

Unidad V

PRUEBAS ESTADÍSTICAS PARA UNA MUESTRA K RELACIONADA

“Es mejor tener una respuesta aproximada a la pregunta correcta que una respuesta exacta a la pregunta equivocada”

John Tukey

Introducción

Si se desea aplicar un diseño experimental y se cuentan con unidades experimentales que no son homogéneas, una forma de reducir el error experimental es formando bloques. De tal manera que estas unidades experimentales sean lo más homogéneas posible dentro de los bloques y lo más heterogéneas entre bloques.

El diseño experimental que se utiliza en este tipo de situaciones es el Diseño de Bloques Completos al Azar (DBCA). Para realizar un análisis de los datos que provienen de este tipo de diseño se puede hacer uso de la prueba F del Análisis de Varianza. Sin embargo, este análisis exige el cumplimiento de ciertos supuestos, como la normalidad de los errores, homogeneidad de varianzas y aleatoriedad de errores.

La prueba F del Análisis de Varianza solo responde la pregunta si al menos uno de los tratamientos presenta una media distinta a los demás. Para determinar que tratamiento(s) es(son) el(los) mejor(es) se deben realizar pruebas de comparación por pares como: DLS, Tukey, Contrastes Ortogonales, etc.

Los resultados de estas pruebas también tendrían validez solo si se cumplen los supuestos que estas exigen. Si esto no sucede una alternativa es recurrir a métodos no paramétricos.

Por otro lado, en estudios organolépticos de áreas como Industrias Alimentarias los bloques son los jueces que evalúan, por ejemplo, el sabor de diferentes marcas de un producto como yogurt. En este tipo de aplicación se pueden rescatar dos puntos importantes: primero, la variable para analizar, por lo general, es de tipo ordinal (1: pésimo a 5 excelente). Segundo, los jueces se podrían considerar como una muestra relacionada que evalúan k marcas. En esta situación, definitivamente se debe utilizar pruebas no paramétricas.

Si bien es cierto que programas estadísticos como SPSS denomina a estas pruebas como métodos para k muestras relacionadas, el término muestras no sería el más adecuado dado que se analiza una muestra a la que se le aplica k distintos tratamientos. Particularmente, pienso que debería denominarse Pruebas para una muestra k veces relacionada.

Este capítulo tiene como objetivo desarrollar algunas de las pruebas no paramétricas que se pueden utilizar para analizar los datos provenientes de una muestra k veces relacionada. También se desarrollarán las respectivas pruebas no paramétricas de comparación.

1. Para una variable cualitativa binaria

A. Prueba Q de Cochran

➤ Aspectos Generales

Evalúa si tres o más grupos de datos difieren significativamente o no entre ellos. El diseño es simple, a n sujetos se les toma k medidas dicotómicas o k medidas dicotómicas son medidas en una muestra.

Estamos, por lo tanto, en un diseño idéntico al DBCA pero aquí la variable respuesta es dicotómica.

➤ Supuestos

- La muestra de tamaño n es aleatoria.
- La variable respuesta es dicotómica o puede ser dicotomizada.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

Si los datos son organizados en una tabla de doble entrada donde en las filas se encuentran los b bloques (individuos o jueces) y en las columnas los k grupos (o tratamientos) y se define a 1 como un éxito y a 0 como fracaso, los pasos a seguir son:

- Calcular los G_j , que es el número total de éxitos de la j -ésima columna.
- Calcular los L_i , que es el número total de éxitos de la i -ésima fila.
- Aplicar el siguiente estadístico de prueba

$$Q = \frac{(k-1) \left[k \sum_{j=1}^k G_j^2 - \left(\sum_{j=1}^k G_j \right)^2 \right]}{k \sum_{i=1}^b L_i - \sum_{i=1}^b L_i^2} \sim \chi^2_{(1-\alpha, k-1)}$$

Las hipótesis son:

$$H_0 : \pi_1 = \pi_2 = \dots = \pi_k$$

$$H_1 : \text{Al menos una } \pi_i \text{ es diferente a las demás } i = 1, 2, \dots, k$$

Si se rechaza H_0 se deben hacer las pruebas de comparaciones por pares de tratamientos o grupos

$$H_0 : \pi_i = \pi_j$$

$$H_1 : \pi_i \neq \pi_j$$

➤ Aplicación

Se desea verificar si 4 preguntas de una prueba de rendimiento poseen o no el mismo nivel de dificultad. Para ello se evalúan a 10 individuos anotando su desacuerdo (0) o acierto (1). Realice la prueba respectiva use $\alpha=0.05$.

Sujeto	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4
1	1	1	1	1
2	1	0	0	0
3	1	0	1	0
4	1	1	1	0
5	0	0	1	0
6	0	1	1	0
7	1	0	1	1
8	0	0	1	0
9	1	1	1	0
10	0	1	1	0

$$H_0 : \pi_1 = \pi_2 = \dots = \pi_k$$

$H_1 : \text{Al menos una } \pi_i \text{ es diferente a las demás } i = 1, 2, \dots, k$

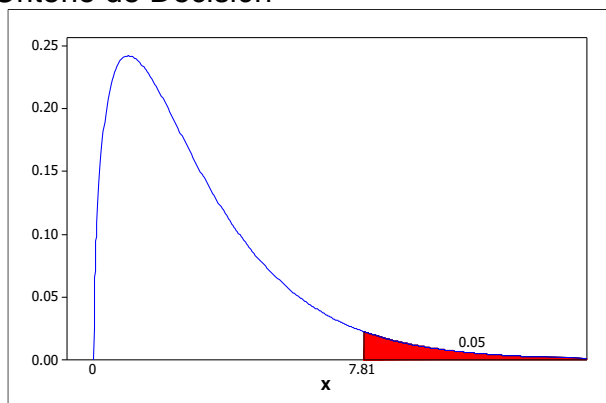
$$\alpha = 0.05$$

Prueba Estadística

$$Q = \frac{(k-1) \left[k \sum_{j=1}^k G_j^2 - \left(\sum_{j=1}^k G_j \right)^2 \right]}{k \sum_{i=1}^b L_i - \sum_{i=1}^b L_i^2} \sim \chi^2_{(1-\alpha, k-1)}$$

$$Q=10$$

Criterio de Decisión



Como $\chi^2_{(0.95,3)} = 7.81 < 10$ se rechaza H_0

Conclusión

Existe suficiente evidencia estadística a un nivel de significación de 0.05, para rechazar H_0 .

Por lo tanto, se puede afirmar que las preguntas de dicha evaluación no poseen igual nivel de dificultad.

➤ **Funciones con R**

Se encuentra la función `cochran.qtest` del paquete `RVAideMemoire`

También en la web se pueden encontrar funciones elaboradas por algunos investigadores quienes no la formalizan dentro de un paquete, como la función elaborada por Marc Schwartz

```
cochran.qtest(respuesta~factor bloque)

cochranq.test <- function(mat)
{
  k <- ncol(mat)

  C <- sum(colSums(mat) ^ 2)
  R <- sum(rowSums(mat) ^ 2)
  T <- sum(rowSums(mat))

  num <- (k - 1) * ((k * C) - (T ^ 2))
  den <- (k * T) - R

  Q <- num / den

  df <- k - 1
  names(df) <- "df"
  names(Q) <- "Cochran's Q"

  p.val <- pchisq(Q, df, lower = FALSE)

  QVAL <- list(statistic = Q, parameter = df, p.value = p.val,
              method = "Cochran's Q Test for Dependent Samples",
              data.name = deparse(substitute(mat)))
  class(QVAL) <- "htest"
  return(QVAL)
}
```

También está la función `cochrans.q` del paquete `nonpar`

➤ **Resultados con R**

Con la función elaborada por Marc Schwartz

```
datos<-matrix(0,10,4)
fix(datos)
```

```
      Cochran's Q Test for Dependent Samples
data:  datos
Cochran's Q = 10, df = 3, p-value = 0.01857
```

```
library(RVAideMemoire)

res<-
c(1,1,1,1,1,0,0,0,1,0,1,0,1,1,1,0,0,0,1,0,0,1,1,0,1,0,1,1,0
,0,1,0,1,1,1,0,0,1,1,0)
tra<-gl(4,1,40,labels=LETTERS[1:4])
blo<-gl(10,4,40,labels=letters[1:10])
cochran.qtest(res~tra|blo, p.method="none")
```

Cochran's Q test

```
data:  res by tra, block = blo
Q = 10, df = 3, p-value = 0.01857
alternative hypothesis: true difference in probabilities is
not equal to 0
sample estimates:
proba in group A proba in group B proba in group C
                0.6                0.5                0.9
proba in group D
                0.2
```

Pairwise comparisons using Wilcoxon sign test

	A	B	C
B	1.000	-	-
C	0.375	0.125	-
D	0.125	0.375	0.01563

P value adjustment method: none

➤ **Algunas consideraciones en R**

- La función `cochran.qtest` del paquete `RVAideMemoire` realiza las pruebas de comparaciones pero utilizando la prueba de signos para todos los pares de tratamientos.

2. Para un parámetro de locación

Las pruebas que se presentarán en esta sección son alternativas no paramétricas a un análisis de dos vías o Diseño de Bloques Completo al Azar (DBCA).

Hipótesis

$$H_0 : Me_1 = Me_2 = \dots = Me_k$$

$$H_1 : \text{Al menos una } Me_i \text{ es diferente a las demás } i = 1, 2, \dots, k$$

Si se rechaza H_0 se deben realizar las pruebas de comparación por pares de tratamientos

Las hipótesis son:

$$H_0 : Me_i = Me_j \quad i, j = 1, 2, \dots, k \quad \forall i \neq j$$

$$H_1 : Me_i \neq Me_j$$

A. Prueba de Friedman

➤ Aspectos Generales

Esta prueba es apropiada siempre que los datos se midan, al menos en una escala ordinal y puedan disponerse significativamente en una clasificación de dos criterios.

Es una prueba muy utilizada en experimentos donde se consideran jueces (bloques) para que evalúen diferentes productos. Por ejemplo, en Industrias Alimentarias muy frecuentemente se desea analizar el sabor de un producto mediante un calificativo medido en una escala del 1 al 5.

➤ Supuestos

- Los b bloques son mutuamente independientes, es decir, los resultados de un bloque no influyen en los resultados de los otros bloques.
- La escala de medida es al menos ordinal, de modo que las observaciones pueden ser ordenadas dentro de cada bloque.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Para cada bloque, los k grupos son ordenados en un rango de 1 a k .
- Los rangos asignados a las observaciones en cada uno de los k grupos se suman por separado para dar k sumas de rangos.
- Aplicar el estadístico de prueba que se basa en estos rangos y es el siguiente:

$$A = \sum_{i=1}^k \sum_{j=1}^b [R(X_{ij})]^2 \quad B = \frac{1}{b} \sum_{i=1}^k R_i^2$$

$$S = \frac{(k-1) \left[bB - \frac{b^2 k (k+1)^2}{4} \right]}{A - \frac{bk(k+1)^2}{4}} \sim \chi^2_{[k-1, 1-\alpha]}$$

Si no hay empates A se simplifica a:

$A = \frac{bk(k+1)(2k+1)}{6}$ y por lo tanto la expresión de T se reduce a

$$S = \frac{12}{bk(k+1)} \sum_{i=1}^k \left(R_i - \frac{b(k+1)}{2} \right)^2 \sim \chi^2_{[(k-1), 1-\alpha]}$$

donde: b es el número de filas (bloques) y k es el número de columnas (tratamientos o grupos).

- Comparar este valor calculado con el valor χ^2 tabulado con $k-1$ grados de libertad y $1-\alpha$

Si $S > \chi^2_{(1-\alpha, k-1)}$ se rechaza H_0 .

Una prueba alternativa es:

$$F = \frac{(b-1)S}{b(k-1)-S} \sim F_{[k-1, (b-1)(k-1), 1-\alpha]}$$

Comparaciones Múltiples

Si la hipótesis nula es rechazada, la prueba de Friedman presenta un procedimiento para comparar los tratamientos por pares.

Se dirá que los tratamientos i y j difieren significativamente si se cumple que:

$$|R_i - R_j| > t_{\left(1-\frac{\alpha}{2}, (b-1)(k-1)\right)} \sqrt{\frac{2b(A-B)}{(b-1)(k-1)}}$$

Este es denominado el criterio de diferencia mínima significativa (LSD)

Otra alternativa es utilizar una prueba de comparación equivalente a la prueba de Tukey

$$|\bar{R}_i - \bar{R}_j| > z_{(\alpha/k(k-1))} \sqrt{\frac{bk(k+1)}{6}}$$

➤ Aplicación

Se está realizando un experimento para analizar el sabor de una nueva marca de gaseosa sabor cola antes de que se lance al mercado. Las marcas de gaseosas colas en comparación fueron dadas a 5 jueces especializados y se estableció una escala de valores de (1-5) donde 1 es el de peor sabor y 5 el de mejor sabor

Los resultados del experimento se muestran a continuación:

Jueces	Marcas de gaseosas colas			
	Cola 1	Cola 2	Cola 3	Nueva Cola
1	5	2	3	2
2	4	1	3	4
3	5	2	2	3
4	5	1	3	2
5	5	2	3	3

Pruebe si no existe igual preferencia por las gaseosas. Use $\alpha=0.05$.

Solución

Si realizamos el ordenamiento dentro de cada bloque se tendría el siguiente cuadro

Jueces	Marcas de gaseosas colas			
	Cola 1	Cola 2	Cola 3	Nueva Cola
1	4	1.5	3	1.5
2	3.5	1	2	3.5
3	4	1.5	1.5	3
4	4	1	3	2
5	4	1	2.5	2.5
Total	$R_1=19.5$	$R_2=6$	$R_3=12$	$R_4=12.5$

$$H_0 : Me_1 = Me_2 = Me_3 = Me_4$$

$$H_1 : \text{Al menos una } Me_i \text{ es diferente a las demás } i = 1, 2, 3, 4$$

$$\alpha = 0.05$$

Prueba Estadística

$$S = \frac{(k-1) \left[bB - \frac{b^2 k (k+1)^2}{4} \right]}{A - \frac{bk(k+1)^2}{4}}$$

Donde:

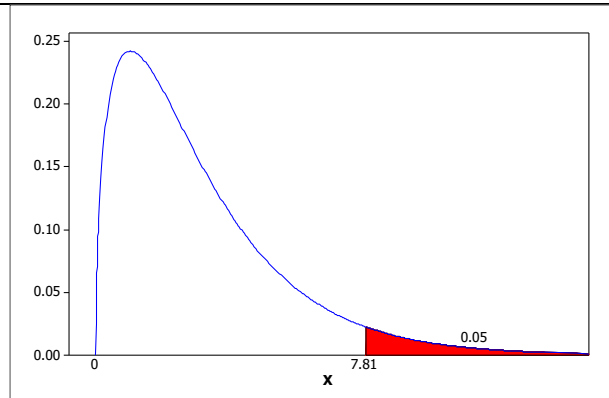
$$A = \sum_{i=1}^k \sum_{j=1}^b [R(X_{ij})]^2 \quad B = \frac{1}{b} \sum_{i=1}^k R_i^2$$

Desarrollo de la prueba

$$A = 4^2 + \dots + 2.5^2 = 148 \quad B = \frac{19.5^2 + 6^2 + 12^2 + 12.5^2}{5} = 143.3$$

$$S = \frac{(4-1) \left[5(143.3) - \frac{5^2 4 (4+1)^2}{4} \right]}{148 - \frac{5(4)(4+1)^2}{4}} = 11.93$$

Criterio de Decisión



Como $\chi^2_{(0.95,3)} = 7.81 < 11.93$ se rechaza H_0

Conclusión

A un nivel de significación de 0.05 se puede afirmar que las gaseosas de sabor cola en estudio no tienen igual preferencia.

Por lo tanto se debe proceder a realizar las pruebas de comparación

$$H_0 : Me_i = Me_j \quad i,j=1,2,3,4 \quad \forall i \neq j$$

$$H_1 : Me_i \neq Me_j$$

$$\alpha = 0.05$$

$$ALS(Fr) = t_{\left(1-\frac{\alpha}{2}, (b-1)(k-1)\right)} \sqrt{\frac{2b(A-B)}{(b-1)(k-1)}} = t_{(0.975, (5-1)(4-1))} \sqrt{\frac{2(5)(148-143.3)}{(5-1)(4-1)}} = 2.18(1.97)$$

$$ALS(Fr) = 4.31$$

Comparaciones	$ R_i - R_j $	ALS(Fr)	Sig
1 vs 2	$ 19.5 - 6 = 13.5$	4.31	*
1 vs 3	$ 19.5 - 12 = 7.5$	4.31	*
1 vs 4	$ 19.5 - 12.5 = 7$	4.31	*
2 vs 3	$ 6 - 12 = 6$	4.31	*
2 vs 4	$ 6 - 12.5 = 5.5$	4.31	*
3 vs 4	$ 12 - 12.5 = 0.5$	4.31	n.s

Resumen

C2 C3 C4 C1

Se puede afirmar a un nivel de significación de 0.05 que la marca de gaseosa de sabor cola de mayor preferencia en cuanto al sabor es la cola 1.

➤ **Funciones en R**

Se cuenta con la función `friedman.test`, el problema de esta función es que no realiza la prueba de comparaciones múltiples por pares.

```
friedman.test(vector de datos, grupos, bloques)
```

También se puede hacer uso de la función `friedman` del paquete `agricolae` que si brinda las pruebas de comparaciones múltiples por pares

```
friedman(bloques, grupos, vector de datos)
```

Esta función requiere que los resultados sean almacenados en un objeto.

Otra función que permite obtener las pruebas de comparación es `friedmanmc` del paquete `pgirmess`

```
friedmanmc(vector de datos, grupos, bloques)
```

➤ **Resultados con R**

```
c1<-c(5,4,5,5,5)
c2<-c(2,1,2,1,2)
c3<-c(3,3,2,3,3)
c4<-c(2,4,3,2,3)
todo<-c(c1,c2,c3,c4)
marca<-rep(1:4,c(5,5,5,5))
jueces<-rep(1:5,4)

friedman.test(todo,marca,jueces)
```

Friedman rank sum test

```
data: todo, marca and jueces
Friedman chi-squared = 11.9348, df = 3, p-value = 0.00761
```

```
library(agricolae)
resul<-friedman(jueces,marca,todo)
resul
```

```
$statistics
      Chisq      p.chisq      F      p.F      LSD
11.93478 0.00760994 15.57447 0.0001939151 4.311995

$parameters
  Df ntr  t.value
  3   4 2.178813

$means
  todo      std r Min Max
1  4.8 0.4472136 5   4   5
2  1.6 0.5477226 5   1   2
3  2.8 0.4472136 5   2   3
4  2.8 0.8366600 5   2   4
```

```
$comparison
NULL
```

```
$groups
  trt Sum of ranks M
1    1          19.5 a
2    4          12.5 b
3    3          12.0 b
4    2           6.0 c
```

```
library(pgirmess)
friedmanmc(todo,marca,jueces)
```

```
Multiple comparisons between groups after Friedman test
p.value: 0.05
Comparisons
  obs.dif critical.dif difference
1-2    13.5    10.77064      TRUE
1-3     7.5    10.77064     FALSE
1-4     7.0    10.77064     FALSE
2-3     6.0    10.77064     FALSE
2-4     6.5    10.77064     FALSE
3-4     0.5    10.77064     FALSE
```

➤ **Algunas consideraciones de R**

- La función `friedman` del paquete `agricolae` es mejor que la función `friedman.test` debido a que la primera brinda las pruebas de comparación por pares de tratamientos.
- La función `friedmanmc` del paquete `pgirmess` brinda resultados distintos a los del paquete `agricolae`. Esto debido a que utiliza una prueba distinta a la de `agricolae`.

B. Prueba de Quade

➤ **Aspectos Generales**

La prueba de Quade fue presentada por primera vez por el bioestadístico estadounidense Dana Quade en la revista científica *Journal of the American Statistical Association* en 1979.

Es utilizada en el análisis de diseños de bloques completos al azar (DBCA) y está considerada dentro de los métodos no paramétricos debido a que no requiere la distribución de los datos en particular en la muestra.

Esta prueba supone que hay k tratamientos experimentales ($k \geq 2$) y que los tratamientos están distribuidos en b bloques.

La prueba implica que los valores entre muestras pareadas y dentro de las muestras sean independientes entre sí. Este análisis se basa en la clasificación de los valores de cada unidad de estudio, por lo que cada conjunto de datos está emparejado, desde el más pequeño hasta el más grande y cada grupo se clasifica por separado. El rango está determinado por la diferencia entre el valor medido más grande y el más pequeño para cada unidad de estudio. Todas las unidades de estudio se ordenan de acuerdo a los rangos, se le asigna el valor de 1 para el de menor rango y el valor del número de unidades de estudio n para el de mayor rango. De esta manera se toman las diferencias de las distancias entre las unidades individuales del análisis. Es importante resaltar que, para poder aplicar la prueba de Quade es necesario que se pueda rankear los bloques según su tamaño.

Se dice que la prueba de Quade es similar a la prueba de Friedman, sin embargo, existen algunas distinciones:

- Para $k = 2$, la prueba de Friedman es equivalente a la prueba de los signos, mientras que la prueba de Quade es equivalente a la prueba de Wilcoxon.
- Según Conover, la prueba de Quade es típicamente más potente para $k < 5$ mientras que la prueba de Friedman tiende a ser más potente para $k \geq 5$.
- La prueba de Friedman sólo requiere datos de escala ordinal (es decir, que puedan ser calificados) mientras que la prueba de Quade requiere al menos que los datos tengan escala de intervalo.

➤ **Supuestos**

- Los b bloques son independientes entre sí. Eso significa que los resultados dentro de un bloque no afectan los resultados dentro de otros bloques.
- Los datos se pueden ordenar significativamente.
- Los datos tienen escala al menos de intervalo.

➤ **Inferencia Estadística**

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

La prueba de Quade asume que hay k tratamientos experimentales ($k \geq 2$) y que las observaciones están dispuestas en bloques, de tamaño b , tal como se muestra en el siguiente esquema:

Bloque	Tratamiento			
	1	2	...	k
1	X_{11}	X_{12}	...	X_{1k}
2	X_{21}	X_{22}	...	X_{2k}
3	X_{31}	X_{32}	...	X_{3k}
...
b	X_{b1}	X_{b2}	...	X_{bk}

Para el desarrollo de la prueba se requiere lo siguiente:

Antes que todo, en esta prueba usaremos el subíndice i , para referirnos a los bloques y el subíndice j , para referirnos a los tratamientos

Primero, asignaremos rangos a cada uno de los bloques de acuerdo a su "tamaño":

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

Ordenamos los tamaños de mayor a menor: $\Delta^{(1)} \leq \Delta^{(2)} \leq \dots \leq \Delta^{(b)}$

Luego se calcula el rango de X_{ij} entre $X_{i1}, X_{i2}, \dots, X_{ik}$, representado por $R_{ij} = R(X_{ij})$ (el rango dentro de su bloque)

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

$$S_{ij} = Q_i \left(R_{ij} - \frac{(k+1)}{2} \right) \quad \text{y} \quad S_j = \sum_{i=1}^b S_{ij}$$

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

Ahora definamos los siguientes valores:

$$A_3 = \sum_{i=1}^b \sum_{j=1}^k S_{ij}^2 \quad \text{y} \quad B_3 = \frac{1}{b} \sum_{i=1}^b S_j^2$$

Donde: A_3 es equivalente a la suma total de cuadrados y B_3 es equivalente a la suma de los cuadrados de tratamiento.

Con estos valores, se podrá calcular el estadístico de prueba, el cual está definido por:

$$T_3 = \frac{(b-1)B_3}{A_3 - B_3} \sim F_{[k-1, (b-1)(k-1)]}$$

Por lo tanto, se rechazará la H_0 si:

$$T_3 > F_{[k-1, (b-1)(k-1), 1-\alpha]}$$

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

Si $A_3 = B_3$ se rechaza la H_0 y el p-valor es igual a:

$$p\text{valor} = \left(\frac{1}{k!} \right)^{b-1}$$

Si no hay empates, el valor de A_3 se obtiene con la siguiente fórmula:

$$A_3 = \frac{b(b+1)(2b+1)k(k^2-1)}{72}$$

Comparaciones Múltiples

Si se rechaza H_0 , a menudo se desea conocer que tratamientos son diferentes entre sí. Para esto utilizamos las pruebas de comparaciones múltiples, empleando el estadístico t , haciendo el siguiente cálculo:

$$|S_i - S_l| > t_{[(b-1)(k-1)1-\alpha/2]} \sqrt{\frac{2b(A_3-B_3)}{(b-1)(k-1)}} \text{ se rechaza } H_0$$

➤ Aplicación

A continuación se presenta el número de unidades vendidas de cinco marcas de loción para las manos en siete tiendas durante una semana.

		Marca				
		A	B	C	D	E
Tienda	1	5	4	7	10	12
	2	1	3	1	0	2
	3	16	12	22	22	35
	4	5	4	3	5	4
	5	10	9	7	13	10
	6	19	18	28	37	58
	7	10	7	6	8	7

Prueba de hipótesis

H_0 : Las cinco marcas de loción para las manos tienen efectos idénticos.

H_1 : Al menos un par de marcas de loción para las manos tienen efectos diferentes entre sí.

$$\alpha = 0.05$$

Desarrollo

Paso 1: Las marcas de loción son los tratamientos y las tiendas serán los bloques. Se calculan los rangos a los bloques de acuerdo a su tamaño (Δ_i), además de los promedios para cada tratamiento y Δ_i .

		Marca = Tratamientos					Rango
		A	B	C	D	E	Δ_i
Tienda = Bloques	1	5	4	7	10	12	8
	2	1	3	1	0	2	3
	3	16	12	22	22	35	23
	4	5	4	3	5	4	2
	5	10	9	7	13	10	6
	6	19	18	28	37	58	40
	7	10	7	6	8	7	4
Promedio		9.43	8.14	10.57	13.57	18.29	12.29

Paso 2: Rankeamos los datos respecto a las bloques (tiendas) (Cálculo de los R_{ij}) y hallamos $Q_i = rango(\Delta_i)$

	R_{ij}	Marca = Tratamiento					Q_i
		A	B	C	D	E	
Tienda = Bloques	1	2	1	3	4	5	5
	2	2.5	5	2.5	1	4	2
	3	2	1	3.5	3.5	5	6
	4	4.5	2.5	1	4.5	2.5	1
	5	3.5	2	1	5	3.5	4
	6	2	1	3	4	5	7
	7	5	2.5	1	4	2.5	3

Donde, por ejemplo:

$R_{23} = 2.5$: El rango de las ventas para la tienda 2 (bloque) de la marca de loción C (tratamiento).

Paso 3: Calculamos el tamaño relativo de cada observación dentro del i-ésimo bloque (S_{ij}) y sumamos estos tamaños por tratamiento (S_j).

	S_{ij}	Marca = Tratamientos				
		A	B	C	D	E
Tienda = Bloques	1	-5	-10	0	5	10
	2	-1	4	-1	-4	2
	3	-6	12	3	3	12
	4	1.5	-0.5	-2	1.5	-0.5
	5	2	-4	-8	8	2
	6	-7	-14	0	7	14
	7	6	-1.5	-6	3	-1.5
S_j		-9.5	-38	-14	23.5	38

Donde, por ejemplo:

$$S_{11} = 5 \left(2 - \left(\frac{5+1}{2} \right) \right) = -5$$

Paso 4: Calculamos A_3 y B_3

$$A_3 = \sum_{i=1}^b \sum_{j=1}^k S_{ij}^2 = (-5)^2 + (-10)^2 + (0)^2 + \dots + (-1.5)^2 = 1366.5$$

$$B_3 = \frac{1}{b} \sum_{i=1}^b S_j^2 = \frac{1}{7} ((-9.5)^2 + (-38)^2 + (-14)^2 + (23.5)^2 + (38)^2) = 532.357143$$

Criterio de decisión

Cálculo el estadístico T_3 y contraste con el valor de tabla F :

$$T_3 = \frac{(7-1)(532.357143)}{(1366.5 - 532.357143)} \sim F_{[5-1, (7-1)(5-1)]}$$

$$T_3 = 3.829252 > F_{[4,24,0.05]} = 2.7762, \text{ se rechaza } H_0$$

$$pvalor = \Pr(T_3 > F) = 0.015189 < \alpha = 0.05, \text{ se rechaza } H_0$$

Conclusión

A un nivel de significación de 0.05 se puede afirmar que el número mediano de unidades vendidas es diferente en al menos una marcas de loción para las manos es diferente a las demás.

Comparaciones Múltiples:

$$H_0: Me_i = Me_j$$

$$H_1: Me_i \neq Me_j; \forall i, j = 1, 2, \dots, 5$$

$$\alpha = 0.05$$

Se calcula el siguiente valor

$$ALS_Q = t_{[(b-1)(k-1), \alpha/2]} \sqrt{\frac{2b(A_3 - B_3)}{(b-1)(k-1)}} = t_{[24, \frac{0.05}{2}]} \sqrt{\frac{2(7)(1366.5 - 532.357)}{(7-1)(5-1)}}$$

$$= 45.53$$

Comparaciones	$ S_i - S_j $		T_s	Significancia
$T_1 \text{ vs } T_2$	28.5	<	45.53	n.s
$T_1 \text{ vs } T_3$	4.5	<	45.53	n.s
$T_1 \text{ vs } T_4$	33	<	45.53	n.s
$T_1 \text{ vs } T_5$	47.5	>	45.53	*
$T_2 \text{ vs } T_3$	24	<	45.53	n.s.
$T_2 \text{ vs } T_4$	61.5	>	45.53	*
$T_2 \text{ vs } T_5$	76	>	45.53	*
$T_3 \text{ vs } T_4$	37.5	<	45.53	n.s
$T_3 \text{ vs } T_5$	52	>	45.53	*
$T_4 \text{ vs } T_5$	14.5	<	45.53	n.s

Representación gráfica de las comparaciones múltiples.

Este esquema se basa en los resultados obtenidos en la tabla anterior:

B	C	A	D	E
-38	-14	-9.5	23.5	38
_____	_____			
	_____	_____		
			_____	_____

Conclusión:

A un nivel de significación de 0.05 se puede afirmar que las marcas de loción D y E son las más vendidas.

➤ **Funciones en R**

Primero cargamos los datos en R:

```
datos<-read.delim("clipboard")
head(datos)
  Unidades Marca Tienda
1         5     A      1
2         1     A      2
3        16     A      3
4         5     A      4
5        10     A      5
6        19     A      6
attach(datos)
```

En R se puede usar la función *quade.test*, la cual se encuentra dentro del paquete stats. Esta nos permite la realizar la prueba global de Quade (que si los tratamientos tienen efectos idénticos o al menos uno es distinto).

Existen también otra función en R que permite la realización de esta prueba, *quadeTest*, la función dentro del paquete scmamp.

```
quade.test(observaciones,tratamientos,bloques)
quade.test(Unidades,Marca,Tienda)
```

Quade test

```
data: Unidades, Marca and Tienda
Quade F = 3.8293, num df = 4, denom df = 24, p-value =
0.01519
```

Se observa que, al igual que en el desarrollo teórico, se rechaza la H_0 , por lo tanto se tendrían que hacer comparaciones múltiples por pares para identificar que tratamientos presentan efectos diferentes.

Para esta tarea, R también cuenta con una función, *posthoc.quade.test* (dentro del paquete PMCMR), la cual proporciona los p-valores de las comparaciones en pares de los tratamientos, para evaluar la significancia.

Para usar esta función se deben definir los siguientes valores:

```
posthoc.quade.test(observaciones, tratamientos, bloques,
distribucion, método de ajuste de p,...)
```

En este problema, definimos el argumento de distribución como *dist="TDist"* debido a que es la empleada para el cálculo del ALS.

En el caso del método de ajuste de p se define como *p.adj="none"* porque no se ha empleado ningún ajuste en esta prueba. Existen algunos métodos de ajusten que se podrían definir en este parámetro como: *bonferroni*, *holm*, etc.

```
library(PMCMR)
posthoc.quade.test(Unidades,Marca,Tienda,dist="TDist",p.adj
="none")
```

Pairwise comparisons using posthoc-Quade test with
TDist approximation

data: Unidades , Marca and Tienda

	A	B	C	D
B	0.2087	-	-	-
C	0.8401	0.2874	-	-
D	0.1477	0.0102	0.1021	-
E	0.0416	0.0021	0.0269	0.5172

P value adjustment method: none

Con estas salidas se pueden corroborar los resultados obtenidos con el desarrollo teórico de las comparaciones, que la marca de colonia A tiene efectos distintos al de la marca E; que la marca B tiene efectos distintos tanto con la marca D y E; y que la marca C tiene a su vez efectos distintos con la marca E.

C. Prueba de Page

➤ Aspectos Generales

La prueba de Friedman evalúa la hipótesis nula de que la mediana de k grupos o k medidas repetidas son las mismas, contra la hipótesis alterna de uno o más grupos son diferentes. En ocasiones, un investigador puede desear mantener una hipótesis alterna más específica en cuanto al orden de los tratamientos o grupos. La prueba de Page para alternativas ordenadas, evalúa la hipótesis de que la mediana en los grupos son los mismos contra la hipótesis alterna de que los grupos están ordenados en una secuencia específica. Es decir, las medianas se encuentran ordenadas por su magnitud.

Es importante notar que para asegurar el uso adecuado de esta prueba, el investigador debe ser capaz de especificar el orden de los grupos a priori.

➤ Supuestos

- Los b bloques son mutuamente independientes, es decir los resultados de un bloque no influyen en los resultados de los otros bloques.
- La escala de medida es al menos ordinal, de modo que las observaciones pueden ser ordenadas dentro de cada bloque.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Para cada bloque, los k grupos son ordenados en un rango de 1 a k .
- Los rangos asignados a las observaciones en cada uno de los k grupos se suman por separado para dar k sumas de rangos.
- Aplicar el estadístico de prueba que se basa en estos rangos y es el siguiente:

$$L = \sum_{i=1}^k iR_i$$

Donde R_i es la suma de los rangos de los tratamientos

En el caso de muestras grandes se puede usar la aproximación asintótica a la distribución normal

$$\mu_L = \frac{bk(k+1)^2}{4} \quad \sigma_L^2 = \frac{bk^2(k^2-1)^2}{144(k-1)}, \text{ si hay empates} \quad \sigma_L^2 = \frac{k(k^2-1)}{12} \sum_{j=1}^b S_{Bj}^2$$

Donde S_{Bj}^2 es la varianza muestral de los rangos en los b bloques.

$$Z = \frac{L - \mu_L}{\sigma_L}$$

Las hipótesis son:

$$H_0 : Me_1 = Me_2 = \dots = Me_k$$

$$H_1 : Me_1 < Me_2 < \dots < Me_k$$

➤ **Aplicación**

En años recientes se ha incrementado el interés por la capacidad de las personas de percibir patrones táctiles. Se han desarrollado dispositivos que convierten caracteres impresos en patrones táctiles, con el propósito, entre otros, de capacitar a las personas con deficiencias visuales a “leer” textos de manera táctil. Uno de los dispositivos desarrollados, el Optacon, contiene una rejilla de pequeños puntos, cada uno de los cuales puede vibrar independientemente. Las letras del alfabeto producen distintos patrones de puntos que vibran. En investigaciones que involucran la integración temporal de tales patrones táctiles, Craig examinó la cantidad de integración entre los elementos individuales. En un estudio subsecuente, el experimentador manipuló los estímulos de inicio asincrónico (EIA) por parte de los patrones y el espaciamiento entre los renglones de puntos vibrantes, los cuales están en contacto con la yema del dedo de los sujetos. La tarea era indicar si había espaciamiento o no en los modelos presentados.

Se evaluaron cuatro sujetos entrenados durante un gran número de ensayos en los cuales el estímulo variaba en la sincronía inicial y el espaciamiento entre los estímulos. A fin de evaluar la precisión de los sujetos para detectar la presencia de un espacio, era necesario determinar que tan exactos eran los sujetos al reportar la ausencia de un espacio entre los sucesivos estímulos, sólo que también se varió el EIA. En cada sujeto se utilizaron seis diferentes EIA. En la tabla siguiente se resume la proporción de respuestas correctas para cada sujeto en cada condición. El principal interés del investigador era evaluar la hipótesis de que la exactitud estaba inversamente relacionada con los EIA

Sujeto	Estímulos de inicio asincrónico (ms)					
	204	104	56	30	13	0
A	0.797	0.873	0.888	0.923	0.942	0.956
B	0.794	0.772	0.908	0.982	0.946	0.913
C	0.838	0.801	0.853	0.951	0.883	0.837
D	0.815	0.801	0.747	0.859	0.887	0.902

$$H_0 : Me_1 = Me_2 = Me_3 = Me_4 = Me_5 = Me_6$$

$$H_1 : Me_1 < Me_2 < Me_3 < Me_4 < Me_5 < Me_6$$

$$\alpha = 0.05$$

Prueba Estadística

$$L = \sum_{i=1}^k iR_i$$

Desarrollo de la Prueba

$$L=342$$

$$Z=3.07 > 1.96 \text{ se rechaza } H_0$$

Conclusión

A un nivel de significación de 0.05 se puede afirmar que la exactitud estaba inversamente relacionada con los EIA

➤ **Secuencia o funciones con programas estadísticos**

En R

Se cuenta con la función `page.trend.test`, del paquete `crank`,
`page.trend.test(matriz de datos, ranks=FALSE)`

También existe la función `PageTest` del paquete `DescTools`
`PageTest(y, grupos, bloques)`

```
b<-4
k<-6
rango<-matrix(0,b,k)
for (i in 1:b){
  rango[i,]<-rank(datos[i,])
}
L<-sum(apply(rango,2,sum)*seq(1:k))
mu<-(b*k*(k+1)^2)/4
s2<-(b*k^2*(k^2-1)^2)/(144*(k-1))
z<-(L-mu)/sqrt(s2)
2*(1-pnorm(z))
```

➤ **Resultados con programas estadísticos**

Resultados con R

```
datos<-
matrix(c(0.797,0.873,0.888,0.923,0.942,0.956,0.794,0.772,0.
908,0.982,0.946,0.913,0.838,0.801,0.853,0.951,0.883,0.837,0.
.815,0.801,0.747,0.859,0.887, 0.902),nrow=4,byrow=TRUE)
library(crank)
page.trend.test(datos,ranks=FALSE)
```

Page test for ordered alternatives L = 342 p(table) <=.001

```
library(DescTools)
datos2<-read.delim("clipboard")
PageTest(datos2[,1],datos2[,2],datos2[,3])
```

Page test for ordered alternatives data: datos2[, 1], datos2[, 2] and datos2[, 3] L = 342, p-value = 0.0005661
--

D. Prueba de Durbin

➤ Aspectos Generales

En el Diseño de Bloques Completo al Azar se aplica cada tratamiento en cada bloque. Sin embargo, a veces es un poco complicado que se aplique todos los tratamientos a cada bloque, sobre todo cuando el número de tratamientos que deben aplicarse a cada bloque, es grande y el tamaño de bloque es limitado. Además, si el diseño es equilibrado de manera que cada bloque contiene t unidades experimentales, cada tratamiento aparece en r bloques, y cada tratamiento aparece con cada otro tratamiento un número igual de veces, el diseño se llama un Diseño de Bloque Incompleto Equilibrado.

Durbin presentó una prueba de rangos que se puede utilizar para probar la hipótesis nula de que no hay diferencia entre los tratamientos en un Diseño de Bloques Incompletos balanceados.

➤ Supuestos

- La escala de medida es al menos ordinal, de modo que las observaciones pueden ser ordenadas dentro de cada bloque.
- Los b bloques son mutuamente independientes, es decir los resultados de un bloque no influyen en los resultados de los otros bloques.
- No debe existir dependencia entre los bloques.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

Los elementos considerados en el diseño son:

- k : número de tratamientos.
- t : número de unidades experimentales por bloque ($t < k$).
- b : número de bloques.
- r : número de bloques utilizados para cada tratamiento ($r < b$).
- λ : número de bloques en los cuales aparecen juntos los k tratamientos.

Para la obtención del estadístico de contraste tenemos lo siguiente:

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 t . Sea la suma de los rangos correspondientes al tratamiento j ,

$$R_j = \sum_{i=1}^b R_{ij}$$

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 $\hat{X}_i = (X_{i1}, \dots, X_{it})$ son independientes.

El esperado y varianza del cada uno de los tratamientos

$$E(R_{ij}) = \frac{t+1}{2} \quad V(R_{ij}) = \frac{t^2-1}{12}$$

En general:

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

Se define el estadístico de Durbin como:

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

Si hay relación dentro de los bloques, se hace uso de los rangos promedios y es necesario hacer un ajuste. La suma de los cuadrados de las filas y filas medias usadas está dado por:

$$A = \sum_{i=1}^t \sum_{j=1}^b R_{ij}^2$$

Para calcular el factor de corrección dado por C para el ajuste para empates será:

$$C = \frac{bt(t+1)^2}{4}$$

Entonces, el estadístico corregido por empates es:

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

Bajo H_0 , T tiene una distribución asintótica χ^2 con k-1 grados de libertad, por lo tanto, se rechaza H_0 si $T > \chi_{(k-1, 1-\alpha)}^2$.

En el caso de que se rechace H_0 , se debe realizar las **comparaciones múltiples**, para verificar cuales son los pares de tratamiento que son significativamente diferentes. A continuación mostramos el procedimiento:

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

En el caso de no existir empates, se reduce a:

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

➤ **Aplicación**

A un grupo de estudiantes se le pidió que clasifique con puntajes del 1 al 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. Cada definición puede ser elegida como máximo por 7 estudiantes.

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

Procedimiento:

1. Los valores que obtenemos antes de realizar el cálculo, tenemos:

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

2. Planteamiento de la hipótesis:

$$H_0: Me_1 = Me_2 = \dots = Me_k, k = 1, \dots, 8$$

$$H_1: \text{Existe al menos un par } (i, j) \text{ tal que } Me_i \neq Me_j$$

3. Cálculo del estadístico:

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

$$= \frac{12(8-1)}{(7)(8)(4^2-1)} \left[\left(11 - \frac{7(4+1)}{2} \right)^2 + \dots \right. \\ \left. + \left(27 - \frac{7(4+1)}{2} \right)^2 \right]$$

$$T = 34.2$$

El valor para $\chi^2_{(7,1-0.05)} = 14.07$

Al ser el valor de $T > \chi^2_{(7,1-0.05)}$, rechazamos H_0 . Entonces a un nivel de significación del 0.05 se encontró suficiente evidencia para afirmar que existe al menos un par de definiciones de integración racial son diferentes.

Comparaciones múltiples

A consecuencia de rechazar H_0 , debemos realizar este procedimiento para identificar los pares significativamente diferentes.

Tenemos:

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

$$t_{35,1-1-0.975} \left[\frac{7 \times 4(4+1)}{6(14 \times 4 - 14 - 8 + 1)} (14(4-1) - 34.2) \right]^{1/2} = 4.63$$

Se obtiene:

	$ R_i - R_j $	Valor de Comparación	Resultado
A-B	15	4.63	*
A-C	4	4.63	n.s.
⋮	⋮	⋮	⋮
F-G	10	4.63	*
F-H	7	4.63	*

Mostramos a través del siguiente cuadro, como se dieron las diferencias significativas indicadas con un (*):

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

➤ **Funciones con R**

Se cuenta con la función *durbin.test(bloques, tratamientos, datos)* del paquete *agricolae* donde se ordena los datos por jueces o bloques estudiantes y por orden de definiciones (letras de la A:F). Obtenemos el siguiente resultado:

```
library(agricolae)
estudiantes<-gl(14,4)
definicion<-
c(1,3,4,6,2,3,5,6,1,2,6,7,4,5,6,7,2,5,7,8,1,4,7,8,3,4,5,8,1
,2,3,8,3,6,7,8,1,5,6,8,2,4,6,8,1,2,4,5,2,3,4,7,1,3,5,7)
puntajes<-
c(3,1,4,2,4,1,3,2,2,4,1,3,2,4,1,3,4,3,2,1,1,2,4,3,1,3,4,2,2
,4,1,3,1,2,3,4,1,4,2,3,4,2,1,3,1,3,2,4,4,2,1,3,2,1,4,3)
```

```
durbin.test(estudiantes,definicion,puntajes,group=T,console
=T)
```

```
Study: puntajes ~ estudiantes + definicion
definicion, Sum of ranks
```

```
sum
1 12
2 27
3 8
4 16
5 26
6 11
7 21
8 19
```

```
Durbin Test
=====
```

```
Value      : 34.2
DF 1       : 7
P-value    : 1.580183e-05
Alpha      : 0.05
DF 2       : 35
t-Student  : 2.030108
```

```
Least Significant Difference
between the sum of ranks: 4.629358
```

```
Parameters BIB
Lambda      : 3
Treatmeans  : 8
Block size  : 4
Blocks      : 14
Replication : 7
```

Treatments with the same letter are not significantly different.

```
Sum of ranks groups
2      27      a
5      26      a
7      21      b
```

8	19	bc
4	16	cd
1	12	de
6	11	e
3	8	e

Como opción, la función `durbin.test(datos, tratamientos, bloques)` del paquete *PMCMR*. Obtenemos los siguientes resultados:

```
library(PMCMR)
durbin.test(puntajes, definicion, estudiantes)
```

```
Durbin rank sum test
data: puntajes, definicion and estudiantes
Durbin chi-squared = 34.2, df = 7, p-value = 1.58e-05
```

Los resultados con ambas funciones de los paquetes *agricolae* y *PMCMR* para la prueba Durbin, coinciden en el valor del estadístico de Durbin, pero la función `durbin.test` no nos da las comparaciones múltiples directamente, pero este paquete nos proporciona la función `posthoc.durbin.test()` para realizarlo.

```
posthoc.durbin.test(puntajes, definicion, estudiantes, p.adj="none")
```

```
Pairwise comparisons using Durbin's test for a two-way
balanced incomplete block design
```

```
data: puntajes, definición and estudiantes
```

	1	2	3	4	5	6	7
2	1.3e-07	-	-	-	-	-	-
3	0.08816	8.0e-10	-	-	-	-	-
4	0.08816	2.7e-05	0.00126	-	-	-	-
5	5.1e-07	0.66370	2.8e-09	0.00010	-	-	-
6	0.66370	3.6e-08	0.19687	0.03507	1.3e-07	-	-
7	0.00036	0.01257	1.9e-06	0.03507	0.03507	0.00010	-
8	0.00412	0.00126	2.7e-05	0.19687	0.00412	0.00126	0.38644

E. Prueba de Skillings-Mack

➤ Aspectos Generales

La prueba “Skillings Mack”, es una prueba no paramétrica que se utiliza como variante de la prueba de Friedman, cuando se está trabajando con un diseño en bloques completamente al azar (DBCA) que tiene “m” datos faltantes de forma arbitraria. Estos datos faltantes son considerados unidades perdidas al azar. La prueba fue presentada por primera vez en “*On the Use of a Friedman-Type Statistic in Balanced and Unbalanced Block Designs*” (Skillings, J.H., Mack, G.A, 1981).

La estadística de prueba de Skillings-Mack es una generalización de la estadística de prueba que se utiliza en la prueba de Friedman y su p-valor está basado en la distribución Chi-cuadrado y el método de Monte Carlo para aproximar el p-valor cuando hay muchos vínculos y pequeños diseños que llevan a cabo los valores perdidos.

➤ Supuestos

- La escala de medida de las variables debe ser al menos ordinal.
- Se tiene un diseño en bloques completamente al azar con “m” datos faltantes.
- Los bloques son mutuamente independientes (los resultados dentro de un bloque no influyen en los resultados dentro de los otros).

➤ Inferencia Estadística

Considerando un diseño de bloques al azar con $j = 1, 2, 3, \dots, k$ tratamientos, $i = 1, 2, \dots, b$ bloques y X_{ij}

		Tratamiento				
		1	...	j	...	k
Bloques	1	X_{11}	...	X_{1j}	...	X_{1k}
				
	i	X_{i1}	...	X_{ij}	...	X_{ik}
				
	b	X_{b1}	...	X_{bj}	...	X_{bk}

Para hallar la estadística de prueba de “Skillings-Mack Test”, se procede de la siguiente manera:

- ❖ Calcular el vector S_i , el cual contendrá el número total de tratamientos para cada bloque i .
- ❖ Determinar el ranqueo R_{ij} entre bloques:
 - ✓ Si la observación está presente, $R_{ij} = \text{Ranqueo de } X_{ij} \text{ en el bloque } i$.
 - ✓ Si es una observación faltante, $R_{ij} = \frac{(S_i + 1)}{2}$.

En cada bloque, S_i = Número de tratamientos en el bloque i . Cuando se obtienen valores iguales, estos se promedian para obtener el ranqueo promedio.

- ❖ Calcular la suma ajustada del j -ésimo tratamiento:

$$A_j = \sum_{i=1}^k \left(\frac{12}{S_i + 1} \right)^{\frac{1}{2}} \left(R_{ij} - \frac{S_i + 1}{2} \right)$$

Luego, armar la matriz $A = (A_j, \dots, A_{k-1})$

- ❖ Determinar la matriz de covarianza para A:

$$\sum_0 = \begin{bmatrix} \sum_{t=2}^k \lambda_{1t} & -\lambda_{12} & -\lambda_{13} & \dots & -\lambda_{1,k-1} \\ -\lambda_{12} & \sum_{t \neq 2}^k \lambda_{2t} & -\lambda_{23} & \dots & -\lambda_{2,k-1} \\ \cdot & \cdot & \cdot & & \cdot \\ -\lambda_{1,k-1} & -\lambda_{2,k-1} & -\lambda_{3,k-1} & \dots & \sum_{t \neq k-1}^k \lambda_{k-1,t} \end{bmatrix}$$

Para $t \neq q = 1, \dots, k$ (k tratamientos), donde

$\lambda_{qt} = \lambda_{tq}$ = al número de bloques donde q y t tratamientos son observados.

- ❖ Calcular la estadística de Skilling-Mack:

$$SM = A \sum_0^{-} A'$$

Donde \sum_0^{-} es la inversa generalizada de \sum_0 (matriz de covarianza de A)

Prueba de Hipótesis:

En la hipótesis nula para la estadística de Skilling-Mack no existe diferencias significativas en los τ_k efectos de los tratamientos versus la hipótesis alterna de que al menos uno de los tratamientos es diferente a los demás:

$$H_0 : Me_1 = \dots = Me_k$$

H_1 : Al menos un Me_k es diferente a los demás.

La hipótesis nula se rechaza si el valor de la estadística es mayor que o igual a un valor crítico. Sin embargo, porque hay un gran número de combinaciones de diseños de bloques aleatorios no $\sum = 0$ balanceados, no hay un conjunto completo de tablas publicadas para estos valores críticos a excepción de un número limitado de casos especiales (Skillings y Mack, 1981). Estas tablas proporcionan una gran aproximación de la muestra al p-valor basado en la distribución Chi-cuadrado con $k-1$ grados de libertad y un nivel de significación α . La hipótesis nula se rechaza si la estadística de $Skills - Mack \geq \chi_{k-1, \alpha}^2$

Skilling y Mack (1981) señalan que el p-valor de la aproximación Chi-cuadrado es en la mayoría de los casos conservador, sobre todo con el error tipo I de 0.01 o menos. Por ello, se recomienda la simulación para obtener los valores críticos para el estadístico Skillings-Mack. Existe un método para simular todas las posibles clasificaciones para la s_i valores observados, preservando al mismo tiempo las filas de las observaciones faltantes. Para cada clasificación simulada, la estadística de Skillings-Mack se calcula y se compara con el valor real de la estadística. El p-valor simulado es la proporción de estadísticas simuladas Skillings-Mack que son mayores o igual a la estadística real, dividido por el número de simulaciones.

Se recomienda el método de Monte Carlo para aproximar el p-valor cuando hay muchos vínculos y pequeños diseños que llevan a cabo los valores perdidos.

➤ **Aplicación**

Los aditivos para gasolina son compuestos formulados para mejorar la calidad y eficiencia de los combustibles en los motores de automóviles. Cuando añadimos un aditivo para gasolina, la mejor combustión incrementará de manera porcentual la potencia del motor. Como resultado final, al tener más potencia del motor se le exige menos al mismo. Por lo tanto, uno de los resultados últimos de utilizar aditivos para la gasolina, es que podremos obtener más kilómetros de nuestro tanque.

Un ingeniero desea saber si 4 marcas de aditivos para gasolina tienen el mismo efecto en el motor. Para ello, en una prueba de carretera, evalúa el rendimiento del motor (en millas) de 9 autos, los cuales son considerados como bloques.

Los resultados son los siguientes.

Autos	Aditivos de Gasolina			
	A	B	C	D
1	3.2	4.1	3.8	4.2
2	3.1	3.9	3.4	4
3	4.3	3.5	4.6	4.8
4	3.5	3.6	3.9	4
5	3.6	4.2	3.7	3.9
6	4.5	4.7	3.7	-
7	-	4.2	3.4	-
8	4.3	4.6	4.4	4.9
9	3.5	-	3.7	3.9

Planteamiento de la Prueba de Hipótesis:

$$H_0 : Me_1 = Me_2 = Me_3 = Me_4$$

$$H_1 : \text{Al menos un } Me_i \text{ es diferente a los demás } i = 1, 2, 3, 4.$$

$$\alpha = 0.05$$

Desarrollo de la prueba estadística:

- ❖ Cálculo del vector S_i = (Número total de observaciones por cada bloque)

$$S_i = \{4, 4, 4, 4, 4, 3, 2, 4, 3\}$$

- ❖ Se procede a rankear por bloques:

- Se ordenan las observaciones de cada bloque de menor a mayor y se les asigna a cada una un número del 1 a S_i según el orden. Por ejemplo, para el bloque 1, el menor valor para el tratamiento A es 3.2, por ello en el siguiente cuadro es remplazado por el número 1.
- En el caso de tener una observación faltante, se reemplazará con el valor resultante de la fórmula $R_{ij} = \frac{(S_i + 1)}{2}$. Por ejemplo, para el bloque 7, existen dos valores faltantes. Reemplazando los valores en la fórmula, obtenemos el valor de ranqueo de 1.5 para cada observación.

❖ Calcular los valores de $A_j = \text{Suma Ajustada para el } j - \text{ésimo tratamiento.} :$

$$A_j = \sum_{i=1}^k \left(\frac{12}{S_i + 1} \right)^{\frac{1}{2}} \left(R_{ij} - \frac{S_i + 1}{2} \right)$$

Bloques	Tratamientos			
	A	B	C	D
1	-2.3238	0.7746	-0.7746	2.3238
2	-2.3238	0.7746	-0.7746	2.3238
3	-0.7746	-2.3238	0.7746	2.3238
4	-2.3238	-0.7746	0.7746	2.3238
5	-2.3238	2.3238	-0.7746	0.7746
6	0	1.7321	-1.7321	0
7	0	1	-1	0
8	-2.3238	0.7746	-0.7746	2.3238
9	-1.7321	0	0	1.7321
A_j	-14.1256	4.2812	-4.2812	14.1256

Por ejemplo, para el tratamiento A ($j = 1$):

$$A_1 = \left(\frac{12}{4+1} \right)^{\frac{1}{2}} \left(1 - \frac{4+1}{2} \right) + \dots + \left(\frac{12}{3+1} \right)^{\frac{1}{2}} \left(1 - \frac{3+1}{2} \right) = -14.1256$$

$$A = [-14.13 \quad 4.28 \quad -4.28 \quad 14.13]$$

La matriz $A = (A_j, \dots, A_{k-1})$, por ello se eligen los 3 primeros elementos de la matriz anterior.

$$A = [-14.13 \quad 4.28 \quad -4.28]$$

❖ Matriz de Covarianza para A:

$$\sum_0 = \begin{bmatrix} \sum_{t=2}^k \lambda_{1t} & -\lambda_{12} & -\lambda_{13} & \dots & -\lambda_{1,k-1} \\ -\lambda_{12} & \sum_{t \neq 2}^k \lambda_{2t} & -\lambda_{23} & \dots & -\lambda_{2,k-1} \\ . & . & . & & . \\ -\lambda_{1,k-1} & -\lambda_{2,k-1} & -\lambda_{3,k-1} & \dots & \sum_{t \neq k-1}^k \lambda_{k-1,t} \end{bmatrix}$$

$\lambda_{12} = \text{Número de bloques en los que los tratamientos } q = 1 \text{ y } t = 2 \text{ son observados} = 7.$

$$\sum_0 = \begin{bmatrix} 22 & -7 & -8 & -7 \\ -7 & 21 & -8 & -6 \\ -8 & -8 & 23 & -7 \\ -7 & -6 & -7 & 20 \end{bmatrix}$$

Final \sum_0 :

$$\sum_0 = \begin{bmatrix} 22 & -7 & -8 \\ -7 & 21 & -8 \\ -8 & -8 & 23 \end{bmatrix}$$

❖ La inversa generalizada de \sum_0 es \sum_0^- :

$$\sum_0^- = \begin{bmatrix} 419 & 225 & 224 \\ 225 & 442 & 232 \\ 224 & 232 & 413 \end{bmatrix}$$

❖ La estadística Skilling-Mack seria:

$$SM = A \sum_0^- A'$$

$$SM = \begin{bmatrix} -14.13 & 4.28 & -4.28 \end{bmatrix} \begin{bmatrix} 419 & 225 & 224 \\ 225 & 442 & 232 \\ 224 & 232 & 413 \end{bmatrix} \begin{bmatrix} -14.13 \\ 4.28 \\ -4.28 \end{bmatrix} = 15.49$$

$$SM = 15.49 \sim \chi^2_{(\alpha=0.05, k-1=3)}$$

$$\chi^2_{(\alpha=0.05, k-1=3)} = 7.814733 < 15.49, \text{ se rechaza } H_0.$$

Conclusión

- A un nivel de significancia del 5%, existe suficiente evidencia estadística para rechazar H_0 . Por lo tanto, se puede afirmar que las 4 diferentes marcas de aditivos para gasolina no tienen el mismo efecto en el motor.

➤ Funciones con R

Para realizar la prueba Skillings-Mack se hará uso de la función Ski.Mack del paquete Skillings.Mack

```
library(Skillings.Mack)
B <- rep(c(1,2,3,4,5,6,7,8,9), rep(4,9))
G <- rep(c('A','B','C','D'),9)
y <- c(3.2,4.1,3.8,4.2, 3.1,3.9,3.4,4.0, 4.3,3.5,4.6,4.8,
3.5,3.6,3.9,4.0, 3.6,4.2,3.7,3.9, 4.5,4.7,3.7, NA,NA
,4.2,3.4,NA , 4.3,4.6,4.4,4.9, 3.5, NA,3.7, 3.9)
Ski.Mack(y,groups = G,blocks = B)
```

```
skillings-mack Statistic = 15.493049 , p-value = 0.001440
Note: the p-value is based on the chi-squared distribution with d.f. =
3

$`nblocks`
```



```
[1] 9

$Ntreatments
[1] 4

$rawdata
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]  3.2  3.1  4.3  3.5  3.6  4.5  NA  4.3  3.5
[2,]  4.1  3.9  3.5  3.6  4.2  4.7  4.2  4.6  NA
[3,]  3.8  3.4  4.6  3.9  3.7  3.7  3.4  4.4  3.7
[4,]  4.2  4.0  4.8  4.0  3.9  NA  NA  4.9  3.9

$rankdata
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1    1    2    1    1    2  1.5    1    1
[2,]    3    3    1    2    4    3  2.0    3    2
[3,]    2    2    3    3    2    1  1.0    2    2
[4,]    4    4    4    4    3    2  1.5    4    3

$varCovarMatrix
      [,1] [,2] [,3] [,4]
[1,]   22   -7   -8   -7
[2,]   -7   21   -8   -6
[3,]   -8   -8   23   -7
[4,]   -7   -6   -7   20

$adjustedSum
      [,1]      [,2]      [,3]      [,4]
[1,] -14.12559752  4.281244146 -4.281244146  14.12559752
```

3. Para un parámetro de locación o escala

A. Prueba de Permutación

➤ Aspectos Generales

Pruebas relacionadas y bloqueo son técnicas de diseños experimentales que son utilizados cuando un investigador desea detectar diferencias entre tratamientos en unidades experimentales que presentan alta variabilidad.

Si se cumplen los supuestos de normalidad y homogeneidad de varianzas se puede hacer uso de el Diseño de Bloques Completos al Azar (DBCA), pero si uno de los supuestos no es verificado se puede utilizar como alternativa la prueba de permutación para k muestras relacionadas.

➤ Supuestos

- Los b bloques son mutuamente independientes, es decir, los resultados de un bloque no influyen en los resultados de los otros bloques.
- La escala de medida es al menos ordinal, de modo que las observaciones pueden ser ordenadas dentro de cada bloque.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Obtener el estadístico F para los datos originales, denotado por F_{obs} .
- Permutar las observaciones dentro de cada bloque, haciendo lo mismo para todos los bloques (no permutar entre bloques). Hay $k!^b$ posibilidades.
- Para cada posibilidad en el paso 2, calcular el estadístico F para el diseño de bloques completos al azar.
- Dado que valores grandes de F indican diferencia entre los tratamientos, obtener el pvalor con cola superior de la siguiente manera

$$p - \text{valor} = \frac{\# F's \geq F_{obs}}{(k!)^b}$$

Si no es posible obtener todas las permutaciones, entonces se podría tomar una muestra aleatoria de ellas

Las hipótesis son:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \text{Al menos una } \mu_i \text{ es diferente a las demás } i = 1, 2, \dots, k$$

Donde μ_i es una medida de posición.

➤ Aplicación

Simule datos para tres tratamientos con cinco repeticiones cada uno. Pruebe si existen diferencias significativas

➤ Funciones con R

```
En R
#Generar los datos
set.seed(10)
T1<-round(rexp(5,1/10),1)
T2<-round(rexp(5,1/13),1)
```

```
T3<-round(rexp(5,1/18),1)
tiempo<-c(T1,T2,T3)
factor<-as.factor(rep(1:3,c(5,5,5)))
bloque<-as.factor(rep(1:5,3))
todo<-cbind(tiempo,factor,bloque)
or<-todo[order(todo[,3]),]
Fobs<-
anova(lm(or[,1]~as.factor(or[,2])+as.factor(or[,3])))$F[1]
Fobs
R<-100
Fper<-rep(0,R)
for (i in 1:R){
  b1<-sample(or[1:3,1],3,F)
  b2<-sample(or[4:6,1],3,F)
  b3<-sample(or[7:9,1],3,F)
  b4<-sample(or[10:12,1],3,F)
  b5<-sample(or[13:15,1],3,F)
  resp<-c(b1,b2,b3,b4,b5)
  Fper[i]<-
anova(lm(resp~as.factor(or[,2])+as.factor(or[,3])))$F[1]
}

pvalor<-mean(ifelse(Fper>=Fobs,1,0))
pvalor
```