

Análisis de la varianza de dos factores

El problema anterior consideraba la comparación de k muestras para detectar diferencias entre las posiciones de las respectivas poblaciones. En el modelo de aleatorización N individuos son asignados al azar a uno de k tratamientos. Sin embargo, las diferencias entre tratamientos podrían verse "oscurecidas" por la variabilidad excesiva entre los individuos dentro de los grupos. Este problema puede resolverse dividiendo a los sujetos en subgrupos homogéneos o bloques y realizando las comparaciones entre los bloques. Se trata de una extensión del problema de muestras apareadas.

El diseño que consideraremos inicialmente es el *diseño en bloques completamente aleatorizado con una observación por celda*.

Se tienen $N=nk$ sujetos divididos en n bloques y los sujetos, dentro de cada bloque, se asignan al azar a uno de los k tratamientos. Otro ejemplo de aplicación de este diseño es el problema de mediciones repetidas, en cuyo caso, el sujeto es el bloque y se realizan k mediciones sobre cada sujeto.

El modelo muestral puede definirse de dos formas:

1. $X_{ij} \sim F_i(x - q_j)$, $F_i \in \Omega_o$ para $i = 1, \dots, n$ y $j = 1, \dots, k$ son todas v.a. independientes. Es decir que F_i es la distribución de las observaciones del i -ésimo bloque y, dentro del bloque q_j es la mediana del j -ésimo tratamiento.
2. $\overset{\cdot}{X}_i = (X_{i1}, \dots, X_{ik})$ son n v.a. independientes con distribución $F_i(x_1 - q_1, \dots, x_k - q_k)$ y $F_i(x_1, \dots, x_k) = F_i(x_{p_1}, \dots, x_{p_k})$ para toda permutación (p_1, \dots, p_k) , o sea que X_1, \dots, X_k son canjeables o intercambiables. Este modelo es apropiado para mediciones repetidas donde no tiene sentido suponer independencia dentro del bloque.

Queremos testear

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k \quad \text{vs} \quad H_1: \text{existe al menos un par } (i, j) \text{ tal que } \theta_i \neq \theta_j$$

Test de Friedman (1973): Sea $R_{ij} = R(X_{ij})$ el rango de X_{ij} entre X_{i1}, \dots, X_{ik} . Es decir que R_{ij} es el rango de X_{ij} dentro de su bloque. Entonces

$$R_{.j} = \sum_{i=1}^n R_{ij}$$

es la suma de los rangos correspondientes al tratamiento j . Esquemáticamente, colocando como columnas los tratamientos y los bloques (sujetos) como filas:

	1	2	k
1	R_{11}	R_{12}	R_{1k}
2	R_{21}	R_{22}	R_{2k}
...
n	R_{n1}	R_{n2}	R_{nk}
	$R_{.1}$	$R_{.2}$	$R_{.k}$

Bajo H_0 , y suponiendo que no hay empates, conocemos la distribución de $R_{.1}, \dots, R_{.k}$ pues

$$P(R_{ij} = s) = \frac{1}{k} \quad 1 \leq s \leq k$$

$$P(R_{ij} = s, R_{il} = t) = \frac{1}{k(k-1)} \quad s \neq t \quad j \neq l$$

Entonces,

$$E(R_{ij}) = \frac{k+1}{2} \quad V(R_{ij}) = \frac{k^2-1}{12} \quad \text{cov}(R_{ij}, R_{il}) = -\frac{k+1}{12} \quad \text{si } j \neq l$$

y por lo tanto

$$E(R_{.j}) = \frac{n(k+1)}{2} \quad \text{Var}(R_{.j}) = \frac{n(k^2-1)}{12} \quad \text{cov}(R_{.j}, R_{.l}) = -\frac{n(k+1)}{12} \quad \text{si } j \neq l$$

Friedman propuso el siguiente estadístico:

$$T_1 = \sum_{j=1}^k c_{jN}^2 \left(\frac{R_{.j} - E(R_{.j})}{\sqrt{\text{Var}(R_{.j})}} \right)^2$$

donde los c_{jN} se eligen de manera que el estadístico converja, bajo H_0 a una distribución C_{k-1}^2 .

El estadístico toma la forma

$$T_1 = \sum_{j=1}^k \left(1 - \frac{1}{k} \frac{R_{.j} - n(k+1)/2}{\sqrt{\frac{n(k^2-1)}{12}}} \right)^2 = \left(\frac{12}{nk(k+1)} \sum_{j=1}^k R_{.j}^2 \right) - 3n(k+1)$$

que, bajo H_0 , tiene distribución asintótica χ^2 con $k-1$ grados de libertad.

Se rechazará H_0 si

$$T_1 > c_{k-1, \alpha}^2$$

En caso en que haya empates, debe hacerse una modificación al estadístico del test de Friedman.

Sean $A_1 = \sum_{i=1}^n \sum_{j=1}^k R_{ij}^2$ $C_1 = \frac{nk(k+1)^2}{4}$

Se define el estadístico modificado:

$$T_1 = \frac{(k-1) \left(\sum_{j=1}^k R_{.j}^2 - nC_1 \right)}{A_1 - C_1} = \frac{(k-1) \sum_{j=1}^k \left(R_{.j} - \frac{n(k+1)}{2} \right)^2}{A_1 - C_1}$$

Si $A_1 = C_1$ se considera que se está en la región crítica y se rechaza H_0 .

Otros estudios (Iman y Davenport, 1980) recomiendan utilizar, no T_1 , sino el estadístico del test clásico de análisis de la varianza calculado sobre los rangos, que se puede expresar como función del estadístico T_1 :

$$T_2 = \frac{(n-1)T_1}{n(k-1) - T_1}$$

Bajo H_0 , T_2 tiene distribución F con $(k-1)$ y $(n-1)(k-1)$ grados de libertad. Se rechaza H_0 si $T_2 > F_{(k-1), (k-1)(n-1), \alpha}$.

Ejemplo: 12 amas de casa son seleccionadas para participar en un experimento de siembra. A cada una de ellas se le pide que seleccione cuatro parcelas idénticas en su jardín y plante 4 tipos distintos de césped, uno en cada parcela. Después de cierto periodo, se les pide que ordenen los 4 tipos de césped por orden de preferencia, asignando el número 1 al césped menos preferido, 2 al siguiente, etc. La hipótesis nula

implica que no hay diferencias entre las preferencias de los tipos de césped. Los resultados obtenidos son los siguientes:

Ama de casa	Césped			
	1	2	3	4
1	4	3	2	1
2	4	2	3	1
3	3	1.5	1.5	4
4	3	1	2	4
5	4	2	1	3
6	2	2	2	4
7	1	3	2	4
8	2	4	1	3
9	3.5	1	2	3.5
10	4	1	3	2
11	4	2	3	1
12	3.5	1	2	3.5
R_j	38	23.5	24.5	34

Como hay empates, calculamos el valor del estadístico modificado

$$A_1 = 356.5 \quad C_1 = 300 \quad T_1 = 8.097$$

y obtenemos p-valor = $P(C_3^2 > 8.097) = 0.044$. Por lo tanto, a nivel 0.05, se rechaza la hipótesis nula.

Calculemos el estadístico T_2 .

$$T_2 = \frac{11(8.097)}{12(3) - 8.097} = 3.19$$

Como la región crítica de nivel 0.05 de la distribución F con 3 y 33 grados de libertad corresponde a valores del estadístico mayores que 2.90, se rechaza la hipótesis nula. El correspondiente p-valor es:

$$\text{p-valor} = P(F_{3,33} > 3.19) = 0.036$$

A continuación se presentan los resultados obtenidos al procesar este ejemplo en R. Este comando utiliza el estadístico T_1 .

```
friedman.test(datos)
```

Friedman rank sum test

```
data: datos
```

```
Friedman chi-squared = 8.0973, df = 3, p-value = 0.04404
```

Comparaciones múltiples: Cuando se rechaza H_0 , es posible realizar tests de comparaciones múltiples en base a R_1, \dots, R_k . En efecto, bajo H_0 ,

$$\frac{R_j - R_i}{\sqrt{\text{Var}(R_j - R_i)}} \xrightarrow{d} N(0,1)$$

siendo $\text{Var}(R_j - R_i) = 2\text{Var}(R_j) - 2\text{cov}(R_j, R_i) = \frac{nk(k+1)}{6}$.

Entonces, diremos que θ_i es significativamente distinto de θ_j a nivel global α si

$$|R_j - R_i| \geq z_{\alpha'/2} \sqrt{\frac{nk(k+1)}{6}} \quad \text{con } \alpha' = \frac{2\alpha}{k(k-1)}$$

Conover sugiere otro método. Diremos que θ_i es significativamente distinto de θ_j a nivel global α si

$$|R_j - R_i| \geq t_{(k-1)(n-1), \alpha/2} \left(\frac{2(nA_1 - \sum R_j^2)}{(k-1)(n-1)} \right)^{1/2}$$

En este caso, el nivel α es el mismo que el utilizado en el test de Friedman. La ecuación anterior puede escribirse en función de T_1 :

$$|R_j - R_i| \geq t_{(k-1)(n-1), \alpha/2} \left[\frac{(A_1 - C_1)2n}{(k-1)(n-1)} \left(1 - \frac{T_1}{n(k-1)} \right) \right]^{1/2}$$

Si no hay empates, $A_1 = nk(k+1)(2k+1)/6$ y $A_1 - C_1 = nk(k+1)(k-1)/12$.

Ejemplo: Dado que en el ejemplo anterior, se rechazó H_0 a nivel 0.05, interesa detectar cuáles son los tipos de césped que difieren en cuanto a la preferencia. Aplicaremos el procedimiento sugerido por Conover. El valor crítico 0.025 de la distribución t con 33 grados de libertad es 2.036, entonces

$$t_{(k-1)(n-1), \alpha/2} \left(\frac{2(nA_1 - \sum R_j^2)}{(k-1)(n-1)} \right)^{1/2} = 11.49$$

El césped tipo 1 resulta ser significativamente diferente de los tipos 2 y 3 y no hay otra diferencia significativa.

Eficiencia: La eficiencia del test de Friedman relativa al test F es

$$e(T_1, F) = 12 S^2 \left(\int f^2(x) dx \right)^2 \frac{k}{k+1}$$

Se observa que el test de Friedman no hereda la eficiencia de los tests de Wilcoxon y Mann-Whitney relativas al test de t, como ocurría con el test de Kuskal-Wallis.

Si la distribución es Normal,

$$e(T_1, F) = \frac{3k}{(k+1)p}$$

y, por lo tanto, si $k=2$, $e(T_1, F) = \frac{2}{p} = 0.64$, que es la eficiencia del test de signo relativa al test de t. Cuando k crece, la eficiencia aumenta.

Ejercicio: Verificar que, cuando $k=2$, el estadístico del test de Friedman es equivalente al del test del signo.

Diseño balanceado: El test de Friedman puede ser extendido al caso de varias observaciones por celda. Bernard y van Elteren (1952) trataron el caso general, pero nosotros consideraremos sólo el caso balanceado, es decir supondremos que en cada celda hay m observaciones.

Sean las observaciones $X_{ijt}, 1 \leq t \leq m, 1 \leq i \leq n, 1 \leq j \leq k$, donde

$$X_{ijt} \sim F(x - m - a_i - q_j), F \in \Omega_o$$

y nos interesa testear

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k \quad \text{vs} \quad H_1: \text{existe al menos un par } (i,j) \text{ tal que } \theta_i \neq \theta_j$$

Sea R_{ijt} el rango de X_{ijt} entre X_{i11}, \dots, X_{ikm} , es decir el rango de X_{ijt} dentro del bloque i. Entonces la suma de los rangos correspondientes al tratamiento j es

$$R_{.j} = \sum_{i=1}^n \sum_{t=1}^m R_{ijt}$$

Bajo H_0 ,

$$E(R_{.j}) = nm \frac{mk+1}{2}$$

$$Var(R_{.j}) = \frac{nm^2(mk+1)(k-1)}{12} \quad \text{si no hay empates}$$

Si hay empates, la varianza se estima por

$$Var(R_{.j}) = \frac{m(k-1)}{k(mk-1)} \left[\sum_{i,j,t} R_{ijt}^2 - \frac{mkn(mk+1)^2}{4} \right]$$

Además,

$$\text{cov}(R_{.j}, R_{.l}) = -\frac{nm^2(mk+1)}{12} \quad \text{si } j \neq l$$

El estadístico del test se define como

$$T_4 = \sum_{j=1}^k \left(1 - \frac{1}{k} \right) \left(\frac{R_{.j} - \frac{nm(mk+1)}{2}}{\sqrt{\text{Var}(R_{.j})}} \right)^2$$

Se puede probar que, bajo H_0 , $T_4 \xrightarrow{d} c_{k-1}^2$ y por lo tanto se rechaza H_0 a nivel α si

$$T_4 > c_{k-1, \alpha}^2$$

Si se rechaza H_0 , es necesario realizar comparaciones múltiples. θ_i es significativamente distinto de θ_j si

$$|R_{.j} - R_{.i}| > t_{mnk-k-n+1, \alpha/2} \left\{ \frac{2kn(mk-1)}{(k-1)(mnk-n-k+1)} \text{Var}(R_{.j}) \left[1 - \frac{T_4}{n(mk-1)} \right] \right\}^{1/2}$$

Alternativas ordenadas: Como en el caso del diseño de un factor puede ser de interés testear

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k \quad \text{vs} \quad H_1: \theta_1 \leq \theta_2 \leq \dots \leq \theta_k \text{ con al menos una desigualdad estricta}$$

Una observación por celda: Page (1963) propuso el siguiente estadístico

$$T_5 = \frac{1}{\sqrt{n}} \sum_{j=1}^k \left(j - \frac{k+1}{2} \right) \left(R_{.j} - \frac{n(k+1)}{2} \right)$$

Bajo H_0 , $E(T_5) = 0$ y $\text{Var}(T_5) = \frac{k^2(k^2-1)(k+1)}{144}$. Como $\frac{T_5}{\sqrt{\text{Var}(T_5)}} \xrightarrow{d} N(0,1)$,

rechazaremos H_0 , si

$$T_5 > z_{\alpha} \sqrt{\text{Var}(T_5)}$$

Ejemplo: Investigadores del área de salud sospechan que la actividad física regular tiende a disminuir el pulso de un individuo en reposo. Para poner a prueba esta teoría, 8 voluntarios sanos que no realizaban ejercicios en forma regular fueron integrados a un programa de ejercicios controlado. Se midió su pulso en reposo al comienzo del programa y una vez al finalizar cada mes durante 4 meses. Si θ_i es la mediana del pulso en reposo en el periodo i , las hipótesis a testear son

$$H_0: \theta_1 = \theta_2 = \dots = \theta_5 \quad \text{vs} \quad H_1: \theta_5 \leq \theta_4 \dots \leq \theta_1 \text{ con al menos una desigualdad estricta}$$

siendo θ_1 la mediana inicial, θ_2 la mediana al final del segundo mes y así sucesivamente.

Los pulsos observados se presentan en la siguiente tabla, juntamente con los rangos correspondientes

Individuo	Mes 4	Mes 3	Mes 2	Mes 1	Inicial
1	79 (3)	76 (1)	77 (2)	84 (5)	82 (4)
2	78 (3)	76 (1.5)	76 (1.5)	80 (4.5)	80 (4.5)
3	72 (1)	74 (2)	77 (4)	78 (5)	75 (3)
4	66 (3)	65 (1.5)	68 (4)	72 (5)	65 (1.5)
5	75 (3.5)	75 (3.5)	72 (1)	74 (2)	77 (5)
6	64 (1)	66 (3)	65 (2)	69 (5)	68 (4)
7	68 (1.5)	70 (3.5)	68 (1.5)	74 (5)	70 (3.5)
8	70 (1)	72 (2)	78 (5)	76 (3)	77 (4)
R_j	17	18	21	34.5	29.5

El valor del estadístico T_5 es 14.67247 y $Var(T_5) = 25$, entonces

$$p - \text{valor} = 1 - \Phi(2.934) = 0.002$$

y, por lo tanto se rechaza H_0 .

Diseño no balanceado: Skillings y Wolfe (1978) propusieron un estadístico basado en el de Jonckheere y Terpstra, utilizado para testear alternativas ordenadas en el diseño de un factor. La idea básica es calcular el estadístico J en cada uno de los n bloques y luego combinar estos valores. Sean

$$X_{ijt} \sim F(x - a_i - q_j) \quad 1 \leq t \leq m_{ij}, 1 \leq i \leq n, 1 \leq j \leq k$$

y sean $m_i = \sum_{j=1}^k m_{ij}$ y J_i el estadístico calculado sobre el i -ésimo bloque, o sea

$$J_i = \sum_{1 \leq l < j \leq k} W_{lj}^i \quad W_{lj}^i = \sum_{u,v} s(X_{iju} - X_{ilv})$$

Definiendo $J^* = \sum_{i=1}^n J_i$, se verifica que

$$E(J^*) = \sum_{i=1}^n \frac{m_i^2 - \sum_{j=1}^k m_{ij}^2}{4}$$

$$Var(J^*) = \frac{1}{72} \sum_{i=1}^n \left\{ m_i^2 (2m_i + 3) - \sum_{j=1}^k m_{ij}^2 (2m_{ij} + 3) \right\}$$

Si $n \rightarrow \infty$ y los m_{ij} permanecen fijos,

$$\frac{J^* - E(J^*)}{\sqrt{Var(J^*)}} \xrightarrow{d} N(0,1)$$

Test de Quade: Para el caso de una observación por celda cuando el número de tratamientos k es pequeño, el test de Quade, extensión del test de Wilcoxon, es más potente que el test de Friedman para testear las hipótesis

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k \quad \text{vs} \quad H_1: \text{existe al menos un par } (i,j) \text{ tal que } \theta_i \neq \theta_j$$

Sea $R_{ij} = R(X_{ij})$ el rango de X_{ij} entre X_{i1}, \dots, X_{ik} , es decir el rango dentro de su bloque.

Asignaremos por otro lado rangos a los bloques de acuerdo a su "tamaño":

$$\Delta_i = \max_{1 \leq j \leq k} (X_{ij}) - \min_{1 \leq j \leq k} (X_{ij}) \quad 1 \leq i \leq n$$

Ordenamos los tamaños de menor a mayor

$$\Delta^{(1)} \leq \Delta^{(2)} \leq \dots \leq \Delta^{(n)}$$

y denominamos $Q_i = \text{rango}(\Delta_i)$ en esta muestra ordenada. Definamos

$$S_{ij} = Q_i \left(R_{ij} - \frac{k+1}{2} \right) \quad S_{.j} = \sum_{i=1}^n S_{ij}$$

Observemos que S_{ij} representa el tamaño relativo de cada observación dentro del i -ésimo bloque, ajustado para reflejar la significación relativa del bloque.

Definiendo $A_3 = \sum_{i=1}^n \sum_{j=1}^k S_{ij}^2$ y $B_3 = \frac{1}{n} \sum_{j=1}^k S_{.j}^2$, el estadístico del test de Quade será

$$T_3 = \frac{(n-1)B_3}{A_3 - B_3}$$

que, bajo H_0 tiene distribución F con $(k-1)$ y $(n-1)(k-1)$ grados de libertad. Por lo tanto, se rechazará H_0 si

$$T_3 > F_{(k-1), (n-1)(k-1), \alpha}$$

El estadístico T_3 es el estadístico F del análisis de la varianza calculado sobre los S_{ij} . Si

$A_3 = B_3$ se rechaza H_0 y el p-valor se iguala a $\left(\frac{1}{k!}\right)^{n-1}$.

Observemos además que, si no hay empates, $A_3 = n(n+1)(2n+1)k(k^2 - 1)/72$.

Observación: Es importante notar que, para poder aplicar el test de Quade es necesario que se pueda ranquear los bloques según su tamaño.

Comparaciones múltiples: Si se rechaza H_0 , θ_i es significativamente distinto de θ_j a nivel global α si

$$|S_{.i} - S_{.l}| > t_{(n-1)(k-1), \alpha/2} \sqrt{\frac{2n(A_3 - B_3)}{(n-1)(k-1)}}$$

Eficiencia: Sólo se ha estudiado la eficiencia del test de Quade relativa al test F para el caso $k=2$, y en este caso coincide con la eficiencia del test de Wilcoxon relativa al test de t, es decir es $3/\pi$ para el caso Normal y nunca es menor que 0.864.

Ejemplo: Siete comercios son seleccionados para una encuesta. En cada uno de ellos se colocan en stands contiguos 5 marcas diferentes de crema de manos. Al fin de una semana se registra el número de envases de crema vendidos, con los siguientes resultados:

Comercio	Crema A	Crema B	Crema C	Crema D	Crema E
1	5 (2)	4 (1)	7 (3)	10 (4)	12 (5)
2	1 (2.5)	3 (5)	1 (2.5)	0 (1)	2 (4)
3	16 (2)	12 (1)	22 (3.5)	22 (3.5)	35 (5)
4	5 (4.5)	4 (2.5)	3 (1)	5 (4.5)	4 (2.5)
5	10 (3.5)	9 (2)	7 (1)	13 (5)	10 (3.5)
6	19 (2)	18 (1)	28 (3)	37 (4)	58 (5)
7	10 (5)	7 (2.5)	6 (1)	8 (4)	7 (2.5)

Los números entre paréntesis son los rangos dentro de cada bloque. En la tabla que sigue se presentan los valores de los tamaños de cada bloque, juntamente con los valores de

$$S_{ij} = Q_i \left(R_{ij} - \frac{k+1}{2} \right).$$

Comercio	Tamaño	Q_i	Crema A	Crema B	Crema C	Crema D	Crema E
1	8	5	-5	-10	0	5	10
2	3	2	-1	4	-1	-4	2
3	23	6	-6	-12	3	3	12
4	2	1	1.5	-0.5	-2	1.5	-0.5
5	6	4	2	-4	-8	8	2
6	40	7	-7	-14	0	7	14
7	4	3	6	-1.5	-6	3	-1.5

Los valores de $S_{.j}$ obtenidos son

$$S_{.1} = -9.5, S_{.2} = -38, S_{.3} = -14, S_{.4} = 23.5 \text{ y } S_{.5} = 38$$

Finalmente, $A_3 = 1366.5$ y $B_3 = 532.4$, entonces

$$T_3 = \frac{6(532.4)}{1366.5 - 532.4} = 3.83$$

este valor es mayor que $F_{4,24,0.05} = 2.78$ y, por lo tanto se rechaza H_0 .

A continuación se presentan los resultados obtenidos al procesar este ejemplo en R.

```
quade.test(datos)
```

Quade test

data: datos

Quade F = 3.8293, num df = 4, denom df = 24, p-value = 0.01519

En la siguiente tabla se indican con un * aquellas diferencias significativamente diferentes al utilizar un nivel global 0.05.

	Crema A	Crema B	Crema C	Crema D	Crema E
Crema A					*
Crema B				*	*
Crema C					*
Crema D					
Crema E					

Otra forma de representar el resultado de las comparaciones múltiples es indicar cuáles marcas no son significativamente diferentes. Para ello conviene listar los tratamientos (marcas) en orden creciente de rangos promedio:

B C A D E
—————

Diseño en bloques balanceado incompleto

En el diseño en bloques aleatorizado completo cada tratamiento es aplicado al menos una vez en cada bloque. Sin embargo, a veces ésto es imposible desde el punto de vista práctico. Por ejemplo, si 20 alimentos deben ser probados y ranqueados en orden de preferencia, a cada juez (bloque) puede resultarle muy difícil ranquear adecuadamente los 20 alimentos. Pero si cada juez probase sólo 5 alimentos y se usasen 4 veces más jueces, el ranqueo sería seguramente más confiable.

En estos casos puede resultar conveniente utilizar un diseño en bloques incompleto balanceado, el cual debe satisfacer las siguientes condiciones:

- cada bloque contiene k unidades experimentales
- cada tratamiento aparece en r bloques
- cada par de tratamientos aparece conjuntamente igual número de veces

Durbin propuso un test de rangos que puede utilizarse en este tipo de diseño y este test se reduce al test de Friedman si el número de tratamientos es igual al número de unidades experimentales por bloque.

Test de Durbin: Usaremos la siguiente notación (Conover, pag. 388)

- t : número de tratamientos
- k : número de unidades experimentales por bloque ($k < t$)
- b : número de bloques
- r : número de veces que aparece cada tratamiento ($r < b$)
- λ : número de bloques en los cuáles aparecen juntos los tratamientos i y j (este número es el mismo para todo par de tratamientos)

Sea X_{ij} el resultado del tratamiento j en el bloque i, si el tratamiento j aparece en el bloque i y sea $R_{ij} = R(X_{ij})$ el rango de X_{ij} dentro de su bloque. Estos rangos toman valores entre 1 y k. Sea la suma de los rangos correspondientes al tratamiento j,

$$R_{.j} = \sum_{i=1}^n R_{ij}$$

donde en realidad hay sólo r sumandos (podríamos suponer que los otros rangos toman el valor 0).

Supondremos que los bloques son independientes, es decir que los vectores aleatorios $\underline{X}_i = (X_{i1}, \dots, X_{it})$ son independientes, con distribución $F_i(x - q_1, \dots, x - q_t)$ y las hipótesis a testear son

$$H_0: \theta_1 = \theta_2 = \dots = \theta_t \quad \text{vs} \quad H_1: \text{existe al menos un par } (i,j) \text{ tal que } \theta_i \neq \theta_j$$

Propiedades: $E(R_{ij}) = \frac{k+1}{2}$ $V(R_{ij}) = \frac{k^2-1}{12}$

$$E(R_{.j}) = r \frac{k+1}{2} \quad V(R_{.j}) = r \frac{k^2-1}{12}$$

Se define el estadístico de Durbin como

$$T = \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left(R_{.j} - \frac{r(k+1)}{2} \right)^2$$

Si hay empates, es necesario hacer un ajuste. Sean

$$A = \sum_{i=1}^b \sum_{j=1}^t R_{ij}^2 \quad C = \frac{bk(k+1)^2}{4}$$

El estadístico corregido por empates es

$$T = \frac{(t-1) \sum_{j=1}^t \left(R_{.j} - \frac{r(k+1)}{2} \right)^2}{A - C} = \frac{(t-1) \left[\sum_{j=1}^t R_{.j}^2 - rC \right]}{A - C}$$

La segunda igualdad resulta de observar que en un diseño incompleto balanceado $rt = bk$.

Bajo H_0 , T tiene distribución asintótica χ^2 con $t-1$ grados de libertad y por lo tanto se rechaza H_0 si $T > c_{t-1, \alpha}^2$.

Un procedimiento alternativo consiste en calcular el estadístico F del análisis de la varianza sobre los rangos. El estadístico T^* así obtenido se relaciona con el estadístico T en la forma

$$T^* = \frac{T/(t-1)}{(b(k-1)-T)/(bk-b-t+1)}$$

Este estadístico tiene distribución F con (t-1) y (bk-b-t+1) grados de libertad. Esta aproximación tiende a dar valores más próximos a los exactos que la obtenida mediante la distribución chi cuadrado.

Comparaciones múltiples: Si se rechaza H_0 , es necesario realizar comparaciones múltiples para identificar los pares significativamente diferentes. Conover sugiere el siguiente procedimiento. θ_i es significativamente distinto de θ_j a nivel global α si

$$|R_{.j} - R_{.i}| > t_{bk-b-t+1, \alpha/2} \left[\frac{(A-C)2r}{bk-b-t+1} \left(1 - \frac{T}{b(k-1)} \right) \right]^{1/2}$$

que, si no hay empates, se reduce a

$$|R_{.j} - R_{.i}| > t_{bk-b-t+1, \alpha/2} \left[\frac{rk(k+1)}{6(bk-b-t+1)} (b(k-1) - T) \right]^{1/2}$$

Ejemplo: A cada uno de un grupo de 14 estudiantes se le pidió que clasifique, con puntajes entre 1 y 4, a 4 definiciones de integración racial de un grupo total de 8 definiciones, asignando el mayor puntaje (4) a la definición que consideraba más aproximada a su propia definición y el menor puntaje (1) a la menos aproximada. Los resultados obtenidos son los siguientes:

Estudiante	Definición							
	F	C	A	D	H	G	E	B
1	2	1	3	4				
2	2	1					3	4
3	1		2			3		4
4	1			2		3	4	
5					1	2	3	4
6			1	2	3	4		
7		1		3	2		4	
8		1	2		3			4
9	2	1			4	3		
10	2		1		3		4	
11	1			2	3			4
12			1	2			4	3
13		2		1		3		4
14		1	2			3	4	
$R_{.j}$	11	8	12	16	19	21	26	27

En este ejemplo,

- t = 8
- k = 4
- b = 14
- r = 7
- $\lambda = 3$

La hipótesis nula es que no hay diferencias en los puntajes medianos asignados a las 8 definiciones y la alternativa es que hay al menos un par de definiciones para los cuáles los puntajes medianos son diferentes.

El valor del estadístico, calculado en la forma

$$T = \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left(R_{.j} - \frac{r(k+1)}{2} \right)^2 = \frac{12(8-1)}{(7)(8)(4^2-1)} \left[\left(11 - \frac{7(4+1)}{2} \right)^2 + \dots + \left(27 - \frac{7(4+1)}{2} \right)^2 \right]$$

es $T = 34.2$ y el valor $c_{7,0.05}^2 = 14.07$ y, por lo tanto se rechaza H_0 . El p-valor es menor que 0.001.

A continuación, se deberían realizar las comparaciones múltiples para identificar los pares significativamente diferentes. Usando el método propuesto por Conover, calculamos

$$t_{35,0.025} \left[\frac{rk(k+1)}{6(bk-b-t+1)} (b(k-1) - T) \right]^{1/2} = 4.63$$

y las diferencias significativas son las indicadas con un * en la siguiente tabla:

	A	B	C	D	E	F	G	H
A		*			*		*	*
B			*	*		*	*	*
C				*	*		*	*
D					*	*	*	
E						*	*	*
F							*	*
G								
H								