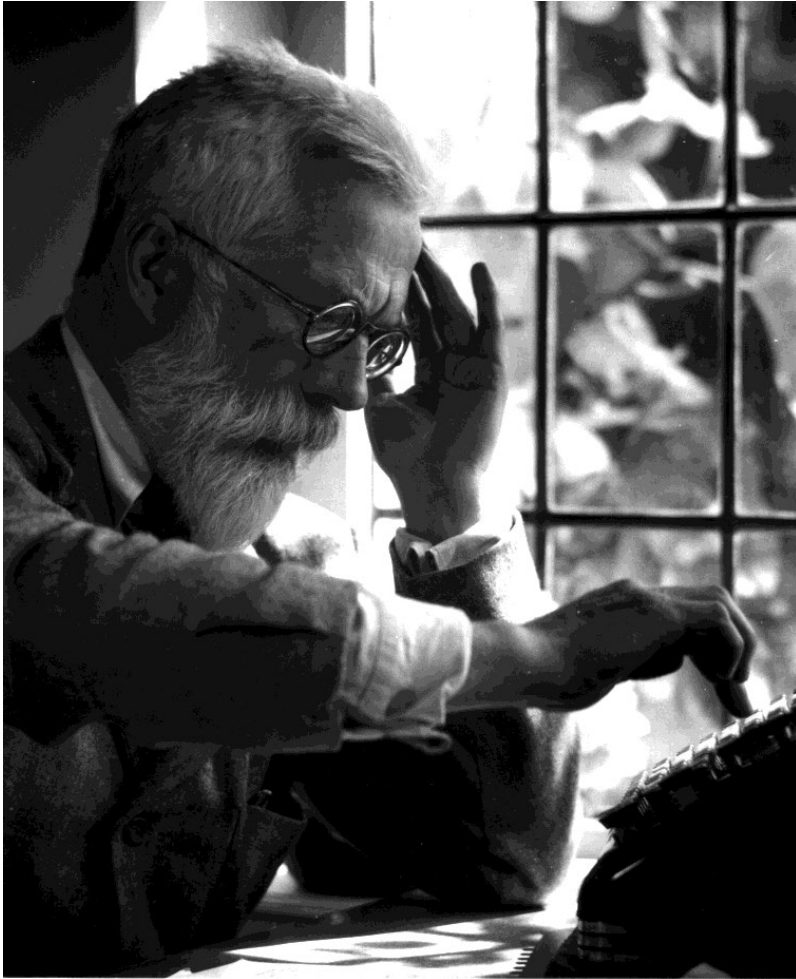
Dadas las operaciones aritméticas y las propiedades de la tabla de contraste  $F$ , se debería cumplir con:

1. Independencia entre réplicas y entre tratamientos
2. Los datos se deben aproximar a una dist. Normal
3. Homocedasticidad de las varianzas

Tema de última clase!!



# ¿Lo reconocen?



**Ronald A. Fisher**  
1890-1962

*The Design of Experiments*, 1935  
*Statistical Methods for Research Workers*, 1925

-Matemático, Biólogo teórico  
(agronomía y genética cuantitativa)

## ANOVA