Bioestadística y uso de software científico

TEMA 7
COMPARACIONES DE MEDIAS:
3 O MÁS GRUPOS

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

Introducción al ANOVA

- 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.

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

$$H_1 : \text{alguna de las medias es distinta}$$

Introducción al ANOVA

- No se puede utilizar la t de Student porque sólo sirve para comparar 2 medias.
- No es correcto hacer tres comparaciones:

$$H_0 : \mu_A = \mu_B$$

$$H_0 : \mu_A = \mu_C$$

$$H_0 : \mu_B = \mu_C$$

- Porque si cada una tiene α=0,05, en total saldrá α=0,15

Introducción al ANOVA

- ANOVA = **AN**alysis **O**f **Va**riance
- NO compara las varianzas
- Utiliza las varianzas para hacer las comparaciones de medias

Cálculo de ANOVA

- 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.

1. Calcular las medias y varianzas de cada grupo y del total

Tratamiento	Número de pacientes	Media	Varianza
A	n_A	m_A	s_A^2
B	n_B	m_B	s_B^2
C	n_C	m_C	s_C^2
Total	**n**	**m**	s^2

Cálculo de ANOVA

2. Calcular la suma de cuadrados total (SCT)

$$SCT = (n-1) \times s^2$$

Cálculo de ANOVA

3. Calcular la suma de cuadrados intragrupo o residual (SCR)

$$SCR = \sum (n_i - 1) \times s_i^{\,2}$$

$$SCR = (n_A - 1) \times s_A^{\,2} + (n_B - 1) \times s_B^{\,2} + (n_C - 1) \times s_C^{\,2}$$

Cálculo de ANOVA

4. Calcular la suma de cuadrados entre grupos (SCE)

$$SCE = SCT - SCR$$

Cálculo de ANOVA

6. Calcular varianzas
   - Varianza residual = SCR/n-k
   - Varianza entre grupos = SCE/k-1

Cálculo de ANOVA

7. Calcular

$$F = \frac{Varianza_{entre\ grupos}}{Varianza_{residual}}$$

8. Buscar el valor p en la tabla F con k-1, n-k grados de libertad

Cálculo de ANOVA

- 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.

1. Calcular las medias y varianzas de cada grupo y del total

Tratamiento	Número de pacientes	Media	Varianza
A	30	140	100
B	40	135	90
C	35	155	110
Total	**105**	**143**	**108**

Cálculo de ANOVA

2. Calcular la suma de cuadrados total (SCT)

$$SCT = (n-1)\times s^2 = 104\times 108 = 11232$$

Cálculo de ANOVA

3. Calcular la suma de cuadrados intragrupo o residual (SCR)

$$SCR = \sum (n_i - 1) \times s_i^{\,2}$$

$$SCR = (n_A - 1) \times s_A^{\,2} + (n_B - 1) \times s_B^{\,2} + (n_C - 1) \times s_C^{\,2} =$$

$$29 \times 100 + 39 \times 89 + 34 \times 110 = 10250$$

Cálculo de ANOVA

4. Calcular la suma de cuadrados entre grupos (SCE)

$$SCE = SCT - SCR =$$

$$= 11232 - 10250 = 982$$

Cálculo de ANOVA

6. Calcular varianzas
   - Varianza residual = SCR/n-k
   - Varianza entre grupos = SCE/k-1

$$Var_{residual} = \frac{SCR}{n-k} = \frac{10250}{105-3} = 100,5$$

$$Var_{entre\ grupos} = \frac{SCE}{k-1} = \frac{982}{3-1} = 491$$

Cálculo de ANOVA

7. Calcular

$$F = \frac{Varianza_{entre\ grupos}}{Varianza_{residual}} = \frac{491}{100,5} = 4,89$$

8. Buscar el valor p en la tabla F con k-1, n-k grados de libertad

$$F_{2,103} = 4,89 \rightarrow p = 0,009$$

Relación entre ANOVA y t de Student

- Si sólo hay dos medias para comparar, el resultado de ANOVA es igual al resultado de t de Student

Tabla del ANOVA

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianza	F
Entre grupos	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

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianza	F
Entre grupos	SCE=982	2	V_e=491	4,89
Residual	SCR=10250	102	V_r=100,5	
Total	SCT=11232	104		

ANOVA: segundo ejemplo

Grupo	N	Media	Varianza
1	15	25.93915	8.332712
2	17	26.58824	5.334213
3	18	27.62018	4.299075
Total	50	26.76501	6.102354

Tabla del ANOVA (segundo ejemplo)

Fuente de variación	Suma de cuadrados	g.l.	Varianza	F
Entre grupos	SCE=23,9257	2	V_e=11,96285	2,04
Residual	SCR=275,08965	47	V_r=5,8529713	
Total	SCT=299,01535	49		

Condiciones de aplicación del ANOVA

- Aplicar ANOVA requiere:
  - Una variable es cuantitativa
  - La otra variable es categórica
  - La variable cuantitativa es normal
  - Las varianzas de la variable cuantitativa en cada categoría de la otra variable son homogéneas (homocedasticidad)

Condiciones de aplicación del ANOVA

- Aplicar ANOVA requiere:
  - Una variable es cuantitativa: tensión arterial
  - La otra variable es categórica: medicamento
  - La tensión arterial es normal
  - La varianza de la tensión arterial en los que toman el medicamento A es similar a la de los que toman el B o el C

Contrastes a posteriori: comparaciones múltiples

- En los ejemplos anteriores se han comparado las medias de 3 medicamentos
- El resultado indica que alguno de los medicamentos funciona mejor que los otros, pero no dice cuál
- A continuación se pueden comparar los medicamentos dos a dos:
  - A frente a B
  - A frente a C
  - B frente a C

Contrastes a posteriori: comparaciones múltiples

- Al aumentar el número de comparaciones, se incrementa el error α:
  - Cada comparación tiene $\alpha=0,05$
  - Tres comparaciones tienen $\alpha=0,05\times3=0,15$
- Para evitarlo, se aplica una corrección

- La más sencilla de utilizar es la corrección de Bonferroni
- Es una corrección demasiado exigente

Contrastes a posteriori: comparaciones múltiples

- Corrección de Bonferroni
  - Con 1 comparación, se considera significativo si $p<0,05$
  - Con 3 comparaciones, se considera significativo si $p<0,05/3=0,0167$
  - En general, con n comparaciones, se considera significativo si $p<0,05/n$

Contrastes a posteriori: comparaciones múltiples

- Corrección de Bonferroni
  - El número posible de comparaciones entre k tratamientos es $k \times (k-1)/2$

Número de grupos	Número de comparaciones	p penalizada con el método de Bonferroni
2	1	0,05/1=0,05
3	3	0,05/3=0,0167
4	6	0,05/6=0,0083
5	10	0,05/10=0,005

Contrastes a posteriori: comparaciones múltiples

- Se comparan tres tratamientos (A, B y C) para el control de la tensión arterial.

Tratamiento	Número de pacientes	Media	Varianza
A	30	140	100
B	40	135	90
C	35	155	110
Total	105	143	108

$p=0,009$

- ¿Cuál de los 3 tratamientos difiere de los otros?

Contrastes a posteriori: comparaciones múltiples

Tratamientos a comparar	t de Student	p	p penalizada
A frente a B	2,13	0,036	0,05/3=0,0167
B frente a C	8,67	<0,0001	0,05/3=0,0167
A frente a C	5,87	<0,0001	0,05/3=0,0167

B frente a C y A frente a C tiene p<p penalizada⟶SÍ hay diferencias significativas

A frente a B tiene p>p penalizada⟶NO hay diferencias significativas

Test de Kruskal-Wallis

- ANOVA es un método paramétrico

- Si no se cumplen las condiciones, hay que usar un método no paramétrico: test de Kruskal-Wallis

Test de Kruskal-Wallis

14 pacientes reciben los fármacos A, B o C para tratar la tensión arterial. Los resultados después del tratamiento son:

Medicamento A	Medicamento B	Medicamento C
3,5	-4	0
3	-4,5	-0,5
2,5	-5	-1
0	-5,5	-31
-2	-7	

Test de Kruskal-Wallis

1. Ordenar los datos

Valor	Tratamiento	Orden	Valor	Tratamiento	Orden
-31	C	1	-1	C	8
-7	B	2	-0,5	C	9
-5,5	B	3	0	A	10
-5	B	4	0	C	11
-4,5	B	5	2,5	A	12
-4	B	6	3	A	13
-2	A	7	3,5	A	14

Test de Kruskal-Wallis

2. Sumar los órdenes de cada tratamiento

Valor	Tratamiento	Orden	Valor	Tratamiento	Orden
-31	C	1	-1	C	8
-7	B	2	-0,5	C	9
-5,5	B	3	0	A	10,5
-5	B	4	0	C	10,5
-4,5	B	5	2,5	A	12
-4	B	6	3	A	13
-2	A	7	3,5	A	14

R(A)=7+10,5+12+13+14=56,5

R(B)=2+3+4+5+6=20

R(C)=1+8+9+10,5=28,5

R(total)=105

Test de Kruskal-Wallis

3. Calcular la media de los órdenes de cada tratamiento

$R(A)=7+10{,}5+12+13+14=56{,}5$

$R(B)=2+3+4+5+6=20$

$R(C)=1+8+9+10{,}5=28{,}5$

$R(total)=105$

$Rmed(A)=56{,}5/5=11{,}3$

$Rmed(B)=20/5=4$

$Rmed(C)=28{,}5/4=7{,}125$

$Rmed(total)=105/14=7{,}5$

Test de Kruskal-Wallis

4. Calcular

$$\chi^2_{k-1} = \frac{\sum n_i \left(Rmed_i - Rmed_{total} \right)^2}{n(n+1)/12}$$

$$\chi^2_{3-1} = \frac{5\times\left(11,3-7,5\right)^2 + 5\times(4-7,5)^2 + 4\times(7,125-7,5)^2}{14\times15/12} = 7,7$$

$$\chi^2_2 = 7,7 \rightarrow p = 0,02$$

Ejercicios de elección del tipo de test

Elección del tipo de test

- Se quiere conocer si los portadores de una mutación en el gen NAT-1 tienen mayor riesgo de cáncer de colon que los no portadores

Elección del tipo de test

- Se quiere conocer si los portadores de una mutación en el gen NAT-1 tienen mayor nivel de colesterol en sangre que los no portadores

Elección del tipo de test

- Se quiere conocer si los portadores de una mutación en el gen NAT-1 tienen distinto nivel de colesterol en sangre que los no portadores

Elección del tipo de test

- ¿Influye el grupo sanguíneo ABo en la aparición de cáncer de estómago?

Elección del tipo de test

- ¿Influye el grupo sanguíneo ABo en el nivel de colesterol en sangre?

Elección del tipo de test

- ¿El nivel medio de colesterol en la sangre de los alumnos de la facultad es diferente de 180 mg/dL?

Elección del tipo de test

- Para conocer si el cáncer de próstata tiene un componente genético, se estudian 1000 gemelos. Se averigua si cuando un gemelo tiene cáncer de próstata, el otro lo tiene también o no.

Estimación de una media

- En una muestra de 150 alumnos de la Facultad de Medicina, la altura media fue 170,0 cm, con desviación estándar 20 cm. La altura media de los alumnos de la Universidad es 173,0. ¿Es menor la altura de los alumnos de Medicina?

$$H_0: \mu = 173$$

$$H_1: \mu < 173$$

$$Z = (m - 173) / \text{Error estándar} = (170 - 173) / 1{,}63 = -1{,}84$$

$$P = 0{,}032$$

T de Student para comparar una media con un valor de referencia

- H_o: $\mu=\mu_o$
- Calcular:

$$t = \frac{m - \mu_0}{s / \sqrt{n}}$$

- Buscar *t* en la tabla t de Student con n-1 grados de libertad

Tabla t de Student

p =	0.60	0.70	0.80	0.85	0.90	0.950	0.975	0.990	0.995	0.999
k=1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	31.821	63.657	318.309
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893
35	0.255	0.529	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.528	0.850	1.049	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.528	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
80	0.254	0.526	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
100	0.254	0.526	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160
Inf	0.253	0.524	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090

Elección del tipo de test

- Se quiere comercializar un nuevo medicamento (X) para tratar la diabetes; para evaluar su eficacia, se compara con el medicamento Y.
- A 200 pacientes se les mide la glucemia basal. A continuación, unos reciben X y otros reciben Y. Al cabo de un tiempo se vuelve a medir la glucemia basal.
- ¿Y si en vez de 200 pacientes solo hubieran sido 20?

Elección del tipo de test

- Se dispone de cuatro tratamientos para la cardiopatía dilatada.
- A 250 pacientes se les organiza en cuatro grupos y cada uno recibe un tratamiento. Al cabo de un año se mide la fracción de eyección del ventrículo izquierdo.

- ¿Y si en vez de 250 pacientes solo hubieran sido 30?

Elección del tipo de test

- Se dispone de cuatro tratamientos para la cardiopatía dilatada.
- A 250 pacientes se les organiza en cuatro grupos y cada uno recibe un tratamiento. Al cabo de un año se mide la fracción de eyección del ventrículo izquierdo.
- Se quiere saber si los cuatro tratamientos son igual de eficaces y si el sexo interfiere en su eficacia.

Elección del tipo de test

- Se dispone de cuatro tratamientos para la cardiopatía dilatada.
- A 250 pacientes se les mide la FEVI; a continuación, se organizan en cuatro grupos y cada uno recibe un tratamiento. Al cabo de un año se vuelve a medir la FEVI.
- ¿Y si en vez de 250 pacientes solo hubieran sido 30?