ANOVA

Análisis de Varianza
(de un factor)

Objetivo

- Probar si existen diferencias significativas entre las medias de más de dos poblaciones

Supuestos

- La variable en estudio tiene distribución normal en cada una de las poblaciones bajo análisis
- Las poblaciones son independientes entre sí
- Las poblaciones son homocedásticas (tienen igual varianza)

Ejemplo de aplicación

- Se registraron los siguientes datos correspondientes a cantidades de café (en gramos) dosificadas por 4 máquinas similares.
- Se tomó una muestra de 5 cafés servidos por cada una de las máquinas.

Observación #	Máquina 1	Máquina 2	Máquina 3	Máquina 4
1	22,5	21,8	22,4	22
2	22,7	22,5	22,3	21,7
3	22,4	22,6	22,3	21,9
4	22,5	22,2	22	21,8
5	22	21,9	21,6	21,6

Procedimiento ANOVA

<u>Planteo de las hipótesis</u>

- H0 : $\mu_1=\mu_2=\mu_3=\mu_4$ (No existen diferencias significativas entre las dosificaciones medias de café en cada una de las máquinas)
- H1 : $\mu_i \neq \mu_j$ $(i \neq j)$ (Alguna de las máquinas tiene una dosificación media de café significativamente distinta a las otras máquinas)

- La variable en estudio presenta dos tipos de variabilidad:

Una variabilidad producida ENTRE cada máquina (Grupo o Tratamiento)

Y

Una variabilidad producida DENTRO de cada máquina (Grupo o Tratamiento)

- Entonces bajo el supuesto de homocedasticidad de las poblaciones, para que la diferencia entre las medias poblacionales sea significativa, la variabilidad ENTRE los grupos o tratamientos debe ser significativamente superior a la variabilidad DENTRO de los grupos.
- Esto indicaría que los grupos no se comportan homogéneamente y por tanto deberían tener medias diferentes.

ESTADISTICO DE PRUEBA

$$F = \frac{\dfrac{\sum_{i=1}^{p} n_i \cdot \left(\bar{X}_i - \bar{X}\right)^2}{p-1}}{\dfrac{\sum_{i=1}^{p} \sum_{t=1}^{n} \left(X_{it} - \bar{X}_i\right)^2}{N-p}} = \frac{\dfrac{SCE}{p-1}}{\dfrac{SCD}{N-p}} = \frac{CME}{CMD}$$

$$F = \frac{\text{promedio de los cuadrados ("entre")}}{\text{promedio de los cuadrados ("dentro")}}$$

N : Tamaño total de la muestra
p : # de grupos o tratamientos
n_i : Tamaño de cada grupo
$\bar{X}$: Media de toda la muestra
$\bar{X}_i$: Media de cada grupo

Distribución probabilística del estadístico

- Por tratarse de un cociente de dos varianzas, la distribución será una F de Fisher-Snedecor

$$F_{(p - 1\,;\, N - p)}$$

Condición de Rechazo de la H0

- Si $F \geq F_{(p-1\,;N-p\,;1-\alpha)}$

(Sólo se obtiene el valor crítico unilateral derecho, pues la H0 se rechazará sólo si la variabilidad entre los grupos es significativamente superior a la variabilidad dentro de los grupos)

Datos y cálculos

- N = 20 observaciones ; p = 4 grupos ; $n_1=n_2=n_3=n_4=5$ observaciones en cada grupo

Estimación puntual de las medias de los grupos			
$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
22,42	22,2	22,12	21,8

$$\boldsymbol{Media\ de\ toda\ la\ muestra}$$
$$\boldsymbol{\bar{X} = 22,135}$$

- **SCE** = $5 * (22{,}42-22{,}135)^2 + 5 * (22{,}2-22{,}135)^2 + 5 * (22{,}12-22{,}135)^2 + 5 * (21{,}8-22{,}135)^2 =$ **0,9895**

- **SCD** =

Obs #	Máq 1	Máq 2	Máq 3	Máq 4		
1	22,5	21,8	22,4	22		
2	22,7	22,5	22,3	21,7		
3	22,4	22,6	22,3	21,9		
4	22,5	22,2	22	21,8		
5	22	21,9	21,6	21,6		
$\bar{X}_i$	22,42	22,2	22,12	21,8		
	0,0064	0,16	0,0784	0,04		
	0,0784	0,09	0,0324	0,01		
$(X_{it} - \bar{X}_i)^2$	0,0004	0,16	0,0324	0,01		
	0,0064	0	0,0144	0		
	0,1764	0,09	0,2704	0,04	Suma de las sumas	
Sumas	0,268	0,5	0,428	0,1	1,296	

Cálculo del promedio de los cuadrados

- $CME = \frac{SCE}{p-1} = \frac{0,9895}{4-1} = 0,329833$

- $CMD = \frac{SCD}{N-p} = \frac{1,296}{20-4} = 0,081$

Cálculo del estadístico de prueba

- $F = \frac{CME}{CMD} = \frac{0,329833}{0,081} = 4,072$

Cálculo del Fractil crítico del estadístico F

- $\mathbf{F_{(p-1\,;N-p;1-\alpha)} = F_{(3\,;16\,;0,95)} = 3,239}$

Conclusión

- Como F = 4,072 es mayor a Fcrítico = 3,239 entonces

  se Rechaza H0 y se concluye que :

  Existe al menos una máquina de café con una dosificación media distinta al resto de las máquinas

  Esta prueba no explora cuál o cuáles de los grupos serían los que presentan tales diferencias respecto del conjunto.

MUCHAS GRACIAS !