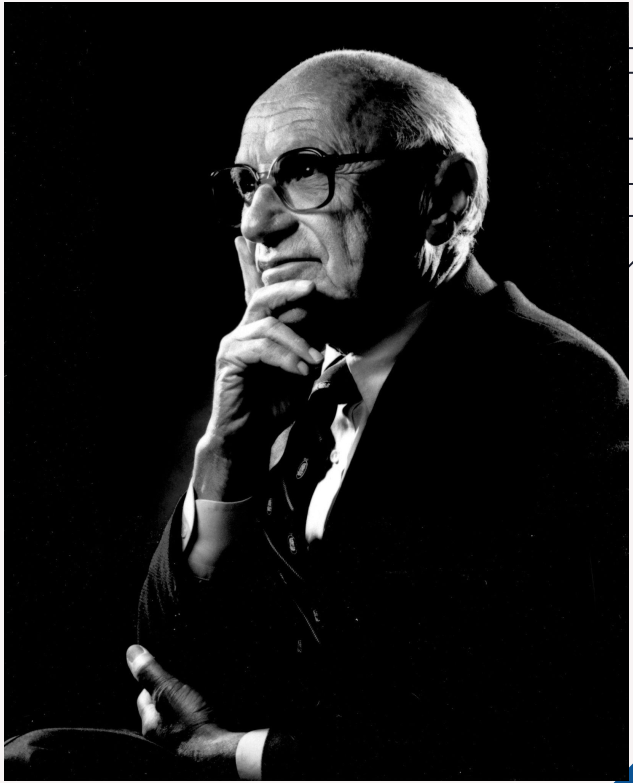
TEST DE FRIEDMAN

Integrantes: Daniel Felipe Castellanos
Diego Phelipe Morales
Jhon Alejandro Ramirez
Natalia Buriticá Nuñez
Miguel Angel Martin

TABLA DE CONTENIDO

Milton Friedman

31 de Julio de 1912 (NY, Nueva York)-
16 de Noviembre de 2006 (CA, San Francisco)



Idea Intuitiva

Supongamos que tenemos 4 marcas de cerveza, de la cuales van a ser degustadas por 10 catadores.
Los cuales van a valorar de 1 a 5 los siguientes aspectos;
Color, Aspecto, Aroma y Sabor.
Donde 1 es la nota y 5 la nota mayor.

Características

- Generaliza el test del Signo de muestras pareadas a más de dos tratamientos.
- Hay independencia entre individuos, pero no entre tratamientos.
- K tratamientos, B bloques.
- Hipótesis: H_0: Igualdad de medianas.

H_a: Diferencia entre al menos un par de medianas.

- Variable: Ordinal (ver Artículo de Friedman).

Variable dependiente	Una muestra (bondad de ajuste)	Muestras relacionadas		Muestras independientes	
		2 muestras	>2 muestras	2 muestras	>2 muestras
Nominal	Binomial Chi-Cuadrado Rachas	McNemar	Cochran	-	-
Ordinal/ Intervalo	Kolmogorov-Smirnov	Signos Wilcoxon	Friedman Kendall	Rachas de Wald-Wolfowitz U de Mann-Whitney Moses Kolmogorov-Smirnov	Mediana Kruskal-Wallis Jonckheere-Terpstra

Muestra	Prueba paramétrica	Prueba no paramétrica
Muestras relacionadas		
2 muestras	t-Student	Wilcoxon
> 2 muestras	ANOVA	Friedman
Muestras independientes		
2 muestras	t-Student	U de Mann-Whitney
> 2 muestras	ANOVA	Kruskal-Wallis

Fuente: Berlanga y Rubio (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. REIRE (5)2, p. 103.

Arreglo de datos

Suponiendo una observación por cada bloque

	Factor		
Bloque	Nivel 1	...	Nivel k
1	y_{11}	$\cdots$	y_{1k}
2	y_{21}	$\cdots$	y_{2k}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_{n1}	$\cdots$	y_{nk}
Mediana	$\tilde{y}_1$	$\cdots$	$\tilde{y}_k$

Arreglo de datos

Suponiendo una observación por cada bloque

Factor

Bloque	Nivel 1	...	Nivel k
1	y_{11}	$\cdots$	y_{1k}
2	y_{21}	$\cdots$	y_{2k}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_{n1}	$\cdots$	y_{nk}
Mediana	$\tilde{y}_1$	$\cdots$	$\tilde{y}_k$

Participante	Mañana	Tarde	Noche
1	9	5	2
2	6	3	1
3	5	5	5
4	11	11	1
5	8	8	3
6	10	10	1
7	7	7	4

Competidor	Juez1	Juez2	Juez3	Juez4
1	8.5	8.6	8.2	8.4
2	9.8	9.7	9.4	9.6
3	7.9	8.1	7.5	8.2
4	9.7	9.8	9.6	9.6
5	6.2	6.8	6.9	6.5
6	8.9	9.2	8.1	8.7
7	9.2	9.2	8.7	8.9
8	8.4	8.5	8.4	8.6
9	9.2	9.6	8.9	9.5
10	8.8	9.2	8.6	9.30

Hipótesis

En términos poblacionales: Suponga que θ es la mediana total

$$H_0 : Y_{ij} = \theta + \epsilon_{ij}, i = 1, \ldots, n; \; j = 1, \ldots, k$$
$$H_a : Y_{ij} = \theta + \beta_i + \lambda_j + \epsilon_{ij},$$

donde las variables ϵ_{ij} son una muestra aleatoria simple (m.a.s.) de una distribución continua con mediana 0.

Definición del estadístico de prueba.

$$S = \frac{12}{nk(k+1)}\left(\sum_{j=1}^{k} R_j^2\right) - 3n(k+1)$$

- n: el tamaño de la muestra.
- k: el número de grupos.
- R_j: el rango de la puntuación del grupo j.
- $\sum_{j=1}^{k}$: la suma de los valores de R_j para todos los grupos, desde $j = 1$ hasta $j = k$.
- $-\frac{3n(k+1)}{2}$: la media de la estadística de Friedman, que es una suma de cuadrados de rangos.

Distribución asintótica.

Esta converge a una distribución normal por el teorema Lindeberg-Lévy.

$$\frac{\sum_{j=1}^{k} R_j - E(R_j)}{\sqrt{Var(R_j)}}$$

La estadística S se construye a partir de la estadística de Friedman. La estadística de Friedman se define como:

$$\sum_{j=1}^{k} R_j$$

donde Rj es el rango de la puntuación del grupo j.

Para construir la estadística S, primero necesitamos calcular la media y la varianza de la estadística de Friedman. La media de la estadística de Friedman es:

$$\sum_{j=1}^{k} E(R_j)$$

La varianza de la estadística de Friedman es:

$$\sum_{j=1}^{k} var(R_j)$$

La media y la varianza de la estadística de Friedman se pueden calcular de la siguiente manera:

$$\sum_{j=1}^{k} \frac{n(k+1)^2}{4} = \frac{3kn(k+1)}{2}$$

$$\sum_{j=1}^{k} \frac{k(k+1)(2k+1)}{6} = \frac{k(k+1)(2k+1)(k)}{6}$$

Una vez que tenemos la media y la varianza de la estadística de Friedman, podemos construir la estadística S de la siguiente manera:

$$S = \left(\frac{\sum_{j=1}^{k} R_j - E(R_j)}{\sqrt{Var(R_j)}}\right)^2$$

$$S = \frac{\sum_{j=1}^{k} R_j^2 - \frac{3kn(k+1)}{2}}{\sqrt{\frac{k(k+1)(2k+1)(k)}{6}}}$$

$$S = \frac{12}{nk(k+1)}\left(\sum_{j=1}^{k} R_j^2\right) - 3n(k+1)$$

Tratamiento de empates.

$$S_{\text{emp}} = \frac{S}{C}, \quad C = 1 - \frac{\sum_{i=1}^{m}(t_i^3 - t_i)}{n(k^3 - k)}$$

- t_i: Número de empates en el i-ésimo grupo de empates
- m: Número de grupos con rangos empatados

Decisión:

Como ya vimos $S \sim \chi^2_{k-1}$

Por lo que rechaza la hipótesis nula de igualdad de medianas cuando:

$$S > q_{1-\alpha}^{\chi^2_{k-1}} \quad P(\chi^2 > S) < \alpha$$

Test de Mack-Skillings: Bloques con varias unidades

Table: Arreglo de datos suponiendo que hay c observaciones por bloque

	Factor		
Bloque	Nivel 1	...	Nivel k
1	y_{111}	$\cdots$	y_{1k1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	y_{11c}	$\cdots$	y_{1kc}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{n11}	$\cdots$	y_{nk1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{n1c}	$\cdots$	y_{nkc}

Test de Mack-Skillings: Bloques con varias unidades

$$S_j = \sum_{i=1}^{n} \left[\sum_{q=1}^{c} r_{ijq}/c \right], \quad \text{for } j = 1, \ldots, k.$$

$$MS = \left[\frac{12}{k(N+n)} \right] \sum_{j=1}^{k} \left[S_j - \frac{N+n}{2} \right]^2,$$

$$= \left[\frac{12}{k(N+n)} \right] \left\{ \sum_{j=1}^{k} S_j^2 \right\} - 3(N+n),$$

Test de Mack-Skillings: Bloques con varias unidades

Donde:

- K número de tratamientos
- n bloques de tamaño c
- N son el total de datos observados.

$$MS > q_{1-\alpha}^{\chi_{k-1}^2}$$

$$P(\chi^2 > MS) < \alpha$$

EJEMPLO : Competencia de salto

En una competencia de salto olímpico, 4 jueces califican una competencia de salto que incluye 10 competidores. La calificación que cada juez le da a los 10 competidores, toma valores entre 1 y 10

Competidor	Juez1	Juez2	Juez3	Juez4
1	8.5	8.6	8.2	8.4
2	9.8	9.7	9.4	9.6
3	7.9	8.1	7.5	8.2
4	9.7	9.8	9.6	9.6
5	6.2	6.8	6.9	6.5
6	8.9	9.2	8.1	8.7
7	9.2	9.2	8.7	8.9
8	8.4	8.5	8.4	8.6
9	9.2	9.6	8.9	9.5
10	8.8	9.2	8.6	9.30

Competidor	Juez1	Juez2	Juez3	Juez4
1	3	4	1	2
2	4	3	1	2
3	2	3	1	4
4	3	4	1.5	1.5
5	1	3	4	2
6	3	4	1	2
7	3.5	3.5	1	2
8	1.5	3	1.5	4
9	2	4	1	3
10	2	3	1	4
R_j	25	34.5	14	26.5

- Se le asignan rangos a las calificaciones dadas, donde se le otorga un 1 a la calificación más baja

- Cuando hay un empate se promedian las posiciones de las puntuaciones empatadas

H0= La mediana de la calificación de los jueces es la misma.
H1= Hay diferencias significativas en la calificación de los jueces.

$$R_1 = 25,\ R_2 = 34.5,\ R_3 = 14,\ R_4 = 26.5$$

$$S_c = \frac{12(25^2 + 34.5^2 + 14^2 + 26.5^2)}{(10)(4)(5)} - 3(10)(5) = 12.81$$

$$C = 1 - \frac{(2^3 - 2) + (2^3 - 2) + (2^3 - 2)}{10(4^3 - 4)} = 0.97$$

$$S_{\text{emp}} = \frac{S}{C} = 13.2$$

$$\chi^2_{0.01,3} = 11.34$$

ALL YOU CAN EAT

Tratamiento	1h	5h	8h	12h
Cupcake 1	8	7	9	6
Cupcake 2	6	5	7	4
Cupcake 3	9	8	8	7
Cupcake 4	5	4	6	3
Cupcake 5	7	8	6	5
Cupcake 6	4	7	9	6
Cupcake 7	6	5	7	4
Cupcake 8	9	8	8	7

Tratamiento	1h	5h	8h	12h
1	3	2.0	4.0	1
2	3	2.0	4.0	1
3	4	2.5	2.5	1
4	3	2.0	4.0	1
5	3	4.0	2.0	1
6	1	3.0	4.0	2
7	3	2.0	4.0	1
8	4	2.5	2.5	1

```
        Friedman rank sum test

data:  Valor and Momento and Tratamiento
Friedman chi-squared = 14.308, df = 3, p-value = 0.002515
```

```r
require(PMCMR)
require(PMCMRplus)
require(Skillings.Mack)
library(reshape2)


datos <- data.frame(
  Tratamiento = rep(c("1", "2", "3", "4", "5", "6", "7", "8"), each = 4),
  Momento = rep(c("1h", "5h", "8h", "12h"), times = 8),
  Valor = c(8, 7, 9, 6, 6, 5, 7, 4, 9, 8, 8, 7, 5, 4, 6, 3, 7, 8, 6, 5, 4, 7, 9, 6, 6, 5, 7, 4, 9, 8, 8, 7)
)

datos$Momento <- factor(datos$Momento, levels = c("1h", "5h", "8h", "12h"))

datos_organizados <- dcast(datos, Tratamiento ~ Momento, value.var = "Valor")

friedman.test(Valor ~ Momento | Tratamiento, data = datos)
frdAllPairsNemenyiTest(Valor ~ Momento | Tratamiento, data = datos)
```

```r
datos2 <- data.frame(
  Tratamiento = rep(c("1", "2", "3", "4", "5", "6", "7", "8"), each = 4),
  Momento = rep(c("1h", "5h", "8h", "12h"), times = 8),
  Valor = c(3, 2, 4, 1, 3, 2, 4, 1, 4, 2.5, 2.5, 1, 3, 2, 4, 1, 3, 4, 2, 1, 1, 3, 4, 2, 3, 2, 4, 1, 4, 2.5, 2.5, 1)
)

datos2$Momento <- factor(datos2$Momento, levels = c("1h", "5h", "8h", "12h"))

datos2_organizados <- dcast(datos2, Tratamiento ~ Momento, value.var = "Valor")

# Sumar todos los valores para cada momento
sumas_por_momento <- aggregate(Valor ~ Momento, data = datos2, FUN = sum)

ran= sumas_por_momento$Valor

S = 12*(sum(ran^2))/(8*4*(5)) - 3*8*5
C = 1 - (2*(2^3-2))/(8*(4^3-4))

S_emp = S/C
1-pchisq(S-emp,3)
```

EJEMPLO POR BLOQUES

TABLE 7.20. Amount of Niacin in Enriched Bran Flakes

	Amount of Niacin Enrichment (Milligrams per 100 g Bran Flakes)		
Laboratory	0	4	8
1	7.58 (3)	11.63 (7)	15.00 (2)
	7.87 (8)	11.87 (11)	15.92 (9)
	7.71 (6)	11.40 (3)	15.58 (4)
2	8.00 (9.5)	12.20 (12)	16.60 (12)
	8.27 (12)	11.70 (8.5)	16.40 (11)
	8.00 (9.5)	11.80 (10)	15.90 (7)
3	7.60 (4)	11.04 (2)	15.87 (6)
	7.30 (1)	11.50 (5.5)	15.91 (8)
	7.82 (7)	11.49 (4)	16.28 (10)
4	8.03 (11)	11.50 (5.5)	15.10 (3)
	7.35 (2)	10.10 (1)	14.80 (1)
	7.66 (5)	11.70 (8.5)	15.70 (5)

Figure: Ver página 331 Hollander & Wolfe.

Table: Respuesta: mg/100gr de Niacina. 4 tratamientos (laboratorios), 3 bloques (enriquecimiento de Niacina) y 3 réplicas

Laboratorio

Bloque	1	2	3	4
0	7.58 (3)	8.00 (9.5)	7.60 (4)	8.03 (11)
0	7.87 (8)	8.27 (12)	7.30 (1)	7.35 (2)
0	7.71 (6)	8.00 (9.5)	7.82 (7)	7.66 (5)
4	11.63 (7)	12.20 (12)	11.04 (2)	11.50 (5.5)
4	11.87 (11)	11.70 (8.5)	11.50 (5.5)	10.10 (1)
4	11.40 (3)	11.80 (10)	11.49 (4)	11.70 (8.5)
8	15.00 (2)	16.60 (12)	15.87 (6)	15.10 (3)
8	15.92 (9)	16.40 (11)	15.91 (8)	14.80 (11)
8	15.58(4)	15.90 (7)	16.28 (10)	15.70 (5)
	$S_1 = 17.6$	$S_2 = 30.5$	$S_3 = 15.8$	$S_4 = 14$

Usando la ecuación 7.57 se tiene: $k = 4$ tratamientos. $n = 3$ bloques. $c = 3$ número de réplicas dentro de bloque

$$S_1 = 17.67,\ S_2 = 30.05,\ S_3 = 15.83,\ S_4 = 14.$$

$$N = 36; c = 3; k = 4, n = 3.$$

$$MS = \left(\frac{12}{4(36+3)}\right)\left[(17.67)^2 + (30.5)^2 + (15.83)^2 + (14)^2\right]$$

$$- 3(36+3) = 12.93.$$

$$\texttt{ValorP} = 1 - \texttt{pchisq}(12.93, 3).$$

$$= 0.0047.$$

Gracias

Referencias.

- https://sci2s.ugr.es/keel/pdf/algorithm/articulo/1937-JSTOR-Friedman.pdf The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance Author(s): Milton Friedman
- Métodos No Paramétricos. Giraldo Ramón. (2020).
- Wikepedia (2013). Portrait of Milton Friedman. Recuperado de https://commons.wikimedia.org/wiki/File:Portrait_of_Milton_Friedman.jpg