Bioestadística y uso de software científico

TEMA 8
ANOVA FACTORIAL
ANOVA DE MEDIDAS REPETIDAS

Hasta ahora…

Tema	**Variable dependiente**	**Variable independiente**	**Test**
Tema 4	Categórica	Categórica	χ^2, McNemar
Tema 5	Continua	Dicotómica	t de Student U de Mann-Whitney
Tema 7	Continua	Categórica (>2 categorías)	ANOVA de una vía Kruskal-Wallis
Tema 8	Continua	Categóricas (dos variables)	ANOVA de dos vías

Algunos ejemplos

- Se comparan tres tratamientos (A, B y C) para el control de la tensión arterial. Se quiere saber si alguno de ellos es más eficaz.
  - Sólo una variable (tratamiento)
  - ANOVA de una vía

$$H_0 : \mu_A = \mu_B = \mu_C$$

Algunos ejemplos

- Se comparan tres tratamientos (A, B y C) para el control de la tensión arterial. Se quiere saber si alguno de ellos es más eficaz <u>y si el sexo del paciente influye en la eficacia</u>
  - Dos variables (tratamiento y sexo)
  - ANOVA de dos vías

$$H_0 : \mu_A = \mu_B = \mu_C$$

$$H_0 : \mu_{\text{varones}} = \mu_{mujeres}$$

$$H_0 : (\mu_A = \mu_B = \mu_C)_{\text{varones}} = (\mu_A = \mu_B = \mu_C)_{\text{mujeres}}$$

Anova de 2 vías

- Tres hipótesis nulas: influencia en la tensión arterial de:
  - El tratamiento:

$$H_0 : \mu_A = \mu_B = \mu_C$$

  - El sexo:

$$H_0 : \mu_{\text{varones}} = \mu_{mujeres}$$

  - La interacción tratamiento-sexo

$$H_0 : (\mu_A = \mu_B = \mu_C)_{\text{varones}} = (\mu_A = \mu_B = \mu_C)_{\text{mujeres}}$$

Tabla del ANOVA de una vía

Fuente de variación	Suma de cuadrados	g.l.	Varianza	F
Tratamiento	$SCE=SCT-SCR$	$k-1$	$V_e=SCE/(k-1)$	V_e/V_r
Residual	$SCR=\Sigma(n_i-1)s_i^2$	$n-k$	$V_r=SCR/(n-k)$	
Total	$SCT=(n-1)s^2$	$n-1$		

Tabla del ANOVA de dos vías

Fuente de variación	**Suma de cuadrados**	**g.l.**	**Varianza**	**F**
Tratamiento	SCTratamiento	t-1	$V_{tratamiento}$	V_t/V_r
Sexo	SCSexo	s-1	V_{sexo}	V_s/V_r
Interacción	SCInteracción	(t-1)(s-1)	$V_{interacción}$	V_i/V_r
Residual	SCResidual		$V_{residual}$	
Total	SCTotal	n-1		

ANOVA de dos vías (factorial)

- Datos necesarios:

Tratamiento	Varones	Mujeres	Total
A	n=20 m=140	n=25 m=135	n=45 m=137
B	n=25 m=135	n=20 m=130	n=45 m=133
C	n=23 m=155	n=25 m=140	n=48 m=147
Total	n=68 m=143	n=70 m=135	n=138 m=139

$$s_{total}^2 = 160$$

ANOVA de dos vías (factorial)

1. Calcular la suma de cuadrados debida al tratamiento ($SC_{tratamiento}$)

$$SC_{tratamiento} = \sum n_i (m_i - m_{total})^2$$

$$SC_{tratamiento} = n_A (m_A - m_{total})^2 + n_B (m_B - m_{total})^2 + n_C (m_C - m_{total})^2$$

$$SC_{tratamiento} = 45(137-139)^2 + 45(133-139)^2 + 48(147-139)^2 = 4872$$

ANOVA de dos vías (factorial)

2. Calcular la suma de cuadrados debida al sexo (SC_{sexo})

$$SC_{sexo} = \sum n_j (m_j - m_{total})^2$$

$$SC_{sexo} = n_{\mathrm{varones}} (m_{\mathrm{varones}} - m_{total})^2 + n_{mujeres} (m_{mujeres} - m_{total})^2$$

$$SC_{sexo} = 68(143 - 139)^2 + 70(135 - 139)^2 = 2208$$

ANOVA de dos vías (factorial)

3. Calcular la suma de cuadrados debida a la interacción ($SC_{\text{interacción}}$)

$$SC_{\text{interacción}} = \sum n_{ij}(m_{ij} + m_{total} - m_i - m_j)^2$$

$$SC_{\text{interacción}} = 20(140+139-143-137)^2 + 25(135+139-135-137)^2 +$$
$$+25(135+139-143-133)^2 + 20(130+139-135-133)^2 +$$
$$23(155+139-143-147)^2 + 25(140+139-135-147)^2 = 833$$

ANOVA de dos vías (factorial)

4. Calcular la suma de cuadrados totales (SC_{total})

$$SC_{\text{total}} = (n-1)s_{total}^2$$

$$SC_{\text{total}} = (138-1)\times 160 = 21920$$

ANOVA de dos vías (factorial)

5. Calcular la suma de cuadrados residual ($SC_{residual}$)

$$SC_{\text{residual}} = SC_{total} - SC_{tratamientos} - SC_{sexo} - SC_{\text{interacción}}$$

$$SC_{\text{residual}} = 21920 - 4872 - 2208 - 833 = 14007$$

ANOVA de dos vías (factorial)

6. Calcular los grados de libertad

Fuente de variación	Categorías	Grados de libertad
Tratamiento	3-1	2
Sexo	2-1	1
Interacción	(3-1)(2-1)	2
Residual	Total-t-s-inter	132
Total	n-1	137

ANOVA de dos vías (factorial)

7. Calcular las varianzas (SC/gl) y construir la tabla de ANOVA

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianza
Tratamiento	4872	2	2436
Sexo	2208	1	2208
Interacción	833	2	416,5
Residual	14007	132	106,1
Total	21920	137	

ANOVA de dos vías (factorial)

8. Calcular F (Varianza del factor / Varianza residual)

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianza	F
Tratamiento	4872	2	2436	23
Sexo	2208	1	2208	20,8
Interacción	833	2	416,5	3,9
Residual	14007	132	106,1	
Total	21920	137		

ANOVA de dos vías (factorial)

9. Buscar los valores p en la tabla F

$$F_{tratamiento} = F_{2,132} = 23 \rightarrow p < 0,001$$

$$F_{sexo} = F_{1,132} = 20,8 \rightarrow p < 0,001$$

$$F_{\text{interacción}} = F_{2,132} = 3,9 \rightarrow p = 0,02$$

ANOVA con medidas repetidas

ANOVA con medidas repetidas

- Se quiere conocer la evolución de la tensión arterial en 30 sujetos. Para ello, se les toma la tensión al iniciar el tratamiento (medida 1), a los 6 meses (medida 2) y al año (medida 3).

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

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

ANOVA con medidas repetidas

- Se quiere conocer la evolución de la tensión arterial en 30 sujetos. Para ello, se les toma la tensión al iniciar el tratamiento (medida 1), a los 6 meses (medida 2) y al año (medida 3).

- Es similar a la t de Student con datos emparejados. Pero la t vale sólo para dos medidas y el ANOVA vale para >2

ANOVA con medidas repetidas

- Datos necesarios:

$$Media\ total : m_{total}$$

$$Varianza\ total : s^2_{total}$$

$$Media\ de\ cada\ medición : m_1, m_2, m_3$$

$$Media\ de\ las\ 3\ mediciones\ en\ cada\ sujeto :$$

$$m_{\_1}, m_{\_2}, \ldots, m_{\_30}$$

ANOVA con medidas repetidas

1. Calcular la suma de cuadrados entre las medidas ($SC_{medidas}$):

$$SC_{medidas} = n\sum(m_i - m_{total})^2$$

$$SC_{medidas} = 30 \times \left[(m_1 - m_{total})^2 + (m_2 - m_{total})^2 + (m_3 - m_{total})^2 \right]$$

ANOVA con medidas repetidas

2. Calcular la suma de cuadrados entre los sujetos ($SC_{sujetos}$):

$$SC_{sujetos} = k\sum(m_{\_i} - m_{total})^2$$

$$SC_{sujetos} = 3\times\left[(m_{\_1} - m_{total})^2 + (m_{\_2} - m_{total})^2 + \cdots + (m_{\_30} - m_{total})^2\right]$$

ANOVA con medidas repetidas

3. Calcular la suma de cuadrados total (SC_{total}):

$$SC_{total} = (kn - 1)s_{total}^2$$

ANOVA con medidas repetidas

4. Calcular la suma de cuadrados residual ($SC_{residual}$):

$$SC_{residual} = SC_{total} - SC_{mediciones} - SC_{sujetos}$$

ANOVA con medidas repetidas

5. Calcular los grados de libertad:

$$g.l._{total} = kn - 1 = 3 \times 30 - 1 = 89$$

$$g.l._{mediciones} = k - 1 = 3 - 1 = 2$$

$$g.l._{sujetos} = n - 1 = 30 - 1 = 29$$

$$g.l._{residual} = \left(kn - 1\right) - (k - 1) - (n - 1) = 58$$

ANOVA con medidas repetidas

6. Calcular las varianzas (suma de cuadrados/g.l.) y construir la tabla ANOVA:

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianza
Mediciones	$SC_{mediciones}$	k-1	$Var_{mediciones}$
Sujetos	$SC_{sujetos}$	n-1	
Residual	$SC_{residual}$	(kn-1)-(k-1)-(n-1)	$Var_{residual}$
Total	SC_{total}	kn-1	

ANOVA con medidas repetidas

7. Calcular F dividiendo las varianzas, y buscar el valor p en la tabla F con k-1,(kn-1)-(k-1)-(n-1) grados de libertad:

$$F_{k-1,(kn-1)-(k-1)-(n-1)} = \frac{Varianza_{medidas}}{Varianza_{residual}}$$

Test de Friedman

- Cuando no se cumplen las condiciones del ANOVA
  - Variable dependiente con distribución normal
  - Homocedasticidad (varianzas homogéneas)

- Es necesario utilizar un método no paramétrico:
  - Test de Friedman=método no paramétrico en lugar del ANOVA de medidas repetidas
  - Test de Kruskal-Wallis=método no paramétrico en lugar del ANOVA de una vía