

II Jornadas Internacionales de Estadística Aplicada 5 y 6 de diciembre de 2019

USO DE CUADRADO LATINO PARA EL DISEÑOS QUE EMPLEAN BLOQUES (EN MACERACIÓN DE CERVEZA)

Autores: Lopez Claudio, Aparicio Sausedo Milton, Farfán Pablo Sebastián

Institución: Facultad de Ingeniería, Universidad Nacional de Salta.

Datos de contacto: williamayapura@gmail.com www.facebook.com/maraycerveza

RESUMEN

En el siguiente trabajo lo que se busca es analizar tres estilos de cervezas (AHUMADA, GINGER BEER, IPUÑA), en la presencia de dos fuentes de variabilidad como el PH y Temperatura; mediante el uso del método estadístico del diseño de cuadrado latino que se utiliza más datos experimentales. En este caso, lo que se analizó fueron los estilos de cerveza frente a dos variables controlables como el ph y la temperaturas con las cuales se armó la matriz de 3x3 y se buscó mediante las densidades que fueron las variables de respuesta para realizar el análisis de datos, es si hay grandes diferencias en los estilo de cerveza en estas condiciones tomadas.

INTRODUCCION

Los datos registrados para esté trabajo los obtuvimos de un microemprendimiento CERVECERIA MARAY (marca de la cerveza) en el cual se elabora cerveza artesanal. Es un emprendimiento que empezó hace casi 2 años y ha crecido bastante, debido la calidad del producto y al gran compromiso de sus creadores. Empezaron con un estilo de cerveza GINGER BEER, pero la necesidad de crecer los llevo a incrementar la producción y también a buscar e innovar con nuevos estilos en los cuales usan elementos de la zona para así tener un sabor que los identifique.

En el proceso de la elaboración de la cerveza los datos fueron tomados en la maceración, que consiste en remojar el grano (mayor mente cebada) para así activar las enzimas de la malta con agua caliente a diversas temperaturas para hacer una infusión, y que en un proceso posterior serán metabolizados por la levadura en alcohol etílico. El tiempo y la temperatura dependerán de la receta y estilo de cerveza que se desea hacer.

Buscamos con el análisis estadístico de cuadrado latino, es el proceso de maceración en donde al tener mismas condiciones de temperaturas (fila) y ph (columnas) variables controlables, aplicamos al azar los 3 estilos de cerveza que se tienen, para así buscar si el estilo de cerveza genera grandes diferencias en estas condiciones planteadas, esas diferencias se ve reflejado en la densidad que es nuestra variable de respuesta, la cual usaremos para llevar a cabo el análisis de varianza.

METODOLOGIA:

Cuadrados latinos: Los diseños en cuadrados latinos son apropiados cuando es necesario controlar dos fuentes de variabilidad. En dichos diseños el número de niveles del factor principal tiene que coincidir con el número de niveles de las dos variables de bloque o factores secundarios y además hay que suponer que no existe interacción entre ninguna pareja de factores. Si consideramos una tabla de doble entrada donde las filas y las columnas representan cada uno de los dos factores de bloque y las celdas los niveles del factor principal o tratamientos, el requerimiento anterior supone que cada tratamiento debe aparecer una vez y sólo una en cada fila y en cada columna.

El procedimiento para construir un diseño en cuadrado latino es el siguiente:

- 1) Se elige aleatoriamente un cuadrado latino de los disponibles.
- 2) Se asigna aleatoriamente el orden de las filas y columnas.
- 3) Se asignan aleatoriamente los tres factores a las filas, columnas y letras, respectivamente.

Podemos decir que un diseño en cuadrado latino tiene las siguientes características:

- Se controlan tres fuentes de variabilidad, un factor principal y dos factores de bloque.
- Cada uno de los factores tiene el mismo número de niveles, K.
- Cada nivel del factor principal aparece una vez en cada fila y una vez en cada columna.
- No hay interacción entre los factores.

Planteamiento del modelo

En un diseño en cuadrado latino intervienen los siguientes factores: un factor principal y dos factores secundarios o variables de bloque. Se supone que no existe interacción entre esos tres factores. Así el modelo empleado es un modelo aditivo.

Si consideramos que los tres factores son de efectos fijos, el modelo estadístico para este diseño es:

$$y_{ijh} = \mu + \tau_i + \beta_j + \gamma_h + \dots + u_{ijh}$$

- $y_{ij(h)}$ representa la observación correspondiente a la i-ésima fila, j-ésima columna y h-ésima letra latina.
- μ es la media global. τ_i es el efecto producido por el i-ésimo nivel del factor fila. Dichos efectos están sujetos a la restricción $\sum_i \tau_i = 0$.
- τ_i es el efecto producido por el i-ésimo nivel del factor fila. Dichos efectos están sujetos a la restricción $\sum_i \tau_i = 0$.
- β_j es el efecto producido por el j-ésimo nivel del factor columna. Dichos efectos están sujetos a la restricción $\sum_j \beta_j = 0$.
- γ_h es el efecto producido por la h-ésima letra latina. Dichos efectos están sujetos a la restricción $\sum_h \gamma_h = 0$.
- $u_{ij(h)}$ son variables aleatorias independientes con distribución $N(0, \sigma)$

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F_{exp}
E. fila	$\frac{1}{K} \sum_i y_{i..}^2 - \frac{y_{...}^2}{K^2}$	$K - 1$	\hat{S}_F^2	$\hat{S}_F^2 / \hat{S}_R^2$
E. columna	$\frac{1}{K} \sum_j y_{.j.}^2 - \frac{y_{...}^2}{K^2}$	$K - 1$	\hat{S}_C^2	$\hat{S}_C^2 / \hat{S}_R^2$
E. letra lat.	$\frac{1}{K} \sum_h y_{..h}^2 - \frac{y_{...}^2}{K^2}$	$K - 1$	\hat{S}_L^2	$\hat{S}_L^2 / \hat{S}_R^2$
Residual	$SCT - SCF$ $SCC - SCL$	$(K - 1)(K - 2)$	\hat{S}_R^2	
TOTAL	$\sum_i \sum_j y_{ij(.)}^2 - \frac{y_{...}^2}{K^2}$	$K^2 - 1$	S_T^2	

Tabla (1): Forma práctica de la tabla ANOVA para el modelo de cuadrado latino

Y se rechaza H_0 , al nivel de significación α , cuando el valor experimental del respectivo estadístico sea mayor que el valor crítico de la distribución F con $K - 1$ y $(K - 1)(K - 2)$ grados de libertad.

Para el siguiente trabajo, cuando visitamos las instalaciones se llevaba a cabo la elaboración de tres estilos de cerveza AHUMADA, GINGER BEER, IPUÑA, a los cuales les tomamos los datos. Para medir la densidad se usó un densímetro, para el PH un peachímetro y la temperatura con un termómetro digital. Los contenedores donde se llevaba a cabo el macerado son de 100 litros.

DESARROLLO

Validación de los supuestos del ANOVA:

El análisis de varianza es sensible a las propiedades estadísticas de los términos de error aleatorio del modelo lineal.

Los supuestos implican errores:

- Normalmente distribuidos,
- Independientes,
- Varianzas homogéneas para todas las observaciones.

a) NORMALIDAD

En un primer momento debemos realizar el análisis de los datos para probar la normalidad de los mismos, puesto que al confirmar esta hipótesis se podrá realizar el estudio estadístico del diseño de cuadrado latino.

Se realizó el análisis de normalidad de los datos con la prueba de Shapiro – Wilk

$$\hat{e}_{ijk} = y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{...} + 2*\bar{Y}_{...} \quad \text{Ecuación de los residuos}$$

T [°C]	pH		
	5,4	5,5	5,6
64	-0,667	-13,333	-3
65	9,667	14	-0,667
66	-9	-0,667	3,667

Tabla N°2: Valores de densidad [g/ cm³] de los residuos para los tres estilos de cerveza.

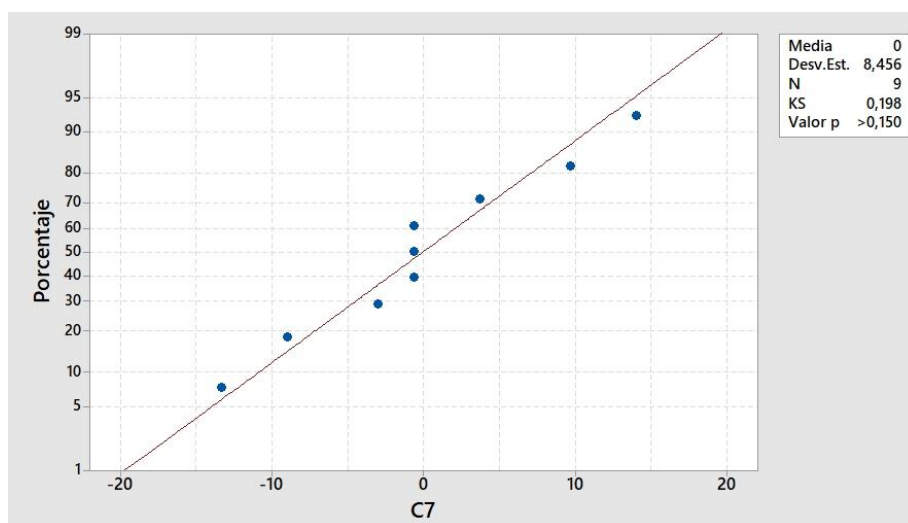


Figura (1): Grafica de normalidad.

Las hipótesis que se someten a prueba son:

H0: los residuos tienen distribución normal

Ha: los residuos no tienen distribución normal.

El valor $p > 0.05$ (nivel de significación nominal de la prueba) por lo que aceptamos la hipótesis nula, implica que los residuos tiene una distribución normal.

b) INDEPENDENCIA

Gráfico de dispersión de los residuos para comprobar la independencia

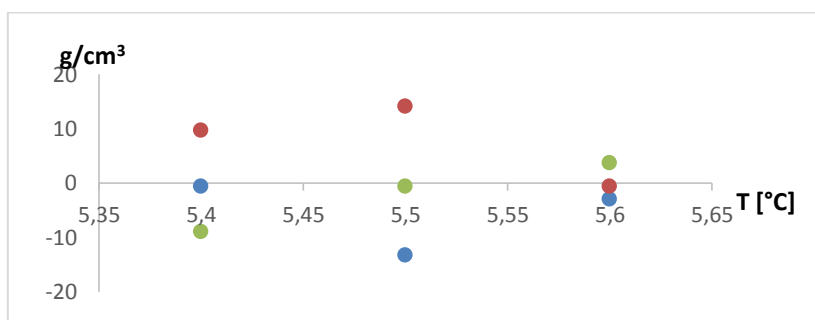


Figura (2): Residuos vs columnas (temperatura)

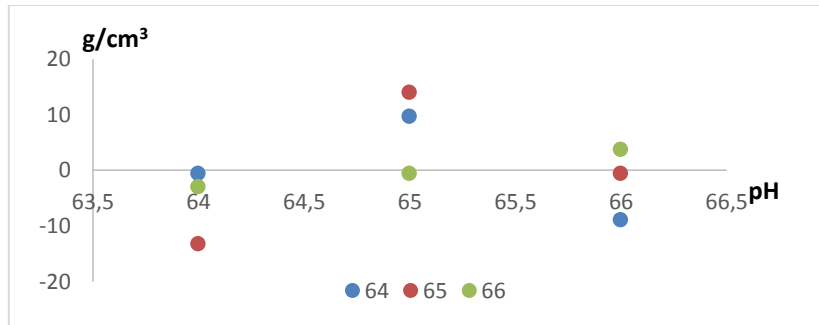


Figura (3): Residuos vs filas (pH)

De los gráficos de dispersión de los residuos no se observa agrupamiento de datos ni tendencias que indiquen falta de independencia.

c) HOMOGENEIDAD

Es bastante complicada y más aún su tratamiento. Existen múltiples pruebas conducentes a comprobar la presencia de homogeneidad. Entre los procedimientos gráficos mostraremos las representaciones de los residuos frente a los valores predichos. Con dicha gráfica se puede observar que la variable está influyendo en algún sentido en la variabilidad de las observaciones.

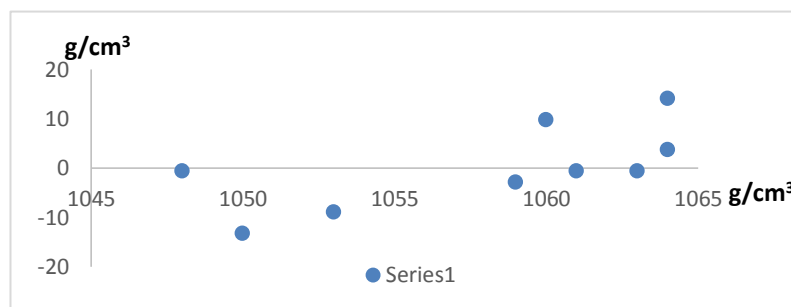


Figura 4: Residuos vs Valores predichos

De la gráfica de dispersión de los residuos se puede observar que no hay un reagrupamiento o tendencia por lo que se puede decir que las varianzas son iguales u homogéneas.

3- ANALISIS DE VARIANZA Y CUADRADO LATINO

Al aceptar que los datos son normales comenzamos con el análisis de cuadrado latino mediante una aleatorización de los mismos (Tabla N°1) dándonos el siguiente orden:

A	B	C
C	A	B
B	C	A

T [°C]	PH		
	5,4	5,5	5,6
64	1063	1050	1059
65	1060	1064	1048
66	1053	1061	1064

Tabla N°3: Valores de densidad [g/ cm³] para los tres estilos de cerveza.

Así, nuestros datos quedaron ordenados en la siguiente tabla para poder realizar su análisis, donde las letras latinas corresponden a las densidades de los 3 estilos de cervezas (**A: Ahumada; B: Ginger Beer; C: Ipuña**) controlando la temperatura (bloque factor fila) y el pH (bloque factor columna):

Prueba de Hipotesis de la media de los distintos estilos de cerveza

Nuestra hipótesis es que $\mu=0$ es decir:

$H_0: \mu=0$ $\alpha = 0,05$

$H_a: \mu \neq 0$

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F observados
Tratamientos	284,667	2	142,333	106,75
Filas	8	2	4	3
Columnas	4,667	2	2,333	1,75
Error	2,667	2	1,333	
Total	300	8		

Tabla N°4: Cuadro de análisis de varianza para cuadrado latino

Comparando el valor observado $F_{obs}=106,75$ contra el valor crítico
 $F_{crit}(\alpha=0,05, v_1=2, v_2=8)=19$

$F_{obs} > F_{cri} \rightarrow 106,75 > 19$

Rechazamos H_0 , es decir que hay diferencias entre las medias de las densidades de los estilos de cervezas.

Como sabemos que no se cumple la hipótesis de que la media de los tratamientos es igual, podemos plantear comparaciones de a pares por el método de Tuckey. N° de comparaciones = $\frac{t*(t-1)}{2} = 3$

La diferencia honesta significativa se calcula DHS = $q * \sqrt{\frac{CME}{r}}$, donde $q(\alpha, t, n-t)$.

$$q(0,05;3;6) = 4,34$$

$$CME = 1,333$$

$$r = 3$$

$$DHS = 4,34 * \sqrt{\frac{1,333}{3}} = 2,893$$

Comparación n°1

$$H_0: \mu_A = \mu_B$$

$$H_a: \mu_A \neq \mu_B$$

Comparación n°2

$$H_0: \mu_A = \mu_C$$

$$H_a: \mu_A \neq \mu_C$$

Comparación n°3

$$H_0: \mu_B = \mu_C$$

$$H_a: \mu_B \neq \mu_C$$

El estadístico de prueba : $|\bar{Y}_i - \bar{Y}_j|$

$$|\bar{Y}_A - \bar{Y}_B|$$

$$|1063,667 - 1050,333|$$

$$|\bar{Y}_A - \bar{Y}_C|$$

$$|1063,667 - 1060|$$

$$|\bar{Y}_B - \bar{Y}_C|$$

$$|1050,333 - 1060|$$

$$|\bar{Y}_A - \bar{Y}_B| = 13,333$$

$$|\bar{Y}_A - \bar{Y}_B| > DHS$$

$$13,333 > 2,893$$

$$|\bar{Y}_A - \bar{Y}_C| = 3,667$$

$$|\bar{Y}_A - \bar{Y}_C| > DHS$$

$$3,667 > 2,893$$

$$|\bar{Y}_B - \bar{Y}_C| = 9,667$$

$$|\bar{Y}_B - \bar{Y}_C| > DHS$$

$$9,667 > 2,893$$

Ninguna de las medias de las comparaciones es igual. Lo que nos dice q ningún estilo de cerveza es semejante al otro en las condiciones planteadas.

Para calcular la eficiencia relativa del método aplicado frente a un diseño por bloques al azar, lo realizamos primero sin tomar en consideración la variación debida a la temperatura (bloque fila) y luego calculamos la eficacia sin considerar la variación debida al pH (bloque columna), para ello también debemos calcular un factor de corrección, procediendo de la siguiente manera:

$$\text{Corrección} = \frac{(gl_{cl}+1)*(gl_{bca}+3)}{(gl_{cl}+3)*(gl_{bca}+1)} = \frac{(2+1)*(4+3)}{(2+3)*(4+1)} = 0,84$$

donde: gl_{cl} son los grados de libertad del cuadrado latino

gl_{bca} son los grados de libertad para el bloque completamente aleatorizados

Eficiencia relativa del bloque por columna:

$$ER_{bc} = \text{corrección} * \frac{s_{bca}^2}{s_{cl}^2}$$

$$\frac{2,333 + (3-1)*1,333}{3}$$

$$s_{bca}^2 = \frac{CM_{columna} + (t-1)*CME}{t} =$$

$$s_{bca}^2 = 1,667$$

$$s_{cl}^2 = CME = 1,333$$

$$ER_{bc} = 0,84 * \frac{1,667}{1,333} = 1,05$$

Esto quiere decir que hay una ganancia de eficiencia del cuadro latino frente al bloque completamente aleatorizado de un 5% cuando no se tiene en consideración la variación aportada por la temperatura(filas) o sea que se necesitarían más replicas en un diseño de bloques completamente aleatorizados para obtener una varianza estimada de la media del tratamiento igual que a la del diseño cuadrado latino.

Eficiencia relativa del bloque por fila:

$$ER_{bf} = \text{corrección} * \frac{s_{bca}^2}{s_{cl}^2}$$

$$\frac{4 + (3-1)*1,333}{3}$$

$$s_{bca}^2 = \frac{CM_{fila} + (t-1)*CME}{t} =$$

$$s_{bca}^2 = 2,222$$

$$s_{cl}^2 = CME = 1,333$$

$$ER_{bf} = 0,84 * \frac{2,222}{1,333} = 1,4$$

Esto quiere decir que hay una ganancia de eficiencia del cuadro latino frente al bloque completamente aleatorizado de un 40% cuando no se tiene en consideración la variación aportada por el pH(columnas) o sea que se necesitarían más replicas en un diseño de bloques completamente aleatorizados para obtener una varianza estimada de la media del tratamiento igual que a la del diseño cuadrado latino.

CONCLUSIONES

Pudimos comprobar que existe una diferencia estadística significativas entre la temperatura y el pH sobre la densidad, aunque los valores de los mismos son muy similares.

También mediante cuadrado latino comprobamos que hay diferencias entre los tratamientos en estas condiciones, esto se debe a que cada cerveza tiene un proceso de elaboración específico dependiendo del porcentaje de malta, cebada y otros agregados que hacen única a cada variedad. Este tipo de análisis fue seleccionado debido a que con un diseño completamente aleatorizado se necesitarían una cantidad de datos más amplia para obtener un resultado igual o similar.

BIBLIOGRAFIA

Robert O. Kuehl. "Principios estadísticos para el diseño y análisis de investigaciones"

García Leal, J. & Lara Porras, A.M. (1998). "Diseño Estadístico de Experimentos. Análisis de la Varianza." Grupo Editorial Universitario.

* Lara Porras, A.M. (2000). "Diseño Estadístico de Experimentos, Análisis de la Varianza y Temas Relacionados: Tratamiento Informático mediante SPSS" Proyecto Sur de Ediciones.