

# ANOVA

- ANOVA se utiliza para probar si dos o más poblaciones tienen la misma media.
- Implica un examen de las varianzas muestrales, de allí el término *análisis de varianza*.
- Su uso se origina en la agricultura, donde se aplicaban *tratamientos* (fertilizantes) a diferentes poblaciones (parcelas) y se medía el impacto en el desempeño promedio.

## **ANOVA – modelos conceptuales**

- El *Modelo de efectos fijos* asume que los datos provienen de poblaciones normales las cuales podrían diferir únicamente en sus medias.

**Ej: grupos de empleados que reciben diferente capacitación, midiendo su productividad.**

- En el *Modelo de efectos aleatorios* los tratamientos utilizados en el estudio se seleccionan aleatoriamente de una población de tratamientos posibles.
- **Ej: La empresa tenía muchos programas de capacitación distintos, y desean saber si éstos tienen efectos diferentes en el desempeño de los empleados.**

## **ANOVA - Supuestos**

- Todas las poblaciones involucradas son normales.
- Todas las poblaciones tienen la misma varianza.
- Las muestras se seleccionan de manera independiente.

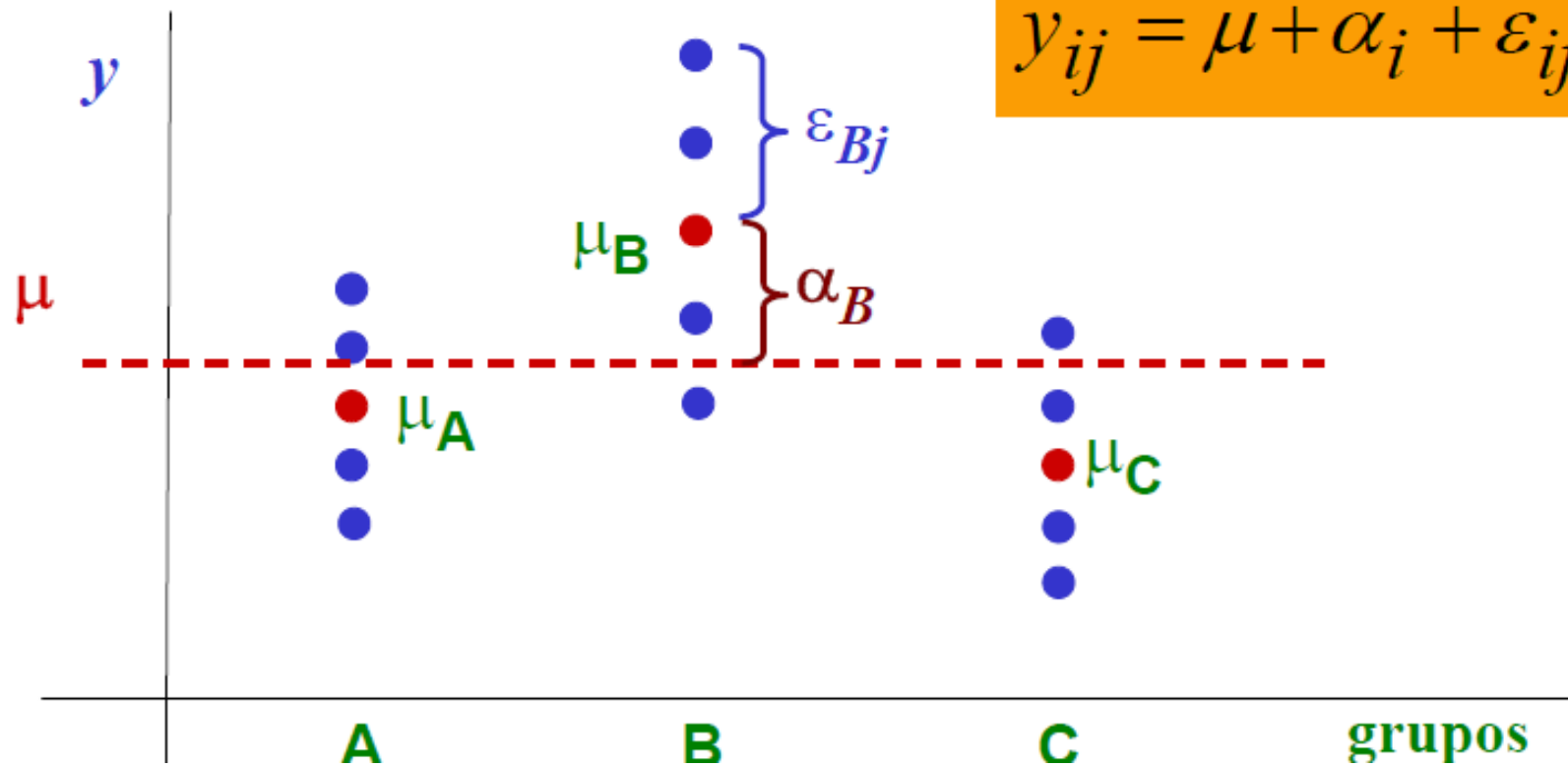
## **ANOVA una vía**

- ANOVA de una vía se utiliza para testear diferencias entre 2 o más grupos independientes.
- ANOVA se utiliza generalmente para testear diferencias entre más de grupos ya que el caso de 2 grupos está cubierto por el test t.
- Se basa en dividir la suma de los cuadrados (varianza) total en una varianza asociada a los tratamientos y otra no explicada.
- El cuociente de estas varianzas se distribuye F y si su valor es grande implica que los tratamientos tienen una influencia significativa en la varianza total.

## ANOVA una vía (2)

Modelo I – efectos fijos

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$



## ANOVA una vía (3)

$$y_{ij} - \mu = (\mu_i - \mu) + (y_{ij} - \mu_i)$$

$\alpha_i$   $\varepsilon_{ij}$  **En la población**

$$y_{ij} - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

**En la muestra**

Elevando al cuadrado:

$$(y_{ij} - \bar{y})^2 = (\bar{y}_i - \bar{y})^2 + (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y})(y_{ij} - \bar{y}_i)$$

Sumando:

$$\sum_{ij} (y_{ij} - \bar{y})^2 = \sum_{ij} (\bar{y}_i - \bar{y})^2 + \sum_{ij} (y_{ij} - \bar{y}_i)^2$$

**SC TOTAL**

**SC ENTRE  
grupos**

**SC DENTRO de  
grupos (residual)**

## ANOVA una vía (4)

- Se cumple entonces:  $SCT = SCTR + SCE$

- SCT: Suma de Cuadrados Totales

$$SCT = \sum_{i=1}^r \sum_{j=1}^c (X_{ij} - \bar{\bar{X}})^2$$

- SCTR: Suma de Cuadrados de Tratamientos ( $r_j = n^\circ$  de obs. en cada fila)

$$SCTR = \sum r_j (\bar{X}_j - \bar{\bar{X}})^2$$

- SCE: Suma de Cuadrados del Error

$$SCE = \sum \sum (X_{ij} - \bar{X}_j)^2$$

## ANOVA una vía (5)

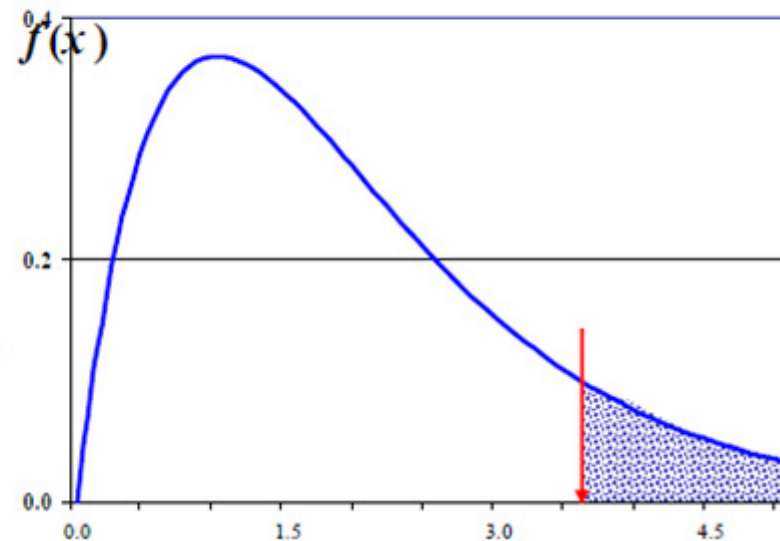
- ANOVA utiliza los cuadrados medios para construir el estadígrafo, que equivalen a la suma de los cuadrados divididos por sus grados de libertad.
- Para calcular SCT se utilizaron  $n$  observaciones y se calculó la gran media:  $n-1$  g.l.
- Para calcular SCTR se utilizaron  $c$  medias de tratamientos y se calculó la gran media:  $c-1$  g.l.
- Para calcular SCE se utilizaron  $n$  observaciones y se calcularon  $c$  medias:  $n-c$  g.l.

## ANOVA una vía (6)

| FUENTE DE VARIACION | SUMA DE CUADRADOS | GL  | MEDIA DE CUADRADOS     | Fcalc                          |
|---------------------|-------------------|-----|------------------------|--------------------------------|
| ENTRE GRUPOS        | SC entre          | c-1 | $SC\ entre / (C - 1)$  | $\frac{MC\ entre}{MC\ dentro}$ |
| DENTRO DE GRUPOS    | SC dentro         | n-c | $SC\ dentro / (n - C)$ |                                |
| TOTAL               | SC total          | n-1 |                        |                                |

El  $F_{\text{calculado}}$  se compara con

el  $F_{\text{tabulado}}$



## **ANOVA una vía (6)**

- Ej: Un gerente de Marketing de un banco desea determinar si existe una diferencia en el nivel de depósitos entre 4 sucursales. Si hubiese alguna diferencia, se propondrán políticas promocionales diferentes a nivel de sucursal, tomando estos depósitos como medida representativa del ingreso.
- Se toma una muestra de  $r=7$  depósitos en cada una de las  $c=4$  sucursales (tratamientos).

## ANOVA una vía (7)

- Ej:

| Depósito    | Sucursal 1 | Sucursal 2 | Sucursal 3 | Sucursal 4 |
|-------------|------------|------------|------------|------------|
| 1           | 5.1        | 1.9        | 3.6        | 1.3        |
| 2           | 4.9        | 1.9        | 4.2        | 1.5        |
| 3           | 5.6        | 2.1        | 4.5        | 0.9        |
| 4           | 4.8        | 2.4        | 4.8        | 1.0        |
| 5           | 3.8        | 2.1        | 3.9        | 1.9        |
| 6           | 5.1        | 3.1        | 4.1        | 1.5        |
| 7           | 4.8        | 2.5        | 5.1        | 2.1        |
| $\bar{X}_j$ | 4.87       | 2.29       | 4.31       | 1.46       |

$$\begin{aligned}
 \bar{\bar{X}} &= \frac{\sum X_{ij}}{n} \\
 &= \frac{(5.1 + 4.9 + 5.6 + \dots + 2.1)}{28} \\
 &= 3.23
 \end{aligned}$$

## ANOVA una vía (8)

- Ej: Se tiene la siguiente hipótesis nula

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H_A$ : No todas las medias son iguales

$$\begin{aligned} SCT &= \sum \sum (X_{ij} - \bar{\bar{X}})^2 \\ &= (5.1 - 3.23)^2 + (4.9 - 3.23)^2 + (5.6 - 3.23)^2 \\ &\quad + \dots + (2.1 - 3.23)^2 \\ &= 61.00 \end{aligned}$$

$$\begin{aligned} SCTR &= \sum r_j (\bar{X}_j - \bar{\bar{X}})^2 \\ &= 7(4.87 - 3.23)^2 + 7(2.29 - 3.23)^2 \\ &\quad + 7(4.31 - 3.23)^2 + 7(1.46 - 3.23)^2 \\ &= 55.33 \end{aligned}$$

## ANOVA una vía (9)

$$\begin{aligned} SCE &= \sum \sum (X_{ij} - \bar{X}_j)^2 \\ &= (5.1 - 4.87)^2 + \dots + (4.8 - 4.87)^2 \\ &\quad + (1.9 - 2.29)^2 + \dots + (2.5 - 2.29)^2 \\ &\quad + (3.6 - 4.31)^2 + \dots + (5.1 - 4.31)^2 \\ &\quad + (1.3 - 1.46)^2 + \dots + (2.1 - 1.46)^2 \\ &= 5.67 \end{aligned}$$

Para el primer tratamiento

Para el segundo tratamiento

Para el tercer tratamiento

Para el cuarto tratamiento

$$CMTR = \frac{55.33}{3}$$

$$= 18.44$$

$$CME = \frac{5.67}{24}$$

$$= 0.236$$

## ANOVA una vía (10)

| Fuente de variación          | Suma de cuadrados | Grados de libertad | Cuadrados medios | Valor $F$ |
|------------------------------|-------------------|--------------------|------------------|-----------|
| Entre muestras (tratamiento) | 55.33             | 3                  | 18.44            | 78.14     |
| Dentro de muestras (error)   | 5.67              | 24                 | 0.236            |           |
| Variación total              | 61.00             | 27                 |                  |           |

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_A$ : No todas las medias son iguales

Regla de decisión: No rechazar si  $F \leq 3.01$ . Rechazar si  $F > 3.01$

Conclusión: Debido a que  $F = 78.14 > 3.01$ , se rechaza la hipótesis nula.