COMPARACION DE TRATAMIENTOS CON UN CONTROL

HARMIN PARRA CASTELLANOS
Profesor Asistente U. N.

INTRODUCCION

Un problema frecuente en investigación aplicada es el de comparación de tratamientos con un control. Esta situación puede surgir por ejemplo en procesos industriales, agronomía, farmacología, etc.

Consideraremos solamente al modelo lineal de efectos fijos :

$$x_{ij} = \mu + \epsilon_j + e_{ij}, \quad i = 1,2, \ldots, n_j = n, \quad j = 1, 2, \ldots, k, \quad (1)$$

en el caso paramétrico y en el no paramétrico; siendo en cualquier caso :

x_{ij} = La observación número i usando el tratamiento j,

μ = La media común a todas las observaciones,

ϵ_j = Al efecto del tratamiento j, y $\sum_{j=1}^{k} \epsilon_j = 0$.

Además, las e_{ij} = Variables aleatorias no observables, mutuamente independientes y distribuidas normalmente con media cero y varianza σ^2, mientras que si el caso es no paramétrico, las e_{ij} = variables aleatorias no observables mutuamente independientes provenientes de la misma población continua.

Para cualquier caso, la hipótesis de interés será la igualdad de los efectos de los tratamientos:

$$H_o: \epsilon_1 = \epsilon_2 = \ldots = \epsilon_k \qquad (2)$$

donde para simplicidad ϵ_1 representará al tratamiento control.

PROCEDIMIENTO

Suponga que el modelo (1) será usado en un experimento con diseño completamente aleatorizado [2], entonces :

a) Caso paramétrico .

La hipótesis (2) es negada o aceptada segun sea el valor de la estadística

$$F = [n \sum_{j=1}^{k} (\bar{x}_{.j} - \bar{x}_{..})^2 / (k-1)] / [\sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{.j})^2 / (nk-k)] \qquad (3)$$

donde $\bar{x}_{.j} = \sum_{i=1}^{n} x_{ij} / n$ y $\bar{x}_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij} / (nk)$

Dado el nivel de significación α, H_o será rechazado si $F > F_{\alpha;k-1, nk-k}$, en caso contrario la aceptaremos y en tal caso nuestra conclusión es que los tratamientos tienen el mismo efecto. Si H_o es rechazada el problema siguiente consistirá en ver si $\epsilon_u \neq \epsilon_1$, $u = 2,3, \ldots, k$ y en qué sentido se da la desigualdad si existe. *Para este propósito se plantean las hipótesis :* $H_{1,0}$: $\epsilon_2 - \epsilon_1 = 0$, $H_{2,0}$: $\epsilon_3 - \epsilon_1 = 0, \ldots, H_{k-1,0}: \epsilon_k - \epsilon_1 = 0$ y usamos el método de Scheffé [3] para establecer las siguientes decisiones al nivel de significación α:

i) *Acepte todas las hipótesis* $\epsilon_j - \epsilon_1 = 0$ *para las que* $0 \in [a_j, b_j]$, $j = 1, 2, \ldots, k$, siendo

$$a_j = (x_{.j} - x_{.1}) - \sqrt{2/n\,(k-1) S^2 F_{\alpha;k-1, nk-k}}\ ,\ b_j = (\bar{x}_{.j} - \bar{x}_{.1}) + \sqrt{2/n\,(k-1) S^2 F_{\alpha;k-1, nk-k}}$$

$$y\ \ S^2 = \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{.j})^2 / (nk-k)$$

ii) Haciendo $\bar{x}_{\cdot j}-\bar{x}_{\cdot 1} = k_{j1}$ y a $\sqrt{2/n\,(k-1)S^2 F_{\alpha;k-1,\,nk-k}} = K_{j2}$, para aquellas hipótesis $\epsilon_j-\epsilon_1$ en las que $0 \notin [a_j, b_j]$ acepte la hipótesis $\epsilon_j > \epsilon_1$ (el efecto del tratamiento j es mayor que el tratamiento control) si $K_{j1} > K_{j2}$, o, a la hipótesis $\epsilon_j < \epsilon_1$ (El efecto del tratamiento j es menor que el del tratamiento control) si $k_{j1} < -k_{j2}$.

b) Caso no paramétrico.

Ilustraremos un procedimiento exacto para el caso $n_1 = n_2 = \ldots = n_k = n = 2$ e incluiremos aproximaciones para nuestras grandes que pueden usarse para configuraciones $(n_1, n_2, \ldots, n_k)$ cualesquieras; para tal propósito ordenaremos las observaciones conjuntamente de menor a mayor. Sea r_{ij} el rango de x_{ij} en este ordenamiento conjunto, entonces para $j = 1, 2, \ldots, k$ definimos a :

$R_j = \sum_{i=1}^{n_j=2} r_{ij}$, $R_{\cdot j} = R_j/n_j$, $y^*(\alpha, k-1,n)$ tal que $P((R_u-R_1) < y^*(\alpha, k-1,n)$, $u = 2, 3, \ldots, k) = 1-\alpha$, y a $y^{**}(\alpha, k-1,n)$ tal que $P(|R_u-R_1| < y^{**}(\alpha,k-1,n)$, $u = 2, 3, \ldots, k) = 1-\alpha$

Seguidamente procederemos como sigue :

1. Calculamos las k-1 diferencias $R_u - R_1$, $u = 2, 3, \ldots, k$

2. Para prueba a dos ramas aceptaremos que $\epsilon_u \neq \epsilon_1$, si $|R_u-R_1| > y^{**}(\alpha, k-1, n)$, (3)

3. Para pruebas a una rama aceptaremos que $\epsilon_u > \epsilon_1$ si

$$R_u - R_1 \not< y^*(\alpha, k-1, n) \qquad (4)$$

o que $\epsilon_u < \epsilon_1$ si $R_1 - R_u \not< y^*(\alpha, k-1,n)$.

Puesto que el valor de $y^*(\alpha, k-1,n)$ o $y^{**}(\alpha, k-1,n)$ hay que encontrarlo, segun sea el caso, ilustraremos a continuación la forma de hallar $y^*(\alpha,k-1,n)$ y $y^{**}(\alpha,k-1,n)$ para $k=3$ y $n=2$; y así, proceder en el evento de carecer de tablas.

Sea k=3 y n = 2, entonces la información requerida es dada en la tabla 1, o la tabla 2 segun el caso

Tabla 1

$R_3 - R_1$

$R_2 - R_1$	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	Totales
-8					1													1
-7				1			1											2
-6			2			1			2									5
-5		1			1			1			1							4
-4	1			1			2			1			1					6
--3			1			2			2			1						6
-2		1			2			2			2			1				8
-1				1			2			2			1					6
0			2			2			6			2			2			14
1					1			2			2			1				6
2				1			2			2			2			1		8
3						1			2			2			1			6
4					1			1			2			1			1	6
5							1			1			1			1		4
6									2			1			2			5
7											1			1				2
8													1					1
Totales	1	2	5	4	6	6	8	6	14	6	8	6	6	4	5	2	1	90

Así :

$Y^*(\alpha, k\text{-}1,n)\ P((R_u\text{-}R_1) < y^*(\alpha, k\text{-}1,n),\ u = 2,3,) = 1 - \alpha$ α

1 $38/90 = 0.42 = 1 - \alpha$ $1 - 0.42 = 0.58$

2 $44/90 = 0.49 = 1 - \alpha$ $1 - 0.49 = 0.51$

3 $54/90 = 0.60$ 0.40

4	$62/90 = 0.69$	0.31
5	$70/90 = 0.78$	0.22
6	$76/90 = 0.84$	0-16
7	$84/90 = 0.93$	0.07
8	$88/90 = 0.98$	0.02

Tabla 2

$|R_2 - R_1|$

$\lvert R_2 - R_1 \rvert$	0	1	2	3	4	5	6	7	8	Totales
0	6			4			4			14
1		4	4		2	2				12
2		4	4		4	2		2		16
3	4			6			2			12
4		2	4		2	2			2	12
5		2	2		2			2		8
6	4			2			4			10
7			2			2				4
8					2					2
Totales	14	12	16	12	12	8	10	4	2	90

De esta manera :

$Y^{**}(\alpha, k\text{-}1, n)$	$P(\lvert R_u - R_1 \rvert < y^{**}(\alpha, k\text{-}1, n),\ u = 2,3) = 1 - \alpha$	α
1	$6/90 = 0.07 = 1 - \alpha$	0.93
2	$10/90 = 0.11$	0.89
3	$22/90 = 0.24$	0.76
4	$36/90 = 0.40$	0.60
5	$50/90 = 0.56$	0.44
6	$62/90 = 0.69$	0.31
7	$78/90 = 0.87$	0.13
8	$86/90 = 0.96$	0.04

Aunque hemos encontrado a $y^*(\alpha,k\text{-}1,n)$ y a $y^{**}(\alpha,k\text{-}1,n)$ para $k=3$ y $n=2$, tales valores se encuentran tabulados para $k=3$ y $n = 2, 3, 4, 5$ y 6 en [1]. En general, los valores de $y^*(\alpha,k\text{-}1,n)$ y $y^{**}(\alpha, k\text{-}1,n)$ se obtienen usando el hecho de que bajo H_o (2) todas las $(nk)!/(n!)^k$ asignaciones de rango son igualmente probables.

Finalmente daremos las reglas de decisión de Dunn para muestras grandes sin la restricción de igual número de observaciones por tratamiento.. Tales reglas son :

1. Para prueba a dos ramas aceptaremos que $\epsilon_u \neq \epsilon_1$ si

$$|R_{\cdot u}-R_{\cdot 1}| \geqslant Z_{\alpha/[2(k-1)]}[N(N+1)/12]^{1/2}[(1/n_u)+(1/n_1)]^{1/2}, \qquad (5)$$

siendo $N = \sum_{j=1}^{k} n_j$.

2. Para pruebas a una rama aceptaremos que $\epsilon_u > \epsilon_1$ si

$$R_{\cdot u} - R_{\cdot 1} \geqslant Z_{\alpha/(k-1)}[N(N+1)/12]^{1/2}[(1/n)] \;,\; [(1/n_u)+(1/n_1)]^{1/2}, \qquad (6)$$

o que $\epsilon_u < \epsilon_1$ si $R_{\cdot 1}-R_{\cdot u} \geqslant Z_{\alpha/(k-1)}[N(N+1)/12]^{1/2}[(1/n_u)+(q/n_1)]^{1/2}$, (7)

Para el caso no paramétrico en la situación de algunos x_{ij} con el mismo valor, asignaremos a éstas el rango promedio de las mismas.

EJEMPLOS

1. Suponga que queremos probar la efectividad de tres métodos de enseñanza vigentes en Bogotá: El método tradicional (ϵ_1), el audiovisual (ϵ_2) y el usado por el canal educativo de la televisora nacional (ϵ_3). Asumiendo el modelo (1), las hipótesis para el caso paramétrico, que dos personas fueron aleatoriamente asignadas a cada tratamiento y que los resultados hipótéticos de una misma evaluación para cada estudiante fué :

	Tratamiento		
	1	2	3
Puntaje	3	3.8	3.2
	3.5	4.6	2.5

Pruebe al nivel de significación $\alpha=0.05$ la hipótesis (2).

Solución : $n_1 = n_2 = n_3 = 2 = n$, $k = 3$, $\bar{x}_{.1} = 3.3$, $\bar{x}_{.2} = 4.2$, $\bar{x}_{.3} = 2.9$, $\bar{x}_{..} = 3.4$; por tanto, $F = 3.9$. Puesto que $F < F_{0.05;\,2,3} = 9.55$, H_o es aceptada.

2. Asumiendo que para $n_1 = n_2 = n_3 = 2 = n$ los resultados de un experimento usando el modelo (1) y los supuestos del caso no paramétrico fueron :

	Tratamiento		
	1	2	3
Valóres Observados	45	58	75
	75	80	84

Pruebe al nivel de significación $\alpha \approx 0.05$ la hipótesis (2).

Solución : $r_{11} = 1$, $r_{21} = 3$, $r_{12} = 2$, $r_{22} = 5$, $r_{13} = 4$, $r_{23} = 6$, $R_1 = 4$, $R_2 = 7$, $R_3 = 10$, $|R_2 - R_1| = 3$, $|R_3 - R_1| = 6$ y de acuerdo a la tabla 2, $P(|R_u - R| < y^{**}(\alpha, 2,2),\ u = 2,3) = 1 - \alpha$ implica que para $\alpha = 0.04$, $y^{**}(0.04,2,2) = 8$. Puesto que $|R_u - R_1| < 8$ para $u = 2,3$, rechazaremos a H_o con $\alpha = 0.04$

Nórese que en nuestros ejemplos, (2) fué siempre aceptada; si este no hubiera sido el caso, debíamos continuar hasta establecer una decisión a una rama.

BIBLIOGRAFIA

[1] Hollander and Wolfe, Nonparametric Statistical Methods, New York, John Wiley and Sons, 1973.

[2] Montgomery, D.C., Design and Analysis of Experiments, New York, John Wiley and Sons, 1976.

[3] Morrison, D.F. Multivariate Statistical Methods, New York, McGraw-Hill, 1976.