FACULTAD DE MATEMÁTICAS

Diseños experimentales

Proyecto 2. Unidad III.

"Café colombiano"

Por:
-Samantha Álvarez Herrera
-Abigail Ciau Puga
-Samantha Sobrino Bermejo

Equipo 8

LM. en C. Salvador Medina Peralta

Licenciatura en Actuaría – Especialización en Estadística.

Fecha: Viernes 11 de Octubre de 2019

TABLA DE CONTENIDOS

1. INTRODUCCIÓN: PLANTEAMIENTO DEL PROBLEMA

Origen de los datos *(Tomado de "12.9 Friedman Rank Test: Nonparametric Analysis for the Randomized Block Design. Análisis de conceptos. Problema 12.107". http://wps.prenhall.com/wps/media/objects/11886/12171343/OnlineTopics/bbs12e_onlinetopic_ch12-9.pdf).*

Nueve expertos calificaron cuatro marcas de café colombiano en un experimento de prueba de sabor. Se otorga una calificación en una escala de 7 puntos (1 = extremadamente desagradable, 7 = extremadamente agradable) para cada una de las siguientes cuatro características: *sabor*, *aroma*, *riqueza* y *acidez*.

La siguiente tabla muestra las clasificaciones sumadas, acumuladas en las cuatro características.

Tabla 1.

Datos de las calificaciones a las cuatro características del café, por marca, por experto.

Experto	Marca			
	A	***B***	***C***	***D***
C.C.	24	26	25	22
S.E.	27	27	26	24
E.G.	19	22	20	16
B.L.	24	27	25	23
C.M.	22	25	22	21
C.N.	26	27	24	24
G.N.	27	26	22	23
R.M.	25	27	24	21
P.V.	22	23	20	19

Se desea determinar, a un nivel de significancia de 0.05, si hay evidencia de una diferencia en las calificaciones sumadas promedio de las cuatro marcas de café colombiano.

2. OBJETIVOS

General: Responder a "¿Existe una diferencia entre las calificaciones sumadas promedio de las cuatro marcas de café colombiano?".

Específicos:

- Realizar un análisis completo de los datos e identificar correctamente la prueba estadística necesaria para resolver el problema.
- Con base a los resultados obtenidos, decidir si existe un ranking entre las marcas analizadas por los expertos.

3. METODOLOGÍA

Los datos recolectados para este análisis se tratan de una sola muestra aleatoria que consta de 9 catadores expertos (bloques) que han probado y calificado cuatro marcas distintas de café colombiano (tratamientos). La escala de medición del experimento es *ordinal*, puesto que son calificaciones.

Modelo

Según las características del problema de estudio, es adecuada la utilización de una prueba de Friedman para un diseño por bloques completamente aleatorizado. Además de lo mencionado con anterioridad, se cumplen otros supuestos requeridos para esta prueba:

- Cada experto asigna una calificación a las marcas independientemente de los otros, es decir los resultados dentro de un bloque no influyen en los resultados dentro de otro bloque.
- No existe un orden específico en que los catadores prueban las marcas, es decir os tratamientos son asignados al azar a unidades experimentales dentro de los bloques.
- Las mediciones se pueden clasificar (asignarles rangos) dentro de los bloques.

Las calificaciones son ejemplos de una *escala ordinal de medición*, por lo que los datos no son adecuados para una prueba paramétrica, no es necesario corroborar que no existe normalidad.

Tal y como fue solicitado, se utilizará un nivel de significancia del 5% ($\propto = 0.05$).

Para saber si las distintas marcas de café tienen calificaciones distintas asignadas por los expertos, nos interesa probar las siguientes *hipótesis*:

H_0: *Las calificaciones de las 4 marcas de café son idénticas.*
(Las marcas de café no influyen en las calificaciones recibidas).
vs
H_1: *Al menos una de las marcas de café tiende a recibir calificaciones mayores que al menos otra marca.*

Formalmente:
Sea θ_j es la mediana poblacional en la j-ésima condición, $(j = 1,2,3,4)$

$$H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 \qquad vs \qquad H_1: \theta_i \neq \theta_j \text{ para al menos un par } (i,j) \text{ con } i \neq j, \text{con } i,j = 1,2,3,4$$

Bajo H_0 cierta, el *estadístico de prueba* es:

$$F_r = \chi_r^2 = \frac{12n}{k(k+1)}\sum_{j=1}^{k}(R_{\bullet j} - R_{\bullet\bullet})^2 = \frac{12}{nk(k+1)}\sum_{j=1}^{k}R_j^2 - 3n(k+1)$$

Donde

$\mathrm{n = b} = 9$ es el número de renglones (bloques), es decir, expertos en café.

$\mathrm{k = a} = 4$ es el número de columnas (tratamientos), es decir, marcas de café evaluadas.

$\mathrm{nk} = 9(4) = 36$ es el total de observaciones.

$R_j = \sum_{i=1}^{n} r_{ij}$ es la suma de los rangos en la j-ésima columna.

$\overline{R_j} = R_{\bullet j} = \frac{R_j}{n}$ es rango promedio del j-ésimo tratamiento.

$\overline{R} = R_{\bullet\bullet} = \frac{(k+1)}{n} = \frac{5}{9}$ es el promedio total de los rangos de las nk observaciones.

En caso de empates entre los rangos de un bloque, el estadístico debe ser corregido para modificar la distribución muestral:

$$F_r = \chi_r^2 = \frac{12\sum_{j=1}^{k}\left(R_{\bullet j} - R_{\bullet\bullet}\right)^2}{nk(k+1) - \left[\frac{1}{(k-1)}\right]\sum_{l=1}^{n}\left[\left(\sum_{j=1}^{g_l} t_{lj}^3\right) - k\right]}$$

Donde:
g_l: es el número de grupos de rangos empatados en el i-ésimo bloque.
t_{jl}: es el tamaño del j-ésimo grupo de rangos empatados en el i-ésimo bloque.

Rechace H_0 con un nivel de significancia $\alpha = 0.05$ si $F_r = \chi_r^2 > \chi_{r;0.05,4,9}^2$ donde la constante $\chi_{r;0.05,4,9}^2$ satisface $P\left(\chi_r^2 > \chi_{r;0.05,4,9}^2\right) = 0.05$ y se obtiene de la tabla de valores de la distribución χ_r^2 de Friedman (Conover, 1999; Siegel y Castellan, 2001; Zar, 2010).

La distribución del estadístico F_r de Friedman cuando H_0 es verdadera, se puede aproximar mediante una distribución ji-cuadrada con $k-1$ grados de libertad mientras $n = b$ sea "grande". La evidencia empírica indica que la aproximación es adecuada si $n = b$ o $k = a$ es mayor que 5 (Wackerly *et al.*, 2010).

Como $\mathrm{n} = \mathrm{b} = 9 > 5$, por lo tanto, la distribución del estadístico de prueba F_r es aproximadamente ji-cuadrada con 4-1 g.l.,así, rechace H_0 con un nivel de significancia $\alpha = 0.05$ si $F_r = \chi_r^2 > \chi_{0.05;3}^2$

Modelo comparaciones múltiples no paramétricas

Con el fin de determinar que pares de tratamientos difieren cuando la prueba estadística sugiere diferencias significativas se llevará a cabo la prueba de comparación múltiple de Dunn.

Comparaciones múltiples, muestras grandes sin tratamiento de control (Siegel y Castellan, 2001)

Hipótesis

$$H_0: \theta_u = \theta_v \qquad vs \qquad H_1: \theta_u \neq \theta_v \; para\ algunos\ grupos\ u \text{ y } v, con\ u \neq v \text{ y } u, v = 1,2,3,4$$

Donde θ_u representa la mediana de la población para el i-ésimo grupo (tratamiento).

Se determinan las diferencias $|R_u - R_v|$ para todos los pares de grupos o condiciones. Cuando el tamaño de las muestras es grande, estas diferencias se distribuyen aproximadamente normal. Sin embargo, ya que hay una cantidad muy grande de diferencias y que las diferencias no son independientes, el procedimiento de comparación debe ajustarse apropiadamente.

Dada la presencia de empates en las observaciones, se rechaza H_0 con un nivel se significación α si:

$$|\bar{R}_u - \bar{R}_v| \geq \left(z_{\alpha/k(k-1)}\right)\sqrt{\left[\frac{N(N+1)}{12} - \frac{\sum_{j=1}^{g}\left(t_j^3 - t_j\right)}{12(N-1)}\right]\left(\frac{1}{n_u} + \frac{1}{n_v}\right)}$$

Donde
$N = \sum_{i=1}^{k} n_i$ Es el total de observaciones y n_i es el número de observaciones del i-ésimo grupo.
$\bar{R}_u$ promedio de los rangos de las observaciones del u-ésimo tratamiento.

g: número de grupos empatados.
t_j: Tamaño del j-ésimo grupo empatado.

4. RESULTADOS

De la tabla 4 del Anexo A se obtuvo que, para la prueba de Friedman, $P < 0.0001 < 0.05 = \alpha$ por lo tanto se rechaza H_0, y se decide por H_1: al menos dos de las medianas poblacionales difieren; es decir, las calificaciones dadas por los expertos entre las marcas de café son significativamente distintas ($H = 20.724, P = 0.0001 < 0.05$).

De la tabla 5 del Anexo A, vemos que para la comparación múltiple entre las calificaciones de las marcas A y D, B y D y B y C, e ($P - Valor_{AD} = P - Valor_{BC} = 0.049, P - Valor_{BD} < 0.0001 < 0.05 = \alpha$), por lo que se rechaza la hipótesis nula y se decide por la alternativa. Esto quiere nos indica que las sumas de calificaciones promedio de dichos pares son significativamente distintos, además, observando la tabla de rangos (tabla 6 del anexo A), se concluye que la marca de café B tiende a presentar calificaciones significativamente mayores que las marcas C y D. Análogamente, la marca de café A tiene calificaciones significativamente superiores a la marca D.

Para las comparaciones múltiples entre las calificaciones de las parejas C y D, A y C, y A y B se obtuvo un p valor mayor al nivel de significancia de 0.05 por lo que no se rechaza la hipótesis nula, es decir, las calificaciones entre estos pares de parejas no difieren significativamente.

5. CONCLUSIONES

Al estudiarse las calificaciones de la prueba de sabor, en torno a las cuatro características mencionadas: sabor, aroma, riqueza y acidez de cuatro marcas de café colombiano, se obtuvo evidencia de que puede elegirse una mejor marca cuando se tiene la opción de escoger entre los pares A-D, B-D o B-C, en donde la A y la B son “mejores” que sus contrastadas, respectivamente. Respecto a las comparaciones entre los otros 3 pares de marcas, no se determinó una diferencia significativa por lo que se es indiferente entre una y otra.

Por lo anterior, NO es posible determinar una marca que se considere la mejor entre el grupo, además, NO se puede determinar ningún tipo de *ranking* de las marcas de café basándose en las calificaciones de sabor (compuesto por *sabor*, *aroma*, *riqueza* y *acidez*) asignadas por los cuatro expertos.

6. REFERENCIAS

Montgomery, D.C. (2004). Diseño y análisis de experimentos. 2ª Ed. Limusa Wiley, México, D.F.

12.9 Friedman Rank Test: Nonparametric Analysis for the Randomized Block Design” Recuperado de:
http://wps.prenhall.com/wps/media/objects/11886/12171343/OnlineTopics/bbs12e_onlinetopic_ch12-9.

ANEXO A: Resultados con el paquete estadístico

Dentro de cada bloque se asignan los rangos a las medidas con base a sus magnitudes relativas

Tabla 2.

Experto	Rangos A	Rangos B	Rangos C	Rangos D
C.C.	3	1	2	4
S.E.	1.5	1.5	2	4
E.G.	3	1	2	4
B.L.	3	1	2	4
C.M.	2.5	1	2.5	4
C.N.	2	1	3.5	3.5
G.N.	1	2	4	3
R.M.	2	1	3	4
P.V.	2	1	3	4
R_j	20	10.5	24	34.5
R_j^2	400	110.25	576	1190.25

Como existen empates entre los rangos de un bloque, el estadístico debe ser corregido para modificar la distribución muestral, tal como se mencionó en la metodología.

Resultados con el paquete estadístico:

Prueba de Friedman

Tabla 3. Rangos

	Rango promedio
A	2.78
B	3.83
C	2.22
D	1.17

Tabla 4. Estadísticos de prueba[a]

n	9
Chi-cuadrado	20.724
gl	3
Sig. asintótica	<0.0001

a. Prueba de Friedman

En la tabla 4 podemos ver que, tal como se menciona en la metodología, se cumple el supuesto de que el número de bloques, n, es mayor que 5, entonces, F_r se distribuye aproximadamente ji-cuadrada con 3 g.l. (Wackerly *et al.*, 2010).

Decisión estadística (prueba de Friedman) con $\alpha = 0.05$

De la tabla 4 observamos que $valor - P < 0.0001 < 0.05 = \alpha$ por lo tanto se rechaza H_0, y se decide por H_1: al menos dos de las distribuciones poblacionales difieren en localización; es decir, al menos una de las marcas de café tiende a recibir calificaciones más altas que las otras.

Debido a que la prueba de Friedman determinó diferencias estadísticas significativas, procederemos a efectuar comparaciones múltiples para determinar qué pares de difieren.

Tabla 5. Comparaciones múltiples

Muestra 1-...	Estadístico de prueba	Error estándar	Estadístico de prueba estándar	Sig.	Sig. ajust.
D-C	1.056	.609	1.734	.083	.497
D-A	1.611	.609	2.647	.008	.049
D-B	2.667	.609	4.382	.000	.000
C-A	.556	.609	.913	.361	1.000
C-B	1.611	.609	2.647	.008	.049
A-B	-1.056	.609	-1.734	.083	.497

En la tabla 5 vemos que para la comparación entre las calificaciones de las marcas D y A $Valor-P=0.049<0.05=\alpha$, por lo que se rechaza la hipótesis nula y se decide por la alternativa. Al comparar las marcas D y B se obtiene un valor-P menor que 0.0001, por lo que, de la misma forma, se rechaza la hipótesis nula. La comparación entre las marcas C y B también resulta significativa, ya que $Valor-P=0.049<0.05=\alpha$, lo que lleva al rechazo de H_0. Esto nos indica que la marca de café B tiende a presentar calificaciones significativamente diferentes que las marcas C y D. Además, las calificaciones de la marca D también muestra diferencias significativas con las calificaciones de la marca A. La dirección de estas diferencias puede identificarse mediante sus rangos promedio:

Tabla 6. Rangos significativos

	Rango promedio
A	2.78 **a b**
B	3.83 **b**
C	2.22 **a c**
D	1.17 **c**

*Rangos promedios con letras distintas difieren significativamente.

De la tabla 6, podemos decir que la marca B tiene una calificación significativamente *mejor* (mayor) que las marcas C y D. La marca A también presenta calificaciones significativamente *mejores* que la marca D.

El valor P de las otras comparaciones fue mayor que el nivel de significancia de 0.05, por lo que no se rechaza la hipótesis nula, es decir, las calificaciones de las marcas A y B, A y C, así como C y D, *no difieren significativamente*. Esto implica que es indistinto escoger entre A y B, A y C, o entre C y D.

De esta manera, NO existe una marca que tenga mejor calificación en la prueba de sabor por encima de todas las demás, además, no se puede establecer un orden (o ranking) para las 4 marcas.