



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA

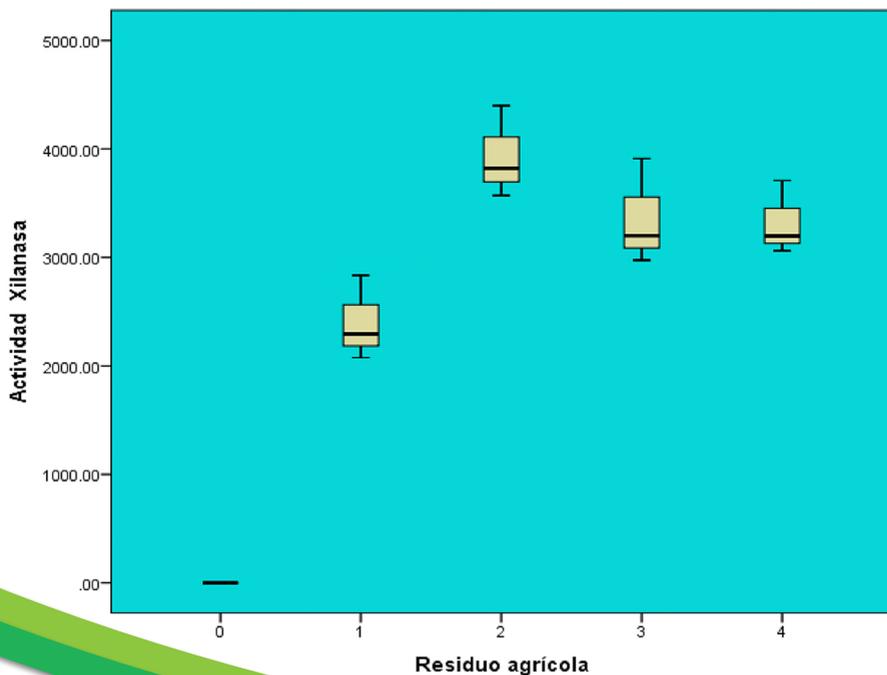
Ejercicios para el Taller de Diseño Experimental



Alberto **Castillo Morales**

Flor de María **Cuervo López**

Rosa Obdulia **González Robles**



UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD IZTAPALAPA

RECTOR GENERAL

Dr. Eduardo Abel Peñalosa Castro

SECRETARIO GENERAL

Dr. José Antonio De Los Reyes Heredia

UNIDAD IZTAPALAPA

RECTOR

Dr. Rodrigo Díaz Cruz

SECRETARIO

M. en B. E. Arturo Leopoldo Preciado López

DIRECTORA DE LA DIVISIÓN DE C.B.S.

Dra. Sara Lucía Camargo Ricalde

COORDINADOR DE EXTENSIÓN UNIVERSITARIA

Mtro. Federico Bañuelos Bárcena

JEFE DE LA SECCIÓN DE PRODUCCIÓN EDITORIAL

Lic. Adrián Felipe Valencia Llamas

Primera edición 2018

ISBN: 978-607-28-1306-9

UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD IZTAPALAPA

Av. Michoacán y La Purísima

Iztapalapa, 09340. México, Ciudad de México

Índice

Introducción.....	5
1. Mejoramiento de producción de xilanas usando <i>Staphylococcus sp.</i> con diferentes fuentes de azúcar	7
1.1. Ejercicio	12
Bibliografía	12
2. Biorremediación de suelos contaminados por hidrocarburos.....	13
2.1. Ejercicio	18
Bibliografía.....	19
3. Biofiltración de vapor de metano	13
3.1. Ejercicio	25
Bibliografía.....	26
4. Uso de residuos agrícolas para la producción de xilanas de <i>Staphylococcus sp.</i>	27
4.1. Ejercicio	39
Bibliografía.....	39
5. Efecto de la estimulación eléctrica, la adición de epinefrina y el tiempo de molido sobre el contenido de ácido tiobarbitúrico en carne de cerdo.....	41
5.1. Ejercicio	46
Bibliografía.....	47
6. Regresión lineal simple para recuentos de células vivas y muertas.....	49
6.1. Ejercicio	55
Bibliografía.....	55
7. Relación del contenido de ácido tiobarbitúrico con el pH en carne molida fresca de cerdo almacenada	57
7.1. Ejercicio.....	67
Bibliografía.....	67
8. Perfil de actividad metanogénica específica en un reactor anaerobio de flujo ascendente con lecho de lodos	69
8.1 Ejercicio.....	80

Bibliografía	81
9. Mejoramiento de la actividad de xilanasa de Staphylococcus sp. usando diferentes fuentes de nitrógeno	83
9.1 Ejercicio	98
Bibliografía	99
10. Producción de biomasa y esporas de Idriella bolleyi	101
10.1 Ejercicio	108
Bibliografía	109
11. Producción de poligalacturonasa por fermentación en medio sólido	110
11.1 Ejercicio	120
Bibliografía	120
12. Efecto del tipo de filtro y velocidad de carga en la eliminación de metanol en reactores de biofiltros.....	123
11.1 Ejercicio	130
Bibliografía	130
13. Optimización del medio de cultivo para producción de estreptomicina	131
13.1 Ejercicio	138
Bibliografía	138
14. Proceso de cocido por extrusión en amaranto	139
14.1 Ejercicio	147
15. Optimización de producción de dextrina	149
15.1 Ejercicio	157
Bibliografía	157
Bibliografía General.....	158

Introducción

El cambio de los programas de estadística en las carreras de Ingeniería de los alimentos e Ingeniería bioquímica industrial, propone que la enseñanza de las materias de estadística se base en el uso de ésta como herramienta para la investigación experimental y en general para el manejo de la información. En alusión a este objetivo aplicativo es que se les ha dado el nombre de: Taller de Bioestadística y Taller de Diseño Experimental. Para esta última UEA es que se han desarrollado los ejercicios de la presente publicación. En este curso, el estudiante debe utilizar la computadora para obtener los resultados de los análisis estadísticos adecuados a cada situación experimental, con el objetivo de poder dedicar más tiempo a la discusión y los fundamentos del análisis. Esto modifica la forma de enseñanza: primero se plantea el problema y se realiza el análisis estadístico de los datos, después, a la vista de los resultados, se hace una discusión que permita incidir en los conceptos estadísticos que la fundamentan, posteriormente, regresando al planteamiento del problema, se verifica la idoneidad del análisis estadístico realizado y finalmente se hace la discusión sobre la solución del problema usando los resultados obtenidos del análisis estadístico.

De acuerdo con esto, los temas contenidos en Taller de Diseño Experimental (TDE): diseños completamente al azar con uno y dos factores y bloques al azar, regresión lineal simple y correlación bivariada de Pearson, se imparten con un enfoque práctico. Sin embargo, los problemas a los que se enfrentará el estudiante, tanto en algunos de los cursos que tomará en los últimos trimestres de la carrera, como en la vida profesional propiamente dicha, son más complejos que las aplicaciones de los temas mencionados. Atendiendo esta realidad, el presente material incluye, además de ejemplos simples, problemas un poco más complejos que se resuelven por completo. Si durante el curso se desea tener más problemas sencillos, solamente se necesita eliminar la información de los factores que dan complejidad a los problemas.

En vista de que el programa de TDE demanda una revisión detallada de los conceptos básicos de diseños experimentales y de regresión, con base en el análisis estadístico de los datos, incluyendo: el proceso de estimación, el planteamiento de hipótesis estadísticas y sus respectivas pruebas, las conclusiones estadísticas y las que se establecen a partir de los objetivos del problema, se espera que el uso de los ejercicios que se presentan promueva la aplicación de todos estos conceptos y procedimientos, a partir de la revisión de los análisis estadísticos. Así, sin pretender que este material sea utilizado como texto del curso, los autores estamos convencidos de que facilitará al profesor incidir en los conceptos y procedimientos de estimación y pruebas de hipótesis y a los alumnos a comprenderlos.

El documento está conformado por una serie de ejemplos en los que se hace, primeramente, una breve descripción del marco teórico, el objetivo científico que se persigue y la metodología experimental realizada. Se continúa con la metodología estadística, el análisis de los resultados estadísticos y la obtención de conclusiones de acuerdo a éstos, para finalmente presentar la conclusión científica del ejemplo. El alumno puede hacer los cálculos y análisis adecuados que le llevan a los resultados de cada problema usando el paquete R con las instrucciones que se le proporcionan. Una vez resuelto, se puede discutir sobre la estructura del problema que lleva al análisis realizado, sobre las bases estadísticas que soportan al análisis, incluyendo conceptos básicos de estadística sobre los modelos que se usan y el significado de cada una de sus partes, sobre estimación y pruebas de hipótesis. Para reafirmar, se pueden plantear problemas similares que permitan profundizar en los conceptos de la estadística para interpretar correctamente los resultados.

Los alumnos de TDE dieron las ideas de los temas al hacer la presentación de sus tareas sobre el análisis de trabajos publicados. Se agradece a todos ellos su dedicación y capacidad de selección de trabajos interesantes. Los datos de los problemas fueron generados por los autores basándose en la bibliografía de cada problema y en su experiencia. Los experimentos y resultados pueden ser consultados por el lector en los artículos originales correspondientes.

Alberto Castillo Morales
Flor de María Cuervo López
Rosa Obdulia González Robles

1. Mejoramiento de producción de xilanasa usando *Staphylococcus sp.* con diferentes fuentes de azúcar

El xilano es un polímero que forma parte desde el 20 y hasta el 40% de la biomasa total de las plantas. Las xilanasas son enzimas hidrolíticas que convierten el xilano en los azúcares que lo constituyen. Debido a esto, las xilanasas son ampliamente utilizadas en diversos procesos para generar papel y productos químicos, así como en el tratamiento de desechos orgánicos para la generación de energía.

Con el objeto de probar el efecto de diferentes fuentes de azúcar sobre la actividad enzimática de las xilanasas, denotada por la variable x_{75} , se hizo un experimento utilizando *Staphylococcus sp.* La unidad experimental fue un matraz Erlenmeyer de 250 ml con 100 ml de medio líquido estándar, que se agitó a 200 rpm y se mantuvo a 37°C con pH inicial de 8.0 durante tres días. Como se tuvieron dos repeticiones se usaron 12 matraces. La actividad enzimática, que corresponde a la variable respuesta, se evalúa en Unidades Internacionales (UI) y se determina a un pH de 7.5. En la Tabla 1.1 se muestran los datos de la actividad enzimática de xilanasas en UI obtenidos al final del cultivo para cada tratamiento ensayado con diferentes fuentes de azúcar.

Tabla 1.1. Datos de actividad enzimática de xilanasas con diferentes fuentes de azúcares.

Fuente de Azúcar 0.5% w/v	UI de Xilanasas (x_{75})
Glucosa	210.7
Glucosa	205.5
Lactosa	133.5
Lactosa	110.9
Galactosa	135.9
Galactosa	168.4
Fructosa	117.7
Fructosa	116.8
Xilosa	379.8
Xilosa	369.5
Sacarosa	110.3
Sacarosa	105.5

Los datos anteriores provienen de la realización de un diseño completamente al azar con un factor, por lo que se plantea el modelo que relaciona a la media de la actividad de xilanasas ($\mu_{x_{75},i,j}$) con una media general (μ) más el efecto de la fuente de azúcar (A_i), que es el factor que define a los tratamientos. Se usa un subíndice para indicar el nivel o tipo de la fuente de azúcar, numerándolas de 1 a 6.

Se tiene:

$$\mu_{x_{75},i,j} = \mu + A_i .$$

El primer término ($\mu_{x75,i,j}$) de la ecuación anterior, es la media de x75 para el nivel i del factor A en la repetición j . El subíndice i en el factor A, identifica al efecto (A_i) que el nivel i del factor A tiene en la media de x75. El subíndice i puede tomar los valores 1,2,3,4,5 y 6. De acuerdo con la tabla de datos (Tabla 1.1), $i = 1$ identifica a glucosa, $i = 2$ a lactosa, $i = 3$ a galactosa, $i = 4$ a fructosa, $i = 5$ a xilosa e $i = 6$ a sacarosa. El subíndice j se refiere a la repetición, que en este problema puede ser 1 o 2. Todo esto se especifica junto al modelo en forma abreviada con $i = 1, 2, \dots, 6$, y $j = 1, 2$. Los otros términos en la ecuación son μ , que representa a la media general, y A_i que es el efecto de la fuente de azúcar i en la media de la actividad de xilanasas.

La variable aleatoria x75 está dada por la media que le corresponde mas un componente aleatorio o error, como se muestra en la siguiente ecuación:

$$x75_{i,j} = \mu_{x75,i,j} + \varepsilon_{i,j} = \mu + A_i + \varepsilon_{i,j}$$

El término $x75_{i,j}$ representa a la actividad de xilanasas, que es una variable aleatoria cuando se tiene el nivel i de la fuente de azúcar y la repetición j , y $\varepsilon_{i,j}$ es el error aleatorio asociado a la observación $x75_{i,j}$, este error se distribuye normal con media cero y varianza σ^2 .

El análisis de varianza para el modelo produjo los resultados que se muestran en la Tabla 1.2. En este análisis se presenta la prueba de la hipótesis nula de que los efectos de las fuentes de azúcar en la media de la actividad de xilanasas son iguales. La hipótesis alternativa niega a la hipótesis nula. Se plantea la pareja de hipótesis:

$$H_0: A_1 = A_2 = A_3 = A_4 = A_5 = A_6$$

$$H_A: A_i \neq A_{i'} \text{ para al menos una pareja de subíndices } i \neq i' \text{ con } i, i' = 1, 2, 3, 4, 5, 6, \text{ o lo que es lo mismo}$$

H_A : Hay al menos una pareja de efectos de las fuentes de azúcar A_i y $A_{i'}$ en la media de actividad de xilanasas que no son iguales

La prueba de la hipótesis de efectos iguales de los niveles de A se lleva a cabo con la F calculada o, mejor aún, con el valor p. Se rechaza la hipótesis nula con un valor p de $3.626E^{-06}$ (que es menor que 0.05) y se concluye que al menos una pareja de fuentes de azúcar tiene efectos diferentes en la media de producción de xilanasas.

Tabla 1.2. Análisis de varianza del diseño completamente al azar con dos repeticiones para la variable x75 con diferentes fuentes de azúcar como factor con seis niveles.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Fuente de Azúcar	5	103860	20771.9	144.59	$3.626E^{-06}$
Residual (Error)	6	862	143.7		

En la Tabla 1.3 se presentan las medias muestrales de actividad de xilanasas para cada fuente de azúcar. Se ve que la mayor media de xilanasas pH 7.5 corresponde a la fuente de azúcar xilosa y vale 374.6, le sigue la media de xilanasas para la fuente de azúcar glucosa, con un valor de 208.1 y la tercera, con valor de 152.2, se obtiene con galactosa. Las otras fuentes de azúcar: lactosa, fructosa y sacarosa tienen medias cercanas, que oscilan entre 107 y 122

Tabla 1.3. Medias de actividad de xilanas con diferentes fuentes de azúcares.

Etiqueta para Fuente de azúcar	Fuente de azúcar	Media de actividad de Xilanas
1	Glucosa	208.1
2	Lactosa	122.2
3	Galactosa	152.2
4	Fructosa	117.3
5	Xilosa	374.6
6	Sucrosa	107.9

Para saber cuáles parejas de fuentes de carbono tienen efectos diferentes se sugiere hacer la prueba de Tukey. En la Tabla 1.4 se presentan los resultados que muestran todas las comparaciones entre las parejas de tratamientos. Recuerde que Tukey usa un nivel de confianza adecuado para el conjunto de las 15 comparaciones de parejas de tratamientos, se le llama nivel de confianza a nivel de experimento. En esta tabla se ve que las medias de actividad de xilanas de las fuentes 5 (etiqueta de xilosa) y 1 (etiqueta de glucosa) son diferentes entre sí y de las otras fuentes de azúcar. En cambio, las medias que corresponden a las fuentes de azúcar 2, 3, 4 y 6 no difieren entre sí. El análisis muestra que la máxima actividad de xilanas se obtiene con xilosa y la segunda mejor producción se obtiene con glucosa. Las otras fuentes de azúcar: lactosa, galactosa, fructosa y sacarosa tienen, para fines prácticos, el mismo efecto en la media de actividad de xilanas.

Tabla 1.4. Resultados de la prueba de Tukey de comparaciones múltiples de actividad de xilanas con diferentes fuentes de azúcar.

Fuentes de azúcar que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
2-1	-85.90	-133.60271	-38.197285	0.0028939
3-1	-55.95	-103.65271	-8.247285	0.0247319
4-1	-90.85	-138.55271	-43.147285	0.0021427
5-1	166.55	118.84729	214.252715	0.0000709
6-1	-100.20	-147.90271	-52.497285	0.0012567
3-2	29.95	-17.75271	77.652715	0.2564268 no
4-2	-4.95	-52.65271	42.752715	0.9976149 no
5-2	252.45	204.74729	300.152715	0.0000051
6-2	-14.30	-62.00271	33.402715	0.8265548 no
4-3	-34.90	-82.60271	12.802715	0.1618098 no
5-3	222.50	174.79729	270.202715	0.0000112
6-3	-44.25	-91.95271	3.452715	0.0680929 no
5-4	257.40	209.69729	305.102715	0.0000046
6-4	-9.35	-57.05271	38.352715	0.9613757 no
6-5	-266.75	-314.45271	-219.047285	0.0000038

En la Tabla 1.5 se muestran las medias de los tratamientos ordenadas de menor a mayor, donde las medias que no presentan diferencia significativa están unidas por una línea recta. Se puede ver que con 5% de significación (de Tukey), las medias de los tratamientos 6,4,2 y 3 no difieren entre sí y son diferentes de las medias de los tratamientos 1 y 5, las cuales difieren entre sí.

Tabla 1.5. Prueba de Tukey para actividad de xilanasa. Medias unidas por una línea no difieren

Tratamiento (Fuente de azúcar)	Sucrosa (6)	Fructuosa (4)	Lactosa (2)	Galactosa (3)	Glucosa (1)	Xilosa (5)
Media muestral	107.9	117.3	122.2	152.2	208.1	374.6

Una gráfica de cajas y ejes (Figura 1.1) –no representa pruebas– muestra las diferencias en los valores de actividad de xilanasa en los datos y complementa lo ya visto. En la gráfica 1.1 no se ven ejes hacia afuera de las cajas, esto se debe a que cada caja se forma con sólo dos observaciones.

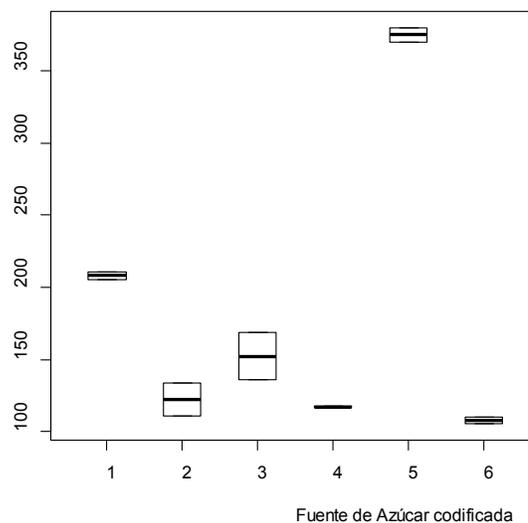


Figura 1.1. Gráfica de cajas para los tratamientos codificados de actividad de xilanasa.

El análisis muestra que la mayor media de actividad de xilanasa se obtiene con xilosa y la segunda mayor media con glucosa, ambas medias son significativamente diferentes entre si y de las otras medias. De aquí se concluye que conviene usar xilosa como fuente de azúcar, aunque si es muy cara, se recomienda usar glucosa.

El análisis de este problema se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares.

Guía para resolver con R

```
datosXilanasaA<-read.table(file = "xilanasaA.csv",header = T,sep = ",")
```

Se lee la tabla (read.table) de datos en Excel con tipo csv. Está en el archivo entre comillas.

Después de file = Se pone la dirección completa del archivo. Si el archivo está en su computadora

```

# Sugiero lo copie de la barra del explorador de Windows y cambie las diagonales \ por diagonales /,
# se usa header = T si los nombres de las columnas están en la primera hilera, y se usa
# sep = ";" porque el tipo csv usa comas como separadores. La tabla se guarda como datosXilanasaA
#Cambie de directorio a la carpeta donde está el archivo xilanasaA.csv y no necesita dar dirección
datosXilanasaA

fA<-datosXilanasaA$fA # Se define el vector de fA

x75<-datosXilanasaA$x75 # Se define el vector de X75

trat<-c(1,1,2,2,3,3,4,4,5,5,6,6) # Se crea el vector de subíndices de los tratamientos

# fA # x75 # trat # Estas instrucciones mostrarían los vectores

baseXilanasaA<-data.frame(Xilanasa = x75,FuenteAzucar = factor(fA),Tratamientos = factor(trat))

# Se crea y guarda el marco baseXilanasaA. Los factores se definen usando factor(),
# la variable respuesta se define directo

baseXilanasaA # Muestra la base

attach(baseXilanasaA) # Conviene indicar al R la base que se usará para el análisis usando aov

x75_fN_AdeV<-aov(Xilanasa~Tratamientos)

# El AdeV se hace con aov. Se escribe el modelo (respuesta~factor)

replications(Xilanasa~Tratamientos,data = baseXilanasaA,na.fail)

# Se verifica el número de repeticiones con replications, se da el modelo y

# se indica que na es faltante con na.fail. aov requiere igual número de repeticiones

summary(x75_fN_AdeV) # Muestra el AdeV

list(x75_fN_AdeV) # Otro resumen, muestra el error estándar del residual (del error)

medias<-model.tables(x75_fN_AdeV,type = "means",se = TRUE)

# Las medias (type = "means") de x75 para cada nivel del factor en el AdeV y el error estándar de

# la diferencia de medias (eedm). Si no se desea el eedm se escribe se = FALSE

medias # Muestra las medias

list(FuenteAzucar) # Para tener a la vista las fuentes de azúcar en el análisis

TukeyHSD(x75_fN_AdeV,which = "Tratamientos",orderer = TRUE, conf.level = 0.95)

```

Prueba de Tukey (TukeyHSD) para el AdeV (x75_fN_AdeV), para el factor tratamientos (wich = "Tratamientos),

ordenados (orderer = TRUE) con nivel de confianza de 95% (conflvel = 0.95)

boxplot(Xilanasa~ Tratamientos,data = baseXilanasaA,outline = TRUE,beside = TRUE,boxwex = .5,

xlab = "Fuente de Azúcar codificada",ylab = "Xilanasa pH 7.5",main = " ")

Gráfica de cajas para los Tratamientos en el modelo

1.1. Ejercicio

Se hizo un experimento utilizando *Staphylococcus sp.* para evaluar la actividad de xilanasa variando la proporción de glucosa y xilosa. Se realizaron tres repeticiones y la unidad experimental fue un matraz Erlenmeyer de 250 ml con 100 ml de medio líquido estándar, se agitó a 200 rpm y se mantuvo a 37°C con pH inicial de 8.0 durante tres días, se usaron 12 matraces. Se midió la actividad de xilanasa a pH de 7.5 (UI). Interesa probar el efecto de las combinaciones de glucosa y xilosa. Los datos están a continuación.

% de Glucosa	% de Xilosa	Actividad de Xilanasa a pH 7.5 (UI)
0.3	0.7	215.2
0.3	0.7	213.1
0.3	0.7	222.0
0.4	0.6	200.5
0.4	0.6	194.6
0.4	0.6	201.3
0.5	0.5	189.5
0.5	0.5	194.8
0.5	0.5	203.6
0.6	0.4	182.3
0.6	0.4	173.9
0.6	0.4	166.7

Haga el análisis similar al del problema 1. Note que sólo debe usar una de las variables explicativas, ya que ambas suman uno.

Bibliografía

Grupa, S., Kunhad, R.C., Bhusmhan, B. and Hoondal, G.S. (2001) Improved xylanase production from a haloalkalophilic *Staphylococcus sp.* SG-13 using inexpensive agricultural residues. *World Journal of Microbiology and Biotechnology*. 17:5-8.

Nota: La referencia se utilizó, de manera fraccionada, para obtener problemas con el nivel y el interés del curso.

2. Biorremediación de suelos contaminados por hidrocarburos

Se realizó un experimento para probar el efecto de la adición de nutrientes, inoculación de microorganismos y oxigenación en el porcentaje de biodegradación de hidrocarburos. Para simular un suelo, se utilizaron 24 columnas cilíndricas de 80 cm de largo por 10 cm de diámetro empacadas con arena de coral y contaminadas con 45000 ppm de hidrocarburos de petróleo. Los nutrientes que se aplicaron disueltos en agua estuvieron compuestos por 20% de NH_4Cl y 2% de K_2HPO_4 ; el inóculo de microorganismos fue de 0.5 g/l de materia seca con 1.2×10^7 CFU/g, y la oxigenación se hizo adicionando 0.1% (p/p) de CaO_2 al agua.

El agua se suministró por aspersión para mantener 11% de humedad. Después de 12 semanas de iniciado el experimento, se determinó la concentración de hidrocarburos presentes en la sección superior de la columna. A fin de calcular la variable de respuesta: el porcentaje de eliminación de hidrocarburos (EH), se hizo una transformación de los datos experimentales, pues se calculó la diferencia entre el contenido inicial y final de hidrocarburos y el resultado se dividió entre el contenido inicial y se multiplicó por 100.

En la Tabla 2.1 se muestran los datos obtenidos en cada tratamiento evaluado. Como se puede observar, los datos de EH provienen de la realización de un experimento con diseño completamente al azar y arreglo factorial de tratamientos 2^3 , con los tres factores antes mencionados, de dos niveles cada uno: ausencia (valor 0) o presencia (valor 1). Se realizaron tres repeticiones.

Tabla 2.1. Porcentaje de eliminación de hidrocarburos obtenido en cada tratamiento.

Tratamiento	Nutrientes (Nutr)	Inóculo (Inoc)	Oxigenación (Oxig)	% Eliminación (EH)
1	0	0	0	47.47
1	0	0	0	44.3
1	0	0	0	37.4
2	0	0	1	64.01
2	0	0	1	57.23
2	0	0	1	54.11
3	0	1	0	59.75
3	0	1	0	65.33
3	0	1	0	62.21
4	0	1	1	61.51
4	0	1	1	64.24
4	0	1	1	76.38
5	1	0	0	60.36
5	1	0	0	69.64
5	1	0	0	59.09
6	1	0	1	63.74
6	1	0	1	59.95
6	1	0	1	76.96
7	1	1	0	59.65
7	1	1	0	63.7
7	1	1	0	63.99
8	1	1	1	75.87
8	1	1	1	80.6
8	1	1	1	65.43

Para probar las hipótesis sobre los efectos de los factores nutrición, inoculación, oxigenación y sus interacciones se plantea el modelo:

$$\mu_{EH/Nutr_i*Inoc_j*Oxig_k} = \mu + Nutri_i + Inoc_j + Oxig_k + (Nutr*Inoc)_{i,j} + (Nutr*Oxig)_{i,k} + (Inoc*Oxig)_{j,k} + (Nut*Inoc*Oxig)_{i,j,k}$$

El significado de cada término en la ecuación anterior es el siguiente: del lado izquierdo de la ecuación el término $\mu_{EH/Nutr_i*Inoc_j*Oxig_k}$ representa la media de EH cuando se tienen los niveles i, j y k (con $i = 0,1, j = 0,1$ y $k = 0,1$) de los factores Nutr, Inoc y Oxig, donde el nivel cero indica que no se aplica el factor, el nivel uno que si se aplica. El producto de los factores $Nutr*Inoc*Oxig$ es la forma abreviada de representar todos los factores e interacciones presentes en el modelo. En el lado derecho de la ecuación, μ es la media general, Nutri es el efecto del nivel i del factor Nutr, Inoc $_j$ es el efecto del nivel j de Inoc, Oxig $_k$ es el nivel k del factor Oxig, los términos entre paréntesis se refieren a los efectos de las interacciones entre dos factores ($Nutr*Inoc$), ($Nutr*Oxig$), ($Inoc*Oxig$) y a la interacción entre los tres factores ($Nutr*Inoc*Oxig$).

La variable aleatoria EH tiene el mismo modelo mas el término de error, como se muestra en la siguiente ecuación. El término $\epsilon_{i,j,k,l}$ es la variable aleatoria ϵ que corresponde a los efectos de los niveles i, j y k de los factores Nutr, Inoc y Oxig, y a la repetición l ($l = 1,2,3$). El término aleatorio o de error ϵ , se distribuye normal con media cero y varianza σ^2 .

$$EH_{i,j,k,l} = \mu_{EH/Nutr_i*Inoc_j*Oxig_k} + \epsilon_{i,j,k,l}$$

Mostrando todos los términos del modelo, la ecuación anterior se escribe:

$$EH_{i,j,k,l} = \mu + Nutri_i + Inoc_j + Oxig_k + (Nutr*Inoc)_{i,j} + (Nutr*Oxig)_{i,k} + (Inoc*Oxig)_{j,k} + (Nut*Inoc*Oxig)_{i,j,k} + \epsilon_{i,j,k,l}$$

$$\text{con } i = 0,1, j = 0,1 \text{ y } k = 0,1 \text{ y } l = 1,2,3 \text{ y } \epsilon_{i,j,k,l} \sim N(0, \sigma^2)$$

Los subíndices del término de error deben ser los mismos que los de la variable EH en la ecuación.

Para el factor nutrientes (Nutr) se plantea la hipótesis de igualdad de efectos de los niveles de Nutr en la media de EH, se escribe $H_0: Nutr_0 = Nutr_1$, con $H_A: Nutr_0 \neq Nutr_1$ y la prueba se efectúa con la F del análisis de varianza, y su correspondiente valor p , que vale 0.003016 (Tabla 2.2). Se rechaza la hipótesis nula y se concluye que hay un efecto de la aplicación de nutrientes en la media de EH.

Para el factor inóculo (Inoc) se plantea la hipótesis de igualdad de efectos de inóculo en la media de EH, se tiene $H_0: Inoc_0 = Inoc_1$ y $H_A: Inoc_0 \neq Inoc_1$. La prueba se efectúa con la F del análisis de varianza. El valor $p = 0.003155$ conduce a rechazar la hipótesis nula y se concluye que hay un efecto de la aplicación de inóculo en la media de EH.

De manera similar, para el factor oxígeno (Oxig) se plantea la hipótesis de igualdad de efectos

$H_0: Oxig_0 = Oxig_1$ con $H_A: Oxig_0 \neq Oxig_1$. La prueba se efectúa con la F del análisis de varianza. El valor $p = 0.002603$ conduce a rechazar la hipótesis nula y se concluye que hay un efecto de la aplicación de oxigenación en la media de % de eliminación de hidrocarburos.

Revisando la tabla de análisis de varianza se ve que, de las interacciones, sólo la de nutrientes por inóculo es significativa, pues tiene valor $p = 0.045059$. La pareja de hipótesis que corresponde a la interacción nutrientes por inóculo es:

$$H_0: (Nutr*Inoc)_{0,0} = (Nutr*Inoc)_{0,1} = (Nutr*Inoc)_{1,0} = (Nutr*Inoc)_{1,1} = 0 \text{ con}$$

$$H_A: (Nutr*Inoc)_{i,j} \neq 0 \text{ para al menos una pareja } (i,j); i = 0,1 \text{ y } j = 0,1$$

y su rechazo indica que los efectos de inóculo en la media de EH son diferentes para los dos niveles de nutrientes.

Tabla 2.2. Análisis de varianza completo del tri-factorial: Nutrientes, inóculo y oxígeno en la eliminación de hidrocarburos

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Nutrientes	1	459.73	459.73	12.1918	0.003016
Inóculo	1	454.14	454.14	12.0437	0.003155
Oxígeno	1	478.29	478.29	12.6841	0.002603
Nutr:Inoc	1	178.21	178.21	4.7262	0.045059
Nutr:Oxig	1	9.25	9.25	0.2453	0.627124
Inc Oxig	1	2.90	2.90	0.0769	0.785151
Nutr:Inoc:Oxig	1	123.03	123.03	3.2629	0.089710
Error	16	603.32	37.71		
Total	23	2308.87			

Los resultados de las pruebas en el análisis de varianza, mostrado en la Tabla 2.2, indican que conviene hacer la prueba de Tukey para conocer a qué pareja de combinaciones de nutrientes e inóculo se debe el rechazo de la prueba de su interacción. Para las interacciones no significativas no se hace la prueba de Tukey.

Antes de mostrar los resultados de Tukey, es importante presentar las medias de eliminación de hidrocarburos (Tabla 2.3) para los diferentes niveles de los factores y de las interacciones, que reflejan la acción de los efectos de los factores. Así, en el resultado: hay un efecto de la aplicación de nutrientes en la media de EH, se muestra con la diferencia de las medias que se estimaron en la realización de este experimento, 66.58 en presencia de nutriente y 57.83 en ausencia de éste, que es de 8.75%.

Tabla 2.3. Medias de % de eliminación de hidrocarburos obtenidas en los diferentes tratamientos.

Nutriente		Inóculo		Oxigenación		Nutriente*Inóculo* Oxigenación	
Nivel	Media	Nivel	Media	Nivel	Media		
0	57.83	0	57.86	0	57.74	Nivel-Tratamiento	Media
1	66.58	1	66.56	1	66.67	0,0,0 - 1	43.06
						0,0,1 - 2	58.45
						0,1,0 - 3	62.43
Nivel	Media	Nivel	Media	Nivel	Media	0,1,1 - 4	67.38
0,0	50.75	0,0	52.74	0,0	53.04	1,0,0 - 5	63.03
0,1	64.90	0,1	62.91	0,1	62.67	1,0,1 - 6	66.88
1,0	64.96	1,0	62.74	1,0	62.44	1,1,0 - 7	62.45
1,1	68.21	1,1	70.43	1,1	70.67	1,1,1 - 8	73.97
Nivel 0 = no se adiciona, Nivel 1 = Se adiciona, dos o más factores se refiere a la interacción entre ellos							

La interacción de nutrientes por inóculo se analiza con la prueba de Tukey (Tabla 2.4) donde se observa que el tratamiento sin nutrientes ni inóculo (0,0) es significativamente diferente de los otros tres, pero estos tres no difieren entre sí. La interpretación se facilita con la gráfica de la interacción (Figura 2.1).

Tabla 2.4. Tukey para efectos de interacción nutrientes-inoculación en la media de eliminación de hidrocarburos.

Efectos de interacción entre Nutr e Inoc que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
(1,0)-(0,0)	14.20333333	4.060119	24.34655	0.0050584
(0,1)-(0,0)	14.15000000	4.006786	24.29321	0.0052158
(1,1)-(0,0)	17.45333333	7.310119	27.59655	0.0007944
(0,1)-(1,0)	-0.05333333	-10.196548	10.08988	0.9999987
(1,1)-(1,0)	3.25000000	-6.893214	13.39321	0.7963725
(1,1)-(0,1)	3.30333333	-6.839881	13.44655	0.7885019

Los tres tratamientos combinados: nutriente sin y con inóculo e inóculo sin nutriente, no difieren significativamente y los tres superan al de no aplicar ninguno de los dos factores. El investigador puede elegir cualquiera de las tres posibilidades. Una forma de mostrar los resultados de la prueba de Tukey se presenta en la Tabla 2.5, donde las medias que no difieren entre sí están unidas por una línea.

Tabla 2.5. Prueba de Tukey. Medias unidas por una línea no difieren

Tratamiento (Nutr-Inoc)	1,1	1,0	0,1	0,0
Media muestral	68.21	64.96	64.90	50.75
	-----			-----

En la siguiente gráfica (Figura 2.1) se ve que las líneas no son paralelas –indicando que hay efecto de interacción–, que las tres medias mayores no son muy diferentes y la menor media, sin nutrientes ni inóculo, es menor que las otras y difiere, como lo mostró la prueba de Tukey. El valor de la media más grande (68.21) se obtiene cuando se aplican nutrientes e inóculo, la cual no difiere de las que toman valores prácticamente iguales (64.96 y 64.90); la diferencia se podría tratar de verificar en un experimento futuro.

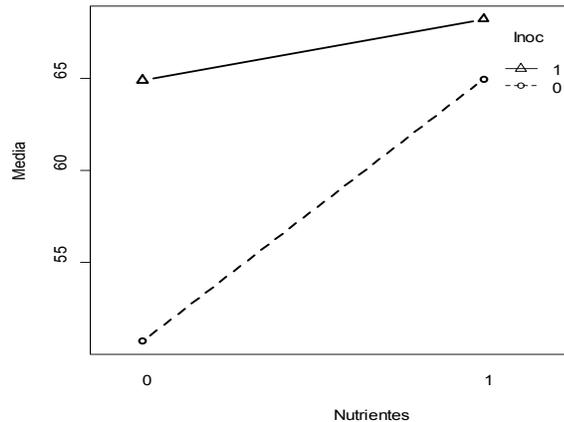


Figura 2.1. Gráfica de la interacción entre nutrientes e inóculo

El análisis se resume en que, para cada factor, la media de adicionarlo es mayor que la media de no adicionarlo, sólo la interacción de nutrientes con inóculo resultó significativa, es decir, la adición de nutrientes tiene diferente efecto si el inóculo está presente o no.

El análisis de eliminación de hidrocarburos se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares, tratando de no repetir anotaciones que están en instrucciones de problemas anteriores.

Guía para resolver con R

```
datosEHemed<-read.table(file = "bioRemed.csv",header = T,sep = ",")
```

```
# Note que los datos en Excel no tienen la identificación de los ocho tratamientos
```

```
# Recuerde que la localización y nombre del archivo que lee, entre comillas luego de file = debe modificarlo
```

```
datosEHemed
```

```
Nu<-datosEHemed$Nutrientes
```

```
In<-datosEHemed$Inoculación
```

```
Ox<-datosEHemed$Oxigenación
```

```
EH<-datosEHemed$BioRemed
```

```
tra<-c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6,7,7,7,8,8,8)
```

```
# Se genera el vector que identifica a los 8 tratamientos
```

```
int<-c(1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4)
```

```
# se genera el vector que identifica a las combinaciones de tratamientos que entran en la interacción
```

```
# nutrientes*inoculación, son (0,0), (0,1), (1,0) y (1,1) para Nu e In
```

```
# Nut # Inoc # Oxig # EH # tra # In

baseEH<-data.frame(Nutr = factor(Nu),Inoc = factor(In),Oxig = factor(Ox),

trat = factor(tra),intNI = factor(int),BioRem = EH)

# Se crea el marco con todos los factores que se usarán en el análisis

baseEH

attach(baseEH)

EHAdEV<-aov(EH~Nutr*Inoc*Oxig,data = baseEH)

replications(EH~Nutr*Inoc*Oxig,data = baseEH,na.fail)

summary(EHAdEV)

SCT<-459.73+454.14+478.29+178.21+9.25+2.90+123.03+603.32

SCT # Cálculo de Suma de Cuadrados Total

medias<-model.tables(EHAdEV,type = "means",se = FALSE)

medias # medias de cada nivel de factores e interacciones

TukeyHSD(EHAdEV, which = "Nutr:Inoc",orderer = TRUE, conf.level = 0.95)

# Prueba de Tukey para la interacción Nutr:Inoc (wich = "Nutr:Inoc")

interaction.plot(Nutr,Inoc,EH,fun = mean,type = "b",lwd = 2,pch = 1:3,xlab = "Nutrientes",ylab = "Media")

# Gráfica de la interacción Nutr:Inoc para EH, se deben escribir en orden los factores que van en los ejes de x,

# y y la respuesta, luego se especifica que se usa la media y los detalles usuales en una gráfica.
```

2.1. Ejercicio

Se realizó un experimento complementario al del problema 2, en el que se utilizó un diseño completamente al azar y arreglo factorial 2^3 con tres repeticiones para probar el efecto de la adición de nutrientes, la inoculación de microorganismos y la oxigenación en el porcentaje de biodegradación de hidrocarburos. De manera similar al ejercicio 2, para simular un suelo, se utilizaron 24 columnas cilíndricas de 80 de largo por 10 de diámetro empacadas con arena de coral y con 45000 ppm de hidrocarburos de petróleo. Los factores, que se aplicaron fueron nutrientes en dos niveles: 1) compuesto de 20% de NH_4Cl y 2% de K_2HPO_4 y 2) compuesto de 40% de NH_4Cl y 4% de K_2HPO_4 , inóculo de microorganismos en dos niveles: 1) 0.5 g/ l de materia seca con 1.2×10^7 CFU/g y 2) 1.0 g/ l de materia seca con 1.2×10^7 CFU/g, y oxigenación en dos niveles, que se hizo adicionando 1) 0.1% (p/p) de CaO_2 y 2) 0.2% (p/p) de CaO_2 al agua. El agua se suministró a las columnas por aspersión para mantener 11% de humedad. Después de 12 semanas de iniciado el experimento, se determinó la concentración de hidrocarburos presentes en la sección superior de la columna y por diferencia con el contenido inicial de hidrocarburos, se calculó el porcentaje de biodegradación de hidrocarburos.

Haga el análisis.

Tratamiento	Nutrientes (Nutr)	Inóculo (Inoc)	Oxigenación (Oxig)	% Biorremediación (BR)
1	1	1	1	57.43
1	1	1	1	54.35
1	1	1	1	45.42
2	1	1	2	64.04
2	1	1	2	57.62
2	1	1	2	57.23
3	1	2	1	61.36
3	1	2	1	65.27
3	1	2	1	58.91
4	1	2	2	63.36
4	1	2	2	62.39
4	1	2	2	71.21
5	2	1	1	60.230
5	2	1	1	69.52
5	2	1	1	58.18
6	2	1	2	63.12
6	2	1	2	64.83
6	2	1	2	72.41
7	2	2	1	60.67
7	2	2	1	64.28
7	2	2	1	59.96
8	2	2	2	74.92
8	2	2	2	81.55
8	2	2	2	79.66

Bibliografía

-  Phelps, T.J., Siegrist, R.L., Korte, N.E., Pickering D.A., Strong-Gunderson, J.M., Palumbo, A.V., Walker, J.F., Morrissey, C.M. and Mackowski R. (1994) Bioremediation of petroleum hydrocarbons in soil column lysimeters from Kwajalein Island. *Applied Biochemistry and Biotechnology*. 45 (1): 835-845.
-  Rios, V. (2005) Tesis de Maestría en Biotecnología. Estudio de estimulación biológica para el tratamiento de residuos de perforación petrolera empleando lisímetros. Universidad Autónoma Metropolitana-Iztapalapa. México.

Nota: Las referencias se utilizaron, de manera combinada para plantear un problema al nivel y el interés del curso

3. Biofiltración de vapor de metano

Se hizo un experimento para conocer la eficiencia de remoción de metano utilizando un reactor tipo columna empacada que consta de cuatro secciones. Dentro de cada columna, que constituyó la unidad experimental, están las cuatro secciones, caracterizadas por su altura, por lo que cada sección es una subunidad. El soporte sólido está conformado por 2 partes de turba esterilizada con vapor por 3 partes de partículas de perlita. Como inóculo se usó un consorcio bacteriano constituido por ocho miembros y que había sido seleccionado anteriormente de un suelo expuesto a metano y que se hizo crecer en 1% de metano. El inóculo se cosechó por centrifugación y se le suspendió en un medio mineral fresco; a la columna se agregó suficiente suspensión para llenar el 50% de los poros del soporte con el líquido.

Se tuvieron quince columnas de 1 m de altura con cuatro secciones de soporte cada una. El factor tratamientos fue la concentración de metano en la corriente de aire que se inyectaba a la columna, con tres niveles de concentración (CMI = concentración de ingreso de metano): 2.6, 4.2 y 6.3 g/m³. La variable de respuesta fue la eficiencia de remoción de metano, que se calculó como el porcentaje de remoción con respecto al valor de ingreso mediante la transformación de datos: $[(\text{concentración final} - \text{concentración inicial}) / (\text{concentración inicial})] * 100$ después de diez días de iniciado el proceso.

El experimento realizado corresponde a un diseño completamente al azar con cinco repeticiones y tres niveles del factor concentración de ingreso de metano: 2.6, 4.2 y 6.3. Dado que cada unidad experimental tiene cuatro subunidades que corresponden a las alturas del biofiltro en las que se hizo la medición del metano eliminado, el diseño se conoce como diseño completamente al azar de parcelas divididas, con tres niveles en el factor de parcelas grandes (unidades) y cuatro niveles en el factor de parcelas chicas (subunidades). En total hay $5 \times 3 \times 4 = 60$ observaciones. Los datos obtenidos se presentan en la Tabla 3.1. Las columnas se denotan por col, los tratamientos son los niveles de concentración de ingreso de metano, las secciones de soporte son las alturas y el porcentaje de remoción de metano se denota como metano.

Tabla 3.1. Arreglo experimental y datos de eliminación de metano obtenidos por tratamiento.

col.	Tratamiento (CMI)	Altura	Metano (%)	col.	Tratamiento (CMI)	Altura	Metano (%)	col.	Tratamiento (CMI)	Altura	Metano (%)
1	2.6	0.25	0.478	6	4.2	0.25	0.44	11	6.3	0.25	0.829
1	2.6	0.5	0.488	6	4.2	0.5	0.474	11	6.3	0.5	0.788
1	2.6	0.75	0.386	6	4.2	0.75	0.53	11	6.3	0.75	0.878
1	2.6	1	0.498	6	4.2	1	0.497	11	6.3	1	0.814
2	2.6	0.25	0.512	7	4.2	0.25	0.543	12	6.3	0.25	0.771
2	2.6	0.5	0.492	7	4.2	0.5	0.517	12	6.3	0.5	0.86
2	2.6	0.75	0.421	7	4.2	0.75	0.554	12	6.3	0.75	0.825
2	2.6	1	0.43	7	4.2	1	0.495	12	6.3	1	0.872
3	2.6	0.25	0.454	8	4.2	0.25	0.461	13	6.3	0.25	0.777
3	2.6	0.5	0.601	8	4.2	0.5	0.446	13	6.3	0.5	0.793
3	2.6	0.75	0.46	8	4.2	0.75	0.485	13	6.3	0.75	0.803
3	2.6	1	0.387	8	4.2	1	0.592	13	6.3	1	0.899
4	2.6	0.25	0.504	9	4.2	0.25	0.588	14	6.3	0.25	0.75
4	2.6	0.5	0.476	9	4.2	0.5	0.583	14	6.3	0.5	0.797
4	2.6	0.75	0.521	9	4.2	0.75	0.541	14	6.3	0.75	0.87
4	2.6	1	0.345	9	4.2	1	0.524	14	6.3	1	0.853
5	2.6	0.25	0.506	10	4.2	0.25	0.534	15	6.3	0.25	0.814
5	2.6	0.5	0.449	10	4.2	0.5	0.49	15	6.3	0.5	0.868
5	2.6	0.75	0.434	10	4.2	0.75	0.596	15	6.3	0.75	0.839
5	2.6	1	0.537	10	4.2	1	0.477	15	6.3	1	0.867

Para probar las hipótesis de igualdad de efectos de tratamientos en las unidades experimentales, de efectos de altura y de la interacción tratamientos por altura en subunidades, se plantea el siguiente modelo completamente al azar con parcelas divididas:

$$m_{i,j,k} = \mu + \text{cim}_j + \varepsilon_{i,j} + \text{alt}_k + (\text{cim} * \text{alt})_{j,k} + \delta_{i,j,k}$$

$$\varepsilon_{i,j} \sim N(0, \sigma_{ue}^2) \quad \delta_{i,j,k} \sim N(0, \sigma_{\text{sub-ue}}^2); i = 1,2,3,4,5, j = 1,2,3 \text{ y } k = 1,2,3,4$$

El significado de los símbolos en la ecuación anterior en el orden en que aparecen es el siguiente. Del lado izquierdo de la ecuación, $m_{i,j,k}$ es la variable aleatoria remoción de metano, el subíndice i se refiere a la columna i , se tienen cinco columnas para cada tratamiento de concentración de ingreso de metano, los cuales se identifican con el subíndice j , cuyos valores son: 1 = 2.6, 2 = 4.2 y 3 = 6.3. El subíndice k identifica a la altura de la columna, que cuenta con cuatro niveles 0.25, 0.5, 0.75 y 1.00, correspondientes a los valores 1 a 4 de este subíndice. Del lado derecho, μ es la media general del experimento, cim_j es el nivel j del factor concentración de ingreso de metano, $\varepsilon_{i,j}$ es una variable aleatoria que corresponde al error de unidad experimental en relación a la columna i y al nivel j de cim , alt_k es el nivel k de la altura del biofiltro a la que se hizo la medición de la variable respuesta remoción de metano, $(\text{cim} * \text{alt})_{j,k}$ es el efecto de la combinación del nivel j de cim con el nivel k de alt , y $\delta_{i,j,k}$ es el error de subunidad. Los subíndices i, j, k identifican a los niveles de columna, concentración de cim y altura del biofiltro al que se hizo la medición de remoción de metano (m). El término aleatorio $\varepsilon_{i,j}$ se distribuye normal con media 0 y varianza desconocida σ_{ε}^2 , y $\sigma_{\text{sub-ue}}^2$, el error de subunidad que corresponde al nivel i de biofiltro en el nivel j de concentración de ingreso de metano y al nivel k de altura de medición, se distribuye normal con media cero y varianza desconocida σ_{ε}^2 .

El análisis de varianza permite probar las hipótesis siguientes. Para el factor concentración de ingreso de metano se tiene:

$$H_0: \text{cim}_1 = \text{cim}_2 = \text{cim}_3, \quad H_A: \text{cim}_1 \neq \text{cim}_2 \text{ ó } \text{cim}_1 \neq \text{cim}_3 \text{ ó } \text{cim}_2 \neq \text{cim}_3, \text{ para altura de la medición:}$$

$$H_0: \text{alt}_1 = \text{alt}_2 = \text{alt}_3 = \text{alt}_4, \quad H_A: \text{alt}_k \neq \text{alt}_{k'}, \text{ para al menos una pareja de subíndices } k = k', k, k' = 1,2,3,4.$$

y para interacción entre concentración de ingreso de metano y altura de la medición:

$$H_0: (\text{cim} * \text{alt})_{1,1} = (\text{cim} * \text{alt})_{1,2} = (\text{cim} * \text{alt})_{1,3} = \dots = (\text{cim} * \text{alt})_{3,4}, \text{ con:}$$

$$H_A: (\text{cim} * \text{alt})_{j,k} \neq (\text{cim} * \text{alt})_{j,k'} \text{ para al menos una pareja de subíndices } (j,k) \neq (j,k'), j = 1,2,3 \text{ y}$$

$$k, k' = 1,2,3,4$$

En la Tabla 3.2 del análisis de varianza se muestra que el valor p para la prueba de igualdad de efectos de ingreso de metano, es $1.906E^{-12}$, al ser un valor menor que 0.05 la hipótesis nula de igualdad de efectos se rechaza. Ante este resultado es conveniente hacer la prueba de Tukey para identificar el nivel que da la mayor media. El valor p para la prueba de igualdad de efectos de altura es 0.91419, es decir; no hay evidencia para rechazar la hipótesis de igualdad de efectos de altura en la media de remoción de metano. Por último, la tabla 3.2 muestra un valor p de 0.07636, que no es significativo al 5%, para la pareja de hipótesis sobre la interacción ingreso de metano y altura. Con base en estos resultados se concluye: no hay efecto de la altura en el biofiltro en la media de remoción de metano, no hay efecto de interacción entre tratamiento y altura y las medias de remoción de metano para los niveles de concentración de ingreso de metano no son iguales.

Tabla 3.2. Análisis de varianza para la eliminación de metano.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
IngMetano	2	1.51806	0.75903	532.86	1.906E ⁻¹²
Residual (unidades)	12	0.01709	0.00142		
Altura	3	0.001329	0.0004429	0.1726	0.91419
IngMetano: Altura	6	0.032446	0.0054077	2.1079	0.07636
Residual (subunidades)	36	0.092356	0.0025654		

En la Tabla 3.3 se presentan las medias muestrales. Por nivel de ingreso de metano, la mayor media, con valor 0.828, corresponde al ingreso de metano 6.3. Entre las medias de remoción de metano correspondientes a los niveles de la interacción concentración-altura, la mayor (con valor de 0.861) corresponde a una concentración de ingreso de metano de 6.3 y una altura en el biofiltro de 1.

Tabla 3.3. Medias de eliminación de metano (a) y medias de eliminación de metano con la interacción concentración*altura del biofiltro (b).

(a) Medias para los niveles de Ingreso de Metano	
Ingreso Metano	Media
2.6	0.469
4.2	0.518
6.3	0.828

(b) Medias de remoción de metano para la interacción concentración-altura				
Nivel de Concentración de Ingreso de metano	Nivel de altura en el Biofiltro			
	0.25	0.50	0.75	1.00
2.6	0.491	0.501	0.444	0.439
4.2	0.513	0.502	0.541	0.517
6.3	0.788	0.821	0.843	0.861

La instrucción de TukeyHSD utilizando el paquete R no se puede aplicar a estos modelos. Sin embargo, la prueba de Tukey se puede hacer utilizando el valor crítico de la q de Tukey para 5% de significación, tres tratamientos y 12 grados de libertad del residual ($q_{0.95,3,12}$), el cuadrado medio del error de unidades y el número de repeticiones. Así, cada diferencia de medias, en valor absoluto, se compara con:

$$q_{0.95,3,12} \sqrt{\frac{\text{cme}(\text{unidades})}{r}} = 3.772929 \sqrt{\frac{0.0014}{5}} = 0.08545497$$

La diferencia para los tratamientos 1 y 2 es 0.049, para los tratamientos 1 y 3 es 0.359 y para los tratamientos 2 y 3 es 0.31. Los tratamientos uno y dos no difieren (su diferencia es menor que 0.0854), pero ambos difieren del tratamiento 3 (tanto 0.359 como 0.31 son mayores que 0.0854). Esto se muestra en la Tabla 3.4, donde las medias unidas por una línea no difieren entre sí.

Tabla 3.4. Prueba de Tukey para tratamientos. Las medias unidas por una línea no difieren

Tratamiento (cim)	3 (6.3)	2 (4.2)	1 (2.6)
Media muestral	0.828	0.518	0.469
	-----		-----

La interacción no resultó significativa al 5% pero si al 10%. En la gráfica de la interacción (Figura 3.1), se puede ver que las líneas no son paralelas en los dos niveles bajos de concentración de ingreso de metano.

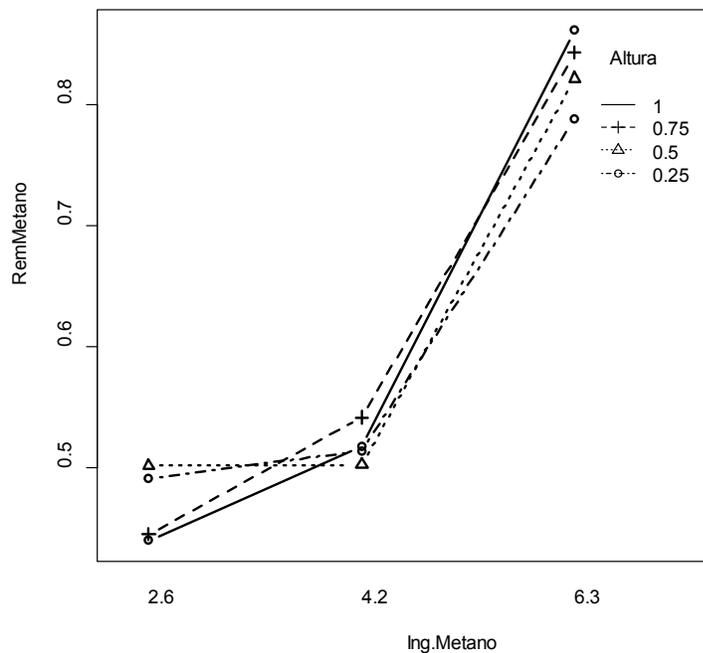


Figura 3.1 Gráfica de interacciones.

El análisis del experimento de biofiltración de vapor de metano se hace con las instrucciones del paquete R que siguen.

Guía para resolver con R

```
datmetanoA<-read.table(file = "metanoA.csv",header = T,sep = ",")
```

```
datmetanoA
```

```
col<-datmetanoA$columna
```

```
bloq<-datmetanoA$rep
```

```
ci<-datmetanoA$conc.ingreso
```

```

ac<-datmetanoA$altura.columna

cm<-datmetanoA$conc.metano

# col # bloq # ci # ac # cm

basemetanoA<-data.frame(BF = factor(col),blo = factor(bloq),IngMetano = factor(ci),
Altura = factor(ac),RemMetano = cm)

basemetanoA

attach(basemetanoA)

metanoAdeV<-aov(RemMetano~IngMetano+Error(BF)+Altura+IngMetano:Altura)

replications(RemMetano~IngMetano+Error(BF)+Altura+IngMetano:Altura,data = basemetanoA,na.fail)

summary(metanoAdeV)

print(model.tables(metanoAdeV,"means"),digits = 3)

q0.95<-qtukey(0.95, nmeans = 3, df = 12, nranges = 1, lower.tail = TRUE, log.p = FALSE)

vCritico<-q0.95*sqrt(0.002565/5)

vCritico

t1_t2<-abs(0.469-0.518)

t1_t3<-abs(0.469-0.828)

t2_t3<-abs(0.518-0.828)

t1_t2

t1_t3

t2_t3

interaction.plot(IngMetano,Altura,RemMetano,type = "b",lwd = 2,pch = 1:3,xlab = "Ing.Metano",
ylab = "RemMetano")

```

3.1. Ejercicio

Se hizo un experimento complementario al problema 3, con las variables que resultaron significativas para conocer la velocidad de remoción de metano en una columna de 1 m formada por cuatro secciones de soporte sólido de 2 partes de turba esterilizada con vapor por 3 partes de partículas de perlita. Se usó un consorcio de bacterias de ocho miembros que se había seleccionado anteriormente de suelo expuesto a metano y que se hizo crecer en 1% de metano, se le cosechó por centrifugación y se le suspendió en un medio mineral fresco; se agregó suficiente suspensión para llenar 50% de los poros del soporte con el líquido.

Se tuvieron quince columnas con cuatro secciones de soporte cada una, el factor tratamientos fue la concentración de metano en la corriente de aire que se inyectaba a la columna, con tres niveles de concentración (CMI = concentración de ingreso de metano): 2.6, 4.2 y 6.3 g/m³. La variable respuesta fue la eficiencia de remoción (de metano), que se obtuvo como el porcentaje de remoción con respecto al valor de ingreso (concentración final-concentración inicial) /concentración inicial) en el día diez después de iniciado el proceso en la parte media de la columna. Los datos fueron:

col.	Tratamiento	Metano	col.	Tratamiento	Metano	col.	Tratamiento	Metano
1	2.6	0.488	6	4.2	0.474	11	6.3	0.788
2	2.6	0.492	7	4.2	0.517	12	6.3	0.86
3	2.6	0.601	8	4.2	0.446	13	6.3	0.793
4	2.6	0.476	9	4.2	0.583	14	6.3	0.797
5	2.6	0.449	10	4.2	0.49	15	6.3	0.868

Se tiene un diseño completamente al azar con cinco repeticiones del factor concentración de ingreso de metano con tres niveles: 2.6, 4.2 y 6.3. En total hay $5 \times 3 = 15$ observaciones. La variable respuesta fue el porcentaje de remoción de metano. Haga el análisis.

Bibliografía

-  Shareefdeen, Z., Baltzis, B.C., Oh, Y.S. and Bartha, R. (1993) Biofiltration of methane vapor. *Biotechnology and Bioengineering*. 41: 512-524.

Nota: La referencia se utilizó para plantear un problema similar pero llevado al nivel del curso

4. Uso de residuos agrícolas para la producción de xilanas de *Staphylococcus sp.*

Las xilanasas son enzimas hidrolíticas que actúan sobre el xilano para convertirlo en los azúcares que lo constituyen. La acción de las xilanasas permite su utilización en diversos procesos industriales y principalmente en la obtención de papel. Una fuente importante de xilanasas es mediante el cultivo de microorganismos.

Con el fin de producir xilanasas a partir de *Staphylococcus sp.*, se hizo un experimento en el que se evaluaron dos factores. Se estudió el efecto de cuatro residuos agrícolas como sustrato, más un control sin residuo agrícola, de forma que el factor residuo tiene 5 niveles. Además, el efecto de dos tipos de base de sustrato (medios de cultivo): agua desionizada y medio Horikoshi. Los residuos orgánicos se obtuvieron en tres meses diferentes y cada mes se hizo una repetición. Esto significa que se realizó un diseño en bloques al azar con dos factores: residuo agrícola y medio de cultivo, donde los meses corresponden a los bloques. La unidad experimental fue un matraz Erlenmeyer de 250 ml con 100 ml de medio líquido con pH inicial de 8.0, se agitó a 200 rpm y se mantuvo a 37°C. Como variable de respuesta se determinó la actividad enzimática de xilanasas (Unidades Internacionales, UI) a dos valores de pH, 7.5 y 9.2. Dado que la variable respuesta se analiza por separado a cada valor de pH y como en ambos casos se mide la actividad enzimática de xilanasas, parte del objetivo de este experimento es saber si es suficiente con evaluar la actividad enzimática a un solo pH. Las mediciones de actividad enzimática de xilanasas se llevaron a cabo cuando la fermentación alcanzó su máxima actividad y se muestran en la Tabla 4.1.

Tabla 4.1. Datos de actividad enzimática de xilanasas obtenidas con dos fuentes de residuo agrícola y dos tipos de base de sustrato.

Residuo agrícola 0.5% w/v+	Mes o bloque	Producción de enzima UI y tiempo a máxima actividad			
		Agua desionizada		Medio de Horikoshi	
		pH 7.5	pH 9.2	pH 7.5	pH 9.2
Control	1	0	0	340	220
Madera de álamo	1	2077	2021	2374	1680
Bagazo de caña	1	3571	3597	3880	3395
Salvado de trigo	1	2974	2406	3435	3120
Carozo de maíz	1	3061	2224	3110	2337
Control	2	0	0	805	582
Madera de álamo	2	2293	1925	2978	2521
Bagazo de caña	2	3822	3419	3979	3100
Salvado de trigo	2	3200	3042	4462	3908
Carozo de maíz	2	3197	3047	3086	2895
Control	3	0	0	520	363
Madera de álamo	3	2833	1971	2877	2317
Bagazo de caña	3	4398	4202	4378	4367
Salvado de trigo	3	3911	3498	4537	4017
Carozo de maíz	3	3708	3166	3589	3462

Para realizar el análisis de varianza, inicialmente se incluye el control, en una segunda etapa se elimina del análisis debido a que produce mucho menos xilanasas que los residuos agrícolas.

Para la actividad enzimática xilanasas a pH 7.5 se plantea el modelo para las medias:

$$\mu_{x75/rA_i, mC_j, b_k, (rA * mC)_{i,j}} = \mu + rA_i + mC_j + b_k + (rA * mC)_{i,j}; i = 0, \dots, 4, j = 1, 2, k = 1, 2, 3$$

que se completa con la suposición de un error aleatorio con distribución normal con media cero y varianza σ^2 desconocida. Ya con el error (la variación aleatoria), el modelo se escribe:

$$x75_{i,j,k} = \mu_{x75/rA_i, mC_j, b_k, (rA * mC)_{i,j}} + \varepsilon_{i,j,k}; i = 0, \dots, 4, j = 1, 2, k = 1, 2, 3$$

ó, incluyendo todos los términos como:

$$x75_{i,j,k} = \mu + rA_i + mC_j + b_k + (rA * mC)_{i,j} + \varepsilon_{i,j,k}; i = 0, \dots, 4, j = 1, 2, k = 1, 2, 3.$$

En las ecuaciones anteriores se debe entender que:

- $x75$ representa a la variable producción de xilanasas pH 7.5
- $\mu_{x75/rA_i, mC_j, b_k, (rA * mC)_{i,j}}$ representa la media de la variable $x75$ en el nivel i de rA , el nivel j de mC y el nivel k de b ,
- rA_i es el efecto del i -ésimo residuo agrícola sobre la media de $x75$,
- mC_j es el efecto del nivel j del medio de cultivo en la media de $x75$,
- b_k es el efecto del nivel k de bloque en la media de $x75$,
- $(rA * mC)_{i,j}$ es el efecto de la interacción de los niveles i de rA y j de mC en la media de $x75$,
- μ es la media general.
- Hay cinco niveles de residuo agrícola, que son los cinco residuos estudiados: (0) control (sin residuo agrícola), (1) Madera de álamo, (2) Bagazo de caña, (3) Salvado de trigo y (4) Carozo de maíz, se escribe $i = 0, 1, 2, 3, 4$
- Hay dos niveles de medio de cultivo: agua desionizada y Medio de Horikoshi, se escribe $j = 1, 2$,
- Hay tres bloques que corresponden con los tres meses, se escribe $k = 1, 2, 3$,
- $x75_{i,j,k}$ es la variable aleatoria producción de xilanasas pH 7.5 que corresponde al nivel i de residuo agrícola, el nivel j de medio de cultivo y en el bloque k ,
- $\varepsilon_{i,j,k}$ es el error o variación aleatoria que corresponde a la variable aleatoria $x75_{i,j,k}$ y
- $\varepsilon_{i,j,k} \sim N(0, \sigma^2)$ y $x75_{i,j,k} \sim N(\mu_{x75/rA_i, mC_j, b_k, (rA * mC)_{i,j}} = \mu + rA_i + mC_j + b_k + (rA * mC)_{i,j}, \sigma^2)$

Utilizando los datos, se hacen las estimaciones de los parámetros del modelo:

$$\mu_{x75/rA_i, mC_j, b_k, (rA * mC)_{i,j}} = \mu + rA_i + mC_j + b_k + (rA * mC)_{i,j} \text{ para } i = 0, \dots, 4, j = 1, 2 \text{ y } k = 1, 2, 3 \text{ y } \sigma^2.$$

En el análisis de varianza (Tabla 4.2) se prueba la igualdad de efectos de residuo agrícola, la igualdad de efectos de medios de cultivo y la no interacción entre los factores residuo agrícola y medio de cultivo.

La hipótesis nula del análisis de varianza en el renglón Residuo Agrícola es que no hay efecto de los residuos agrícolas en la actividad enzimática de xilanasa a pH de 7.5. Los efectos de residuo agrícola en la media de actividad enzimática de xilanasa pH 7.5 se representaron con rA_i , la hipótesis nula de no efecto de residuo agrícola es:

$$H_0: rA_0 = rA_1 = rA_2 = rA_3 = rA_4$$

Puede decirse que la hipótesis nula establece que los efectos de los residuos agrícolas, incluyendo el control, son iguales en la media de producción de xilanasa pH 7.5. La hipótesis alternativa niega a la hipótesis nula y establece que los efectos de los diferentes residuos agrícolas no son todos iguales en la media de xilanasa pH 7.5; puede escribirse como:

HA: Al menos una pareja de efectos de los residuos agrícolas rA_i y rA_j , con $i, j = 0, 1, 2, 3, 4$ no son iguales.

Se rechaza la hipótesis nula con un valor p de $1.989E^{-15}$, se concluye que al menos una pareja de residuos agrícolas tiene efectos diferentes. Más adelante se hace la prueba de Tukey para conocer cuáles parejas de efectos de residuo agrícola son diferentes.

En el renglón correspondiente a medios de cultivo se prueba la hipótesis de que los efectos de los medios de cultivo en la media de actividad enzimática de xilanasa a pH 7.5 son iguales. Los efectos de medio de cultivo en la media de actividad enzimática de xilanasa a pH 7.5 se representaron con mC_i , la hipótesis nula de no efecto de medio de cultivo es:

$$H_0: mC_1 = mC_2$$

Se puede decir que la hipótesis nula establece que los efectos de los medios de cultivo en la media de actividad enzimática de xilanasa a pH 7.5 son iguales. La hipótesis alternativa niega a la hipótesis nula y establece que los efectos de los dos medios de cultivo son diferentes y puede escribirse como:

$$H_A: mC_1 \neq mC_2$$

Se rechaza la hipótesis nula con un valor p de 0.0006394 y se concluye que el efecto del medio Horikoshi es diferente del efecto del agua desionizada. Como se cuenta con sólo dos niveles para medio de cultivo, no es necesario realizar la prueba de Tukey para identificar el nivel que produce un efecto mayor, basta con observar los valores de las medias. De hecho, el efecto del medio Horikoshi es mayor, pues las medias son 2603.0 y 2956.7 para agua desionizada y medio Horikoshi respectivamente.

Aunque no hay interés en probar si los efectos de los bloques son iguales o no, en el análisis de varianza se ve que no son iguales. El valor p de 0.0001045 confirma la idea del investigador de que usar los meses como bloques fue acertada.

En el renglón de la interacción se prueba la hipótesis nula de que no hay interacción entre el residuo agrícola y el medio de cultivo. Una forma sencilla de entender esta hipótesis es plantear que las diferencias entre los efectos de los residuos agrícolas cuando todos se prueban usando agua desionizada, son iguales que las diferencias entre los efectos de los residuos agrícolas cuando todos se prueban usando el medio Horikoshi. También se puede interpretar que las diferencias de los efectos de los medios de cultivo en la media de actividad enzimática de xilanasa a pH 7.5 en un residuo agrícola, son iguales que las diferencias para el otro residuo agrícola. Se puede escribir:

$$H_0: (rA * mC)_{ij} = 0 \text{ para } i = 0,1,2,3,4; j = 1, 2.$$

La hipótesis alternativa niega a la nula y se plantea como:

$$H_A: (rA * mC)_{ij} \neq 0 \text{ para al menos una pareja } i = 0,1,2,3,4 \text{ y } j = 1, 2.$$

Si la hipótesis de interacción se plantea sobre las medias se tienen las siguientes hipótesis:

$$H_0: \mu_{i1} - \mu_{i2} = \mu_{i'1} - \mu_{i'2} \text{ para } i \neq i', i, i' = 0,1,2,3,4 \text{ y}$$

HA: $\mu_{i1} - \mu_{i2} \neq \mu_{j1} - \mu_{j2}$ para al menos una pareja $i \neq j$, $i, j = 0, 1, 2, 3, 4$.

El último valor de la columna de valores de p, de la Tabla 4.2, muestra que la prueba de interacción se rechaza con nivel de significación de 5% ya que dicho valor de p es 0.0474130, esto quiere decir que los efectos de medio de cultivo no son los mismos para los diferentes residuos agrícolas.

Tabla 4.2. Análisis de varianza de la actividad enzimática determinada a pH 7.5 y obtenida con diferentes residuos agrícolas y base de medio de cultivo.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Residuo Agrícola	4	54087632	13521908	244.9664	1.989E ⁻¹⁵
Medio de Cultivo	1	938101	938101	16.9949	0.0006394
Bloque	2	1757736	878868	15.9218	0.0001045
Interacción RAgr:MedioCul	4	657629	164407	2.9784	0.0474130
Error	18	993583	55199		

La interacción se comprenderá en la gráfica de medias de x75 para las combinaciones de niveles de rA y mC (Figura 4.1).

Las medias de actividad enzimática de xilanasa a pH de 7.5 para los residuos agrícolas están en la Tabla 4.3, en la cual se observa que la media más alta, con valor de 4005, se obtiene con Bagazo de caña y la segunda más alta (3753), con Salvado de trigo.

Tabla 4.3. Medias de actividad enzimática de xilanasa obtenidas con diferente residuo agrícola.

Residuo agrícola codificado	Residuo agrícola	Media de xilanasa pH = 7.5
0	Control	277
1	Madera de álamo	2572
2	Bagazo de caña	4005
3	Salvado de trigo	3753
4	Carozo de maíz	3292

Para saber cuáles parejas de efectos de residuos agrícolas en la actividad enzimática de xilanasa a pH de 7.5 son diferentes, se sugiere hacer la prueba de Tukey. En la Tabla 4.4 está el resultado de la prueba de Tukey que muestra cuáles parejas de efectos de residuo agrícola son diferentes. Recuerde que Tukey usa un nivel de confianza adecuado para el conjunto de las 10 comparaciones de parejas de residuos agrícolas, se le llama nivel de confianza a nivel de experimento, los valores p también se obtienen para las pruebas conjuntas, a nivel de experimento.

En la columna del valor p, se ve que sólo la pareja 3-2 de residuos agrícolas (bagazo de caña y salvado de trigo) no da evidencia para rechazar la hipótesis de igualdad de efectos, las demás sí presentan diferencia de efectos.

Tabla 4.4. Resultados de la prueba de Tukey de comparaciones múltiples de actividad de xilanas con diferentes residuos agrícolas y evaluada a pH 7.5.

Efectos de residuos agrícolas que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
1-0	2294.5000	1884.3360	2704.66402	0.0000000
2-0	3727.1667	3317.0026	4137.33069	0.0000000
3-0	3475.6667	3065.5026	3885.83069	0.0000000
4-0	3014.3333	2604.1693	3424.49736	0.0000000
2-1	1432.6667	1022.5026	1842.83069	0.0000000
3-1	1181.1667	771.0026	1591.33069	0.0000007
4-1	719.8333	309.6693	1129.99736	0.0004088
3-2	-251.5000	-661.6640	158.66402	0.3753739
4-2	-712.8333	-1122.9974	-302.66931	0.0004554
4-3	-461.3333	-871.4974	-51.16931	0.0232567

La Tabla 4.5 muestra las medias que no difieren entre sí unidas por una línea. Los residuos agrícolas que producen mayor media de actividad enzimática de xilanas pH 7.5 son bagazo de caña y salvado de trigo, sus efectos no difieren significativamente y ambos producen efectos diferentes de los demás residuos agrícolas.

Tabla 4.5. Prueba de Tukey para actividad de xilanas. Las medias unidas por una línea no difieren entre sí.

Tratamiento (Residuo agrícola)	Control (0)	Madera de álamo (1)	Carozo. de maíz (4)	Salvado de trigo (3)	Bagazo de caña (2)
Media muestral	277	2572	3292	3753	4005
	-----	-----	-----	-----	-----

Lo anterior se manifiesta en la gráfica de cajas (Figura 4.1) donde las cajas de los residuos agrícolas 2 y 3 están a la misma altura, más arriba que las demás cajas.

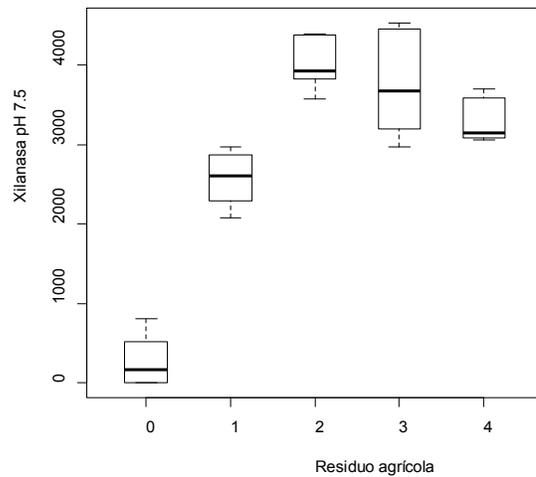


Figura 4.1. Gráfica de cajas de actividad enzimática de xilanas.

La interacción de medio de cultivo con residuos agrícolas se comprende mejor con la gráfica que muestra las medias de actividad enzimática de xilanas de los residuos agrícolas para cada uno de los medios de cultivo (Figura 4.2). El medio de cultivo Horikoshi es la línea sólida y el medio de cultivo de agua desionizada es la línea quebrada. La interacción se manifiesta en la falta de paralelismo entre las dos líneas. En este caso, las líneas son casi paralelas para los residuos agrícolas 0, 1 y 2, pero en los residuos agrícolas 3 y 4 no lo son.

Debido a que la línea del medio de cultivo Horikoshi está arriba de la que corresponde a agua desionizada, el medio Horikoshi produce más xilanas que el agua desionizada para cada uno de los residuos agrícolas, excepto para el 4. Con el medio de cultivo Horikoshi los residuos agrícolas 3 y 4 presentan mayor media (y parece que no difieren) y son las dos combinaciones de mayor producción: bagazo de caña y salvado de trigo en medio Horikoshi.

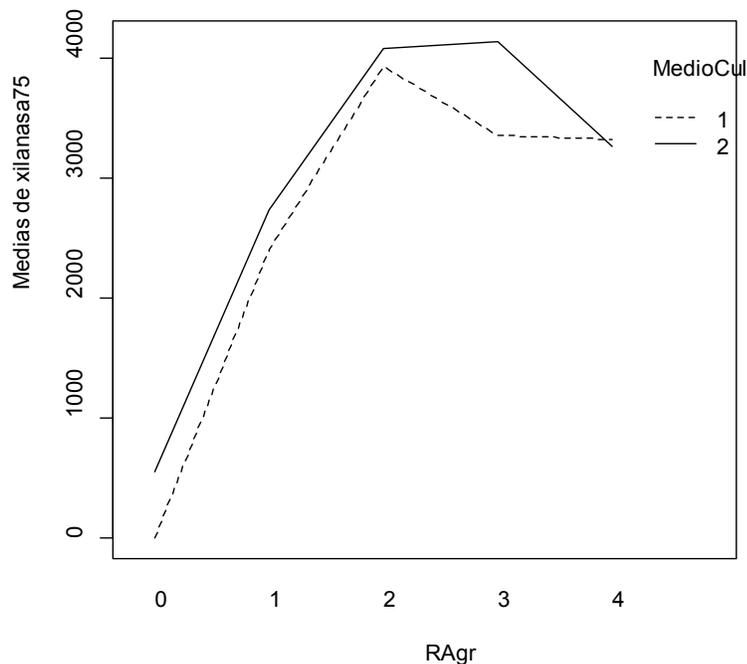


Figura 4.2. Medias de actividad enzimática de xilanas para cada uno de los medios de cultivo

Lo anterior, que se mostró para su mejor comprensión en la figura 4.2, se corrobora con la prueba de Tukey para las 10 combinaciones de tratamientos que resultan de los cinco residuos agrícolas (incluyendo el control) y los dos medios de cultivo. Son 45 comparaciones y las parejas de tratamientos combinados que no presentan diferencia significativa se muestran en la Tabla 4.5, las parejas de tratamientos combinados que no están en la Tabla 4.5 resultaron significativas en la prueba de Tukey.

Tabla 4.5. Resultados de la prueba de Tukey de comparaciones múltiples para la interacción entre residuo agrícola y medio de cultivo. Se muestran las parejas no significativas evaluadas a pH 7.5.

Efectos de residuos agrícolas que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
(0,2)-(0,1)	555.00000	-132.79105	1242.79105	0.1779595
(1,2)-(1,1)	342.00000	-345.79105	1029.79105	0.7372705
(3,1)-(2,1)	-568.66667	-1256.45772	119.12439	0.1575851
(4,1)-(2,1)	-608.33333	-1296.12439	79.45772	0.1092380
(2,2)-(2,1)	148.66667	-539.12439	836.45772	0.9981475
(3,2)-(2,1)	214.33333	-473.45772	902.12439	0.9761465
(4,2)-(2,1)	-668.66667	-1356.45772	19.12439	0.0606322
(4,1)-(3,1)	-39.66667	-727.45772	648.12439	1.0000000
(1,2)-(3,1)	-618.66667	-1306.45772	69.12439	0.0990008
(4,2)-(3,1)	-100.00000	-787.79105	587.79105	0.9999215
(1,2)-(4,1)	-579.00000	-1266.79105	108.79105	0.1435040
(4,2)-(4,1)	-60.33333	-748.12439	627.45772	0.9999990
(4,2)-(1,2)	518.66667	-169.12439	1206.45772	0.2425624
(3,2)-(2,2)	65.66667	-622.12439	753.45772	0.9999978

En la Tabla 4.6 se presentan las medias de las 10 combinaciones de tratamientos, unidas con una línea las medias que no resultaron con diferencia significativa al 5% en la prueba de Tukey. Como puede observarse, las combinaciones de tratamientos de los residuos agrícolas 2 y 3 con medio Horikoshi y el residuo agrícola 2 con agua desionizada no difieren entre sí y corresponden a los tratamientos identificados como (2,2), (3,2) y (2,1), en que se obtuvieron las medias muestrales más altas. A su vez los residuos agrícolas 2 y 3 con medio Horikoshi (2,2 y 3,2) difieren de los demás tratamientos excepto el formado por residuo agrícola 2 y agua desionizada (2,1). Esto permite concluir que las mejores combinaciones son bagazo de caña en medio Horikoshi, salvado de trigo en medio Horikoshi y bagazo de caña con medio de agua desionizada.

Tabla 4.6. Prueba de Tukey para la interacción sobre la actividad de xilanasa.
Las medias unidas por una línea no difieren entre sí

Interacción (A,C)	(0,1)	(0,2)	(1,1)	(1,2)	(4,2)	(4,1)	(3,1)	(2,1)	(2,2)	(3,2)
Media muestral	0	555	2401	2743	3262	3322	3362	3930	4079	4145

La tendencia anterior se manifiesta en la gráfica de cajas de los tratamientos combinados (Figura 4.3), donde se ve que las cajas de las combinaciones de tratamientos (2,1), (2,2), y (3,2), están más arriba que las demás.

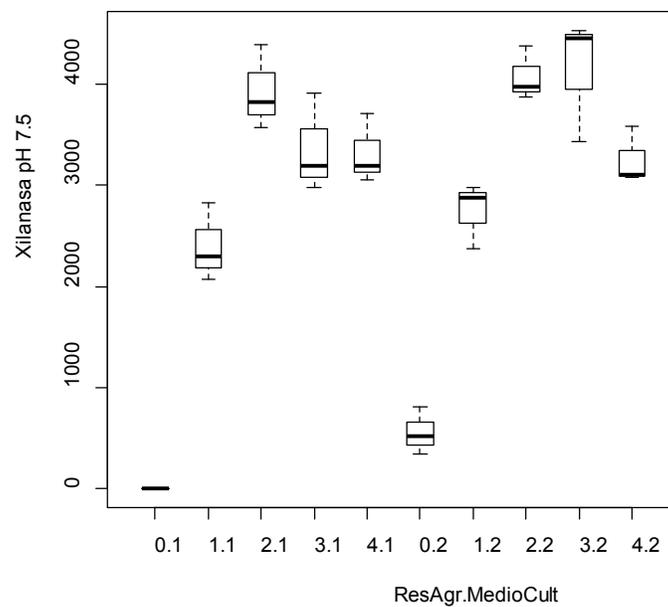


Figura 4.3. Gráfica de cajas de actividad enzimática de xilanasa para la interacción Residuo Agrícola y Medio de cultivo, ambos codificados.

La relación entre los valores de actividad enzimática de xilanasa determinadas a pH 7.5 y 9.2 se puede ver en la recta de regresión entre estas dos variables y su diagrama de dispersión (Figura 4.4). Los puntos de las observaciones se distribuyen alrededor de la línea de regresión. Dado que ambas variables miden la actividad enzimática de la xilanasa, pero utilizan diferente valor de pH, deberán producir una línea con pendiente uno. Si esto ocurre se puede elegir a cualquiera de las dos variables como variable respuesta.

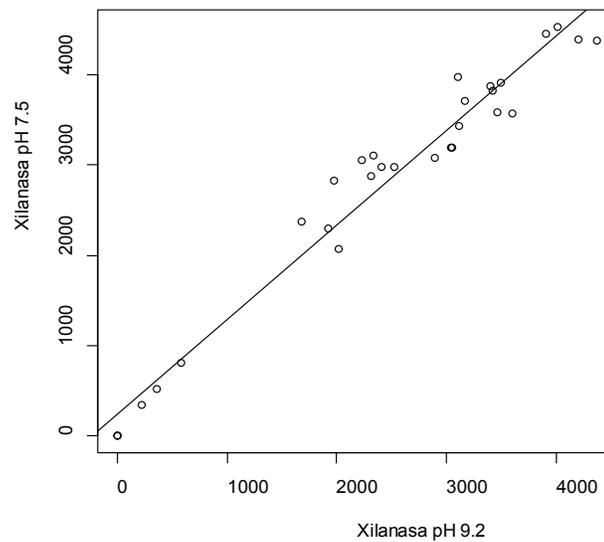


Figura 4.4. Diagrama de dispersión y regresión de las actividades enzimáticas de xilanasa determinadas a pH 7.5 y 9.2.

El resultado del análisis de regresión se presenta en la Tabla 4.7.

Tabla 4.7. Estimación de la regresión de xilanasa pH 7.5 sobre xilanasa pH 9.2

	Estimación	Error estándar	t calculada	valor p
β_0	232.74473	107.57429	2.164	0.0392
β_{dc}	1.04960	0.03904	26.882	$<2E^{-16}$
R cuadrada	0.9627		GL error	28

Se escribe la ecuación de la media estimada:

$$\mu'_{x75/x92} = 232.7447 + 1.0496 x_{92}.$$

El coeficiente de regresión es significativo (para la prueba de $H_0: x_{92} = 0$) con valor $p < 2E^{-16}$.

La pareja de hipótesis de interés, de que el coeficiente de regresión vale uno es $H_0: \beta_{x_{92}} = 1$,

HA: $\beta_{x_{92}} \neq 1$. Para obtener el valor de t calculada se utiliza

$$t_c = \frac{1.04960 - 1}{0.03904} = 1.270492$$

El valor de t calculada se representa como $t_c = 1.27$ con 28 grados de libertad y su valor p es 0.2143. No se rechaza la hipótesis nula y para fines prácticos $\beta_{x_{92}} = 1$. Esto significa que las mediciones de actividad enzimática obtenidas a los dos diferentes pH, cambian de manera directamente proporcional con tasa de cambio igual a 1; por lo que se concluye que la variable de respuesta se puede medir a cualquiera de los dos valores de pH. El valor de $R^2 = 0.9627$ indica que la regresión explica el 96.27% de la variación de la actividad enzimática de xilanasa a pH 7.5.

La gráfica de residuales vs predichos (Figura 4.5.) y de probabilidad normal (Figura 4.6), muestran que pudiera haber alguna curvatura y que los datos difieren de la normal en los cuantiles pequeños y en los grandes, aunque el ajuste es bueno en los intermedios.

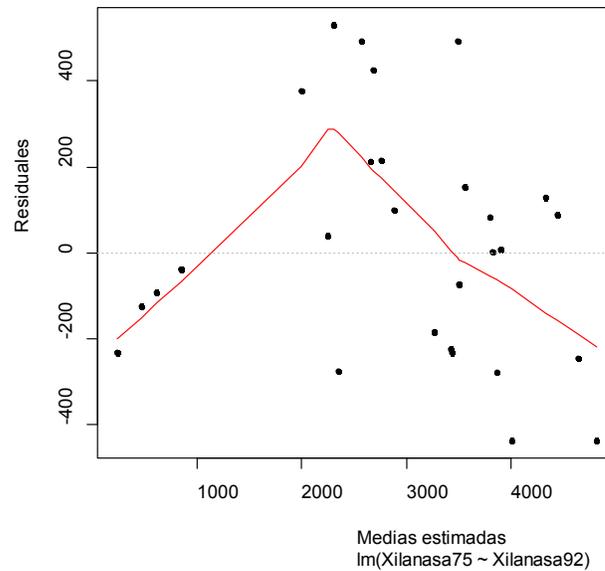


Figura 4.5. Valores residuales y predichos para actividad enzimática de xilanasas.

La prueba de normalidad de Shapiro-Wilk produce un valor p de 0.1169, por lo que con 10% de significación no se rechaza la hipótesis nula de que los residuales siguen la distribución normal.

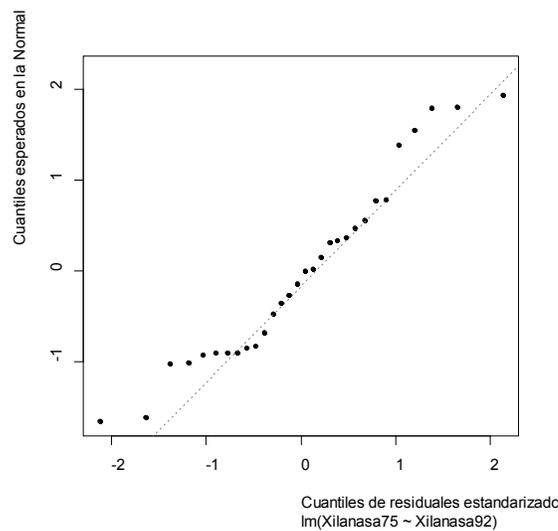


Figura 4.6. Gráfica de probabilidad normal para actividad enzimática de xilanasas.

El análisis de este problema se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares.

Guía para resolver con R

```

datosXilanasaRA<-read.table(file = "xilanasaRA.csv",header = T,sep = »,»)

# Se leen los datos de Excel tipo csv. Si tiene los datos en otra carpeta sugierocopiar el nombre completo del archivo
# de la barra del explorador, si las diagonales están así \, cámbielas a /.

datosXilanasaRA # Para ver sus datos

rC<-datosXilanasaRA$rC # Se definen los vectores de las variables en datosXilanasaRA usando $
mC<-datosXilanasaRA$mC
rep<-datosXilanasaRA$repet
x75<-datosXilanasaRA$x75
x92<-datosXilanasaRA$x92

# rC # mC # rep # x75 # x92      # Si desea ver los vectores de datos

baseXilanasaRA<-data.frame(RAgr = factor(rC),MedioCul = factor(mC),bloque = factor(rep),
Xilanasa75 = x75,Xilanasa92 = x92)

# Se forma el marco para el AdeV, = factor() indica que la variable será un factor en el análisis
baseXilanasaRA

attach(baseXilanasaRA) # Necesario para que el porcedimiento aov reconozca los datos

x75AdeV<-aov(Xilanasa75 ~ RAgr+MedioCul+bloque+RAgr:MedioCul)

# Especificando todos los términos en el modelo
anova(x75AdeV)

# x75AdeV<-aov(Xilanasa75 ~ bloque+RAgr*MedioCul)

# Otro modo de especificar: * forza los tres términos RAgr, MedioCul y RAgr:MedioCul
replications(Xilanasa75 ~ RAgr+MedioCul+bloque+RAgr*MedioCul,data = baseXilanasaRA,na.fail)

# Para asegurarse de que el modelo es balanceado, es un requisito para aov
summary(x75AdeV)

# x75AdeV      # list(x75AdeV) # Si desea el error estándar del residual

medias<-model.tables(x75AdeV,type = »means»,se = FALSE) # Medias de X75 para los niveles de los factores
medias

```

```

TukeyHSD(x75AdeV, which = "RAgr",orderer = TRUE, conf.level = 0.95)

# Prueba de Tukey para los niveles de RAgr (which = "RAgr")

boxplot(Xilanasa75~RAgr,data = baseXilanasaRA,outline = TRUE,beside = TRUE,boxwex = .5,
xlab = "Residuo agrícola",ylab = "Xilanasa pH 7.5",main = " ")

# Gráfica de cajas para RAgr

TukeyHSD(x75AdeV, which = "RAgr:MedioCul",orderer = TRUE, conf.level = 0.95)

# Prueba de Tukey para la interacción ( which = "RAgr:MedioCul")

interaction.plot(RAgr,MedioCul,Xilanasa75,ylab = "Medias de xilanasa75")

# Gráfica de medias de Xilanasa75 para la interacción RAgr:MedioCul

# Se escribe un factor, otro factor, variable respuesta.

boxplot(Xilanasa75~RAgr:MedioCul,data = baseXilanasaRA,outline = TRUE,beside = TRUE,boxwex = .5,
xlab = "ResAgr.MedioCult",ylab = "Xilanasa pH 7.5",main = "")

# Gráfica de cajas para las combinaciones de tratamientos de la interacción RAgr:MedioCul

plot(Xilanasa75~Xilanasa92,data = baseXilanasaRA,xlab = "Xilanasa pH 9.2",ylab = "Xilanasa pH 7.5",
main = " ") # Diagrama de dispersión para Xilanasa75 sobre Xilanasa92 con la línea de regresión

abline(reg) # Agrega la línea de regresión al diagrama de dispersión . Solo en regresión lineal simple

reg<-lm(Xilanasa75~Xilanasa92) # Regresión lineal simple de Xilanasa75 sobre Xilanasa92

reg

summary(reg)

valort<-(1.04960 -1)/0.03904

#Calcula el valor de t para probar si el coeficiente de regresión es 1. Valores en consola de R

valorp<- 2*(1-pt(abs(valort), 28, 0, lower.tail = TRUE, log.p = FALSE)) # Se calcula el valor p

# pt() da la probabilidad de la t. Se usa el valor absoluto de valort (abs(valort)),

# la t tiene 28 grados de libertad , 0 es el parámetro de no centralidad,

# probabilidad de valores menores que abs(valort) (lower.tail = TRUE) y no se usa log.p (log.p = FALSE)

valort

valorp

```

```

par(mfrow = c(1,1))

plot(reg,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Medias estimadas",ylab = "Residuales",col.lab = 1)

# Gráfica de residuales sobre medias estimadas (wich = 1). Tiene su sintaxis específica pra los títulos

plot(reg,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Cuantiles de residuales estandarizados",
ylab = "Cuantiles esperados en la Normal",col.lab = 1) # Gráfica de Cuantiles

# esperados en la Normal sobre residuales estandarizados (wich = 2). Sintaxis específica para los títulos

residuo<-residuals(reg)

shapiro.test(residuo)

```

4.1. Ejercicio

Considere los datos del problema anterior sin el control:

Residuo agrícola 0.5% w/v	Repetición	Producción de enzima U/l y tiempo a máxima actividad			
		Agua desionizada		Medio de Horikoshi	
		pH 7.5	pH 9.2	pH 7.5	pH 9.2
Madera de álamo	1	2077	2021	2374	1680
Bagazo de caña	1	3571	3597	3880	3395
Salvado de trigo	1	2974	2406	3435	3120
Carozo de maíz	1	3061	2224	3110	2337
Madera de álamo	2	2293	1925	2978	2521
Bagazo de caña	2	3822	3419	3979	3100
Salvado de trigo	2	3200	3042	4462	3908
Carozo de maíz	2	3197	3047	3086	2895
Madera de álamo	3	2833	1971	2877	2317
Bagazo de caña	3	4398	4202	4378	4367
Salvado de trigo	3	3911	3498	4537	4017
Carozo de maíz	3	3708	3166	3589	3462

Haga el análisis de actividad enzimática de xilanas a pH 7.5 como en el problema anterior, vea la relación entre las dos variables respuesta xilanas a pH 7.5 y xilanas a pH 9.2. Si desea ejercitarse más, haga el análisis de la variable actividad enzimática de xilanas a pH 9.2 como se hizo el de actividad enzimática de xilanas a pH 7.5.

Bibliografía

-  Grupa, S., Kunhad, R.C., Bhusmhan, B. and HoodaL, G.S. Improved xylanase production from a haloalkalophilic *Staphylococcus* sp. SG-13 using inexpensive agricultural residues. World Journal of Microbiology and Biotechnology. 2001, 17: 5-8.

Nota: La referencia se utilizó, de manera fraccionada, para plantear problemas similares pero llevados al nivel y el interés del curso

5. Efecto de la estimulación eléctrica, la adición de epinefrina y el tiempo de molido sobre el contenido de ácido tiobarbitúrico en carne de cerdo

El ácido tiobarbitúrico (TBO1A) se puede encontrar presente en la carne de cerdos destinada al consumo humano. A fin de estudiar la generación de este compuesto, se estudió en cerdos el efecto de tres factores: la administración de epinefrina antes de la matanza, la estimulación eléctrica después de la matanza y el tiempo de desagrado, con dos niveles cada uno (diseño 2³). Para llevar a cabo el experimento se seleccionaron 8 cerdos de peso similar (95-120 kg) que se encontraban en cada uno de 4 corrales, de forma que se realizó un diseño de bloques completos al azar y arreglo factorial de tratamientos. Los cuatro bloques corresponden a los cuatro corrales.

Cada factor se trabajó con dos niveles de la siguiente forma: cuatro horas antes de la matanza se aplicó a los cerdos una inyección subcutánea que podía contener una dosis de epinefrina de 25 mg/kg (proveniente de una solución 1:1000), o bien agua destilada, de forma que se evaluó la adición y ausencia de epinefrina. Para estudiar el efecto de la estimulación eléctrica o ausencia de ésta, después de 30 minutos de la matanza, se aplicó a lo largo de 2 minutos e intercalando un reposo de 3 segundos, una corriente de 480 V y 3 A durante dos segundos; luego de una pausa de dos minutos, nuevamente se aplicó la corriente eléctrica hasta observar contracción muscular. Como tiempos de desagrado se tomaron 1 y 24 horas.

Al término del tiempo de desagrado, se tomó al azar de uno de los lados de cada cerdo y en éste el músculo del tríceps, al que se le quitó la grasa visible y se molió. Se tomó una muestra de 100 g, se le adicionó 1% de neosporin en polvo, se mezcló, se acomodó en una charola envuelta en material permeable al oxígeno e impermeable a la humedad y se almacenó durante 12 días a 3 °C. La variable de respuesta fue el aumento del contenido de TBA durante el tiempo de almacenamiento. Debido a esto, el análisis se hace sobre la transformación de la variable TBA, de forma que $TBA = TBA_{12} - TBA_0$, donde TBA_{12} y TBA_0 representan el ácido tiobarbitúrico encontrado en cada muestra de carne al inicio y a los 12 días de almacenamiento. Los datos se presentan en la Tabla 5.1.

Tabla 5.1. Valores de ácido tiobarbitúrico (TBA) obtenidos en cada tratamiento

Cerdo	Corral (C)	Estimulación eléctrica (EE)	Epinefrina (Ep)	Tiempo de desagrado (TD)	TBA ₀ (Día cero)	TBA ₁₂ (Día 12)
1	1	0	0	1	0.2	0.49
2	1	0	0	24	0.2	0.48
3	1	0	1	1	0.2	1.05
4	1	0	1	24	0.2	1.06
5	1	1	0	1	0.23	1.39
6	1	1	0	24	0.24	1.43
7	1	1	1	1	0.2	0.89
8	1	1	1	24	0.2	0.86
9	2	0	0	1	0.19	0.46
10	2	0	0	24	0.19	0.46
11	2	0	1	1	0.19	1.09
12	2	0	1	24	0.18	1.07
13	2	1	0	1	0.2	1.52
14	2	1	0	24	0.2	1.58
15	2	1	1	1	0.21	0.86
16	2	1	1	24	0.21	0.89

17	3	0	0	1	0.19	0.29
18	3	0	0	24	0.19	0.29
19	3	0	1	1	0.19	3.35
20	3	0	1	24	0.19	3.36
21	3	1	0	1	0.19	3.63
22	3	1	0	24	0.2	3.54
23	3	1	1	1	0.2	2.51
24	3	1	1	24	0.2	2.45
25	4	0	0	1	0.19	0.29
26	4	0	0	24	0.2	0.28
27	4	0	1	0.75	0.2	3.55
28	4	0	1	24	0.2	3.53
29	4	1	0	0.75	0.2	3.39
30	4	1	0	24	0.2	3.48
31	4	1	1	0.75	0.2	2.54
32	4	1	1	24	0.2	2.55

Se evalúa el efecto de todas las interacciones entre los factores de tratamientos, por lo que el modelo planteado es:

$$\mu_{TBA/i,j,k,h} = \mu + c_i + ee_j + epi_k + td_h + (ee * epi)_{j,k} + (ee * td)_{j,h} + (epi * td)_{k,h} + (ee * epi * td)_{j,k,h}$$

con:

$$TBA_{i,j,k,h} = \mu_{TBA/i,j,k,h} + \varepsilon_{i,j,k,h} \text{ y } \varepsilon_{i,j,k,h} \sim N(0, \sigma_\varepsilon^2) \text{ y } i = 1,2,3,4, j = 0,1, k = 0,1 \text{ y } h = 0,1.$$

El resultado del análisis de varianza se muestra en la Tabla 5.2, donde están las pruebas de que los efectos son cero para cada factor e interacción.

Se plantean las hipótesis siguientes, y se incluye el valor p cuando es menor que 0.05, ya que en este caso se rechaza la hipótesis nula de igualdad de efectos. De cualquier forma, todos los valores de p -se rechace o no Ho- se presentan en la tabla 5.2.

Para estimulación eléctrica: H0: $ee_0 = ee_1$, HA: $ee_1 \neq ee_0$, valor p = 0.002,

para aplicación de epinefrina: H0: $epi_0 = epi_1$, HA: $epi_1 \neq epi_0$, valor p = 0.021,

para tiempo de desangrado: H0: $td_0 = td_1$, HA: $td_0 \neq td_1$,

para interacción entre estimulación eléctrica y aplicación de epinefrina:

$$H0: (ee * epi)_{j,k} = 0 \text{ para } j = 0,1 \text{ y } k = 0,1;$$

$$HA: (ee * epi)_{j,k} \neq 0 \text{ para al menos una pareja } j = 0,1, k = 0,1 \text{ con valor } p = 0.000004,$$

para la interacción entre estimulación eléctrica y tiempo de desangrado:

$$H0: (ee * td)_{j,h} = 0 \text{ para } j = 0,1 \text{ y } h = 0,1;$$

$$HA: (ee * td)_{j,h} \neq 0 \text{ para al menos una pareja } j = 0,1, h = 0,1,$$

para la interacción entre aplicación de epinefrina y tiempo de desangrado:

$H_0: (\text{epi} * \text{td})_{k,h} = 0$ para $k = 0,1$ y $h = 0,1$;

$H_A: (\text{epi} * \text{td})_{k,h} \neq 0$ para al menos una pareja $k = 0,1$, $h = 0,1$

y para la interacción entre los tres factores se tiene:

$H_0: (\text{ee} * \text{epi} * \text{td})_{j,k,h} = 0$ para $j = 0,1$, $k = 0,1$ y $h = 0,1$;

$H_A: (\text{ee} * \text{epi} * \text{td})_{j,k,h} \neq 0$ para al menos una tripleta de subíndices (j, k, h) con $j = 0,1$, $k = 0,1$ y $h = 0,1$

Tabla 5.2. Análisis de varianza para la producción de TBA.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Corral	3	17.3418	5.7806	15.2975	1.665E ⁻⁰⁵
Estimulación Eléctrica (EE)	1	4.6742	4.6742	12.3695	0.002048
Epinefrina (Epi)	1	2.3382	2.3382	6.1877	0.021346
Tiempo de Desangrado (TD)	1	0.0000	0.0000	0.0000	0.997733
Interacción EE:Epi	1	14.2979	14.2979	37.8374	4.203E ⁻⁰⁶
Interacción EE:TD	1	0.0002	0.0002	0.0004	0.984130
Interacción Epi:TD	1	0.0004	0.0004	0.0010	0.975063
Interacción EE:Epi:TD	1	0.0007	0.0007	0.0019	0.966001
Error	21	7.9354	0.3779		

Del análisis mostrado en la Tabla 5.2 se desprende que hay efecto sobre la media de TBA de: 1) la aplicación de estimulación eléctrica, 2) la aplicación de epinefrina y 3) la interacción estimulación eléctrica y aplicación de epinefrina. Para identificar el mayor efecto en 1) y 2) basta con revisar las medias que se presentan en la Tabla 5.3, así, el nivel 1 de EE produce el mayor efecto en la media de TBA ($1.8894 > 1.1250$), y en el caso de la epinefrina, el mayor efecto corresponde también al nivel 1, ya que $1.7775 > 1.2369$. Para mostrar la diferencia entre las medias de interacción se debe recurrir a la prueba de Tukey (o cualquier otra de comparaciones múltiples), concluir con base en el valor de las medias muestrales puede ser errático. Por ejemplo, si se ordenan las medias muestrales de interacción (Tabla 5.3) en forma ascendente, la menor (0.1863) corresponde a no estimulación y no epinefrina, siguiéndole la de aplicación de ambos (1.4913), y, por último, con muy poca diferencia numérica entre ellas, las medias sin estimulación con epinefrina (2.0638) y con estimulación y sin epinefrina (2.2875). Aparentemente, el efecto de aplicar estímulo eléctrico es diferente cuando se aplica epinefrina que cuando no se aplica, siendo mejor no aplicarla que hacerlo. Esta apreciación, basada sólo en los valores estimados de las medias de TBA, no se corrobora con la prueba de Tukey, como se muestra más adelante.

Tabla 5.3. Medias de TBA obtenidas en cada tratamiento.

Medias de TBA para los niveles de EE		Medias de TBA para los niveles de Epinefrina		Medias de TBA para los niveles de la interacción EE-Epinefrina		
EE	Media	Epinefrina	Media	EE	Epinefrina	Media
0	1.1250	0	1.2369	0	0	0.1863
1	1.8894	1	1.7775	0	1	2.0638
				1	0	2.2875
				1	1	1.4913

Una representación gráfica de las medias de de interacción se muestra en la Figura 5.1. La línea cortada muestra la unión entre las medias de TBA para estimulación 0 y 1 cuando no se aplica epinefrina. La línea sólida muestra la unión entre las medias de TBA para estimulación 0 y 1 cuando se aplica epinefrina. La línea cortada aumenta del nivel 0 al 1 de estimulación eléctrica mientras que la línea sólida disminuye. Aparentemente, el efecto de aplicar estímulo eléctrico es diferente cuando se aplica epinefrina que cuando no se aplica. Por la ubicación de los extremos de las líneas, que representan a las medias, al aplicar sólo el estímulo eléctrico (sin epinefrina), se obtiene la media de TBA más grande y, la segunda más grande, se obtiene aplicando solo epinefrina. Si se aplican ambos, la media de TBA es menor que las anteriores. Como se ha mencionado, las apreciaciones anteriores no fueron corroboradas con Tukey.

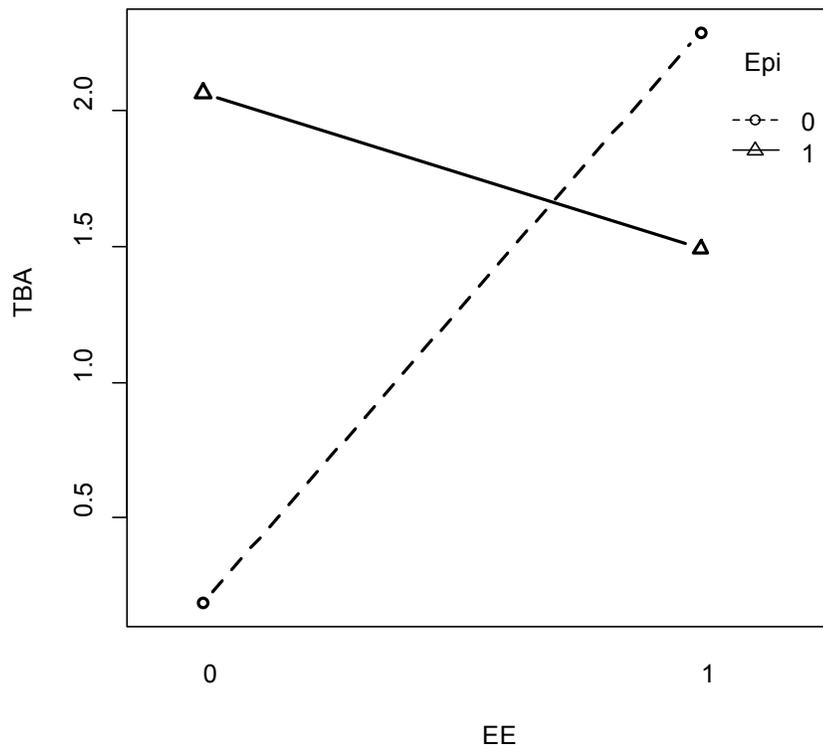


Figura 5.1. Interacción entre los niveles y factores.

En la Tabla 5.4 se muestran los resultados de aplicar Tukey para las interacciones EE-Epinefrina. En las primeras tres filas se evidencia que la interacción (0,0) difiere de las demás con valores p menores que .05. y, en las tres últimas filas, no se probó que las combinaciones (0,1), (1,0) y (1,1) difieran entre sí, los tres valores de p están por arriba de .05.

Tabla 5.4. Resultados de la prueba de Tukey de comparaciones múltiples para la interacción entre estimulación eléctrica y aplicación de epinefrina en la media de TBA.

Efectos de interacción entre EE y Epinefrina que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
(1,0)-(0,0)	2.10125	1.244541	2.95795905	0.0000052
(0,1)-(0,0)	1.87750	1.020791	2.73420905	0.0000258
(1,1)-(0,0)	1.30500	0.448291	2.16170905	0.0018954
(0,1)-(1,0)	-0.22375	-1.080459	0.63295905	0.8847897
(1,1)-(1,0)	-0.79625	-1.652959	0.06045905	0.0745210
(1,1)-(1,1)	-0.57250	-1.429209	0.28420905	0.2736882

En la Tabla 5.5 se presentan las medias de las combinaciones de factores EE-Epinefrina, donde las medias unidas por una línea no difieren en la prueba de Tukey. La combinación sin estimulación eléctrica y sin epinefrina (0,0), que tiene la media menor, es diferente de las otras tres combinaciones [(0,1), (1,0) y (1,1)] que no difieren entre sí. En la práctica, es conviene elegir el más barato o más fácil de aplicar entre ellos, ya sea sólo epinefrina o sólo estimulación eléctrica.

Tabla 5.5. Prueba de Tukey para la interacción de los efectos de EE-epinefrina sobre la media de TBA. Medias unidas por una línea no difieren.

Tratamiento (EE, Epinefrina)	(0,0)	(1,1)	(0,1)	(1,0)
Media muestral	0.1863	1.4913	2.0637	2.2875

El análisis del experimento de TBA con estimulación eléctrica, epinefrina y tiempo de desangrado se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares, procurando no repetir anotaciones que ya están en los ejemplos anteriores.

Guía para resolver con R

```
datosTBA<-read.table(file = "TBA.csv",header = T,sep = ";")

# Antes de leer sus datos cambie directorio a la carpeta donde están los datos

datosTBA<-read.table(file = "C:/Users/Alberto/aacmB_D/R/problemaarios2014/TDE/ddExcel/TBA.csv",header = T,sep = ";")

# Se lee y guarda la tabla de datos

datosTBA

bk<-datosTBA$Corral # Recuerde que nombre de base$nombre de variable se refiere al vector de la variable en la base

ee<-datosTBA$EE
```

```

ep<-datosTBA$Ep
td<-datosTBA$TD
TBA<-(datosTBA$D12-datosTBA$D0)

# Se crea la variable TBA con el incremento de TBA del inicio a los 12 días de almacenamiento

# bl # ee # ep # td # TBA

baseTBA<-data.frame(Corral = factor(bl),EE = factor(ee),Epi = factor(ep),TD = factor(td),TBA = TBA)

# Se crea el marco para el AdeV usando aov. Los factores se crean con factor

baseTBA
attach(baseTBA)

TBAAdeV<-aov(TBA~Corral+EE*Epi*TD)

anova(TBAAdeV)

# EE*Epi*TD crea los términos del modelo EE, Epi, TD, EE:Epi, EE:TD, Epi:TD y EE:Epi:TD

replications(TBA~Corral+EE*Epi*TD,data = baseTBA,na.fail)

# Compruebe que hay igual número de repeticiones para cada factor e interacción

# summary(TBAAdeV)

# list(TBAAdeV)

medias<-model.tables(TBAAdeV,type = "means",se = FALSE)

medias # Medias TBA para niveles de factores e interacciones

interaction.plot(EE,Epi,TBA,type = "b",lwd = 2,pch = 1:3,xlab = "EE",ylab = "TBA")

# Gráfica de la interacción EE: Epi para TBA

TukeyHSD(TBAAdeV, which = "EE:Epi",orderer = TRUE, conf.level = 0.95)

# Prueba de Tukey para la interacción EE:Epi (wich = "EE:Epi")

```

5.1. Ejercicio

Considere que se hace un experimento como el del problema 5, pero se fija el tiempo de desangrado a 24 horas. Se tiene un experimento con diseño de bloques completos (corrales) al azar y arreglo factorial de tratamientos; dos factores cada uno con dos niveles (2^2). Son cuatro bloques, pero en cada bloque hay dos repeticiones. Como en el problema 5, interesa el aumento de TBA del inicio del almacenamiento al día 12. Con los siguientes datos haga el análisis.

Cerdo	Corral (C)	Estimulación Eléctrica (EE)	Epinefrina (Ep)	TBA0 Día 0	TBA12 Día 12
1	1	0	0	0.2	0.51
2	1	0	0	0.2	0.49
3	1	0	1	0.2	1.12
4	1	0	1	0.2	1.16
5	1	1	0	0.23	1.28
6	1	1	0	0.24	1.39
7	1	1	1	0.2	0.92
8	1	1	1	0.2	0.97
9	2	0	0	0.19	0.54
10	2	0	0	0.19	0.50
11	2	0	1	0.19	1.19
12	2	0	1	0.18	1.13
13	2	1	0	0.2	1.52
14	2	1	0	0.2	1.44
15	2	1	1	0.21	0.83
16	2	1	1	0.21	0.88
17	3	0	0	0.19	0.29
18	3	0	0	0.19	0.27
19	3	0	1	0.19	3.32
20	3	0	1	0.19	3.29
21	3	1	0	0.19	3.53
22	3	1	0	0.2	3.48
23	3	1	1	0.2	2.56
24	3	1	1	0.2	2.49
25	4	0	0	0.19	0.32
26	4	0	0	0.2	0.27
27	4	0	1	0.2	3.65
28	4	0	1	0.2	3.51
29	4	1	0	0.2	3.49
30	4	1	0	0.2	3.33
31	4	1	1	0.2	2.42
32	4	1	1	0.2	2.49

Bibliografía

-  Yasosky, J.J., Aberle, E.D., Peng, I.C., Mills, E.W. and Judge, M.D. (1984) Effects of pH and time of grinding on lipid oxidation of fresh ground pork. *Journal of Food Science*. 49: 1510-1512.

Nota: La referencia se utilizó, para plantear un problema similar pero llevado al nivel del curso.

6. Regresión lineal simple para recuentos de células vivas y muertas

El recuento de células en un campo de visualización es una manera de medir la concentración de biomasa bacteriana. Debido a las características de las membranas celulares, al usar colorantes específicos y hacer observaciones mediante microscopía de epifluorescencia, las células vivas se colorean de verde y las muertas (membrana permeable), de rojo (microscopía de epifluorescencia). El recuento está sujeto a variación debido a diferencias en el número de células en los diferentes campos de visualización. El investigador desea probar si existe una relación lineal entre el recuento de células vivas y el de células muertas para, en estudios posteriores, medir solo uno de los dos tipos de células. Para lograr esto, se obtendrá la ecuación de regresión de células vivas sobre células muertas.

La revisión de diez campos de visualización produjo los datos de número de células por ml. En la Tabla 6.1 se reportan divididos entre 10^6 para facilitar su escritura y manejo.

Tabla 6.1. Recuento del número de células vivas y muertas.

Campo de visualización	1	2	3	4	5	6	7	8	9	10
Células muertas (#/ml)/ 10^6	3.6	3.9	3.1	5.2	3	3.2	4.4	2.9	5	4.4
Células vivas (#/ml)/ 10^6	4.4	3.9	3	4.5	3	2.7	4.5	2.6	5.1	4.2

Para una revisión inicial de los datos se hace un diagrama de dispersión (Figura 6.1). Se ve que ambos conteos aumentan juntos, pero con variación.

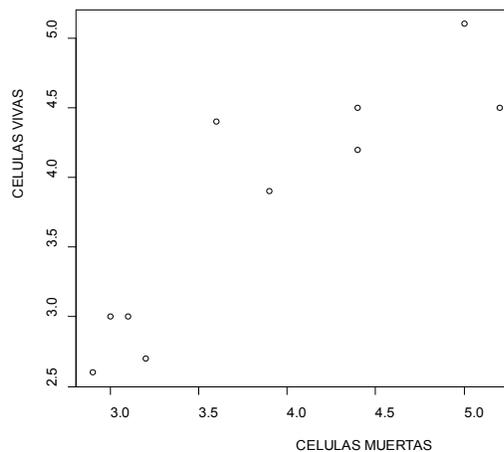


Figura 6.1. Diagrama de dispersión de células vivas y células muertas.

La regresión lineal estimará la línea que relaciona el aumento del conteo de células muertas con el aumento del conteo de células vivas y dará una estimación de la varianza de las dispersiones de los datos alrededor de la línea estimada. Se plantea la distribución:

$$cv_{cm} \sim N(\mu_{cv/cm} = \beta_0 + \beta_{cm} cm, \sigma_{cv/cm}^2)$$

La expresión anterior cv_{cm} representa a la variable aleatoria células vivas, que cuando la variable células muertas vale cm se distribuye (\sim) normal (N), con una media que depende del valor de cm ($\mu_{cv/cm}$) y que se relaciona con el valor de cm por

medio de la ecuación $\mu_{cv/cm} = \beta_0 + \beta_{cm} cm$, que es la ecuación de regresión, en este caso una línea recta. En esta ecuación β_0 es una constante (parámetro) desconocida que representa a la ordenada al origen, β_{cm} es otra constante desconocida que denota al coeficiente de regresión de cv sobre cm (por facilidad, se acostumbra escribir 1 en lugar de cm) y $\sigma_{cv/cm}^2$ es la varianza desconocida de la distribución de cv sobre cm en forma de regresión lineal simple.

Por la forma como se desarrolló la estadística y su notación, se dice que la distribución de cv_{cm} tiene implícita la ecuación de regresión cuando se escribe como sigue:

$$cv = \beta_0 + \beta_{cm} cm + \varepsilon \text{ con } \varepsilon \sim N(0, \sigma_{cv/cm}^2).$$

En esta ecuación cv es la variable que arriba se llamó cv_{cm} , dejando que el lector entienda que la distribución de la variable cv depende del valor que tome cm y ε es un término aleatorio, que se conoce como error (alrededor de la línea) que se distribuye normal con media cero y varianza $\sigma_{cv/cm}^2$ como se mostró antes.

En cualquiera de los casos, pero con mayor claridad en el primer modelo, se obtiene la ecuación con los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$. Para esto se utiliza el método de mínimos cuadrados, que consiste en minimizar la suma de los cuadrados de las distancias de los puntos observados a la línea estimada con respecto a β_0 y β_1 y cuyos resultados produce las ecuaciones para calcular los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$.

$$\hat{\mu}_{cv/cm} = \hat{\beta}_0 + \hat{\beta}_1 cm,$$

donde el símbolo ^ (acento circunflejo) indica que son variables aleatorias, esto es: el valor que tomen depende de la muestra que se use para calcular sus valores. Una vez que se usan los datos de la muestra para calcular los valores de los estimadores, se obtiene el modelo de estimación (también se le llama modelo ajustado) $\mu'_{cv/cm} = \beta'_0 + \beta'_1 cm$, en el que el β' indica que ya no es variable aleatoria, sino un valor calculado con los datos. Debido a que el valor de $\mu'_{cv/cm}$ depende del valor de cm, no se calcula su valor, sino que se escribe la ecuación de regresión estimada:

$$\mu'_{cv/cm} = 0.1514 + 0.9402 cm$$

También por defectos de notación debidos al desarrollo histórico de la estadística, se acostumbra escribirlo como sigue:

$$c\hat{v} = 0.1514 + 0.9402 cm$$

donde $c\hat{v}$ es el valor que toma la estimación de la media de cv_{cm} con los valores de los coeficientes calculados a partir de los datos experimentales.

De acuerdo con la Tabla 6.2, la estimación de la media de células vivas en la línea de regresión es $\mu'_{cv/cm} = 0.1514 + 0.9402 cm$. El coeficiente de regresión de cm vale 0.9402, lo que significa que por cada unidad que aumente la variable cm, la estimación de la media de cv en la línea aumenta 0.9402.

Tabla 6.2. Análisis de regresión del conteo de células vivas sobre el conteo de células muertas.

Coefficiente de la variable	Estimación	Error estándar de la estimación	T para probar si el coeficiente es cero	Valor p
Constante ($\hat{\beta}_0$)	0.1514	0.6647	0.228	0.825488
Células muertas ($\hat{\beta}_1$)	0.9402	$0.1682 = S_{\hat{\beta}_1}$	5.590	0.000516

Como se esperaba en el caso de cambios directamente proporcionales, el coeficiente de regresión estimado $\hat{\beta}_{cm} = 0.9402$ es un valor cercano a uno. En la Figura 6.2 se muestra el diagrama de dispersión de los datos y la recta de estimación de la media

de la regresión.

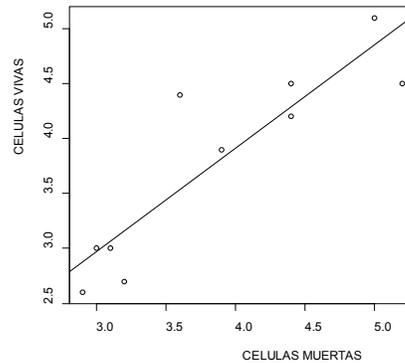


Figura 6.2. Diagrama de dispersión con la recta de estimación de la media la regresión.

En la Tabla 6.3 se prueba que el coeficiente de regresión vale cero, que atende a las hipótesis nula y alternativa: $H_0: \beta_{cm} = 0$ y $H_A: \beta_{cm} \neq 0$, respectivamente. Como el valor $p = 0.000516$ es menor que 0.05, se rechaza la hipótesis nula de que el coeficiente β_{cm} vale cero.

Antes de seguir adelante con la información que proporciona la Tabla 6.3, es importante aclarar que en este problema, en particular, no interesa la prueba de que el coeficiente de regresión β_{cm} sea cero, pero sí interesa probar que su valor es 1, ya que un coeficiente de regresión igual a 1 indica que, por cada unidad que aumente el valor de cm , el valor de $\mu'_{cv/cm}$ también aumenta una unidad. Se plantea la pareja de hipótesis nula y alternativa siguientes $H_0: \beta_{cm} = 1$, $H_A: \beta_{cm} \neq 1$. Para realizar la prueba se calcula un valor de t y un valor p usando la t con 8 grados de libertad como se muestra a continuación:

$$t_c = \frac{\hat{\beta}_{cm} - 1}{S_{\hat{\beta}_{cm}}} = -0.3555291 \quad \text{y} \quad 2P [t_8 > |t_c|] = 0.7313854$$

El valor p de 0.73 no da evidencia para rechazar la hipótesis de que β_{cm} sea igual a uno, el valor p mayor que 0.5 permite que, para fines prácticos, se pueda considerar que el coeficiente de regresión es uno.

La variación hacia arriba y abajo de la recta de estimación (línea predicha) se manifiesta en el valor del coeficiente de determinación, denominado R cuadrada (R^2) que se reporta en la última fila de la Tabla 6.3. En los resultados del paquete R se obtiene un $R^2 = 0.7962$, que indica que la regresión explica el 79.62% de la variación de la variable cm (en el sentido de sumas de cuadrados). R^2 se calcula dividiendo la suma de cuadrados de regresión entre la suma de cuadrados total.

Tabla 6.3. Análisis de varianza de la regresión del conteo de células vivas sobre el conteo de células muertas.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Regresión	1	5.676	5.676	31.25	0.0005161
Residual	8	1.453	0.1816		
Total	9	7.129			
$R^2 = 0.7962$					

La variación descrita anteriormente también se puede ver en la gráfica de residuales (Figura 6.3), donde se ve que el residual más grande en valor absoluto está por arriba de 0.8 pero no llega a uno.

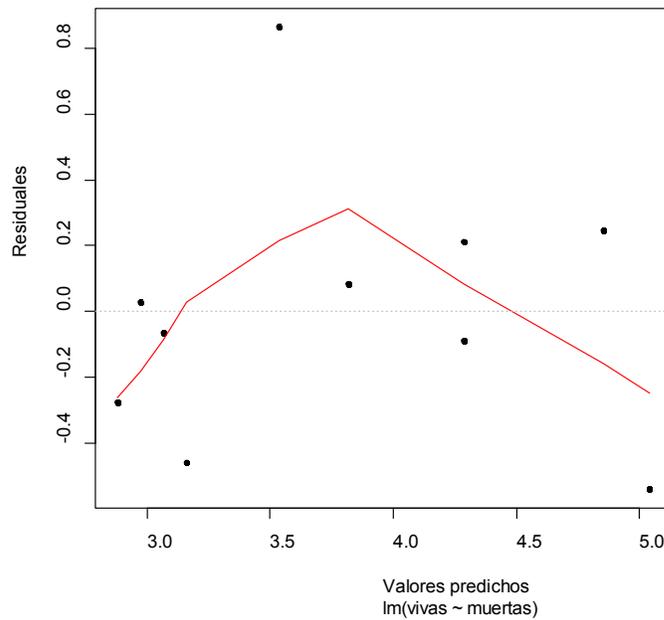


Figura 6.3. Residuales y valores predichos.

La gráfica de probabilidad normal (Figura 6.4) muestra la cercanía de la distribución de los residuales estandarizados a la distribución normal estándar. Entre más alineados estén los puntos de la gráfica, más parecida es su distribución a la normal.

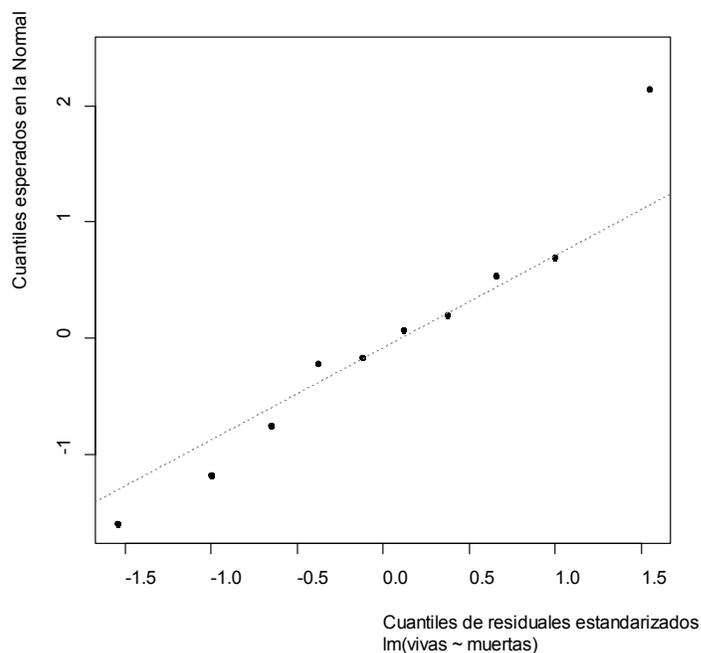


Figura 6.4. Probabilidad normal.

En cualquiera de las dos figuras anteriores se puede mostrar si alguna pareja de valores (cv, cm) se comporta de manera diferente que las demás. El valor de la derecha en la gráfica de probabilidad normal indica que hay un campo de visualización que da un valor de cv alejado de la línea de regresión estimada, hacia arriba de ella, es el campo de visualización número 1. Su residual por arriba de 0.8, es el mayor y también es el mayor cuantil de los residuales estandarizados (en el eje horizontal). Además, el punto en la gráfica no está en la línea formada por los otros nueve puntos en la gráfica.

Suponga que se desea conocer el intervalo con 90% de confianza para el valor estimado de la media de células vivas cuando el valor de células muertas vale 4, y el intervalo de predicción con 90% de confianza para el valor de células vivas cuando el de células muertas vale 4.5. Los resultados se muestran en la Tabla 6.4.

Tabla 6.4. Estimación de la media y predicción de un valor de células vivas con un intervalo de 90% de confianza

Intervalo con 90% de confianza para la media de células vivas cuando la de células muertas vale 4		
Estimación del valor de la media	Límite inferior del intervalo	Límite superior del intervalo
3.912226	3.658338	4.166113
Intervalo con 90% de confianza para un valor de células vivas cuando células muertas vale 4.5		
Estimación del valor de la media	Límite inferior del intervalo	Límite superior del intervalo
4.382324	3.528106	5.236541

En el ejercicio también se obtiene el coeficiente de correlación de Pearson (r), cuyo valor es $r = 0.892289$. El coeficiente es una medida estandarizada entre -1 y 1 de la variación proporcional entre las variables. Si el coeficiente de correlación está entre -1 y cero, la variación entre las variables es inversa, esto es, cuando una crece la otra decrece. Esta relación es directa cuando el coeficiente de correlación es positivo (entre cero y 1) y si vale cero, se dice que no hay correlación lineal entre las variables. Para la estandarización se usan las desviaciones estándar de las variables. En este problema ya se analizó el coeficiente de determinación, ahora el de correlación indica una variación proporcional directa.

El coeficiente de determinación que vale 1 0.7962 apoya a la conclusión práctica de que se puede contar sólo uno de los dos tipos de células, el otro será similar en tamaño y se pueden hacer intervalos de predicción para la variable que no se mide.

El análisis de este problema se hace con las instrucciones del paquete R que siguen: se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares.

Guía para resolver con R

```
muertas<-c(3.6,3.9,3.1,5.2,3.0,3.2,4.4,2.9,5.0,4.4)
# Se crea el vector de observaciones de conteo de células muertas

vivas<-c(4.4,3.9,3.0,4.5,3.0,2.7,4.5,2.6,5.1,4.2)

baseCelulas<-data.frame(vivas,muertas)
# Se crea la base de datos (marco) con los dos vectores de observaciones

baseCelulas
attach(baseCelulas)

plot(vivas~muertas,xlab = »CELULAS MUERTAS»,ylab = »CELULAS VIVAS»,main = » «)
# Diagrama de dispersión de vivas sobre muertas

vivas_muertas<- lm(vivas~muertas)# se calcula la regresión y se le llama vivas_muertas
```

```

vivas_muertas

summary(vivas_muertas) # Resumen de la regresión en vivas_muertas

# anova(vivas_muertas) # Muestra el análisis de varianza de la regresión vivas_muertas

plot(vivas~muertas,xlab = »CELULAS MUERTAS»,ylab = »CELULAS VIVAS»,main = »»)

abline(vivas_muertas) # Agrega la línea de regresión lineal simple al diagrama de dispersión

tcal<- (0.9402-1)/.1682

tcal

abstcal<-abs(tcal)

valorp<- 2*(1-pt(abstcal, 8, 0, lower.tail = TRUE, log.p = FALSE))

valorp

plot(vivas_muertas,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" " ),xlab = "Valores predichos",
ylab = "Residuales",col.lab = 1)

# Diagrama de dispersión de residuales sobre valores estimados de la media (predichos)

plot(vivas_muertas,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" " ),xlab = "Cuantiles de residuales estandarizados",
ylab = "Cuantiles esperados en la Normal",col.lab = 1)

# Gráfica de probabilidad normal para la regresión

# vivas_muertasPred<- predict(vivas_muertas)

# Obtiene los valores estimados de la media

# vivas_muertasPred

# Muestra el vector con valores estimados de la media

casosPred<- data.frame(muertas = c(4)) # Se crea el marco de valores para estimar a la media

predict(vivas_muertas,newdata = casosPred,interval = "confidence",level = 0.90)

# Se estima la media usando la regresión para los datos en casosPred, interval = "confidence"

casosPred<- data.frame(muertas = c(4.5)) # Se crea otro marco para predicciones

predict(vivas_muertas,newdata = casosPred,interval = "prediction",level = 0.90)

```

Se hace un intervalo de predicción usando la regresión, datos en casosPred,interval = "prediction"

cor(vivas,muertas) # Obtiene la correlación de las variables vivas y muertas

6.1. Ejercicio

En la tesis de licenciatura de Alejandra Nivón Pellón, en la Facultad de Química de la Universidad Autónoma de Querétaro, "Evaluación fisicoquímica y sensorial de hidromiel elaborado con dos diferentes formulaciones", 2007, se presenta la relación entre turbidez y pH en hidromiel.

Haga un análisis como el del primer ejercicio a los datos de turbidez (variable respuesta), medida con un refelómetro 2100N HACH y pH (variable explicativa). Los datos se aproximaron a partir de una gráfica de la tesis.

	1	2	3	4	5	6	7
Turbidez (NTU)	9.2	6.2	4.4	1.8	1.7	1.5	1.7
pH	3.20	3.11	3.06	2.58	2.46	2.38	2.22

Haga un análisis como el del ejercicio 6.

Bibliografía

-  Pozo Z.G.A. Optimización de la operación de un reactor aerobio de biopelícula bacteriana inmovilizada: incidencia de los nutrientes en la biosíntesis de polihidroxialcanoato (PHA) como producto del tratamiento de efluentes de celulosa Kraft. Tesis de Maestría en Microbiología. Universidad de Concepción, 2009. (Ver sección 4.6. Recuento viable y acumulación de PHA por la biomasa bacteriana).

Nota: El ejercicio se acondiciona a las necesidades del curso a partir de la aplicación en la literatura.

7. Relación del contenido de ácido tiobarbitúrico con el ph en carne molida fresca de cerdo almacenada

Con el objetivo de conocer la relación entre el pH y el contenido de ácido tiobarbitúrico (TBA), se tomó una muestra de 100 g de cada uno de 36 músculos de tríceps de cerdo desgrasado y molido, que se mantuvo envuelto en material permeable al oxígeno e impermeable a la humedad y almacenado en charolas a 3 °C durante 24 días. Las 36 parejas de datos se presentan en la Tabla 7.1.

Tabla 7.1. Valores de pH y contenido de ácido tiobarbitúrico (TBA) en muestras de carne de cerdo.

pH	6.51	6.11	5.78	6.45	6.79	6.27	5.78	5.93	5.73	6.12	5.89	6.31	5.98	5.67	6.29	6.64	6.08	5.71
TBA	0.2	0.65	1.47	0.27	0.34	0.39	1.41	0.69	1.92	0.53	0.86	0.22	0.78	2.49	0.41	0.27	0.51	2.07
pH	5.88	5.66	6	5.75	6.5	6.79	5.94	5.59	6.12	6.31	6.02	5.68	5.78	5.58	5.93	5.69	6.5	6.79
TBA	1.19	2.24	0.75	1.67	0.21	0.23	0.42	3.35	0.43	0.32	0.43	2.18	1.49	3.67	1.1	2.16	0.42	0.3

Una primera aproximación a los datos con un diagrama de dispersión revela que la relación tiene forma de curva, como una parábola.

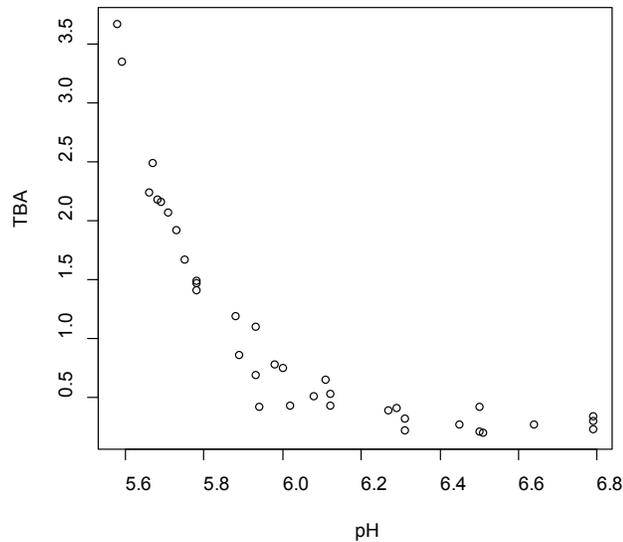


Figura 7.1. Diagrama de dispersión

Se plantea el modelo de regresión cuadrática:

$$\mu_{TBA/pH,pH^2} = \beta_0 + \beta_{pH} pH + \beta_{pH^2} pH^2.$$

Suponiendo que la variación alrededor de la media es normal con varianza desconocida $\sigma_{TBA/pH,pH^2}^2$, la distribución de TBA_{ph,ph^2} es:

$$TBA_{ph,ph^2} \sim N(\mu_{TBA/pH,pH^2} = \beta_0 + \beta_{pH} pH + \beta_{pH^2} pH^2, \sigma_{TBA/pH,pH^2}^2).$$

La regresión cuadrática de TBA sobre pH produce la siguiente ecuación estimada para la media de TBA sobre pH y pH²

$$\mu'_{TBA/pH,pH^2} = 162.9638 - 50.6624 pH + 3.9394 pH^2,$$

usando la ' para indicar que $\mu'_{TBA/pH,pH^2}$ no es variable aleatoria sino un valor calculado usando los datos de la muestra.

En el análisis de varianza de la regresión se prueba la hipótesis nula de que los coeficientes de regresión β_{pH} y β_{pH^2} son ambos cero: $H_0: \beta_{pH} = \beta_{pH^2} = 0$, $H_A: \beta_{pH} \neq 0$ ó $\beta_{pH^2} \neq 0$, esta prueba se conoce como prueba de regresión y se hace con la distribución F con 2 y 33 grados de libertad. En la última fila de la Tabla 7.2 se muestra que el valor calculado de F vale 180.4 y se obtiene un valor p de 2.2E-16, dando clara evidencia para rechazar la hipótesis nula de que los dos coeficientes de regresión son iguales a cero.

Para probar por separado cada coeficiente se dispone de las pruebas parciales para cada uno de ellos basadas en la t (Tabla 7.2). Para el coeficiente de regresión de pH se plantea la pareja de hipótesis $H_0: \beta_{pH} = 0$, $H_A: \beta_{pH} \neq 0$, y se prueba con la t, que da un valor p menor que 2E-12, rechazando la hipótesis nula. Para el coeficiente de pH^2 se plantea la pareja de hipótesis $H_0: \beta_{pH^2} = 0$, $H_A: \beta_{pH^2} \neq 0$. De acuerdo a su prueba de t, se obtiene el valor p de 6.19E-12, lo que nos lleva a rechazar la hipótesis nula.

Tabla 7.2 Pruebas de hipótesis parciales en la regresión cuadrática.

Coeficiente	Estimación	Error estándar	t calculada	valor p
β_0	162.9638	14.4169	11.3037	6.904E ⁻¹³
β_{pH}	-50.6621	4.6827	-10.8190	2.175E ⁻¹²
β_{pH^2}	3.9384	0.3791	10.3881	6.186E ⁻¹²
Estadístico F = 180.4 con 2 y 33 grados de libertad, valor p = 2.22E ⁻¹⁶ . R cuadrada = 0.9162				

Para entender la importancia del modelo en la prueba sobre cada coeficiente de regresión, conviene hacer el análisis secuencial. Para el análisis secuencial de la regresión cuadrática de TBA sobre pH se ajustan tres regresiones, se comienza con el modelo más sencillo y se agrega una variable cada vez. El primer modelo es la regresión de TBA sobre la constante, que en el paquete R se representa con un 1, el segundo modelo es la regresión lineal simple de TBA sobre pH y el tercer modelo es la regresión cuadrática de TBA sobre pH y pH^2 .

Primer modelo: $\mu_{TBA} = \beta_0$,

Segundo modelo: $\mu_{TBA/pH} = \beta_0 + \beta_{pH}pH$, y

Tercer modelo: $\mu_{TBA/pH,pH^2} = \beta_0 + \beta_{pH}pH + \beta_{pH^2}pH^2$.

Como se ilustra en la Tabla 7.3, cada vez que se agrega un coeficiente de regresión al modelo, la suma de cuadrados residual disminuye. La diferencia de sumas de cuadrados residual entre dos modelos consecutivos se conoce como suma de cuadrados extra, tiene un grado de libertad y se usa para probar la hipótesis de que el coeficiente que se agregó al modelo vale cero. Para la prueba se utiliza el cuadrado medio residual del modelo con más coeficientes. La relación de cuadrados medios da como resultado un estadístico F. La Tabla 7.3 presenta las sumas de cuadrados extra y las pruebas secuenciales de F para los coeficientes que se agregan al modelo.

Si el lector observa en la Tabla 7.3 el valor p para el último coeficiente que se agregó: $\beta_{pH^2/pH}$, puede darse cuenta que es igual que el valor p obtenido en la prueba de t del coeficiente de regresión parcial de β_{pH^2} en el modelo completo (Tabla 7.2). Es decir, el resultado de las pruebas es el mismo. En la Tabla 7.2 con un valor t de 10.39 y en la Tabla 7.3 con un valor de F de 107.95, que es el cuadrado de la t y con un valor p de 6.186E-12 en ambos casos. Para las pruebas se usa el cuadrado medio del error con 33 grados de libertad.

Para el coeficiente de pH, β_{pH} las pruebas difieren. En la Tabla 7.3 se prueba que el coeficiente β_{pH} es cero utilizando la suma de cuadrados asignada al coeficiente (con valor 19.2118), que se obtiene en el modelo de regresión simple con sólo pH, o como suma de cuadrados al agregar la variable pH a la regresión con solo la constante. Como estimación del error se usa el cuadrado medio residual del modelo cuadrático.

Tabla 7.3. Análisis de varianza de las pruebas secuenciales para el modelo cuadrático, agregando pH2 al final.

Modelos ajustados			Prueba de hipótesis por diferencia de sce de modelos				
Modelo	g.l. del error	SCE	Hipótesis nula	g.l.	SCEExtra	F	valor p
1: TBA ~ 1	35	29.9190		1			
2: TBA ~ pH	34	10.7072	$H_0: \beta_{pH} = 0$	1	19.2118	252.84	$< 2.17E^{-16}$
3: TBA ~ pH, pH ²	33	2.5075	$H_0: \beta_{pH^2/pH} = 0$	1	8.1997	107.91	$6.186E^{-12}$

Nota: el símbolo - se lee "sobre"

En cambio, la prueba parcial del coeficiente β_{pH} en el modelo cuadrático, que se presenta en la Tabla 7.2, coincide con la prueba del coeficiente β_{pH} en el análisis secuencial, **cuando el último término que se agrega a la regresión es el de pH**. Para esta prueba se usan los modelos siguientes y las pruebas se presentan en la Tabla 7.4.

Primer modelo: $\mu_{TBA} = \beta_0$,

segundo modelo: $\mu_{TBA/pH^2} = \beta_0 + \beta_{pH^2} pH^2$ y

tercer modelo: $\mu_{TBA/pH, pH^2} = \beta_0 + \beta_{pH} pH + \beta_{pH^2} pH^2$

La prueba parcial de β_{pH} en el modelo cuadrático (Tabla 7.2) utiliza $t = -10.819$ y la prueba secuencial en la Tabla 7.4 usa $F = 117.05$, que es el cuadrado de la t .

Tabla 7.4. Análisis de varianza de las pruebas secuenciales para el modelo cuadrático agregando el término de pH al final.

Modelos ajustados			Prueba de hipótesis por diferencia de sce de modelos				
Modelo	g.l. del error	SCE	Hipótesis nula	g.l.	SCEExtra	F	valor p
1: TBA ~ 1	35	29.9190		1			
2: TBA ~ pH ²	34	11.4017	$H_0: \beta_{pH^2} = 0$	1	18.5173	243.70	$2.2E^{-16}$
3: TBA ~ pH, pH ²	33	2.5075	$H_0: \beta_{pH/pH^2} = 0$	1	8.8942	117.05	$2.175E^{-12}$

En este ejemplo, ambos coeficientes (β_{pH} y β_{pH^2}), deben permanecer en la ecuación. Al hacer las pruebas parciales de regresión en la regresión cuadrática, las hipótesis nulas acerca de que cada uno de los coeficientes β_{pH} y β_{pH^2} vale cero, producen valores p que conducen al rechazo de dichas hipótesis nulas.

Cuando se usa el método secuencial se acostumbra especificar en el coeficiente de una variable que se agrega al modelo, las variables que había en el modelo anterior. Por ejemplo, en la Tabla 7.3 se usa $\beta_{pH^2/pH}$ para el coeficiente de regresión de pH² en la regresión cuadrática de TBA sobre pH y pH² luego de haber obtenido la regresión de TBA sobre pH solamente.

Para ver la cercanía de los datos a la curva de regresión que se estimó, se pide la gráfica que se muestra en la figura 7.2. Los datos observados (representados con +) se ajustan muy bien hasta $pH = 6.1$ a los valores estimados para la media (representada con °). Después de este valor la media estimada pasa por debajo de los datos observados y en el último dato por arriba. La R cuadrada vale 0.9162 (Tabla 7.2), explica 91.6% de la variación de TBA, es un porcentaje arriba de 90%, pero quizá pueda mejorarse. Conviene intentar con una regresión cúbica, que podría dar mejor ajuste.

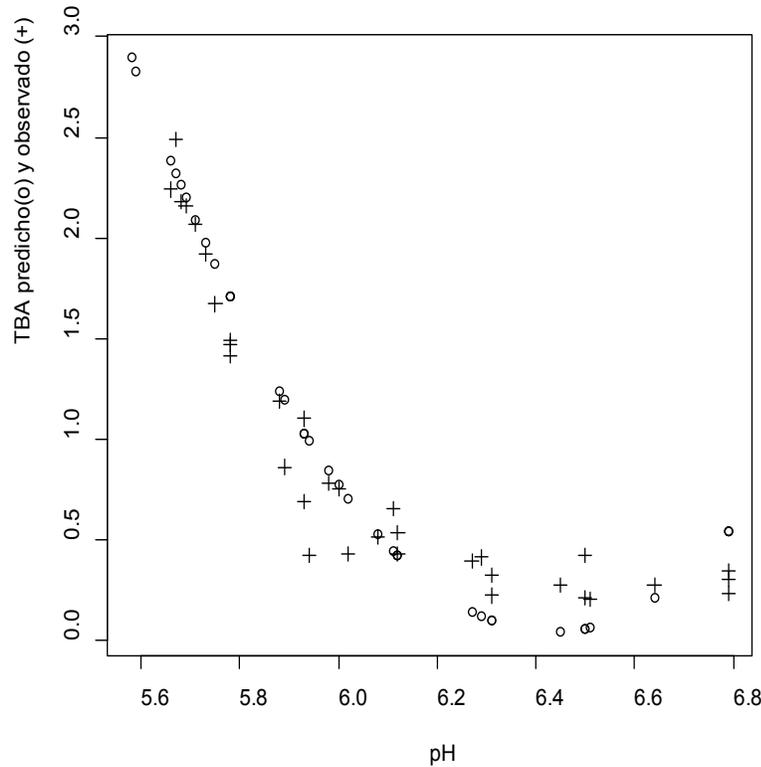


Figura 72. Diagrama de dispersión de los datos observados (+) y con los datos ajustados por la ecuación cuadrática (o).

Para la ecuación cúbica se utiliza la distribución:

$$TBA_{Or1,Or2,Or3} \sim N(\mu_{TBA/Or1,Or2,Or3} = \beta_0 + \beta_{Or1}Or1 + \beta_{Or2}Or2 + \beta_{Or3}Or3, \sigma^2_{TBA/Or1,Or2,Or3}).$$

En este caso, conviene usar el método de ajuste de polinomios ortogonales. Or1, Or2 y Or3 se refieren a los polinomios ortogonales de orden 1, 2 y 3 de pH. La ecuación de estimación de la media en el modelo cúbico usando polinomios ortogonales es:

$$\mu'_{TBA/Or1,Or2,Or3} = 1.057 - 4.383 Or1 + 2.864 Or2 - 1.296 Or3.$$

Para probar si el coeficiente de regresión del término cúbico β_{Or3} vale cero se plantea la pareja de hipótesis $H_0: \beta_{Or3} = 0$ y $H_A: \beta_{Or3} \neq 0$. Debido a que en el análisis de varianza se obtiene un valor p de 3.396E-09, se rechaza la hipótesis nula $H_0: \beta_{Or3} = 0$.

En la Tabla 7.5 el coeficiente del polinomio ortogonal de tercer orden β_{or3} se identifica como $\beta_{or3/or1,or2}$ porque se usó el método secuencial para obtener el análisis. La prueba de la hipótesis nula de que su valor es cero $H_0: \beta_{or3} = 0$; $H_A: \beta_{or3} \neq 0$ se rechaza, puesto que en el análisis de varianza se obtiene un valor p de 3.396E-09. Debido a que las variables independientes son ortogonales, el resumen del modelo cúbico coincide con las pruebas parciales de los coeficientes de regresión del modelo cúbico.

Tabla 7.5. Análisis de varianza de las pruebas secuenciales para el modelo cúbico usando polinomios ortogonales.

Modelos ajustados			Prueba de hipótesis por diferencia de s.e. de modelos				
Modelo	g.l.	SCE	Hipótesis nula	g.l.	SC	F	valor p
1: TBA ~ 1	35	29.9190		1			
2: TBA ~ Or1	34	10.7072	$H_0: \beta_{Or1} = 0$	1	19.2118	742.124	$< 2.2E^{-16}$
3: TBA ~ Or1, Or2	33	2.5075	$H_0: \beta_{Or2/Or1} = 0$	1	8.1997	316.742	$< 2.2E^{-16}$
4: TBA ~ Or1, Or2, Or3	32	0.8284	$H_0: \beta_{Or3/Or1, Or2} = 0$	1	1.6791	64.861	$3.396E^{-09}$
R Cuadrada = 0.9723118							

El diagrama de dispersión con la media estimada y los datos observados muestra que el ajuste es satisfactorio en la zona explorada de pH. La R cuadrada sube a 0.9723.

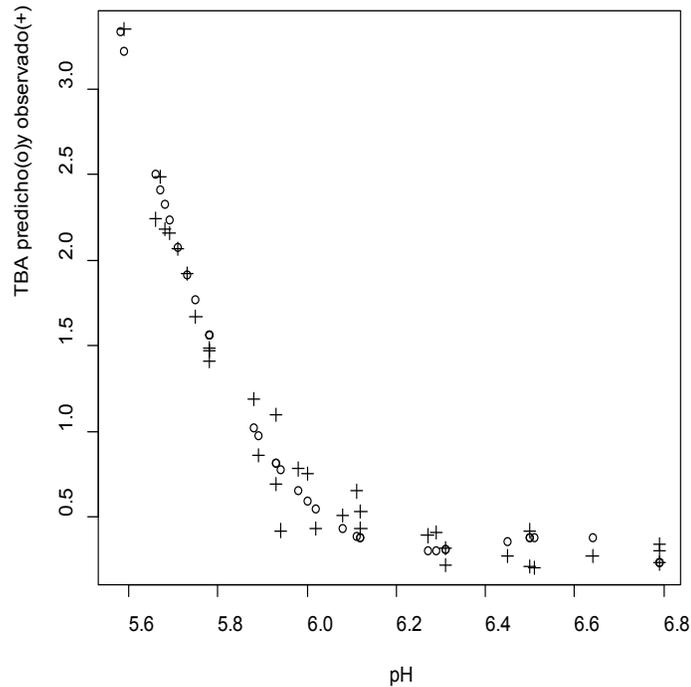


Figura 7.3. Diagrama de dispersión de los datos observados (+) y con los datos ajustados por la ecuación cúbica (o).

Se puede hacer un intervalo con 90% de confianza para $\mu_{TBA/pH, pH^2, pH^3}$, la media de TBA en la regresión cúbica para cualquier valor que se desee de pH, como se muestra en la Tabla 7.6.

Tabla 7.6. Estimación de medias del contenido de TBA a dos valores de pH con un intervalo de 90% de confianza y utilizando el modelo cúbico.

Valor de pH	Estimación de la media	Límite inferior del intervalo	Límite superior del intervalo
6	0.5959730	0.5240488	0.6678972
6.5	0.3762535	0.2775584	0.4749486

En la Tabla 7.6 se puede leer que el intervalo con 90% de confianza para la media de TBA usando el modelo cúbico, cuando pH vale 6 es (0.5240488, 0.6678972).

También se puede hacer un intervalo con 90% de confianza para una observación de TBA basado en la regresión cúbica para valores dados de pH, por ejemplo, para 6 y 6.5, como se muestra en la Tabla 7.7.

Tabla 7.7. Predicción de un valor del contenido de TBA a dos valores de pH con un intervalo de 90% de confianza y utilizando el modelo cúbico.

Valor de pH	Estimación del valor	Límite inferior del intervalo	Límite superior del intervalo
6	0.5959730	0.31410195	0.8778441
6.5	0.3762535	0.08639332	0.6661137

El intervalo con 90% de confianza para una observación de TBA, que se conoce como intervalo de predicción, usando el modelo cúbico, cuando pH vale 6 es (0.31410195, 0.8778441).

El diagrama de dispersión de TBA sobre pH también pudiera sugerir una curva logarítmica. El modelo a probar es la ecuación que representa la media de la siguiente distribución:

$$\ln TBA_{pH} \sim N(\mu_{\ln TBA/pH} = \beta_0 + \beta_{pH} pH, \sigma_{\ln TBA/pH}^2) .$$

La estimación de la regresión de lnTBA sobre pH es $\mu'_{\ln TBA/pH} = 12.8733 - 2.1717pH$, la hipótesis de que el coeficiente de regresión de pH es cero se rechaza con un valor p de $5.37E-14$, el valor de R2 es de 0.8147, como se ve en la Tabla 7.8.

Tabla 7.8. Pruebas de hipótesis parciales en la regresión de log(TBA) sobre pH.

Coefficiente	Estimación	Error estándar	t calculada	valor p
β_0	12.8733361	1.0802771	11.91670	$1.0911E^{-15}$
β_{pH}	-2.1716821	0.1776355	-12.22549	$5.3659E^{-14}$
Estadístico F = 149.4627 con 1 y 34 grados de libertad, valor p = $5.365872E^{-14}$, R cuadrada = 0.8146762				

El diagrama de dispersión de la curva estimada y los datos observados muestran que el ajuste no es adecuado (Figura 7.4), es mejor el ajuste que se obtuvo con el modelo cúbico.

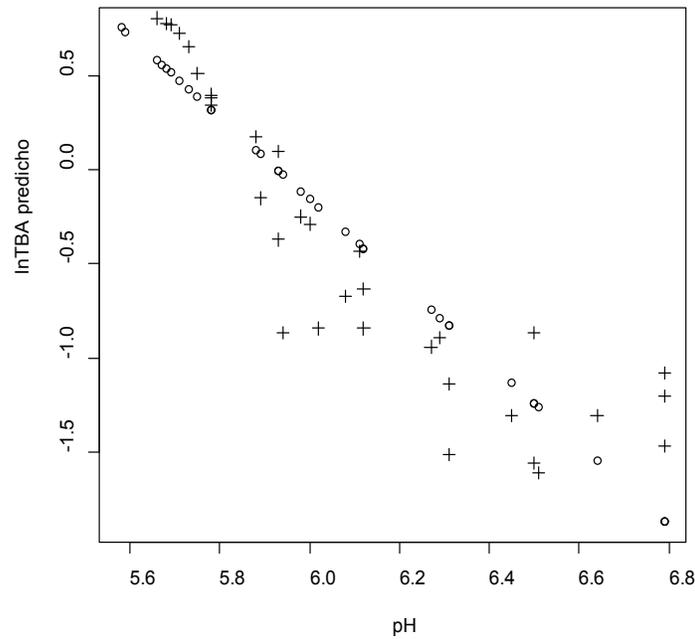


Figura 7.4. Diagrama de dispersión de los datos observados (+) y con los datos ajustados por la ecuación logarítmica (o).

Para terminar, se puede ver en la Figura 7.5 la representación gráfica de residuales estandarizados sobre los valores de la estimación de la media en el modelo de regresión cúbico (valores predichos), que muestra que los residuales son pequeños, puesto que los valores absolutos son menores que 0.4. Note que se muestra la instrucción para obtener el modelo con R: `lm(TBA~poly(pH,3))`.

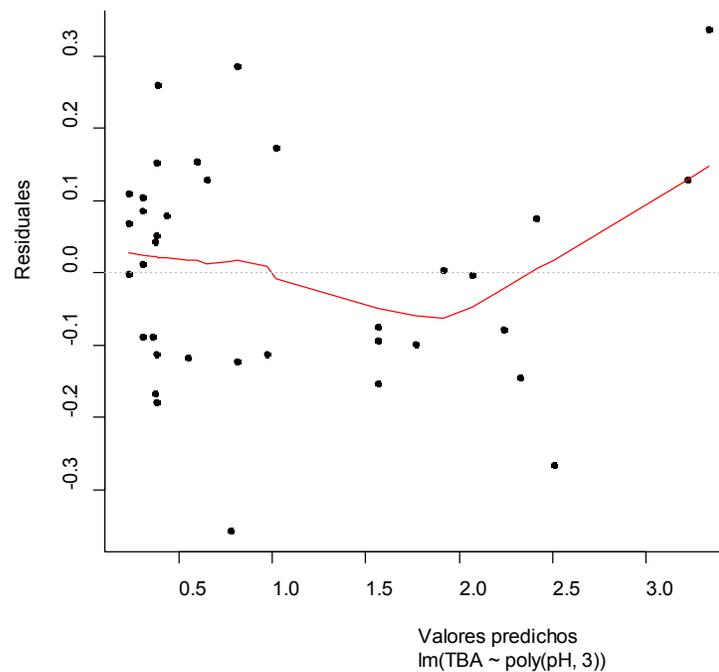


Figura 7.5. Residuales y valores predichos de acuerdo al modelo cúbico.

En la gráfica de probabilidad normal (Figura 7.6) se ve que los datos no se separan mucho de la línea recta, lo que indica que siguen de manera aproximada a la distribución normal.

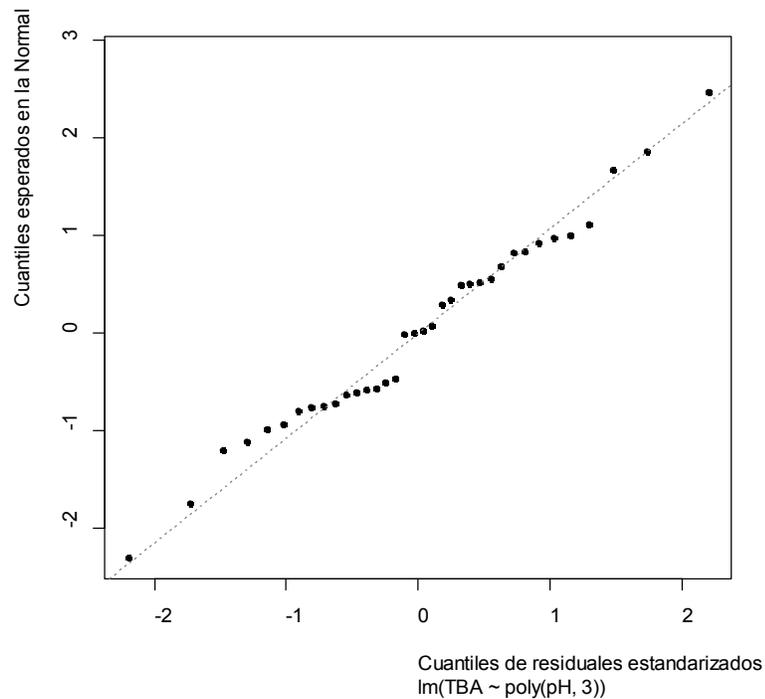


Figura 7.6. Gráfica de probabilidad normal usando el modelo cúbico.

En este caso, las separaciones de los cuantiles esperados en la normal de la línea en la Figura 7.6 pudiera sugerir hacer la prueba de si los residuales siguen la distribución normal. Para esto usamos la prueba de Shapiro-Wilk. Las hipótesis nula y alternativa son:

H_0 : La distribución de los residuales es normal y H_A : La distribución de los residuales no es normal.

La prueba arroja un valor p de 0.7757, mayor que el nivel de significación ($\alpha = 0.10$), por lo que no se rechaza la hipótesis nula y para fines prácticos, la distribución de los residuales es normal.

El análisis de este problema se hace con las instrucciones del paquete R que se indican a continuación. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares, evitando repetir las anotaciones incluidas en el análisis del problema anterior, enfatizando las que se refieren a los análisis estadísticos.

Guía para resolver con R

```
baseTBaPH<-read.table(file = "TBaPH.csv",header = T,sep = ",")
```

```
# Se lee la tabla (read.table) de datos en Excel con tipo csv que están en el archivo entre comillas.
```

```
# Después de file = Se pone la dirección completa del archivo.
```

```
# Si tiene su información en la computadora, sugiero copie la trayectoria de la barra del explorador de Windows
```

```
# y cambie las diagonales \ por diagonales /. Se usa header = T si los nombres de las columnas están en la
```

```
# primera hilera, se usa sep = "," porque el tipo csv usa comas como separadores.
```

```
#Cambie directorio a la carpeta donde está el archivo de datos y la primera instrucción lee los datos
```

```
# Se le llama baseTBaPH y se almacena.
```

```

baseTBApH # Se muestra la base de datos

TBA<- baseTBApH$TBA # el vector con los datos de TBA en la base (baseTBApH$TBA) queda en TBA

pH<-baseTBApH$pH # El vector con pH queda en pH.

# Luego del nombre de la base $ toma al vector cuyo nombre le sigue

# TBA          # pH

plot(TBA~pH,data = baseTBApH,xlab = "pH",ylab = "TBA",main = " ")

# Diagrama de dispersión de TBA sobre pH. TBA eje vertical, pH eje horizontal

pH2<-pH*pH

# Se crea el vector con pH al cuadrado, se le llama pH2

bReg<-data.frame(TBA,pH,pH2)

# Se define la base de datos para las regresiones usando pH sin modificar

RDir1y2<-lm(TBA~pH+pH2,data = bReg) # La regresión cuadrática se llama RDir1y2

summary(RDir1y2) # Muestra el resumen de la regresión cuadrática

anova(RDir1y2) # Muestra el anova de la regresión cuadrática

RDir0<-lm(TBA~1) # Regresión usando como variable sólo a la constante beta cero

# summary(RDir0)

# anova(RDir0)

RDir1<-lm(TBA~pH,data = bReg) # Regresión lineal simple de TBA sobre pH

# summary(RDir1)

# anova(RDir1)

anova(RDir0,RDir1,RDir1y2) # Análisis de varianza (AdeV) secuencial para cuadrático

RDir2<-lm(TBA~pH2,data = bReg) # Regresión lineal simple de TBA sobre pH2

anova(RDir0,RDir2,RDir1y2)

# AdeV secuencial para cuadrático cambiando el orden de entrada de las variables

cuadPred<- predict(RDir1y2) # Vector de estimaciones de medias (predichos) del cuadrático

# cuadPred # Si se desea ver el vector

plot(cuadPred~pH,pch = 1,xlab = "pH",ylab = "TBA predicho(o) y observado (+)",main = " ")

```

```

# Diagrama de dispersión de medias estimadas en el modelo cuadrático

points(TBA~pH, pch = 3)

# points agrega los valores observados de TBA (TBA~pH)

Rpoli3<-lm(TBA~poly(pH,3),data = bReg) # Regresión del modelo cúbico con polinomios ortogonales

summary(Rpoli3)          # Pruebas parciales del modelo de tercer orden de polinomios ortogonales

# coincide con las pruebas parciales del modelo cúbico con las variables originales

Rpoli1<-lm(TBA~poly(pH,1),data = bReg)

Rpoli2<-lm(TBA~poly(pH,2),data = bReg)

anova(RDir0,Rpoli1,Rpoli2,Rpoli3)

cubPred<- predict(Rpoli3) # Estimaciones de medias (predichos) de cúbica

# cubPred      # Si se desea ver el vector

plot(cubPred~pH,xlab = "pH",ylab = "TBA predicho(o)y observado(+)",main = " ")

points(TBA~pH,pch = 3)

# Gráfica de medias estimadas con modelo cúbico sobre pH, points agrega los valores observados de TBA (TBA~pH)

casosPred<- data.frame(pH = c(6,6.5)) # Se define el marco con los datos de pH para hacer intervalos de confianza

predict(Rpoli3,newdata = casosPred,interval = "confidence",level = 0.90)

# Intervalos de confianza. Predict pide el modelo (Rpoli3), el marco donde están los datos de pH (newdata = casosPred),

# para intervalos con 90% (level = 0.90) para la media (interval = "confidence")

predict(Rpoli3,newdata = casosPred,interval = "prediction",level = 0.90)

# Intervalos de confianza usando predict. Intervalos para una observación (interval = "prediction")

lnTBA<-log(TBA) # Se crea la variable logaritmo natural de TBA: lnTBA

# lnTBA      # Si desea ver el vector

lnTBA_pH<- lm(lnTBA~pH) # regresión lineal usando ln(TBA) como variable respuesta

summary(lnTBA_pH)

lnTBA_pHPred<- predict(lnTBA_pH) # Se crea el vector de valores estimados de la media de lnTBA

# lnTBA_pHPred

plot(lnTBA_pHPred~pH,xlab = "pH",ylab = "lnTBA predicho",main = " ")

```

```

points(lnTBA~pH,pch = 3)

# Gráfica de medias estimadas para lnTBA sobre pH. Points agrega los valores observados de lnTBA (lnTBA~pH)

plot(Rpoli3,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Valores predichos",ylab = "Residuales",col.lab = 1)

# Gráfica de residuales sobre medias estimadas (predichos) (wich = 1)

plot(Rpoli3,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Cuantiles de residuales estandarizados",
ylab = "Cuantiles esperados en la Normal",col.lab = 1)

# Grafica de probabilidad normal: Cuantiles esperados bajo la normal sobre cuantiles de residuales estandarizados

ResTBAreCub<-residuals(Rpoli3) # Se obtienen los residuales y se denominan ResTBAreCub

shapiro.test(ResTBAreCub) # Se pide la prueba de Shapiro-Wilk

```

7.1. Ejercicio

En un segundo trabajo para relacionar el contenido de ácido tiobarbitúrico con el pH en carne molida fresca de cerdo almacenada se obtuvieron los siguientes datos de pH y TBA:

pH	5.94	6.10	5.78	6.45	6.77	6.17	5.72	5.67	5.71	6.11	5.87	6.21	5.95	5.57	6.24	6.62	6.07
TBA	0.41	0.67	1.45	0.29	0.35	0.36	1.43	2.14	1.93	0.50	0.85	0.24	1.15	2.35	0.44	0.25	0.53

Resuelva de manera similar al problema 7.

Bibliografía

 Yasosky, J.J., Aberle, E.D., Peng, I.C., Mills, E.W. and Jugde, M.D. (1984) Effects of pH and time of grinding on lipid oxidation of fresh ground pork. *Journal of Food Science*. 49: 1510-1512.

Nota: La referencia se utilizó para plantear un problema similar pero llevado al nivel y el interés del curso.

8. Perfil de actividad metanogénica específica en un reactor anaerobio de flujo ascendente con lecho de lodos

Cuando se trabaja con reactores anaerobios de lecho de lodos con flujo ascendente (UASB), la actividad metanogénica específica (AME) y el contenido de sólidos suspendidos volátiles (SSV), de un lodo anaerobio, son algunas de las variables que se evalúan por la relación directa que tienen con la eficiencia del proceso para el tratamiento de efluentes cloacales. A mayor actividad metanogénica específica (AME) y mayor contenido de sólidos suspendidos volátiles (SSV), se tiene mayor actividad del lodo activo. Sin embargo, al hacer purgas de un reactor, se debe cuidar de no extraer el lodo con mayor actividad metanogénica.

En un reactor metanogénico se midieron la AME y los SSV a diferentes alturas, a fin de conocer la altura más conveniente en la que se pueda extraer lodo activo que pueda ser inoculado en nuevos reactores destinados al tratamiento de líquidos cloacales. Con este propósito se representa el valor de la media de AME como función de la altura, por medio de una ecuación de regresión, a fin de obtener el valor de altura que da la media mínima de AME. Se siguió la misma técnica para obtener el valor de la altura a la que produce la media máxima de SSV. Así pues, la variable explicativa fue la altura del reactor, que tomó valores entre 0 y 2.55 m, mientras que las variables de respuesta fueron la actividad metanogénica específica (AME = volumen de metano producido por día) y el contenido de SSV, que es la concentración de sólidos suspendidos volátiles (g/l). Se hicieron cuatro repeticiones separadas en el tiempo, no hay razón para pensar que las repeticiones pudieran afectar a las medias de las variables de respuesta. Los datos se muestran en la Tabla 8.1.

Tabla 8.1. Datos obtenidos de actividad metanogénica a diferentes alturas del reactor UASB.

Repetición	Altura (m)	SSV (g/l)	AME (vol biogás/d)
1	0.19	20	0.11
2	0.19	22	0.13
3	0.19	24	0.09
4	0.19	23	0.07
1	0.64	31	0.09
2	0.64	36	0.1
3	0.64	33	0.13
4	0.64	32	0.08
1	1.09	36	0.12
2	1.09	44	0.1
3	1.09	41	0.12
4	1.09	42	0.09
1	1.39	25	0.08
2	1.39	27	0.13
3	1.39	31	0.12
4	1.39	28	0.1
1	1.69	8	0.16
2	1.69	10	0.15
3	1.69	9	0.14
4	1.69	13	0.18

Los diagramas de dispersión de las variables respuesta SSV y AME sobre altura, muestran curvas que pueden ajustarse a una cuadrática (Figuras 8.1 y 8.2). El diagrama de AME muestra la presencia de una zona con valor mínimo, mientras que la de SSV muestra una con un valor máximo.

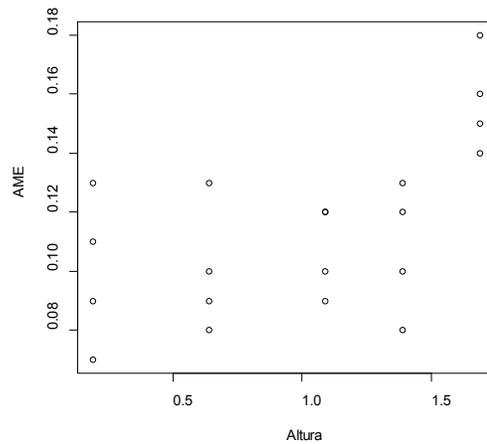


Figura 8.1. Diagrama de dispersión de la variable respuesta AME sobre altura.

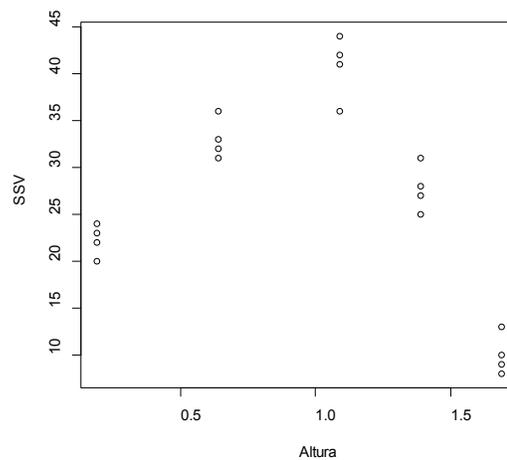


Figura 8.2. Diagrama de dispersión de la variable respuesta SSV sobre altura.

La distribución de AME como función cuadrática de la altura se escribe:

$$AME \sim N(\mu_{AME/ALT,ALT^2} = \beta_0 + \beta_{ALT}ALT + \beta_{ALT^2}ALT^2, \sigma_{AME/ALT,ALT^2}^2).$$

El modelo cuadrático de AME sobre altura se acostumbra escribir como:

$$AME = \beta_0 + \beta_{ALT}ALT + \beta_{ALT^2}ALT^2 + \varepsilon, \text{ con } \varepsilon \sim N(0, \sigma_{AME/ALT,ALT^2}^2).$$

El modelo de estimadores es:

$$\hat{\mu}_{AME/ALT,ALT^2} = \hat{\beta}_0 + \hat{\beta}_{ALT}ALT + \hat{\beta}_{ALT^2}ALT^2, A$$

donde $\hat{\mu}_{AME/ALT,ALT^2}$, $\hat{\beta}_0$, $\hat{\beta}_{ALT}$ y $\hat{\beta}_{ALT^2}$ y son variables aleatorias.

Las estimaciones de los coeficientes de regresión se presentados en la Tabla 8.2. producen el modelo de estimación:

$$\mu'_{AME/ALT,ALT^2} = 0.11395 - 0.06537 ALT + 0.05132ALT^2.$$

Las pruebas parciales sobre los coeficientes de regresión, son las pruebas de las hipótesis nulas de que cada coeficiente en la ecuación de regresión es cero contra que es diferente de cero. En el caso de la pareja $H_0: \beta_{ALT} = 0$, $H_A: \beta_{ALT} \neq 0$, se puede ver en la Tabla 8.2 que no hay evidencia para rechazar la hipótesis nula porque el valor p vale 0.1181. Por el contrario, para el coeficiente de altura al cuadrado $H_0: \beta_{ALT^2} = 0$, $H_A: \beta_{ALT^2} \neq 0$ con un valor $p = 0.0241$, se rechaza la hipótesis nula.

Tabla 8.2. Pruebas de hipótesis parciales de la regresión cuadrática para AME-altura

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	0.11395	0.01579	7.217	1.44E ⁻⁰⁶
Altura (β_{ALT})	-0.06537	0.03971	-1.646	0.1181
Altura cuadrada (β_{ALT^2})	0.05132	0.02072	2.477	0.0241

Para decidir si conviene o no dejar el término de altura en la regresión, se puede ver en el análisis de varianza de la regresión lineal, que hace la prueba de F para la hipótesis $H_0: \beta_{ALT} = 0$, $H_A: \beta_{ALT} \neq 0$. A este análisis, a diferencia de las pruebas parciales, se le conoce como pruebas secuenciales. Como se muestra en la Tabla 8.3, el valor $p = 0.00351$ conduce a rechazar la hipótesis nula. El modelo cuadrático con los dos términos (altura y altura cuadrada), produce una R^2 de 0.5087.

Tabla 8.3. Análisis de varianza de la regresión cuadrática para AME-altura

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Altura	1	0.0052679	0.0052679	11.4677	0.00351
Altura cuadrada	1	0.0028179	0.0028179	6.1343	0.02406
Residual	17	0.0078092	0.0004594	R cuadrada	0.5087

Aun así, el ajuste de la estimación a los datos no es muy bueno, pues la regresión cuadrática sólo explica el 50.87% de la variación de AME. Esto también se ve en la Figura 8.3, que muestra que la curva de estimación (símbolos +) pasa por en medio de los puntos observados (símbolo °). La menor media estimada entre los valores de altura ensayados se encuentra a 0.64 de altura. El ajuste es bueno, pero como la variación de los datos es muy grande, el investigador tiene que encontrar la causa de esa variación usando otras variables explicativas, o logrando tener mayor control durante el trabajo de campo del experimento.

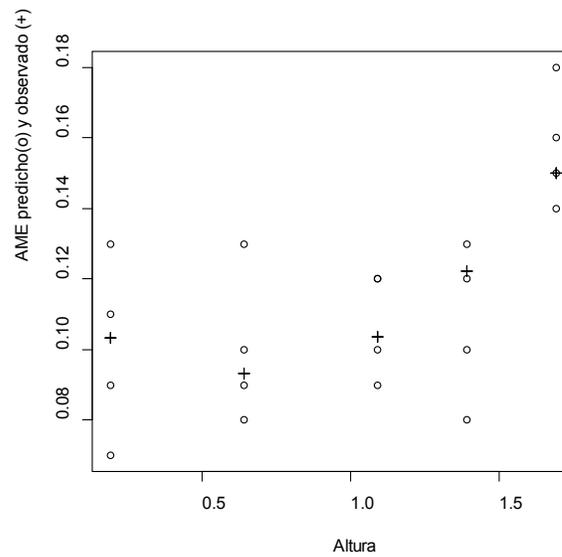


Figura 8.3. Diagrama de dispersión de los datos de AME observados (o) y estimados (+) con el modelo cuadrático.

La distribución de SSV en función de la altura es:

$$SSV \sim N(\mu_{SSV/ALT,ALT^2} = \beta_0 + \beta_{ALT}ALT + \beta_{ALT^2}ALT^2, \sigma_{SSV/ALT,ALT^2}^2).$$

El modelo cuadrático de SSV en términos de la altura se acostumbra escribirlo como sigue:

$$SSV = \beta_0 + \beta_{ALT}ALT + \beta_{ALT^2}ALT^2 + \varepsilon, \text{ con } \varepsilon \sim N(0, \sigma_{SSV/ALT,ALT^2}^2).$$

El modelo de estimadores es:

$$\hat{\mu}_{SSV/ALT,ALT^2} = \hat{\beta}_0 + \hat{\beta}_{ALT}ALT + \hat{\beta}_{ALT^2}ALT^2,$$

donde $\mu_{SSV/ALT,ALT^2}$, $\hat{\beta}_0$, $\hat{\beta}_{ALT}$ y $\hat{\beta}_{ALT^2}$ son variables aleatorias que se evalúan con los valores de los datos. Los estimadores que se muestran en la Tabla 8.4 producen el siguiente modelo de estimación:

$$\hat{\mu}_{SSV/ALT,ALT^2} = 8.817 + 70.693 ALT - 41.077 ALT^2.$$

Las pruebas parciales sobre los coeficientes de regresión, son las pruebas de las hipótesis nulas de que cada coeficiente en la ecuación de regresión es cero, la alternativa es bilateral.

En el caso de la pareja $H_0: \beta_{ALT} = 0$, $H_A: \beta_{ALT} \neq 0$ se rechaza la hipótesis nula porque, como se muestra en la Tabla 8.4, el valor $p = 8.68E-09$ es menor que 0.05. Para el coeficiente de altura al cuadrado, ocurre lo mismo, en la pareja de hipótesis $H_0: \beta_{ALT^2} = 0$, $H_A: \beta_{ALT^2} \neq 0$ también se rechaza la hipótesis nula, el valor $p = 1.73E-09$ es menor que 0.05.

Tabla 8.4. Pruebas de hipótesis parciales de la regresión cuadrática para SSV y altura.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	8.817	2.703	3.262	0.00459
Altura (β_{ALT})	70.693	6.797	10.401	8.68E-09
Altura cuadrada (β_{ALT^2})	-41.077	3.547	-11.582	1.73E-09

En el análisis secuencial de la Tabla 8.5 también se rechazan las hipótesis de que los coeficientes de altura y altura cuadrada valen cero.

Tabla 8.5. Análisis de varianza de la regresión cuadrática para SSV y altura

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Altura	1	203.83	203.83	15.147	0.001172
Altura cuadrada	1	1805.14	1805.14	134.138	1.728E-09
Residual	17	228.77	13.46	R cuadrada	0.8978

El modelo cuadrático explica el 89.78% de la variación (tiene un $R^2 = 0.8978$).

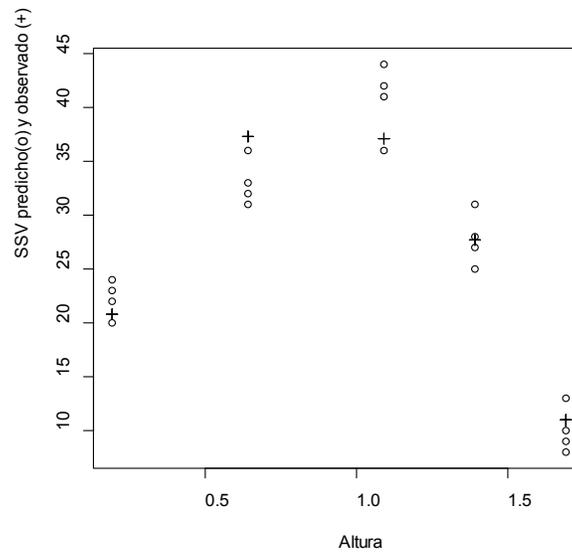


Figura 8.4. Diagrama de dispersión de los datos de SSV observados (o) y estimados (+) con el modelo cuadrático.

El ajuste de la estimación a los datos es bueno, aunque en la Figura 8.4 se ve que la estimación de la media para altura = 0.64 está arriba de las observaciones, por lo que conviene hacer la estimación cúbica y ver si es mejor que la cuadrática. Para SSV sobre altura se tiene el modelo cúbico que produce la distribución:

$$SSV \sim N(\mu_{SSV/ALT,ALT^2,ALT^3} = \beta_0 + \beta_{ALT}ALT + \beta_{ALT^2}ALT^2 + \beta_{ALT^3}ALT^3, \sigma_{SSV/ALT,ALT^2,ALT^3}^2).$$

Se acostumbra escribir:

$$SSV = \beta_0 + \beta_{ALT}ALT + \beta_{ALT^2}ALT^2 + \beta_{ALT^3}ALT^3 + \varepsilon, \text{ con } \varepsilon \sim N(0, \sigma_{SSV/ALT,ALT^2,ALT^3}^2),$$

el modelo de estimadores es:

$$\hat{\mu}_{SSV/ALT,ALT^2,ALT^3} = \hat{\beta}_0 + \hat{\beta}_{ALT} ALT + \hat{\beta}_{ALT^2} ALT^2 + \hat{\beta}_{ALT^3} ALT^3.$$

donde $\hat{\mu}_{SSV/ALT,ALT^2,ALT^3}$, $\hat{\beta}_0$, $\hat{\beta}_{ALT}$, $\hat{\beta}_{ALT^2}$ y $\hat{\beta}_{ALT^3}$ son variables aleatorias, se evalúa con los valores de los datos para obtener el modelo de estimación (Tabla 8.6):

$$\mu'_{SSV/ALT,ALT^2,ALT^3} = 17.320 + 21.085 ALT + 22.603 ALT^2 - 22.413 ALT^3$$

Las pruebas parciales sobre los coeficientes de regresión son las pruebas de las hipótesis nulas de que cada coeficiente en la ecuación de regresión es cero, la alternativa es bilateral.

En la pareja $H_0: \beta_{ALT} = 0$, $H_A: \beta_{ALT} \neq 0$ no se rechaza la hipótesis nula porque, como se puede ver en la tabla 8.6, el valor p es: 0.249937. Para el coeficiente de altura al cuadrado $H_0: \beta_{ALT^2} = 0$, H_A :

$\beta_{ALT^2} \neq 0$ se tiene el valor p = 0.312825. No hay evidencia para rechazar la hipótesis nula de coeficiente cero, y para la hipótesis sobre el coeficiente de altura cúbica $H_0: \beta_{ALT^3} = 0$, $H_A: \beta_{ALT^3} \neq 0$ el valor p es 0.009151 y se rechaza la hipótesis nula.

Tabla 8.6. Pruebas de hipótesis parciales de la regresión cúbica para SSV y altura.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	17.320	3.639	4.759	0.000213
Altura (β_{ALT})	21.085	17.661	1.194	0.249937
Altura cuadrada (β_{ALT^2})	22.603	21.689	1.042	0.312825
Altura cúbica (β_{ALT^3})	-22.413	7.563	-2.963	0.009151

Conviene revisar las pruebas secuenciales en el análisis de varianza que se muestra en la Tabla 8.7. Los resultados evidencian que las hipótesis nulas secuenciales sobre los tres coeficientes se rechazan, con valores p de 0.0002414, 2.178E-10 y 0.0091510.

Tabla 8.7. Análisis de varianza de la regresión cúbica para SSV y altura

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Altura	1	203.83	203.83	22.0801	0.0002414
Altura cuadrada	1	1805.14	1805.14	195.5414	2.178E ⁻¹⁰
Altura cúbica	1	81.07	81.07	8.7819	0.0091510
Residual	16	147.70	9.23	R cuadrada	0.934

El modelo cúbico explica el 93.4 % de la variación (tiene un $R^2 = 0.934$). El ajuste de la estimación a los datos es bueno, se ve en la Figura 8.5, donde los + son las estimaciones y los ° son los datos experimentales. El máximo de SSV en los tratamientos ensayados ocurre en 1.09 de altura.

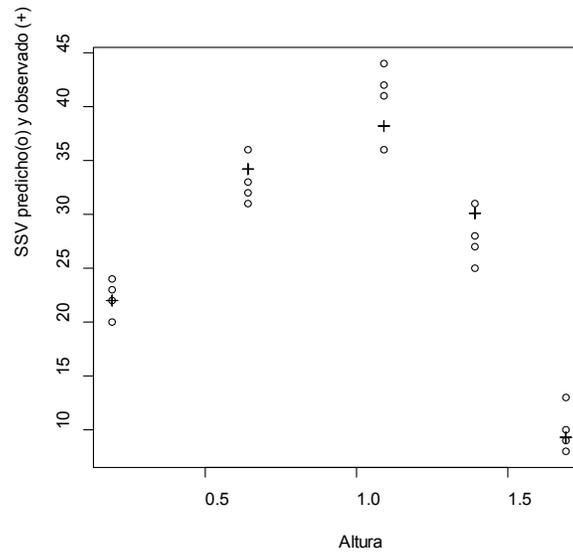


Figura 8.5. Diagrama de dispersión de los datos de SSV observados (o) y estimados (+) con el modelo cúbico.

Por otro lado, el diagrama de dispersión de AME sobre SSV (Figura 8.6), muestra una relación que parece lineal, aunque también pudiera ser cuadrática.

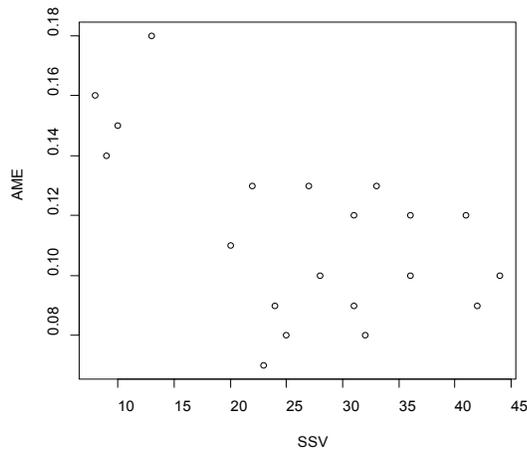


Figura 8.6. Diagrama de dispersión de la AME sobre SSV.

Para el análisis se plantea la distribución:

$$AME \sim N(\mu_{AME/SSV,SSV^2} = \beta_0 + \beta_{SSV}SSV + \beta_{SSV^2}SSV^2, \sigma_{AME/SSV,SSV^2}^2),$$

que se acostumbra representar como:

$$AME = \beta_0 + \beta_{SSV}SSV + \beta_{SSV^2}SSV^2 + \varepsilon \text{ con } \varepsilon \sim N(0, \sigma_{AME/SSV,SSV^2}^2).$$

El modelo con los estimadores (los parámetros con $\hat{\cdot}$), que son variables aleatorias normales es:

$$\hat{\mu}_{AME/SSV,SSV^2} = \hat{\beta}_0 + \hat{\beta}_{SSV}SSV + \hat{\beta}_{SSV^2}SSV^2,$$

y con los datos se obtiene el modelo estimado que es (Tabla 8.8):

$$\mu'_{AME/SSV,SSV^2} = 0.2045 - 0.006282SSV + 0.00009432SSV^2.$$

Las parejas de hipótesis para las pruebas parciales sobre los coeficientes de las variables explicativas del modelo son:

para β_{SSV} : $H_0: \beta_{SSV} = 0$, $H_A: \beta_{SSV} \neq 0$

y

para β_{SSV^2} : $H_0: \beta_{SSV^2} = 0$, $H_A: \beta_{SSV^2} \neq 0$.

Los valores p para las pruebas son 0.0112 y 0.0406 (Tabla 8.8), y conducen al rechazo de las hipótesis nulas correspondientes.

Tabla 8.8. Pruebas de hipótesis parciales de la regresión cuadrática para AME sobre SSV.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	0.2045	0.02625	7.789	5.24E ⁻⁰⁷
SSV (β_{SSV})	-0.006282	0.002208	-2.845	0.0112
SSV cuadrada (β_{SSV^2})	0.00009432	0.00004255	2.216	0.0406

Debido a que ya se rechazó la hipótesis de que el coeficiente de SSV es cero, la Tabla 8.9 del análisis de varianza secuencial tiene poco interés.

Tabla 8.9. Análisis de varianza de la regresión cuadrática para AME sobre SSV.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
SSV	1	0.0050376	0.0050376	10.1670	0.005378
SSV cuadrada ²	1	0.0024342	0.0024342	4.9128	0.040585
Residuales	17	0.0084232	0.0004955	R cuadrada	0.4701

En el diagrama de dispersión con los datos y los valores estimados para las medias (Figura 8.7), se ve que el ajuste es adecuado, aunque el $R^2 = 0.4701$ indica que la regresión sólo explica el 47.01% de la variación. Esto último se debe a la variación que presentan los datos.

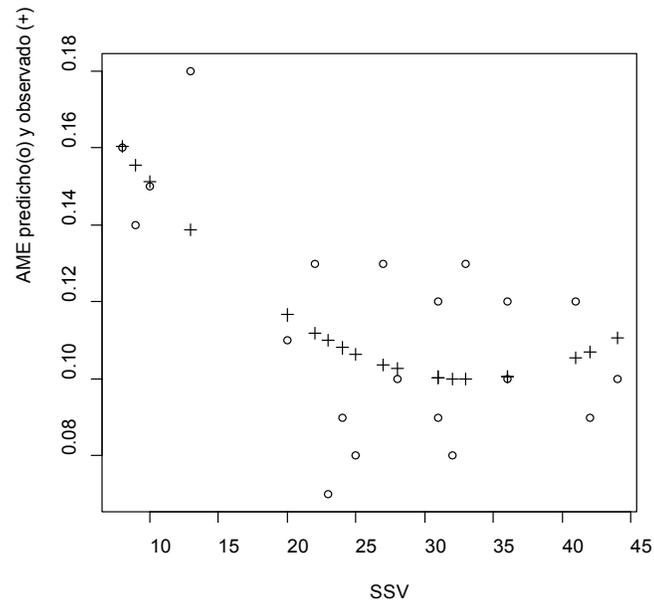


Figura 8.7. Diagrama de dispersión de los datos de AME observados (o) y estimados (+) con el modelo cuadrático.

La Figura 8.8 de los residuales confirma que los datos presentan mucha variación, mientras que la Figura 8.9 de probabilidad normal muestra que la suposición de distribución normal parece no cumplirse, pues los datos tienen patrones que se alejan de la línea recta. Conviene hacer la prueba de normalidad de Shapiro-Wilk sobre los residuales. Al hacer la prueba, se encuentra un valor p de 0.799, por lo que no se rechaza la hipótesis nula y se puede decir que los residuales son normales.

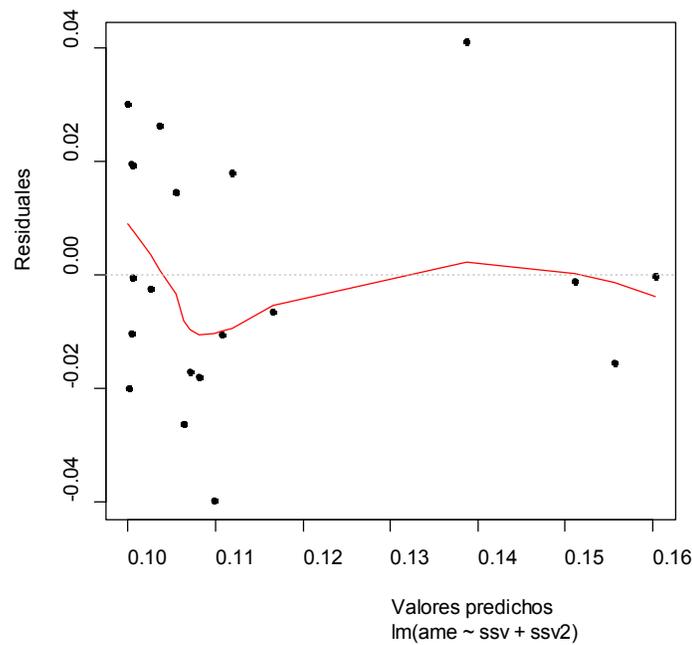


Figura 8.7. Residuales contra predichos de ame sobre ssv cuadrática.

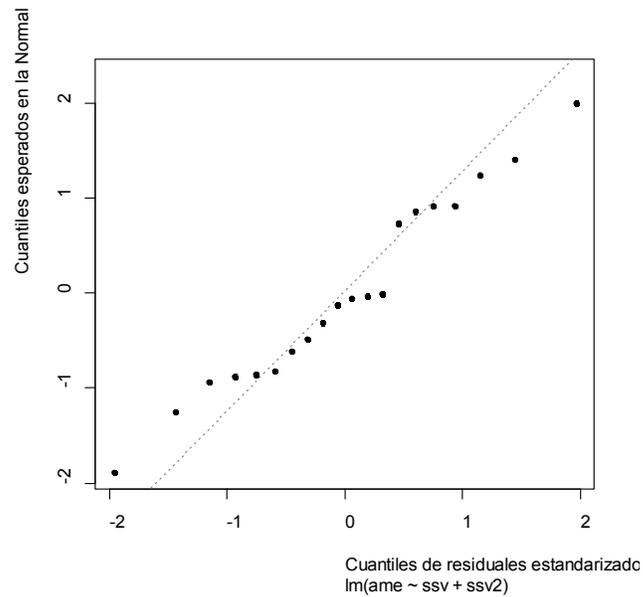


Figura 8.9. Gráfica de probabilidad normal de ame sobre ssv cuadrática.

El análisis de la actividad metanogénica y concentración de SSV se hace con las instrucciones del paquete R que siguen.

Guía para resolver con R

```
baseClo<-read.table(file = "cloaca.csv",header = T,sep = ";")

baseClo

alt<- baseClo$Altura.m

ssv<- baseClo$SSV.g.l

ame<- baseClo$AME.g

alt2<-alt*alt

# alt # ssv # ame

par(mfrow = c(1,1))

plot(ame~alt,xlab = "Altura",ylab = "AME",main = " ")

plot(ssv~alt,xlab = "Altura",ylab = "SSV",main = " ")

ameR2<- lm(ame~alt+alt2)

ameR2

summary(ameR2)

anova(ameR2)
```

```

ameR2Pred<- predict(ameR2)

# ameR2Pred

plot(ame~alt,pch = 1,xlab = "Altura",ylab = "AME predicho(o) y observado (+)",main = " ")

points(ameR2Pred~alt, pch = 3)

# ssvR2p<- lm(ssv~poly(alt,2))

# ssvR2p

# summary(ssvR2p)

# anova(ssvR2p)

ssvR2<- lm(ssv~alt+alt2)

ssvR2

summary(ssvR2)

anova(ssvR2)

ssvR2Pred<- predict(ssvR2)

# ssvR2Pred

plot(ssv~alt,pch = 1,xlab = "Altura",ylab = "SSV predicho(o) y observado (+)",main = " ")

points(ssvR2Pred~alt, pch = 3)

alt3 = alt2*alt

# alt3

ssvR3<- lm(ssv~alt+alt2+alt3)

ssvR3

summary(ssvR3)

anova(ssvR3)

ssvR3Pred<- predict(ssvR3)

# ssvR3Pred

plot(ssv~alt,pch = 1,xlab = "Altura",ylab = "SSV predicho(o) y observado (+)",main = " ")

points(ssvR3Pred~alt, pch = 3)

plot(ame~ssv,xlab = "SSV",ylab = "AME",main = " ")

```

```

ssv2<-ssv*ssv

ameRssv2<- lm(ame~ssv+ssv2)

ameRssv2

summary(ameRssv2)

anova(ameRssv2)

ameRssv2Pred<- predict(ameRssv2)

# ameRssv2Pred

plot(ame~ssv,pch = 1,xlab = "SSV",ylab = "AME predicho(o) y observado (+)",main = " ")

points(ameRssv2Pred~ssv, pch = 3)

# par(mfrow = c(2,2))

# plot(ameRssv2)

par(mfrow = c(1,1))

plot(ameRssv2,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Valores predichos",ylab = "Residuales",col.lab = 1)

plot(ameRssv2,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Cuantiles de residuales estandarizados",

ylab = "Cuantiles esperados en la Normal",col.lab = 1)

Residuales<-residuals(ameRssv2)

shapiro.test(Residuales)
    
```

8.1 Ejercicio

En un experimento similar al del problema 8, se usaron líquidos cloacales de otra localidad. La variable explicativa fue la altura del reactor (0-2.55 m) y la variable respuesta fue AME, donde AME = gas metano producido por día, se hicieron cuatro repeticiones separadas en el tiempo, no hay razón para pensar que las repeticiones pudieran afectar a las medias de las variables respuesta. Haga el análisis como en el problema 8, pero ahora sólo hay una variable respuesta. Los datos son:

Repetición	Altura m	AME Volumen de gas/d
1	0.19	0.08
2	0.19	0.11
3	0.19	0.12
4	0.19	0.09
1	0.64	0.08
2	0.64	0.12
3	0.64	0.11
4	0.64	0.10
1	1.09	0.15
2	1.09	0.17
3	1.09	0.16
4	1.09	0.14
1	1.39	0.17
2	1.39	0.14
3	1.39	0.13
4	1.39	0.19
1	1.69	0.15
2	1.69	0.13
3	1.69	0.14
4	1.69	0.17

Haga el análisis.

Bibliografía

-  Guerra, R.G., González, S.M., Trupiano, A.P., Figueroa, M.E., Seghezzi, L. y Cuevas, C.M. (2000) Perfiles de actividad metanogénica específica en un reactor UASB (Reactor anaerobio de lecho de lodos con flujo ascendente) utilizado para el tratamiento de líquidos cloacales pre-sedimentados. *Revista de la Asociación Argentina de Energías Renovables y Ambiente, ASADES*. 4(2): 06.25-06.30.

Nota: La referencia se utilizó, para plantear un problema similar pero llevado al nivel y el interés del curso.

9. Mejoramiento de la actividad de xilanasas de *Staphylococcus sp.* usando diferentes fuentes de nitrógeno

Las xilanasas son enzimas hidrolíticas que actúan para convertir el xilano en los azúcares que lo constituyen, por lo que se pueden utilizar en la digestión ruminal, tratamiento de desechos orgánicos, generación de energía, de productos químicos y de papel.

El presente ejemplo, utilizando *Staphylococcus sp.*, tiene dos objetivos: 1) probar el efecto de diferentes fuentes de nitrógeno sobre la densidad celular (determinada mediante la densidad óptica a 500 nm), y la actividad de las xilanasas (UI), y 2) conocer la relación entre la densidad celular y la actividad de xilanasas. Para lograrlos se plantea un diseño completamente al azar, donde el factor de estudio corresponde a la fuente de nitrógeno y cuenta con 13 niveles, incluyendo Xilosa que se usa como control. Se realizaron dos repeticiones y se usaron dos variables respuesta: densidad celular y actividad de las xilanasas (UI).

La unidad experimental fue un matraz Erlenmeyer de 250 ml con 100 ml de medio líquido estándar, se agitó a 200 rpm y se mantuvo a 37°C con pH inicial de 8.0 durante tres días. Posteriormente se midió la densidad celular y se determinó la actividad enzimática de la xilanasas a un pH de 7.5 (Unidades Internacionales, UI). Los datos de la actividad enzimática de xilanasas en UI obtenidos al final del cultivo para cada tratamiento ensayado y los valores de densidad celular se muestran en la Tabla 9.1. Se incluye la columna de Fuente de nitrógeno codificada que se usa en los análisis. Primero se hará el análisis para densidad celular.

Tabla 9.1. Datos de actividad enzimática de xilanasas y densidad celular con diferentes fuentes de nitrógeno.

Fuente de nitrógeno 0.5% w/v	Fuente de nitrógeno codificada	Densidad celular		Actividad de Xilanasas (UI)	
Xilosa	1	0.28	0.235	380.2	352.3
NH ₄ Cl	2	0.217	0.163	302.6	279.5
(NH ₄) ₂ SO ₄	3	0.259	0.235	257.3	245.6
(NH ₄) ₂ HPO ₄	4	0.5	0.481	558.1	569.7
(NH ₄)H ₂ PO ₄	5	0.272	0.264	314.1	325.1
(NH ₄) ₂ S ₂ O ₈	6	0.443	0.483	173	193.1
NH ₂ CONH ₂	7	0.148	0.217	159.4	143.5
NaNO ₃	8	0.314	0.307	436	431.2
HONH ₃ Cl	9	0.208	0.233	201.7	180.3
NH ₄ NO ₃	10	0.066	0.019	127.7	95.6
Peptona	11	0.629	0.586	369.5	346.9
Extracto de levadura	12	0.397	0.407	223.8	202.6
Licor de maíz macerado	13	1.043	1.075	1084.1	1107.3

Para analizar el efecto de la fuente de nitrógeno sobre la densidad celular se plantea el modelo:

$$\mu_{dC/fNi,j} = \mu + fN_i ; i = 1,2, \dots, 13 \text{ y } j = 1,2.$$

En la ecuación anterior, dC representa a la variable aleatoria densidad celular, fN representa al factor fuente de nitrógeno, fNi representa al efecto del nivel i de la fuente de nitrógeno en la media de la densidad celular, $\mu_{dC/fNi,j}$ es la media de dC para el nivel i de fuente de nitrógeno (fNi) en la repetición j. El subíndice i identifica a la fuente de nitrógeno y toma los valores i = 1, 2, ..., 13. En la tabla de datos se ve que i = 1 cuando solo se usa xilosa, i = 2 para NH₄Cl, etc. hasta llegar a i = 13 cuando la fuente de nitrógeno es Licor de maíz macerado. El subíndice j identifica a las repeticiones y toma los valores 1 y 2. μ es la media general y fNi es el

efecto de la fuente nitrógeno i en la media de densidad celular. Las observaciones están dadas por:

$$dC_{i,j} = \mu_{dC/fNi,j} + \varepsilon_{i,j} = \mu + fN_i + \varepsilon_{i,j}; i = 1,2, \dots, 13 \text{ y } j = 1,2,$$

$dC_{i,j}$ es la variable dC para la fuente de nitrógeno i y la repetición j , y $\varepsilon_{i,j}$ es el error aleatorio asociado a la observación $dC_{i,j}$. $\varepsilon_{i,j}$ tiene distribución normal con media cero y varianza desconocida σ^2 .

El análisis de varianza para el modelo produjo la Tabla 9.2.

Tabla 9.2. Análisis de varianza para densidad celular con diferentes fuentes de nitrógeno

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Fuente de Nitrógeno	12	1.58768	0.132307	189.44	1.515E-12
Residual (Error)	13	0.00908	0.000698		

La hipótesis nula del análisis de varianza de que no hay efecto de las fuentes de nitrógeno en la media de la densidad celular es:

$$H_0: fN_1 = fN_2 = fN_3 = fN_4 = fN_5 = fN_6 = fN_7 = fN_8 = fN_9 = fN_{10} = fN_{11} = fN_{12} = fN_{13},$$

y a esta hipótesis nula le corresponde la hipótesis alternativa:

$$H_A: fN_i \neq fN_{i'} \text{ para al menos una pareja de subíndices } i \neq i'; i, i' = 1,2, \dots, 13$$

La hipótesis nula establece que los efectos de las fuentes de nitrógeno en la media de la densidad celular son todos iguales. La hipótesis alternativa niega a la hipótesis nula y establece que los efectos de las diferentes fuentes de nitrógeno no son todos iguales; dicho de otro modo: hay al menos una pareja de efectos de las fuentes de nitrógeno fN_i y $fN_{i'}$ que no son iguales.

La hipótesis nula se rechaza con un valor p de 1.515E-12, por lo que se concluye que al menos una pareja de fuentes de nitrógeno tiene efectos diferentes en la media de densidad celular.

Las medias de densidad celular para las fuentes de nitrógenos están a continuación (Tabla 9.3).

Tabla 9.3. Medias de densidad celular obtenidas con las diferentes fuentes de nitrógeno.

Fuente de N codificada	Fuente de Nitrógeno	Media de Densidad celular
1	Xilosa	0.2575
2	NH_4Cl	0.1900
3	$(\text{NH}_4)_2\text{SO}_4$	0.2470
4	$(\text{NH}_4)_2\text{HPO}_4$	0.4905
5	$(\text{NH}_4)_2\text{H}_2\text{PO}_4$	0.2680
6	$(\text{NH}_4)_2\text{S}_2\text{O}_8$	0.4630
7	NH_2CONH_2	0.1825
8	NaNO_3	0.3105
9	HONH_3Cl	0.2205
10	NH_4NO_3	0.0425
11	Peptona	0.6075
12	Extracto de levadura	0.4020
13	Licor de maíz macerado	1.0590

Para saber cuáles parejas de fuentes de nitrógeno tienen efectos diferentes se sugiere hacer la prueba de Tukey. Antes de mostrar los resultados es importante mencionar que se buscan las fuentes de nitrógeno que maximicen la densidad celular.

En la Tabla 9.4 se muestra parte de los resultados de las comparaciones múltiples de la prueba de Tukey, de hecho, sólo incluye las parejas de tratamientos cuyas medias de densidad celular son diferentes. Recuerde que Tukey usa un nivel de confianza adecuado para el conjunto de las 78 comparaciones de parejas de tratamientos, se le llama nivel de confianza a nivel de experimento.

Se puede ver que con 95% de confianza a nivel de experimento, la media de densidad celular de la fuente 1 (xilosa) es diferente de las medias de las fuentes de nitrógeno 4, 6, 10, 11, 12 y 13, etc. Es de notar que la fuente 13 y la fuente 11 que produjeron las medias más grandes, difieren entre sí y también de las otras fuentes de nitrógeno. Así, el lector puede corroborar que la media de densidad celular para la fuente de nitrógeno 13, que es licor de maíz macerado, es el máximo que el investigador estaba buscando.

Tabla 9.4. Resultados de la prueba de Tukey de comparaciones múltiples de la media densidad celular con diferentes fuentes de nitrógeno. Se presentan sólo las parejas significativas.

Fuentes de nitrógeno que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
4-1	0.2330	0.12787798	0.33812202	0.0000364
6-1	0.2055	0.10037798	0.31062202	0.0001415
10-1	-0.2150	-0.32012202	-0.10987798	0.0000874
11-1	0.3500	0.24487798	0.45512202	0.0000003
12-1	0.1445	0.03937798	0.24962202	0.0043462
13-1	0.8015	0.69637798	0.90662202	0.0000000
4-2	0.3005	0.19537798	0.40562202	0.0000020
6-2	0.2730	0.16787798	0.37812202	0.0000061
8-2	0.1205	0.01537798	0.22562202	0.0191379
10-2	-0.1475	-0.25262202	-0.04237798	0.0036258

11-2	0.4175	0.31237798	0.52262202	0.0000000
12-2	0.2120	0.10687798	0.31712202	0.0001016
13-2	0.8690	0.76387798	0.97412202	0.0000000
4-3	0.2435	0.13837798	0.34862202	0.0000223
6-3	0.2160	0.11087798	0.32112202	0.0000831
10-3	-0.2045	-0.30962202	-0.09937798	0.0001490
11-3	0.3605	0.25537798	0.46562202	0.0000002
12-3	0.1550	0.04987798	0.26012202	0.0023167
13-3	0.8120	0.70687798	0.91712202	0.0000000
5-4	-0.2225	-0.32762202	-0.11737798	0.0000603
7-4	-0.3080	-0.41312202	-0.20287798	0.0000015
8-4	-0.1800	-0.28512202	-0.07487798	0.0005524
9-4	-0.2700	-0.37512202	-0.16487798	0.0000069
10-4	-0.4480	-0.55312202	-0.34287798	0.0000000
11-4	0.1170	0.01187798	0.22212202	0.0238242
13-4	0.5685	0.46337798	0.67362202	0.0000000
6-5	0.1950	0.08987798	0.30012202	0.0002449
10-5	-0.2255	-0.33062202	-0.12037798	0.0000521
11-5	0.3395	0.23437798	0.44462202	0.0000005
12-5	0.1340	0.02887798	0.23912202	0.0082630
13-5	0.7910	0.68587798	0.89612202	0.0000000
7-6	-0.2805	-0.38562202	-0.17537798	0.0000044
8-6	-0.1525	-0.25762202	-0.04737798	0.0026875
9-6	-0.2425	-0.34762202	-0.13737798	0.0000233
10-6	-0.4205	-0.52562202	-0.31537798	0.0000000
11-6	0.1445	0.03937798	0.24962202	0.0043462
13-6	0.5960	0.49087798	0.70112202	0.0000000
10-7	-0.1400	-0.24512202	-0.03487798	0.0057155
11-7	0.4250	0.31987798	0.53012202	0.0000000
12-7	0.2195	0.11437798	0.32462202	0.0000699
13-7	0.8765	0.77137798	0.98162202	0.0000000
10-8	-0.2680	-0.37312202	-0.16287798	0.0000075
11-8	0.2970	0.19187798	0.40212202	0.0000023
13-8	0.7485	0.64337798	0.85362202	0.0000000
10-9	-0.1780	-0.28312202	-0.07287798	0.0006174
11-9	0.3870	0.28187798	0.49212202	0.0000001
12-9	0.1815	0.07637798	0.28662202	0.0005084
13-9	0.8385	0.73337798	0.94362202	0.0000000
11-10	0.5650	0.45987798	0.67012202	0.0000000
12-10	0.3595	0.25437798	0.46462202	0.0000002
13-10	1.0165	0.91137798	1.12162202	0.0000000
12-11	-0.2055	-0.31062202	-0.10037798	0.0001415
13-11	0.4515	0.34637798	0.55662202	0.0000000
13-12	0.6570	0.55187798	0.76212202	0.0000000

El resultado de la comparación de parejas de medias que produce la prueba de Tukey se presenta resumido en la Tabla 9.5. A la lista de las medias de los tratamientos en orden creciente, se anexa una columna con letras, en la que de acuerdo con la prueba de Tukey, los tratamientos que reciben la misma letra no difieren significativamente.

Tabla 9.5. Prueba de Tukey para la media de densidad celular con diferentes fuentes de nitrógeno. Medias de fuentes de nitrógeno con la misma letra no difieren con 5% de significación de Tukey

Tratamiento	Fuente de Nitrógeno	Media	
10	NH_4NO_3	0.0425	a
7	NH_2CONH_2	0.1825	b
2	NH_4Cl	0.1900	b
9	HONH_3Cl	0.2205	b d
3	$(\text{NH}_4)_2\text{SO}_4$	0.2470	b d
1	Xilosa	0.2575	b d
5	$(\text{NH}_4)_2\text{H}_2\text{PO}_4$	0.2680	b d
8	NaNO_3	0.3105	d e
12	Extracto de levadura	0.4020	e f
6	$(\text{NH}_4)_2\text{S}_2\text{O}_8$	0.4630	f
4	$(\text{NH}_4)_2\text{HPO}_4$	0.4905	f
11	Peptona	0.6075	i
13	Licor de maíz macerado	1.0590	j

La gráfica de cajas (Figura 9.1) ayuda a entender los resultados de la prueba de Tukey. En la figura 9.1 se ve que la fuente de nitrógeno 13 (Licor de maíz macerado) tiene la mayor mediana de densidad celular y su caja está más arriba que las de las demás fuentes de nitrógeno.

La segunda mayor media corresponde a la peptona (etiqueta 11) y en la tabla de Tukey se ve que difiere de las otras doce fuentes de variación. Asimismo, no se ven ejes saliendo de las cajas, esto se debe a que cada caja se forma con solo dos observaciones.

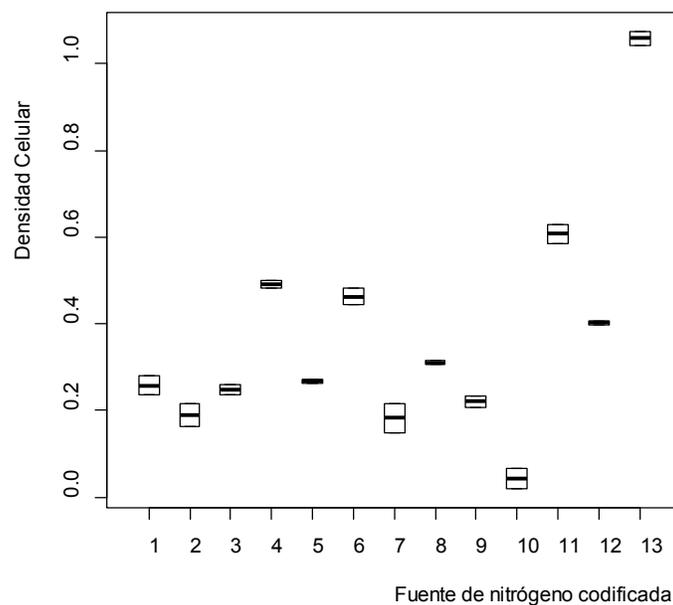


Figura 9.1. Gráfica de cajas para densidad celular.

De la misma forma en que se hizo el análisis de densidad celular se hace el análisis para la actividad enzimática de xilanas. Como en el caso de densidad celular, se tiene un diseño completamente al azar con 13 tratamientos y dos repeticiones. Se busca la o las fuentes de nitrógeno que produzcan las mayores medias de actividad enzimática de xilanas. Se plantea el modelo:

$$\mu_{x75/fNi,j} = \mu + fN_i, i = 1,2, \dots,13 \text{ y } j = 1,2$$

El término $x75$ representa la actividad enzimática de xilanas pH 7.5 medida, fN representa al factor Fuente de nitrógeno que tiene 13 niveles (puede usarse el nombre del tratamiento o el valor codificado que es más apropiado para los paquetes), $\mu_{x75/fNi,j}$ es la media de actividad enzimática de xilanas para la iésima fuente de nitrógeno fNi en la repetición j . En la tabla de datos (Tabla 9.1) se ve la correspondencia de los nombres de las fuentes de nitrógeno con sus valores codificados, así $i = 1$ cuando la fuente de nitrógeno es xilosa, $i = 2$ para NH_4Cl , etc. hasta llegar a $i = 13$ cuando la fuente de nitrógeno es Licor de maíz macerado. El subíndice $j = 1,2$ identifica a las repeticiones, μ es la media general y fNi es el efecto de la fuente nitrógeno i en la media de producción de xilanas pH 7.5. La variable aleatoria para la actividad enzimática de xilanas tiene media que depende del nivel de la fuente de nitrógeno, entonces:

$$x75_{i,j} = \mu_{x75/fNi,j} + \varepsilon_{i,j} = \mu + fN_i + \varepsilon_{i,j}; i = 1, 2, \dots,13 \text{ y } j = 1,2$$

$x75_{i,j}$ es la variable aleatoria $x75$ cuando la fuente de nitrógeno es el nivel i y la repetición es la j , y $\varepsilon_{i,j}$ es el error aleatorio asociado a la variable aleatoria $x75_{i,j}$.

El análisis de varianza para el modelo de actividad enzimática de xilanas produjo la información de la Tabla 9.6. Este análisis, con la F y el valor p , permite probar la hipótesis de que no hay efecto de las fuentes de nitrógeno en la actividad enzimática de xilanas. Note que el efecto se representa igual que en el caso de densidad celular, pues se refiere a las mismas fuentes de nitrógeno.

$$H_0: fN_1 = fN_2 = fN_3 = fN_4 = fN_5 = fN_6 = fN_7 = fN_8 = fN_9 = fN_{10} = fN_{11} = fN_{12} = fN_{13}$$

Se puede decir que la hipótesis nula establece que los efectos de las fuentes de nitrógeno en la media de actividad enzimática de xilanas son iguales. La hipótesis alternativa niega a la hipótesis nula y establece que los efectos de las diferentes fuentes de nitrógeno en la media de actividad enzimática de xilanas no son todos iguales; se escribe: H_A : Al menos una pareja de efectos de las fuentes de nitrógeno fNi y fNi' no son iguales.

Como se muestra en la Tabla 9.6, se rechaza la hipótesis nula con un valor p de $5.884E-16$ y se concluye que al menos una pareja de fuentes de nitrógeno tiene efectos diferentes en la actividad enzimática de xilanas.

Tabla 9.6. Análisis de varianza para actividad enzimática de xilanas con diferentes fuentes de nitrógeno.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Fuente de Nitrógeno	12	1582646	131887	638.42	$5.884E^{-16}$
Residual (Error)	13	2686	207		

Las medias de actividad enzimática de xilanas para las fuentes de nitrógeno están en la Tabla 9.7. En ésta se puede observar que la mayor media de actividad enzimática de xilanas (con valor 1095.7) corresponde a la fuente 13 y la segunda mayor a la fuente 4.

Tabla 9.7. Medias de actividad enzimática de xilanas obtenidas con las diferentes fuentes de nitrógeno.

Fuente de N codificada	Fuente de Nitrógeno	Media de actividad enzimática de xilanas (UI)
1	Xilosa	366.3
2	NH ₄ Cl	291.1
3	(NH ₄) ₂ SO ₄	251.5
4	(NH ₄) ₂ HPO ₄	563.9
5	(NH ₄)H ₂ PO ₄	319.6
6	(NH ₄) ₂ S ₂ O ₈	183.1
7	NH ₂ CONH ₂	151.5
8	NaNO ₃	433.6
9	HONH ₃ Cl	191.0
10	NH ₄ NO ₃	111.7
11	Peptona	358.2
12	Extracto de levadura	213.2
13	Licor de maíz macerado	1095.7

Para saber cuáles parejas de fuentes de nitrógeno tienen efectos diferentes en la actividad enzimática de xilanas se realizó la prueba de Tukey. En la Tabla 9.8 se muestra la parte del resultado de la prueba de Tukey que incluye sólo a las parejas de tratamientos que son significativamente diferentes. Recuerde que Tukey usa un nivel de confianza adecuado para el conjunto de las 78 comparaciones de parejas de tratamientos, se le llama nivel de confianza a nivel de experimento. En las primeras filas de la Tabla 9.8 se puede observar que la media de actividad enzimática de xilanas de la fuente 1 (xilosa) es diferente de las medias obtenidas con las demás fuentes de nitrógeno excepto la 5 y la 11.

Debido a que se tiene interés en la mayor actividad enzimática, el investigador puede comenzar revisando las diferencias de la fuente de nitrógeno con la mayor media de actividad enzimática de xilanas, que es la 13 –licor de maíz macerado–, la cual difiere de las otras doce fuentes de nitrógeno, luego seguir con las diferencias de la fuente de nitrógeno con segunda mayor media, es la 4 y corresponde a (NH₄)₂HPO₄, la cual también difiere de las otras 12 fuentes de nitrógeno restantes.

Tabla 9.8. Resultados de la prueba de Tukey de comparaciones múltiples de la media de actividad de xilanas con diferentes fuentes de nitrógeno. Se muestran sólo las parejas significativas.

Fuentes de nitrógeno que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
2-1	-75.20	-132.371682	-18.028318	0.0063537
3-1	-114.80	-171.971682	-57.628318	0.0001064
4-1	197.65	140.478318	254.821682	0.0000002
6-1	-183.20	-240.371682	-126.028318	0.0000005
7-1	-214.80	-271.971682	-157.628318	0.0000001
8-1	67.35	10.178318	124.521682	0.0155354
9-1	-175.25	-232.421682	-118.078318	0.0000009

10-1	-254.60	-311.771682	-197.428318	0.0000000
12-1	-153.05	-210.221682	-95.878318	0.0000043
13-1	729.45	672.278318	786.621682	0.0000000
4-2	272.85	215.678318	330.021682	0.0000000
6-2	-108.00	-165.171682	-50.828318	0.0002027
7-2	-139.60	-196.771682	-82.428318	0.0000123
8-2	142.55	85.378318	199.721682	0.0000097
9-2	-100.05	-157.221682	-42.878318	0.0004441
10-2	-179.40	-236.571682	-122.228318	0.0000007
11-2	67.15	9.978318	124.321682	0.0158962
12-2	-77.85	-135.021682	-20.678318	0.0047191
13-2	804.65	747.478318	861.821682	0.0000000
4-3	312.45	255.278318	369.621682	0.0000000
5-3	68.15	10.978318	125.321682	0.0141729
6-3	-68.40	-125.571682	-11.228318	0.0137725
7-3	-100.00	-157.171682	-42.828318	0.0004463
8-3	182.15	124.978318	239.321682	0.0000006
9-3	-60.45	-117.621682	-3.278318	0.0343521
10-3	-139.80	-196.971682	-82.628318	0.0000121
11-3	106.75	49.578318	163.921682	0.0002288
13-3	844.25	787.078318	901.421682	0.0000000
5-4	-244.30	-301.471682	-187.128318	0.0000000
6-4	-380.85	-438.021682	-323.678318	0.0000000
7-4	-412.45	-469.621682	-355.278318	0.0000000
8-4	-130.30	-187.471682	-73.128318	0.0000267
9-4	-372.90	-430.071682	-315.728318	0.0000000
10-4	-452.25	-509.421682	-395.078318	0.0000000
11-4	-205.70	-262.871682	-148.528318	0.0000001
12-4	-350.70	-407.871682	-293.528318	0.0000000
13-4	531.80	474.628318	588.971682	0.0000000
6-5	-136.55	-193.721682	-79.378318	0.0000158
7-5	-168.15	-225.321682	-110.978318	0.0000014

8-5	114.00	56.828318	171.171682	0.0001147
9-5	-128.60	-185.771682	-71.428318	0.0000309
10-5	-207.95	-265.121682	-150.778318	0.0000001
12-5	-106.40	-163.571682	-49.228318	0.0002367
13-5	776.10	718.928318	833.271682	0.0000000
8-6	250.55	193.378318	307.721682	0.0000000
10-6	-71.40	-128.571682	-14.228318	0.0097746
11-6	175.15	117.978318	232.321682	0.0000009
13-6	912.65	855.478318	969.821682	0.0000000
8-7	282.15	224.978318	339.321682	0.0000000
11-7	206.75	149.578318	263.921682	0.0000001
12-7	61.75	4.578318	118.921682	0.0295839
13-7	944.25	887.078318	1001.421682	0.0000000
9-8	-242.60	-299.771682	-185.428318	0.0000000
10-8	-321.95	-379.121682	-264.778318	0.0000000
11-8	-75.40	-132.571682	-18.228318	0.0062121
12-8	-220.40	-277.571682	-163.228318	0.0000001
13-8	662.10	604.928318	719.271682	0.0000000
10-9	-79.35	-136.521682	-22.178318	0.0039928
11-9	167.20	110.028318	224.371682	0.0000015
13-9	904.70	847.528318	961.871682	0.0000000
11-10	246.55	189.378318	303.721682	0.0000000
12-10	101.55	44.378318	158.721682	0.0003820
13-10	984.05	926.878318	1041.221682	0.0000000
12-11	-145.00	-202.171682	-87.828318	0.0000080
13-11	737.50	680.328318	794.671682	0.0000000
13-12	882.50	825.328318	939.671682	0.0000000

El resultado de la revisión de diferencias en la tabla anterior (9.8) se resume en la Tabla 9.9, en la que se muestran las medias de los tratamientos en orden creciente y que une con una misma letra a las medias que no difieren en la comparación múltiple de Tukey.

Tabla 9.9. Prueba de Tukey para la media de densidad celular con diferentes fuentes de nitrógeno. Medias de fuentes de nitrógeno con la misma letra no difieren con 5% de significación de Tukey

Fuente de N codificada	Fuente de Nitrógeno	Media de actividad enzimática de xilanasas (UI)	
10	NH_4NO_3	111.7	a
7	NH_2CONH_2	151.5	a b
6	$(\text{NH}_4)_2\text{S}_2\text{O}_8$	183.1	b c
9	HONH_3Cl	191	b c
12	Extracto de levadura	213.2	c d
3	$(\text{NH}_4)_2\text{SO}_4$	251.5	d e
2	NH_4Cl	291.1	e f
5	$(\text{NH}_4)_2\text{H}_2\text{PO}_4$	319.6	f g
11	Peptona	358.2	g
1	Xilosa	366.3	g
8	NaNO_3	433.6	h
4	$(\text{NH}_4)_2\text{HPO}_4$	563.9	i
13	Licor de maíz macerado	1095.7	j

El diagrama de cajas (Figura 9.2) permite identificar las fuentes de nitrógeno que interesan, que son la 4 y la 13.

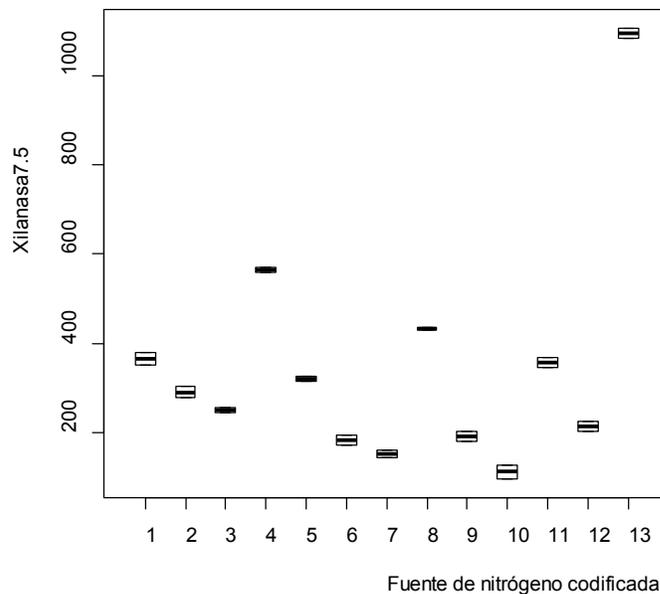


Figura 9.2. Gráfica de cajas de las medias de actividad enzimática de xilanasas.

Los resultados para densidad celular y actividad enzimática de xilanasas coinciden en la fuente de nitrógeno que tiene el mayor efecto positivo en la media, en ambos casos es licor de maíz macerado (etiqueta 13), pero no coinciden en la fuente de nitrógeno que tiene el segundo mejor efecto positivo en la media, que para densidad celular es la peptona y para actividad enzimática de xilanasas pH 7.5 es $(\text{NH}_4)_2\text{HPO}_4$ (etiqueta 4). Esto indica que las variables, aunque tienen comportamiento parecido, no miden lo mismo.

El siguiente diagrama de dispersión muestra la relación entre densidad celular y actividad enzimática de xilanasa (Figura 9.3). Se colocó a la densidad celular en el eje de abscisas (horizontal) y a la actividad enzimática de la xilanasa en el eje de ordenadas (vertical) porque la actividad enzimática de la xilanasa depende de la densidad celular. Se observa que al aumentar de valor la densidad celular aumenta la actividad enzimática de la xilanasa, pero hay dispersión de los valores de actividad enzimática de xilanasa para valores de 0.18 a 0.7 de densidad celular. Esto se deberá de manifestar en los resultados de la regresión lineal.

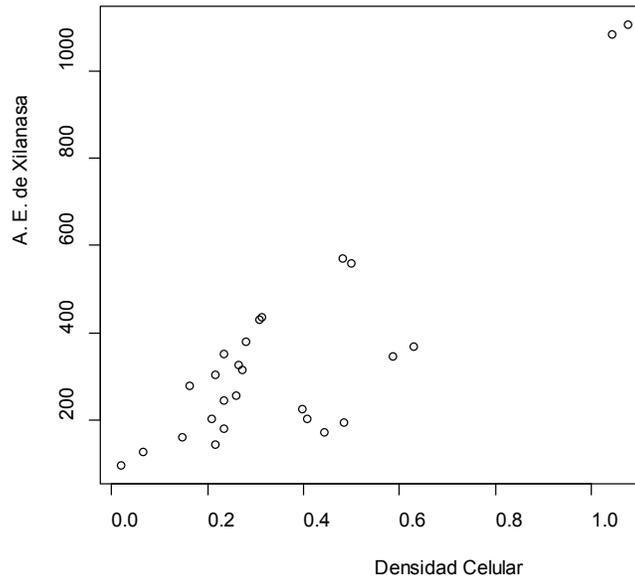


Figura 9.3. Diagrama de dispersión de la relación de densidad celular y actividad enzimática de xilanasa.

Se plantea el modelo de regresión:

$$\mu_{x75/dC} = \beta_0 + \beta_{dC}dC$$

En la ecuación, $\mu_{x75/dC}$ es la media de actividad enzimática de xilanasa para un valor específico de densidad celular (dC), β_0 es la ordenada al origen o constante, β_{dC} es el coeficiente de regresión que corresponde a la variable dC . La distribución de la variable $x75dC$ se supone normal y se tiene:

$$x75_{dC} \sim N(\mu_{x75/dC} = \beta_0 + \beta_{dC}dC, \sigma_{x75/dC}^2)$$

Se acostumbra, por razones de desarrollo histórico escribir a la variable aleatoria $x75$ como su media mas una variable aleatoria ε , que se conoce como error:

$$x75_{dC} = \mu_{x75/dC} + \varepsilon = \beta_0 + \beta_{dC}dC + \varepsilon$$

El error ε se supone tiene distribución normal con media cero y varianza desconocida $\sigma_{x75/dC}^2$.

El modelo produce la relación de estimadores, que son variables aleatorias cuyo valor depende de la muestra que se seleccione: $\hat{\mu}_{x75/dC} = \hat{\beta}_0 + \hat{\beta}_{dC}dC$.

Ya con los valores obtenidos de los datos se obtienen los valores de los coeficientes de regresión que se muestran en la Tabla 9.10.

Tabla 9.10. Estimación de la regresión de actividad enzimática de la xilanasa sobre densidad celular (dC)

	Estimación	Error estándar	t calculada	valor p
β_0	37.54	46.40	0.809	0.426
β_{dC}	852.68	105.23	8.103	2.52e-08
Rcuadrada	0.7323		GL residual	24

Se escribe la ecuación de la media estimada:

$$\mu'_{x75/dC} = 37.54 + 852.68 \text{ dC}.$$

En esta ecuación se usa ' en lugar de ^ para enfatizar que $\mu'_{x75/dC}$ no es una variable aleatoria, sino los valores estimados para la media $\mu_{x75/dC}$ usando las ecuaciones que se obtuvieron para los estimadores.

La prueba usual de que el coeficiente de regresión β_{dC} vale cero se rechaza con valor p de 2.52E-08 (Tabla 9.10), pero la hipótesis que interesa probar es la de que β_{dC} vale 1000 (Por cada unidad de densidad celular la xilanasa 75 crece 1000 unidades), se plantea la pareja de hipótesis $H_0: \beta_{dC} = 1000$,

$H_A: \beta_{dC} \neq 1000$, note que no se usa uno sino mil porque dC y x75 se miden en diferentes escalas. La t calculada usando

$$t_c = \frac{852.88 - 1000}{105.23} = -1.3999.$$

Para $t_c = -1.3999$ se obtiene un valor p de 0.17433. Como 0.17433 es mayor que 0.05, la hipótesis nula no se rechaza, esto indica, para fines prácticos que, por cada unidad de densidad celular, la actividad enzimática de xilanasa crece 1000 unidades.

A pesar de que se manifiesta la relación de crecimiento esperado, la variación alrededor de la línea estimada es muy grande: $R^2 = 0.7323$ (Tabla 9.8), esto es, la regresión sólo explica el 73.23% de la variación. Se puede concluir que no conviene usar una de las variables para explicar el comportamiento de la otra. La variación se ve en la Figura 9.4, donde se grafican los valores de los residuales contra los valores predichos.

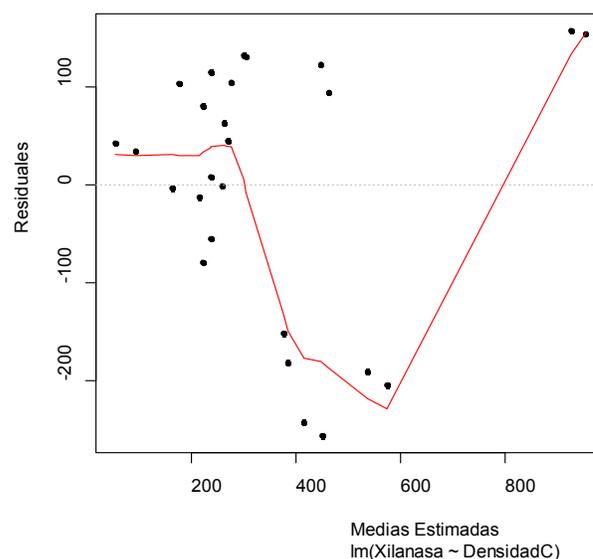


Figura 9.4. Valores de los residuales contra los valores predichos.

También es posible que los datos no sigan la distribución normal. De hecho, se confirma en la gráfica de probabilidad normal (Figura 9.5), pues los puntos presentan curvaturas, una hacia abajo de la línea y otra hacia arriba de la línea.

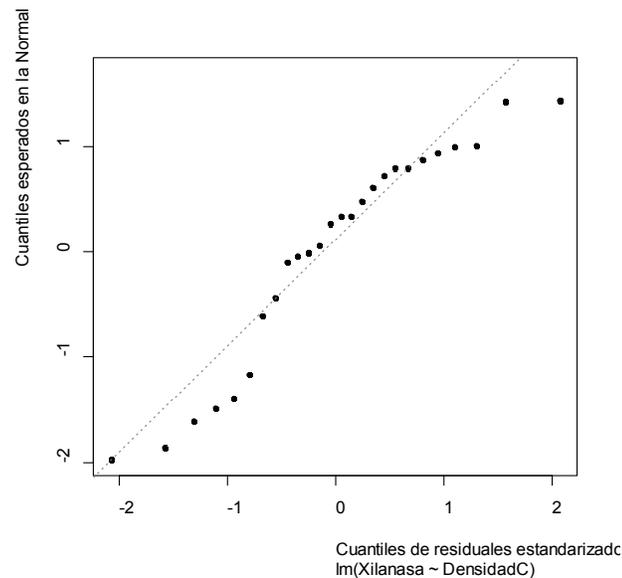


Figura 9.5. Gráfica de distribución de probabilidad normal.

La prueba de normalidad de Shapiro-Wilk realizada en los residuales, que plantea H_0 : Los residuales tienen distribución normal; H_A : Los residuales no tienen distribución normal, produce un valor p de 0.01112, por lo que la hipótesis de distribución normal se rechaza. Los valores de Densidad Celular y de Actividad Enzimática de Xilanasa que corresponden al tratamiento de licor de maíz macerado están muy alejados del resto de las observaciones, como se observó desde el inicio en la Tabla 9.1 y, más adelante en los puntos ubicados en la extrema derecha de las figuras 9.3 y 9.5. Debido a esto, se sugiere realizar un análisis sin considerar este tratamiento.

El análisis de este problema se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares.

Guía para resolver con R

```
datosXilanasa<-read.table(file = "xilanasa.csv",header = T,sep = ",")

# Se lee el archivo de Excel con tipo csv y se almacena en forma de tabla de R en datosXilanasa

# No olvide cambiar directorio a la carpeta donde está su archivo de datos

datosXilanasa

fN<-datosXilanasa$fN          # Se definen los vectores de observaciones de las variables usando $

fNcod<-datosXilanasa$fNcod

dC<-datosXilanasa$dC

x75<-datosXilanasa$x7_5

# fN # fNcod # dC #x75
```

```

baseXilanasa<-data.frame(DensidadC = dC,Xilanasa = x75,FuenteNitrógeno = factor(fN),FuenteN = factor(fNcod))

# Se define el marco baseXilanasa con las variables respuesta y los factores fN original y codificado

baseXilanasa

attach(baseXilanasa)      # Paso conveniente para usar aov

x75_fN_AdeV<-aov(DensidadC~FuenteN, data = baseXilanasa) # Completamente al azar con el factor FuenteN

# Se llama x75_fN_AdeV al análisis aunque se refiere a densidad celular

# esto permite usar la misma nomenclatura para las dos variables respuesta

replications(DensidadC~FuenteN,data = baseXilanasa,na.fail)      # Para comprobar que es diseño balanceado

summary(x75_fN_AdeV) # El resumen del análisis en la consola de R

# x75_fN_AdeV

list(x75_fN_AdeV)          # Para el error estándar de la diferencia de medias

medias<-model.tables(x75_fN_AdeV,type = "means",se = FALSE)

medias                    # Las medias de la variable respuesta para cada nivel del factor

list(FuenteNitrógeno)     # La identificación de los tratamientos

TukeyHSD(x75_fN_AdeV, which = "FuenteN",orderer = TRUE, conf.level = 0.95)

# Prueba de Tukey para las parejas de niveles del factor con confianza a nivel de experimento

# plot(TukeyHSD(x75_fN_AdeV,"FuenteN")) # Gráfica con los intervalos de confianza(Tukey) de las diferencias

## library(gplots)      # Se carga el paquete gplots (library(gplots), se necesita para la gráfica de medias

## plotmeans(DensidadC~FuenteN,data = baseXilanasa,n.label = FALSE,bars = TRUE,connect = FALSE,

## main = "MEDIAS CON 95% IC")

## Gráfica de medias de variable respuesta para los niveles del factor FuenteN

boxplot(DensidadC~FuenteN,data = baseXilanasa,outline = TRUE,beside = TRUE,boxwex = .5,

xlab = "Fuente de nitrógeno codificada",ylab = "Densidad Celular",main = " ")

# Gráfica de cajas de la variable respuesta para los niveles del factor. boxwex = .5 define el ancho de la caja

xilanasa75AdeV<-aov(x75~FuenteN)

# Análisis como el de arriba para la variable respuesta Xilanasa7.5. Cambie DensidadC por Xilanasa7.5

replications(Xilanasa7.5~FuenteN,data = baseXilanasa,na.fail)

```

```

summary(xilanasa75AdeV)

xilanasa75AdeV

list(xilanasa75AdeV)

medias<-model.tables(xilanasa75AdeV,type = "means",se = FALSE)

medias

list(fn)

TukeyHSD(xilanasa75AdeV, which = "FuenteN",orderer = TRUE, conf.level = 0.95)

# plot(TukeyHSD(xilanasa75AdeV,"FuenteN"))

### library(gplots)

## plotmeans(x75~FuenteN,data = baseXilanasa,n.label = FALSE,connect = FALSE,
## main = "MEDIAS CON 95% IC",xlab = "Fuente de Nitrógeno",ylab = "Xilanasa con pH = 7.5")

boxplot(x75~FuenteN,data = baseXilanasa,outline = TRUE,beside = TRUE,boxwex = .5,
xlab = "Fuente de nitrógeno codificada",ylab = "Xilanasa7.5",main = " ")

plot(Xilanasa~DensidadC,data = baseXilanasa,xlab = "Densidad Celular",
ylab = "A. E. de Xilanasa",main = " ")

    # Diagrama de dispersión de Xilanasa sobre Densidad

Xilanasa7.5_DensidadC<- lm(Xilanasa~DensidadC,data = baseXilanasa)

# regresión lineal simple de Xilanasa sobre (~) DensidadC

Xilanasa7.5_DensidadC # Reporta algunos valores de interés

summary(Xilanasa7.5_DensidadC) # Resumen del análisis de regresión

anova(Xilanasa7.5_DensidadC) # Presenta el análisis de varianza de la regresión

valort<-(852.68-1000)/105.2343

valort

# Se calcula la t para la prueba de coeficiente = 1000, datos en consola de R

valorp<- 2*(1-pt(abs(valort), 24, 0, lower.tail = TRUE, log.p = FALSE)) # Se calcula el valor p

# pt() da la probabilidad de la t. Se usa el valor absoluto de valort (abs(valort)),

# la t tiene 24 grados de libertad , 0 es el parámetro de no centralidad,

```

```

# probabilidad de valores menores que abs(valort) (lower.tail = TRUE) y no se usa log.p (log.p = FALSE)

valorp

# par(mfrow = c(2,2))    # Cuatro gráficas por hoja

# plot(Xilanasa7.5_DensidadC)    # Da las gráficas de residuales (en inglés)

par(mfrow = c(1,1))          # Una gráfica por hoja

plot(Xilanasa7.5_DensidadC,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Medias Estimadas",ylab = "Residuales",col.lab = 1)

# Gráfica de residuales sobre medias estimadas (predichos) con wich = 1

plot(Xilanasa7.5_DensidadC,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Cuantiles de residuales estandarizados",

ylab = "Cuantiles esperados en la Normal",col.lab = 1)

# Gráfica de Prob. Normal. Cuantiles esperados de Z sobre Cuantiles de residuales estandarizados con wich = 2

errores<-residuals(Xilanasa7.5_DensidadC)

shapiro.test(errores)
    
```

9.1. Ejercicio

Se hizo un experimento utilizando *Staphylococcus sp.* para producir xilanasa variando las fuentes de nitrógeno: el factor fuentes de nitrógeno tuvo trece niveles. Se tuvieron dos repeticiones, la unidad experimental fue un matraz Erlenmeyer de 250 ml con 100 ml de medio líquido estándar, se agitó a 200 rpm y se mantuvo a 37°C con un pH inicial de 8.0 durante tres días. Posteriormente se determinó la actividad enzimática de xilanasa a un pH de 9.2 (UI) y el contenido de proteína. Interesa probar el efecto de las fuentes de nitrógeno sobre la actividad enzimática de xilanasa y conocer el efecto de las fuentes de nitrógeno en la media de contenido de proteína. Los datos en la tabla siguiente incluyen la columna de Fuente de nitrógeno codificada, que se usa en los análisis.

Fuente de nitrógeno 0.5% w/v	Fuente de nitrógeno codificada	Xilanasas pH 9.2 UI ⁻¹		Contenido de proteína mg/ml	
Xilosa	1	376.5	374.5	376.5	374.5
NH ₄ Cl	2	305.9	314.2	305.9	314.2
(NH ₄) ₂ SO ₄	3	263.3	266.8	263.3	266.8
(NH ₄) ₂ HPO ₄	4	628.9	621.8	628.9	621.8
(NH ₄)H ₂ PO ₄	5	299.4	252.5	299.4	252.5
(NH ₄) ₂ S ₂ O ₈	6	144.7	192.4	144.7	192.4
NH ₂ CONH ₂	7	145.8	95.7	145.8	95.7
NaNO ₃	8	404.6	417.2	404.6	417.2
HONH ₃ Cl	9	182.1	167	182.1	167
NH ₄ NO ₃	10	120.1	95.3	120.1	95.3
Peptona	11	373.9	349.9	373.9	349.9
Extracto de levadura	12	223.5	226.4	223.5	226.4
Licor de maíz macerado	13	1186.6	1164.1	1186.6	1164.1

Haga un análisis como el del problema con densidad celular y actividad enzimática de xilanasas a un pH 7.5 eliminando los datos del tratamiento de licor de maíz macerado.

Bibliografía

-  Grupa, S., Kunhad, R.C., Bhusmhan, B. and Hoondal, G.S. (2001) Improved xylanase production from a haloalkalophilic *Staphylococcus* sp. SG-13 using inexpensive agricultural residues. *World Journal of Microbiology and Biotechnology*. 17:5-8.

Nota: La referencia se utilizó, de manera fraccionada, para plantear problemas similares pero llevados al nivel y el interés del curso.

10. Producción de biomasa y esporas de *Idriella bolleyi*

El hongo *Idriella bolleyi* se puede usar en la raíz de los cereales como agente de control biológico de patógenos. Se hizo un experimento para probar el efecto de la adición de suplementos de vitaminas a un medio de cultivo mínimo sobre las variables: crecimiento de biomasa en peso seco (mg/100 ml) y producción de esporas (número de esporas/ml) del hongo *Idriella bolleyi*. Por facilidad de presentación se usa el número de esporas multiplicado por 10^4 . La línea de *I. bolleyi* que se usó provenía de cebada cultivada en Rothamsted, Inglaterra, se obtuvo seca y congelada y se multiplicó en agar de papa dextrosa. Se usaron matraces Erlenmeyer de 250 ml conteniendo un medio básico con 10 g/l de glucosa y las mediciones se hicieron después de siete días.

Se probaron tres tratamientos con tres repeticiones cada uno: adición de Tiamina (T/0.4m g/l), adición de biotina (B0.005 mg/l) y adición de ambos tiamina y biotina. El medio en los matraces no se agitó. Los datos se muestran en la Tabla 10.1.

Tabla 10.1. Biomasa y número de esporas obtenidas.

Tiamina	Biotina	Peso seco mg/100 ml	# esporasx10 ⁴ /ml
0	1	321.66	49.78
0	1	287.71	64.03
0	1	364.02	39.75
1	0	34.46	0.07
1	0	16.99	0.24
1	0	45.35	0.04
1	1	308.97	27.84
1	1	303.85	66.29
1	1	315.12	39.3

Se tiene un diseño completamente al azar con tres repeticiones y tres niveles de tratamientos: biotina, tiamina y ambos, biotina y tiamina. En cuanto a los factores biotina y tiamina, se parece a un diseño factorial de tratamientos 2^2 , pero no lo es porque no se ensayó y por lo tanto no se probó el tratamiento sin biotina y sin tiamina. Así que no se puede analizar como factorial sino como un diseño con un factor con tres niveles o tratamientos: biotina, tiamina y biotina+tiamina. El modelo correspondiente es:

$$\mu_{\text{pesoS}/\text{trati},j} = \mu + \text{trat}_i, \text{ con } \text{pesoS}_{i,j} = \mu_{\text{pesoS}/\text{trati},j} + \varepsilon_{i,j} = \mu + \text{trat}_i + \varepsilon_{i,j} \text{ con } \varepsilon_{i,j} \sim N(0, \sigma^2).$$

Los símbolos usados en las ecuaciones tienen los significados siguientes: $\mu_{\text{pesoS}/\text{trati},j}$ es la media del peso seco de biomasa (pesoS) para el nivel i de tratamientos y la repetición j , μ es la media general, trat_i es el efecto del tratamiento i de vitaminas, $\text{pesoS}_{i,j}$ es la variable aleatoria de peso seco que corresponde al nivel i de tratamiento y repetición j , y $\varepsilon_{i,j}$ es el término aleatorio o error que hace que $\text{pesoS}_{i,j}$ sea variable aleatoria. $\varepsilon_{i,j}$ se distribuye normal con media cero y varianza desconocida. El subíndice i identifica al tratamiento, son tres, y el subíndice j identifica a la repetición, que también son tres.

Se plantea la pareja de hipótesis $H_0: \text{trat}_1 = \text{trat}_2 = \text{trat}_3$, H_A : Al menos una pareja de efectos de tratamientos en la media de peso seco es diferente. En este problema con tan solo tres niveles de tratamientos se puede escribir $H_A: \text{trat}_1 \neq \text{trat}_2$ ó $\text{trat}_1 \neq \text{trat}_3$ ó $\text{trat}_2 \neq \text{trat}_3$.

Del análisis de varianza (Tabla 10.2) se obtiene el valor $p = 8.603E-06$ (basado en F) para la prueba de las hipótesis nula y alternativa sobre tratamientos. Se rechaza la hipótesis nula de igualdad de efectos, por lo que conviene realizar una prueba de Tukey para conocer las parejas de efectos de tratamiento que son diferentes.

Tabla 10.2. Análisis de varianza para la producción de biomasa de *Idriella bolleyi*.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Tratamiento	2	162363	81182	143.41	8.603E-06
Error	6	3396	566		

Los datos para biotina+ fiamina y tiamina+ biotina, dan las medidas e 324.5, 32.3 y 309.3 respectivamente. A reserva de hacer la prueba de Tukey, parece que el efecto de tiamina sola (tratamiento 2) produce una media más pequeña que la de los otros dos tratamientos.

En la Tabla 10.3 se muestran las comparaciones realizadas con Tukey. Las comparaciones con el tratamiento 2 (tiamina) dan evidencia para rechazar la igualdad de efectos (valores p de 0.0000131 y 0.0000181), pero la comparación de los tratamientos 1 y 3 (biotina vs biotina+tiamina) no es significativa ($p = 0.7280352$), estos dos tratamientos son los que dan mayores medias y no difieren entre sí. El efecto del tratamiento 2 = tiamina es diferente de los efectos de los otros dos tratamientos, que no difieren entre sí.

Tabla 10.3. Parejas de efectos de tratamientos (1 = biotina, 2 = tiamina, 3 = biotina+tiamina) en la media de peso seco

Efectos de tratamientos que se comparan	Diferencia de medias	Límite inferior del intervalo para la diferencia	Límite superior del intervalo para la diferencia	Valor p (Tukey)
2-1	-292.1967	-351.80197	-232.59136	0.0000131
3-1	-15.1500	-74.75531	44.45531	0.7280352
3-2	277.0467	217.44136	336.65197	0.0000181

Por el costo de adicionar al medio de cultivo dos vitaminas con respecto a solo agregar una, se prefiere el tratamiento de biotina. La gráfica de cajas (Figura 10.1) muestra que el tratamiento de biotina+tiamina presenta menos variación, esto puede ser importante para el investigador.

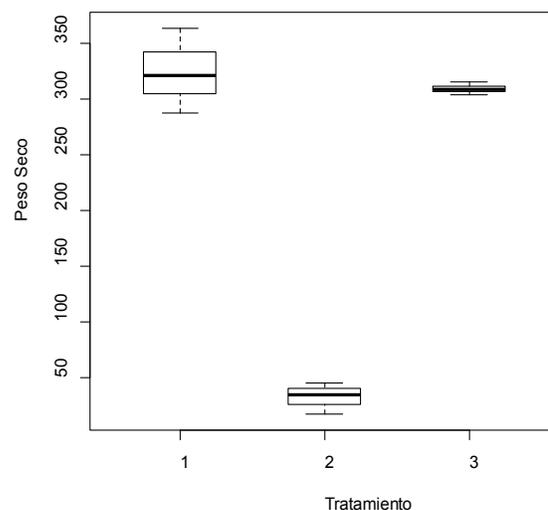


Figura 10.1. Gráfica de cajas para la producción de biomasa.

La relación de la producción de peso seco y la de número de esporas se pueden analizar usando regresión. El diagrama de dispersión se muestra en la Figura 10.2.

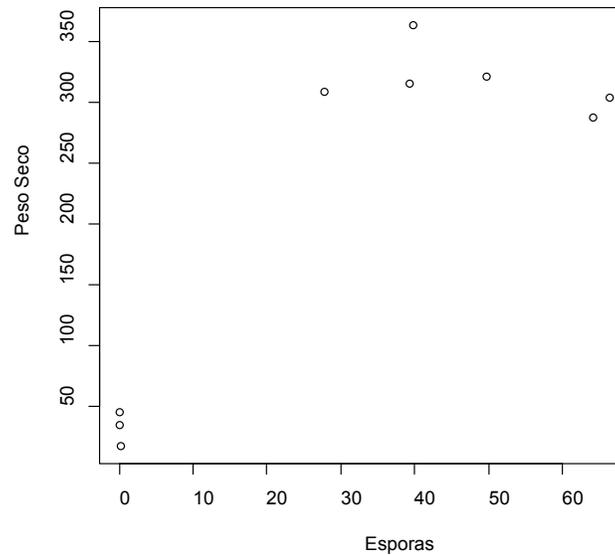


Figura 10.2. Diagrama de dispersión de pesos seco y número de esporas.

Inicialmente se plantea el modelo de regresión lineal:

$$\mu_{pS/Esp} = \beta_0 + \beta_{Esp} Esp, \text{ con:}$$

$$pS_{Esp} = \mu_{pS/Esp} + \varepsilon = \beta_0 + \beta_{Esp} Esp + \varepsilon \text{ y } \sigma_{pS/Esp}^2$$

los significados de los símbolos en las ecuaciones son los usuales. El subíndice de μ especifica la variable a la que se refiere la media y la variable que se usa para explicar los cambios en la media. pS representa peso seco y Esp a esporas, β_0 es la ordenada al origen o constante, β_{Esp} es el coeficiente de regresión de Esp y ε es el término aleatorio o error que se distribuye normal con media 0 y varianza σ^2 . La ecuación con los estimadores utiliza el acento circunflejo $\hat{}$ sobre los parámetros para indicar que son estimadores: $\hat{\mu}_{pS/Esp} = \hat{\beta}_0 + \hat{\beta}_{Esp} Esp$.

La estimación está dada por el resultado de la regresión lineal

$$\hat{\mu}'_{pS/Esp} = 74.953 + 4.606Esp,$$

se usa $\hat{\mu}'_{pS/Esp}$ para indicar que no es variable aleatoria sino la media estimada para cada valor de Esp .

La prueba de la hipótesis de que el coeficiente de regresión β_{Esp} vale cero se rechaza con un valor $p = 0.00338$. La regresión explica el 72.94% de la variación ($R^2 = 0.7294$). En el diagrama de dispersión se ve que hay una curvatura en la tendencia de los datos, los valores de peso seco suben rápidamente y luego se estabilizan y bajan. Los residuales de la regresión lineal sugieren que se haga la estimación usando el modelo cuadrático.

Se plantea el modelo para polinomios ortogonales:

$$\mu_{pS/or1,or2} = \beta_0 + \beta_{or1} or1 + \beta_{or2} or2, \text{ con}$$

$$\mu_{pS/or1,or2} + \varepsilon = \beta_0 + \beta_{or1} or1 + \beta_{or2} or2 + \varepsilon, \text{ y } \varepsilon \sim NN(0, \sigma_{pS/or1,or2}^2)$$

El proceso de estimación produce el modelo con los estimadores $\hat{\mu}_{pS/or1,or2} = \hat{\beta}_0 + \hat{\beta}_{or1}or1 + \hat{\beta}_{or2}or2$, que una vez que se calcula su valor con los datos da lugar al modelo estimado:

$$\mu'_{pS/or1,or2} = 222.0 + 347.7or1 - 205.3or2,$$

recuerde que or1 y or2 son los polinomios ortogonales de orden 1 y 2 para, esporas, los valores estimados para la media y las pruebas para los coeficientes de regresión coinciden con los correspondientes para la regresión de pS sobre Esporas y Esporas², pero se evita la posibilidad de multicolinealidad.

La prueba de hipótesis de regresión, de que ambos coeficientes de regresión de Esp y Esp² son cero: $H_0: \beta_{Esp} = \beta_{Esp^2} = 0$, $H_A: \beta_{Esp} \neq 0$ ó $\beta_{Esp^2} \neq 0$, se prueba con una F calculada de 179 con 2 y 6 grados de libertad, que da un valor p de 4.478E-06 y conduce al rechazo de la hipótesis nula.

Las pruebas parciales de que los coeficientes de los polinomios de primer y segundo orden, por separado, son cero, se refieren a las parejas de hipótesis:

$$H_0: \beta_{Esp} = 0, H_A: \beta_{Esp} \neq 0, \text{ y } H_0: \beta_{Esp^2} = 0, H_A: \beta_{Esp^2} \neq 0.$$

Ambas hipótesis nulas se rechazan, con valores p de 3.401E-06 y 7.227E-05, respectivamente, como lo muestra la Tabla 10.4. El ajuste de la curva de segundo orden (cuadrática) tiene un $R^2 = 0.9835$, explica el 98.35% de la variación de peso seco de la biomasa del hongo. Esto se ve, tanto en la Tabla 10.4, como en el diagrama de dispersión (Figura 10.3) que presenta los datos originales (como +) y los valores estimados (como círculo vacío) usando la regresión cuadrática.

Tabla 10.4. Pruebas de las hipótesis parciales en la regresión cuadrática de peso seco sobre número de esporas.

Coefficiente	Estimación	Error estándar	t calculada	valor p
β_0	222.014	7.113	31.212	7.18E ⁻⁰⁸
β_{Esp}	347.706	21.339	16.294	3.40E ⁻⁰⁶
β_{Esp^2}	-205.251	21.339	-9.619	7.23E ⁻⁰⁵
R cuadrada	0.9835			

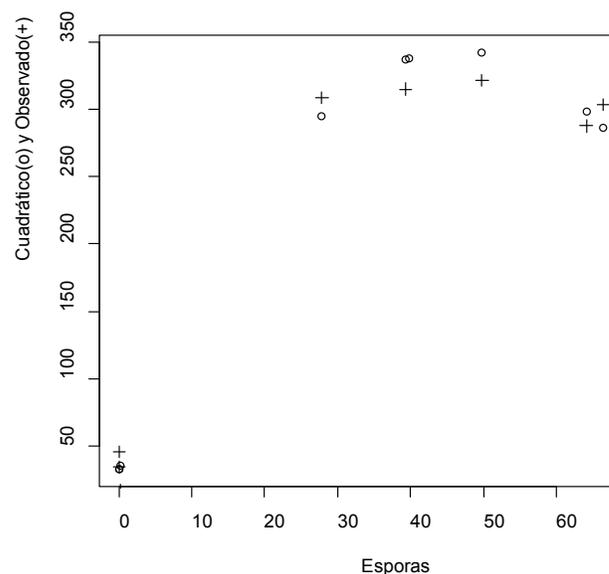


Figura 10.3. Diagrama de dispersión de datos experimentales (+) y predichos con el modelo cuadrático (o).

Las gráficas de residuales (Figura 10.4) no presentan desviaciones grandes de la suposición de errores normales, con excepción del menor y el mayor cuantil, que se salen de la línea en la gráfica de probabilidad normal (Figura 10.5). En la práctica, basta con medir esporas para tener una estimación de peso seco usando la regresión cuadrática.

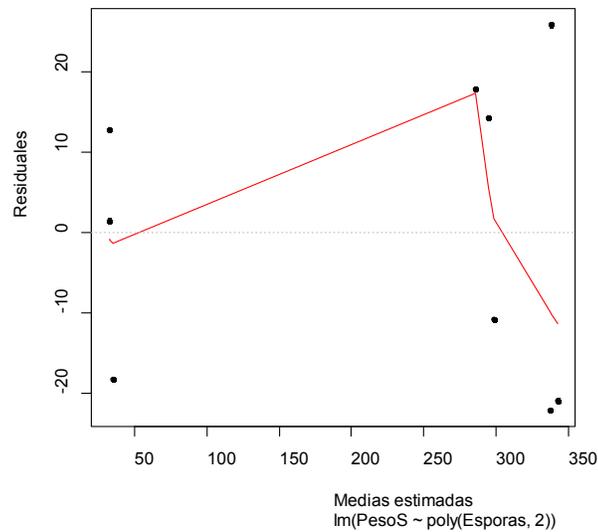


Figura 10.4. Residuales y valores predichos para peso seco y número de esporas.

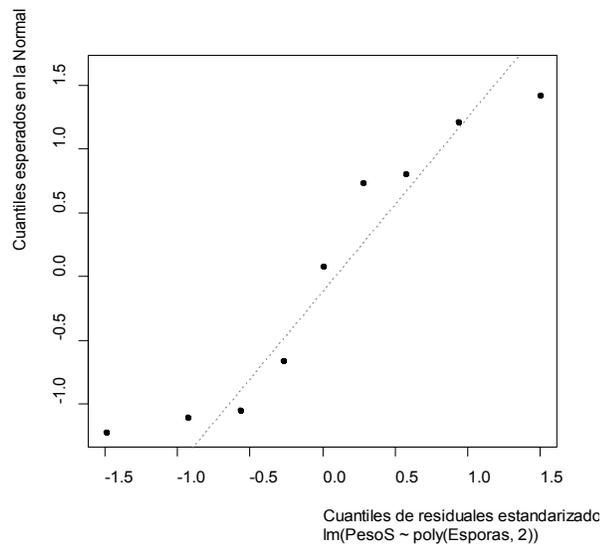


Figura 10.5. Gráfica de probabilidad normal para peso seco y número de esporas.

El intervalo de confianza para la media en un valor dado de Esporas y el intervalo de confianza para un valor de PesoS se ilustra en la Tabla 10.5 con los valores de Esporas = 60 y 65.

Tabla 10.5. Intervalo de confianza para peso seco y número de esporas.

Valor de Esporas	Media estimada	Límite inferior del I.C. 95% para la media	Límite superior del I.C. 95% para la media
60	316.9528	289.4485	344.4571
65	293.3142	257.9394	328.6889
Valor de Esporas	Media estimada	Límite inferior del I.C. 95% para una observación	Límite superior del I.C. 95% para una observación
60	316.9528	257.9369	375.9687
65	293.3142	230.2447	356.3836

El análisis de este problema se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares, sin repetir todos los detalles ya anotados en ejercicios anteriores.

Guía para resolver con R

```

datosIdriella<-read.table(file = "Idriella.csv",header = T,sep = ",")

#Recuerde los datos están en tipo csv de Excel. Debe poner la identificación completa del archivo

# o cambiarse al directorio donde está el archivo de datos

datosIdriella

peso<-datosIdriella$Peso.seco.mg.100ml

espor<-datosIdriella$esporas

tratam<-c(1,1,1,2,2,2,3,3,3) # Se genera el vector que identifica a los tratamientos

# tiam # biot # peso # espor # tratam

baseldriella<-data.frame(trat = factor(tratam),PesoS = peso,Esporas = espor)

# La instrucción aov de AdeV requiere que se tenga un marco con las variables , especificando las que son factores con factor

baseldriella

attach(baseldriella) # Otro requisito de aov, el marco debe estar activo

pesoTB<-aov(PesoS ~ trat,data = baseldriella)

replications(pesoTB ~ trat,data = baseldriella,na.fail) # Revise que el diseño es balanceado

summary(pesoTB)

medias<-model.tables(pesoTB,type = "means",se = FALSE)

# Se piden las medias (type = "means"), se pueden pedir las desviaciones estándar con se = TRUE

medias
    
```

```

TukeyHSD(pesoTB, which = "trat", orderer = TRUE, conf.level = 0.95) # Prueba de Tukey para trat

boxplot(PesoS~trat,data = baseldriella, outline = TRUE, beside = TRUE, boxwex = .5,

xlab = "Tratamiento", ylab = "Peso Seco", main = " ")

# Gráfica de cajas para los tratamientos

reg<-lm(PesoS~Esporas) # Regresión de peso seco sobre número de esporas

reg

summary(reg)

par(mfrow = c(1,1)) # Una gráfica por hoja

plot(PesoS~Esporas,data = baseldriella,xlab = "Esporas",ylab = "Peso seco",

main = " ")

# abline(reg) # Diagrama de dispersión de Peso seco sobre Esporas

# par(mfrow = c(2,2))

# plot(reg)

# par(mfrow = c(1,1))

# plot(reg,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

# title(main = expression("Residuales vs Predichos"),xlab = "Valores predichos",

#ylab = "Residuales",col.lab = 1)

# plot(reg,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

# title(main = expression("Gráfica de Probabilidad Normal"),

#xlab = "Cuantiles de residuales estandarizados",

#ylab = "Cuantiles esperados en la Normal",col.lab = 1)

# PesoSE0<- lm(PesoS~ 1,data = baseldriella)

# pesoSE1<- lm(PesoS~ poly(Esporas,1),data = baseldriella)

pesoSE2<- lm(PesoS~ poly(Esporas,2),data = baseldriella) # Regresión cuadrática de PesoS sobre Esporas

pesoSE2

summary(pesoSE2) # Dan los resultados de la regresión cuadrática

# anova(pesoSE0,pesoSE1,pesoSE2) # Tabla de AdeV para la regresión cuadrática

```

```

pesoSE2pred<- predict(pesoSE2) # Estimaciones de las medias (predichos) para los valores de Esporas en los datos

par(mfrow = c(1,1))

plot(pesoSE2pred~ Esporas,data = baseldriella, pch = 1,xlab = "Esporas",
ylab = "Cuadrático(o) y Observado(+)",main = " ")

points(PesoS~Esporas,data = baseldriella, pch = 3)

# Diagrama de dispersión de peso seco sobre esporas con los puntos de las estimaciones

# par(mfrow = c(2,2))

# plot(pesoSE2) # Las gráficas de residuales en inglés

par(mfrow = c(1,1))

plot(pesoSE2, which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression("Residuales vs Medias estimadas"),xlab = "Medias estimadas",
ylab = "Residuales",col.lab = 1) # Gráfica de residuales (wich = 1)

plot(pesoSE2,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = " ",xlab = "Cuantiles de residuales estandarizados",ylab = "Cuantiles esperados en la Normal",col.lab = 1)

# Gráfica de probabilidad normal (wich = 2)

#errores<-residuals(pesoSE2)

#shapiro.test(errores)

casosIC<- data.frame(Esporas = c(60,65)) # Define el marco con los datos de Esporas para hacer intervalos de confianza

predict(pesoSE2,newdata = casosIC,interval = "confidence",level = 0.95)

# Intervalos de confianza usando predict. Pide el modelo (pesoSE2), el marco donde están

# los datos de pH (newdata = casosIC), para intervalos con 95% (level = 0.95) para la media #(interval = "confidence")
predict(pesoSE2,newdata = casosIC,interval = "prediction",level = 0.95)

# Intervalos de confianza usando predict. Intervalos para una observación (interval = "prediction")

```

10.1. Ejercicio

En un segundo experimento para probar el efecto de la adición de suplementos de vitaminas a un medio de crecimiento con contenido mínimo de nutrientes sobre el crecimiento de biomasa en peso seco mg/100 ml, se modificó y agitó el medio para que los hongos crecieran dispersos en lugar de presentar micelios agregados y se tuvieron cinco tratamientos en cuatro bloques completos. Los niveles de tratamiento fueron: adición de biotina, adición de asparagina (Asp 1g/l), adición de base de levadura sin nitrógeno (LSN) 1.7 g/l, adición de biotina+asparagina y adición de asparagina+LSN. Los bloques se generaron porque en la agitadora solo cabían cinco matraces que correspondían a los cinco tratamientos.

Los datos están en la tabla siguiente.

Tratamiento codificado	Bloque	Tratamiento	Peso seco mg/100 ml
1	1	Biotina	309
2	1	Asparagina	23
3	1	Levadura sin N	288
4	1	Biotina+Asparagina	363
5	1	Asparagina+Levadura sin N	345
1	2	Biotina	317
2	2	Asparagina	20
3	2	Levadura sin N	305
4	2	Biotina+Asparagina	346
5	2	Asparagina+Levadura sin N	339
1	3	Biotina	337
2	3	Asparagina	14
3	3	Levadura sin N	320
4	3	Biotina+Asparagina	340
5	3	Asparagina+Levadura sin N	360
1	4	Biotina	307
2	4	Asparagina	15
3	4	Levadura sin N	308
4	4	Biotina+Asparagina	339
5	4	Asparagina+Levadura sin N	350

Haga el análisis del experimento.

Bibliografía

-  Lascaris, D. and Deacon, J.W. (1994) In vitro growth and microcycle conidiacion of *Idriella bolleyi*, a biocontrol agent of cereal pathogens. *Mycological Research*. 98(10):1200-1206.

Nota: La referencia se utilizó, para plantear un problema similar pero llevado al nivel y el interés del curso.

11. Producción de poligalacturonasa por fermentación en medio sólido

Para optimizar los valores de los factores tiempo de fermentación y adición de celobiosa en la obtención de la enzima poligalacturonasa por fermentación en medio sólido, se hicieron tres experimentos. Se planteó utilizar una secuencia de experimentos debido a que no se conoce la zona de exploración de las variables explicativas que lleva al óptimo de producción.

El primer experimento tuvo un diseño factorial con dos factores (tiempo y concentración de celobiosa), cada uno con tres niveles (3^2) sin repeticiones, y su objetivo fue ajustar una regresión cuadrática que permitiera identificar la segunda zona de exploración. En el segundo experimento se tuvo un arreglo de 2^2 con dos puntos centrales. En el tercer experimento también se tuvo un arreglo 2^2 con dos puntos centrales, pero en otra zona de exploración. Todas las corridas se hicieron en un reactor de acero inoxidable de 3x10x13 cm. Como sustrato se usó cáscara de mango, se adicionaron 40 g de medio de fermentación estéril con el inóculo de *Aspergillus niger* (105 conidios/g). Se incubó a 32 °C y la medición se hizo de acuerdo con el tratamiento de tiempo de fermentación. Los datos se presentan en la Tabla 11.1; se adiciona el valor codificado de los factores tiempo y concentración de celobiosa.

Tabla 11.1. Arreglos experimentales codificados para cada etapa de experimentación y poligalacturonasa obtenida por tratamiento.

Experimento 1, (3^2)					
Corrida	Tiempo (t), h	Concentración de Celobiosa (CC), w/w	t codificado	CC codificada	Poligalacturonasa
1	24	0	-1	-1	3.3
2	24	0.1	-1	0	2.8
3	24	0.2	-1	1	3.23
4	48	0	0	-1	3.17
5	48	0.1	0	0	3.34
6	48	0.2	0	1	3.18
7	72	0	1	-1	3.35
8	72	0.1	1	0	3.58
9	72	0.2	1	1	3.5

Experimento 2, (2^2) centrado en (0,0), punto central repetido					
Corrida	Tiempo (t), h	Concentración de Celobiosa (CC), w/w	t codificado	CC codificada	Poligalacturonasa
10	42	0	-0.25	-1	3.87
11	42	0.2	-0.25	1	3.72
12	48	0.1	0	0	3.28
13	48	0.1	0	0	3.31
14	54	0	0.25	-1	3.13
15	54	0.2	0.25	1	3.3

Experimento 3, (2 ²) centrado en (1,0), punto central repetido					
Corrida	Tiempo (t), h	Concentración de Celobiosa (CC), w/w	t codificado	CC codificada	Poligalacturonasa
16	66	0	0.75	-1	3.43
17	66	0.2	0.75	1	3.37
18	72	0.1	1	0	3.43
19	72	0.1	1	0	3.37
20	78	0	1.25	-1	3.97
21	78	0.2	1.25	1	4.2

El primer experimento es un 3², sin repeticiones. Este experimento se planeó para hacer el análisis de regresión cuadrática y localizar la zona de exploración para el óptimo de producción. La Figura 11.1 muestra que va a ser difícil obtener una tendencia debido a la variación en los datos, para 24 horas la respuesta de poligalacturonasa con 0.1 de concentración de celobiosa es menor y para 48 y 72 horas es mayor.

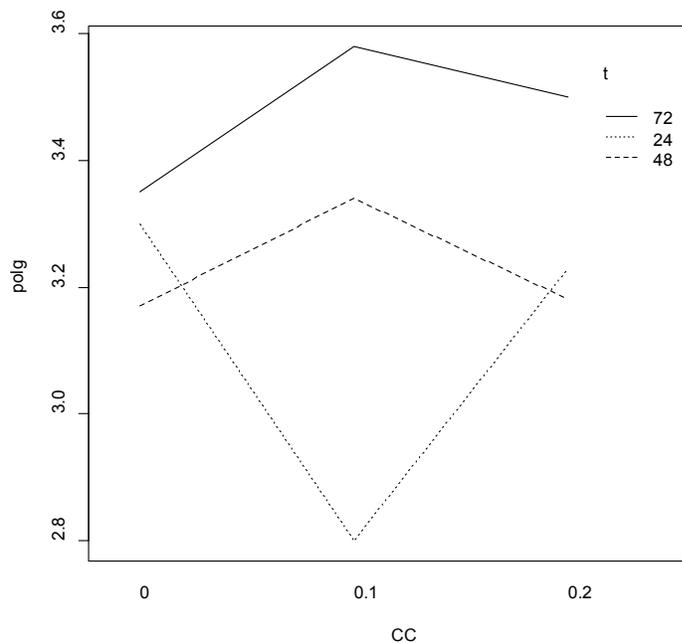


Figura 11.1. Gráfica de interacción entre tiempo y concentración de celobiosa para la producción de poligalacturonasa obtenida en el primer experimento.

La regresión cuadrática de poligalacturonasa sobre el tiempo codificado y concentración de celobiosa codificada confirma lo que mostró antes la Figura 11.1, que ninguno de los coeficientes del modelo cuadrático resulta significativo en las pruebas parciales (Tabla 11.2) ni en las pruebas secuenciales del análisis de varianza (Tabla 11.3), puesto que el valor p del coeficiente de tiempo vale 0.1590 (mayor que 0.05), y los del resto de las variables son todavía más grandes, se concluye que los coeficientes del modelo valen cero y las variables tiempo y concentración de celobiosa no afectan a la media de producción de poligalacturonasa.

Tabla 11.2. Pruebas de hipótesis parciales de la regresión cuadrática obtenida en el experimento 1 para la producción de poligalacturonasa.

Coficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	3.19778	0.17948	17.817	0.000386
Tiempo (β_1)	0.18333	0.09830	1.865	0.159044
C.Celobiosa (β_2)	0.01500	0.09830	0.153	0.888408
Tiempo cuadrado ($\beta_{1,1}$)	0.06333	0.17027	0.372	0.734639
C.Celobiosa cuadrada ($\beta_{2,2}$)	0.04833	0.17027	0.284	0.794975
Tiempo*C. Celobiosa ($\beta_{1,2}$)	0.05500	0.12040	0.457	0.678846

El análisis de varianza da los mismos resultados que los obtenidos para las pruebas parciales debido a que el diseño es ortogonal y se utilizan los valores codificados, como lo muestran los resultados de la Tabla 11.3.

Tabla 11.3. Análisis de varianza de la regresión cuadrática obtenida en el experimento 1 para la producción de poligalacturonasa.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Tiempo	1	0.201667	0.201667	3.4781	0.1590
C.Celobiosa	1	0.001350	0.001350	0.0233	0.8884
Tiempo cuadrado	1	0.008022	0.008022	0.1384	0.7346
C.Celobiosa cuadrada	1	0.004672	0.004672	0.0806	0.7950
Tiempo*C.Celobiosa	1	0.012100	0.012100	0.2087	0.6788
Residual	3	0.173944	0.057981		

Dado que no se obtuvo una respuesta significativa, se revisó nuevamente el área de exploración concentrando los valores de tiempo alrededor de 48 (codificado vale 0) y dejando los mismos valores para concentración de celobiosa.

La Figura 11.2 muestra que la variación en los puntos centrales es pequeña, los valores de poligalacturonasa para el tiempo 42 son los mayores y para el tiempo 54 los menores. También se ve que en el tiempo 42, el valor decrece a mayor contenido de celobiosa, pero en el tiempo 54 el valor se incrementa.

En el experimento 2 se plantea el modelo cuadrático. Igual que en el primer experimento, debido a la configuración de los factores y sus niveles, los términos de Tc^2 y CCc^2 se confunden en el proceso de estimación y sólo se incluye el primer término Tc^2 . Se pone el signo - al final del subíndice para indicar que el modelo cuadrático no está completo.

$$PG \sim N(\mu_{PG/Tc, CCc, cuadrático-} = \beta_0 + \beta_1 Tc + \beta_2 CCc + \beta_{1,1} Tc^2 + \beta_{1,2} Tc * CCc, \sigma_{PG/Tc, CCc, cuadrático-}^2),$$

que da lugar al modelo de estimadores:

$$\hat{\mu}_{PG/Tc, CCc, cuadrático-} = \hat{\beta}_0 + \hat{\beta}_1 Tc + \hat{\beta}_2 CCc + \hat{\beta}_{1,1} Tc^2 + \hat{\beta}_{1,2} Tc * CCc,$$

la estimación con los datos resulta:

$$\mu'_{PG/Tc,CCc, cuadrático} = 3.295 - 1.160 Tc + 0.005 CCc + 3.360 Tc^2 + 0.320 TcCCc$$

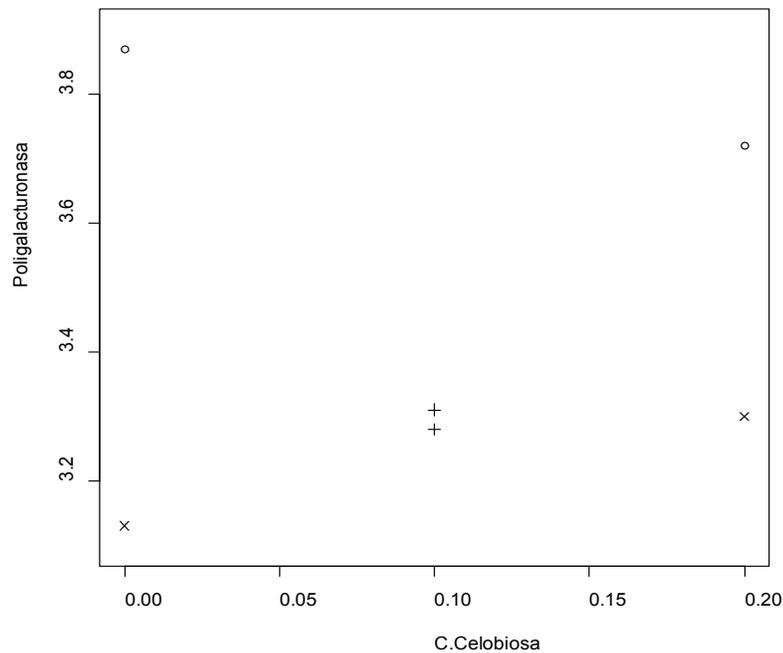


Figura 11.2. Diagrama de dispersión para la producción de poligalacturonasa sobre concentración de celobiosa obtenido en el experimento 2: o para tiempo = 42, + para tiempo = 48 y x para tiempo 54

Las parejas de hipótesis para las pruebas parciales de los coeficientes de regresión en el modelo, y los valores p para la prueba de la hipótesis nula son (Tabla 11.4):

$H_0: \beta_{Tc} = 0$, $H_A: \beta_{Tc} \neq 0$, que da un valor p de 0.0233,

$H_0: \beta_{CCc} = 0$, $H_A: \beta_{CCc} \neq 0$, que da un valor p de 0.7196,

$H_0: \beta_{Tc^2} = 0$, $H_A: \beta_{Tc^2} \neq 0$, que da un valor p de 0.0556

y

$H_0: \beta_{TcCCc} = 0$, $H_A: \beta_{TcCCc} \neq 0$, que da un valor p de 0.0839.

No hay evidencia para rechazar la hipótesis nula para el coeficiente β_{CCc} de concentración de celobiosa, puesto que su valor p vale 0.7196. Las hipótesis nulas sobre los coeficientes tiempo y tiempo cuadrado se rechazan, y la hipótesis nula sobre el coeficiente de tiempo por concentración de celobiosa no se rechaza al 5%, pero su valor p de 0.0839 permite pensar que con cambios en el modelo puede mostrarse como significativa. Se concluye que la variable explicativa tiempo influye en la media de producción de poligalacturonasa.

Tabla 11.4. Pruebas de hipótesis parciales de la regresión cuadrática obtenida en el experimento 2 para la producción de poligalacturonasa.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	3.29500	0.01500	219.667	0.0029
Tiempo (β_1)	-1.16000	0.04243	-27.341	0.0233
C.Celobiosa (β_2)	0.00500	0.01061	0.471	0.7196
Tiempo cuadrado ($\beta_{1,1}$)	3.36000	0.29394	11.431	0.0556
Tiempo*C. Celobiosa ($\beta_{1,2}$)	0.32000	0.04243	7.542	0.0839

Debido a lo anterior se decide utilizar el modelo reducido sin el término de CCc:

$$PG \sim N(\mu_{PG/Tc, Tc^2, Tc*CCc} = \beta_0 + \beta_1 Tc + \beta_{1,1} Tc^2 + \beta_{1,2} Tc * CCc, \sigma_{PG/Tc, Tc^2, Tc*CCc}^2),$$

con

$$\hat{\mu}_{PG/Tc, Tc^2, Tc*CCc} = \hat{\beta}_0 + \hat{\beta}_1 Tc + \hat{\beta}_{1,1} Tc^2 + \hat{\beta}_{1,2} Tc * CCc$$

y

$$\mu'_{PG/Tc, Tc^2, Tc*CCc} = 3.295 - 1.160 Tc + 3.360 Tc^2 + 0.320 TcCCc.$$

Las hipótesis para las pruebas parciales dan valores p que llevan al rechazo de las hipótesis nulas (Tabla 11.5):

H0: $\beta_{Tc} = 0$, HA: $\beta_{Tc} \neq 0$, que da un valor p de 0.000816,

H0: $\beta_{Tc^2} = 0$, HA: $\beta_{Tc^2} \neq 0$, que da un valor p de 0.004644

y

H0: $\beta_{TcCCc} = 0$, HA: $\beta_{TcCCc} \neq 0$, que da un valor p de 0.010572.

Tabla 11.5. Pruebas de hipótesis parciales de la regresión cuadrática con el modelo reducido obtenida en el experimento 2 para la producción de poligalacturonasa.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	3.29500	0.01173	280.999	1.27E ⁻⁰⁵
Tiempo (β_1)	-1.16000	0.03317	-34.975	0.000816
Tiempo cuadrado ($\beta_{1,1}$)	3.36000	0.22978	14.623	0.004644
Tiempo*C. Celobiosa ($\beta_{1,2}$)	0.32000	0.03317	9.648	0.010572

El modelo explica el 99.89% de la variación, ya que $R^2 = 0.9989$.

Las derivadas del modelo estimado con respecto a las variables Tc y CCc producen los valores de Tc = 0 y CCc = 3.625. El valor de CCc está fuera de la zona de exploración y conviene tomar CCc = 1.0. Para estos valores la estimación de la media de poligalacturonasa es 3.295 y el intervalo de confianza para la media es (3.2445, 3.3454).

Para confirmar los resultados se hizo un tercer experimento modificando los valores de tiempo, ahora centrado en 72.

La gráfica es parecida a las del experimento 2 (Figura 11.3). Se plantea al modelo cuadrático, sin el término CCc2 que se confunde con Tc2:

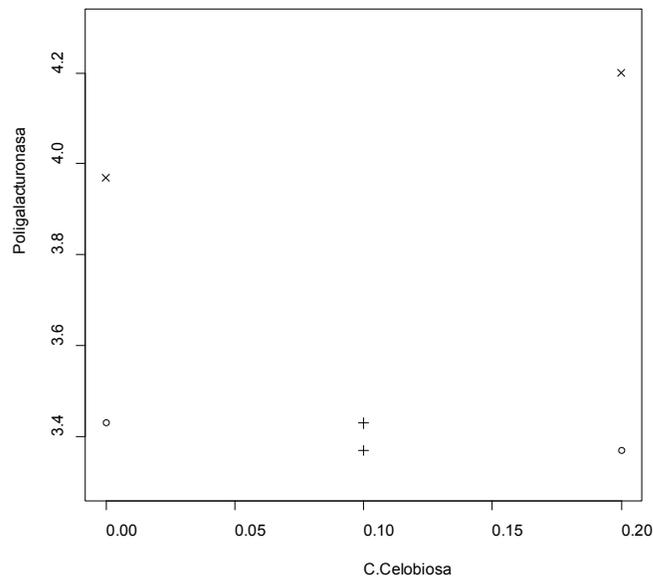


Figura 11.3. Diagrama de dispersión de la producción de poligalacturonasa sobre concentración de celobiosa obtenidos en el experimento 3, con o para tiempo 66, + para tiempo 72 y x para tiempo 78.

En la Tabla 11.6 están las pruebas parciales del modelo cuadrático. Las hipótesis nulas de que el coeficiente de Tc y el de Tc² son cero se rechazan con 10% de significación ya que presentan valores p de 0.0779 y 0.0680 respectivamente. En cambio, la hipótesis de que el coeficiente de regresión de C. Celobiosa es cero no se rechaza, ya que el valor de la significación muestral (o valor p) es 0.2163. Lo mismo ocurre con el coeficiente de tiempo por C. Celobiosa, que produce un valor p de 0.1812

Tabla 11.6. Pruebas de hipótesis parciales de la regresión cuadrática obtenida en el experimento 3 para la producción de poligalacturonasa.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	7.51000	0.57000	13.175	0.0482
Tiempo (β_1)	-9.59000	1.17881	-8.135	0.0779
C.Celobiosa (β_2)	-0.24750	0.08746	-2.830	0.2163
Tiempo cuadrado ($\beta_{1,1}$)	5.48000	0.58788	9.322	0.0680
Tiempo*C. Celobiosa ($\beta_{1,2}$)	0.29000	0.08485	3.418	0.1812

Los resultados anteriores llevan a proponer un modelo que se reduce a un cuadrático en la variable Tc:

$$PG \sim N(\mu_{PG/Tc, Tc^2} = \beta_0 + \beta_1 Tc + \beta_{1,1} Tc^2, \sigma_{PG/Tc, Tc^2}^2), \text{ con}$$

$$\hat{\mu}_{PG/Tc, Tc^2} = \hat{\beta}_0 + \hat{\beta}_1 Tc + \hat{\beta}_{1,1} Tc^2, \text{ y}$$

$$\mu'_{PG/Tc, Tc^2} = 7.51 - 9.59 Tc + 5.48 Tc^2$$

Los coeficientes de las variables de este modelo resultan significativos en las pruebas parciales, con valores p por debajo de 0.05, como se muestra en la última columna de la Tabla 11.7.

Tabla 11.7. Pruebas de hipótesis parciales de la regresión cuadrática con el modelo reducido obtenida en el experimento 3 para la producción de poligalacturonasa.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	7.51000	1.345	5.585	0.0113
Tiempo (β_1)	-9.59000	2.781	-3.449	0.0410
Tiempo cuadrado (β_{11})	5.48000	1.387	3.952	0.0289

La ecuación cuadrática en Tc es creciente hacia la derecha (Figura 11.5). El valor máximo se debe tomar en el valor extremo de la zona de exploración y corresponde a Tc = 1.25. Para este valor se obtiene una media estimada de 4.085 y el intervalo con 95% de confianza es (3.8598, 4.3102).

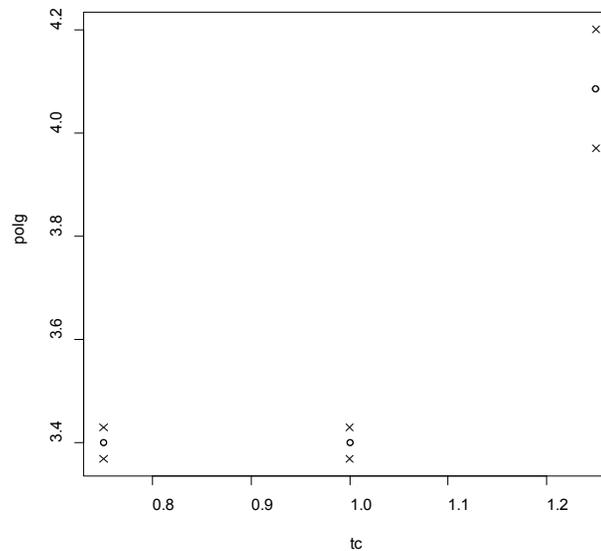


Figura 11.5. Diagrama de dispersión de poligalacturonasa sobre tiempo codificado (x) y los valores de las medias estimadas con el modelo cúbico (o).

El análisis de los experimentos para optimizar la producción de poligalacturonasa se hace con las instrucciones del paquete R que siguen.

Guía para resolver con R

```
datosPlg1<-read.table(file = "poligal1.csv",header = T,sep = ",")
```

```
datosPlg2<-read.table(file = "poligal2.csv",header = T,sep = ",")
```

```
datosPlg3<-read.table(file = "poligal3.csv",header = T,sep = ",")
```

```
# Se leen las bases de datos de los tres experimentos, se trabajarán por separado
```

```
datosPlg1
```

```
datosPlg2
```

```
datosPlg3
```

```
## Se analiza el experimento 1
```

```
t<-datosPlg1$Tiempo.h.
```

```
CC<-datosPlg1$C.Celobiosa.w.w.
```

```
tc<-datosPlg1$Tcodificado
```

```
CCc<-datosPlg1$CCcodificada
```

```
polg<-datosPlg1$Poligalacturonasa
```

```
# t # CC # tc # CCc # polg
```

```
tc2<-tc^2
```

```
CCc2<-CCc^2
```

```
tcCCc<-tc*CCc
```

```
interaction.plot(CC,t,polg,ylab = "polg")
```

```
polgR2<-lm(polg~tc+CCc+tc2+CCc2+tcCCc)
```

```
summary(polgR2)
```

```
anova(polgR2)
```

```
## Se analiza el experimento 2. Se usan los mismos nombres de variables
```

```
t<-datosPlg2$Tiempo.h.
```

```
CC<-datosPlg2$C.Celobiosa.w.w.
```

```
tc<-datosPlg2$Tcodificado
```

```
CCc<-datosPlg2$CCcodificada
```

```
polg<-datosPlg2$Poligalacturonasa
```

```
# t # CC # tc # CCc # polg
```

```
tc2<-tc^2
```

```
CCc2<-CCc^2
```

```

tcCCc<-tc*CCc

cc1<-c(0,.2)

pol<-c(3.87,3.72)

cc2<-c(0.1,0.1)

po2<-c(3.28,3.31)

cc3<-cc1

po3<-c(3.13,3.3)

plot(pol~cc1,main = " ",xlab = "C.Celobiosa",ylab = "Poligalacturonasa",ylim = c(3.1,3.9))

points(po2~cc2,pch = 3)

points(po3~cc3,pch = 4)

polgR2<-lm(polg~tc+CCc+tc2+tcCCc)

summary(polgR2)

anova(polgR2)

polgR2$coef

polgR2a<-lm(polg~tc+tc2+tcCCc)

summary(polgR2a)

anova(polgR2a)

polgR2a$coef

# Se obtiene  $dpolg/dtc = -1.16 + 3.36/2 tc + .32tc CCc = 0$  y  $.32 tc = 0$ 

#  $tc = 0$  y  $CCc = 3.625$ 

dPred1<-data.frame(tc = c(0),tc2 = c(0),tcCCc = c(0))

predict(polgR2a,dPred1,interval = "confidence")

### Se analiza el experimento 3. Se usan los mismos nombres de variables

t<-datosPlg3$Tiempo.h.

CCc<-datosPlg3$C.Celobiosa.w.w.

tc<-datosPlg3$Tcodificado

CCc<-datosPlg3$CCcodificada

```

```

polg<-datosPlg3$Poligalacturonasa

# t # CC # tc # CCc # polg

tc2<-tc^2

CCc2<-CCc^2

tcCCc<-tc*CCc

cc1<-c(0,.2)

pol<-c(3.43,3.37)

cc2<-c(0.1,0.1)

po2<-c(3.43,3.37)

cc3<-cc1

po3<-c(3.97,4.2)

plot(pol~cc1,main = " ",xlab = "C.Celobiosa",ylab = "Poligalacturonasa",ylim = c(3.3,4.3))

points(po2~cc2,pch = 3)

points(po3~cc3,pch = 4)

polgR3<-lm(polg~tc+CCc+tc2+CCc2+tcCCc)

summary(polgR3)

anova(polgR3)

polgR3$coef

polgR3a<-lm(polg~tc+tc2)

summary(polgR3a)

anova(polgR3a)

polgR3a$coef

predict(polgR3a)

vpre<-predict(polgR3a) #Para ver la gráfica de la cuadrática

plot(polg~tc,pch = 4)

points(vpre~tc)

dPred2<-data.frame(tc = c(1.25),tc2 = c(1.5625))

```

```
predict(polgR3a,dPred2,interval = "confidence")
```

11.1 Ejercicio

Considerando que las variables fueron iguales en los tres experimentos del Ejercicio 11, y suponiendo que se realizaron bajo las mismas condiciones, haga un análisis conjunto para los datos de los tres experimentos usando los valores sin codificar.

Bibliografía

 Couri, S., da Costa, T.S., Saavedra, P.G.A., Pereira, F.S. and da Costa, A.C.A. (2000) Hydrolitic enzyme production in solid-state fermentation by *Aspergillus niger* 3T5B8. *Process Biochemistry*. 36:255-261.

Nota: La referencia se utilizó, para plantear un problema similar pero llevado al nivel y el interés del curso. El artículo tenía datos.

12. Efecto del tipo de filtro y velocidad de carga en la eliminación de metanol en reactores biofiltros

Se evaluó un sistema de biofiltros para eliminar metanol del aire, para lo cual se comparó la capacidad de eliminación de cinco biofiltros de corriente ascendente y otros cinco de corriente descendente, alimentados con diferentes velocidades de carga de metanol en el aire suministrado. Como variables de respuesta se midió la capacidad de eliminación de metanol y el conteo de bacterias presentes en los biofiltros diez días después de iniciado el proceso.

Los biofiltros constaban de cuatro secciones en las que se utilizó composta de hongos y anillos de prolipropileno en proporción 8:2 como soporte e inoculados con 10% de lodos activados, cosechado de un sistema de alcantarillado. Se mantuvo entre 40 y 60% de humedad en el biofiltro agregando agua (y nutrientes), y el gas entrante (100 g/(m³h)) tenía 90% de humedad. Debido a que cada corrida de los tratamientos combinados duraba 10 días, se tomaron dos muestras de lodo activado, dando lugar a dos bloques, cada uno con una repetición completa. El diseño en bloques al azar tiene dos bloques, cada uno con una repetición completa de las diez combinaciones de tratamientos que son: 1) velocidad de carga de metanol suministrado (MS) con cinco niveles (0, 50, 100, 150 y 175 g/(m³ h)) y tipo de biofiltro con dos niveles (corriente ascendente y corriente descendente). Dado que el factor velocidad de carga de metanol suministrado tiene niveles cuantitativos, también conviene estudiar el tipo de respuesta que presenta, es decir, si es lineal, cuadrático, etc.

La capacidad de eliminación de metanol g/(m³h) se obtuvo como el porcentaje de la diferencia de concentración de metanol en el aire suministrado y el extraído, ambas determinadas mediante cromatografía. Se inoculó una muestra compuesta por 1g del fondo y 1g de la parte superior del biofiltro en medio de cultivo de agar y se contaron las colonias formadas para determinar las bacterias presentes. Para facilidad de representación, se reporta el conteo de bacterias dividido entre 10⁷. Los datos se presentan en la Tabla 12.1.

Tabla 12.1. Eliminación de metanol y conteo de bacterias.

Bloque	Metanol suministrado (MS) g/(m ³ h)	Eliminación de metanol (EM) %		Conteo de Bacterias (CB)	
		Corriente ascendente	Corriente descendente	Corriente ascendente	Corriente descendente
1	0	0	0	8	1.9
1	50	49	50	5	2
1	100	100	91	10	3.5
1	150	100	95	13	11
1	175	91	90	16	14
2	0	0	0	8.5	1.8
2	50	46	54	6	2.5
2	100	91	95	12	4
2	150	95	96	13.5	12
2	175	87	86	15	15

Se trata de un modelo de bloques al azar con dos factores e interacción para la eliminación de metanol:

$$\mu_{EM/B_i,MS_j,TB_k,(MS*TB)_{j,k}} = \mu + B_i + MS_j + TB_k + (MS*TB)_{j,k}; i = 1,2, j = 1,2,3,4,5 \text{ y } k = 1,2.$$

Los símbolos en la ecuación se leen cambiando las abreviaturas por su significado completo, así: $\mu_{EM/B_i,MS_j,TB_k,(MS*TB)_{j,k}}$ es la media de la eliminación de metanol para el bloque i (B_i), el nivel j de velocidad de carga de metanol suministrado (MS_j) y el nivel k del tipo de biofiltro (TB_k), esa media es igual a una suma de términos. Del lado derecho de la ecuación, la media general μ más los efectos de los niveles de los factores, el efecto del bloque i (B_i), el efecto del nivel j de la velocidad de carga de metanol

(MS_j), mas el efecto del nivel k del tipo de biofiltro (TB_k), mas el efecto de la interacción entre los niveles j de velocidad de carga de metanol y k de tipo de biofiltro (MS*TB)_{j,k}.

La relación entre la media de eliminación de metanol y la variable aleatoria eliminación de metanol (EM), se refleja en la ecuación siguiente:

$$EM_{i,j,k} = \mu_{EM/B_i,MS_j,TB_k,(MS*TB)_{j,k}} + \varepsilon_{i,j,k}; i = 1,2, j = 1,2,3,4,5 \text{ y } k = 1,2,$$

en la que se especifica que la variable aleatoria eliminación de metanol para los niveles i, j y k de bloque, velocidad de carga de metanol y tipo de biofiltro, respectivamente, es la media que corresponde a esos niveles de efectos mas un término aleatorio (ε), que se distribuye normal con media cero y varianza desconocida, la varianza del error σ^2 . Los investigadores prefieren sustituir la media por la forma que incluye los efectos, y se tiene:

$$EM_{i,j,k} = \mu + B_i + MS_j + TB_k + (MS * TB)_{j,k} + \varepsilon_{i,j,k}; i = 1,2, j = 1,2,3,4,5 \text{ y } k = 1,2, \varepsilon_{i,j} \sim N(0, \sigma^2).$$

Se va a probar si hay efecto de velocidad de carga, de tipo de biofiltro y de su interacción en la media de eliminación de etanol. Para cada prueba se establecerán las hipótesis nula y alternativa y la prueba se hará usando los resultados del análisis de varianza y los valores p que produce.

El análisis de varianza permite probar la hipótesis de si cada uno de los factores en el modelo, por separado, tiene niveles cuyos efectos son iguales. Al considerar las restricciones que se imponen para resolver las ecuaciones que permiten calcular los estimadores, se representa como si los efectos valieran cero.

De acuerdo a los valores p, que se muestran en la tabla 12.2, se tiene lo siguiente:

La hipótesis de que los efectos de bloques son iguales no se rechaza porque tiene un valor p = 0.2529, lo que sugiere que no es necesario considerar bloques en trabajos posteriores. La pareja de hipótesis sobre la igualdad de efectos para velocidad de carga de metanol suministrado es:

H₀: MS₁ = MS₂ = MS₃ = MS₄ = MS₅, H_A: al menos una pareja de MS₁ ó MS₂ ó MS₃ ó MS₄ ó MS₅ es diferente.

De acuerdo al valor p = 1.807E-11 para la prueba, se rechaza la hipótesis nula y conviene analizar la forma como se relaciona la velocidad de carga de metanol suministrado con la media de metanol eliminado, recuerde que el suministro de metanol es una variable cuantitativa.

La hipótesis de efectos iguales para los dos tipos de biofiltros produce la pareja:

H₀: TB₁ = TB₂, H_A: TB₁ ≠ TB₂.

En la tabla de análisis de varianza está el valor p = 0.8820 para la prueba. Los datos no dan evidencia para rechazar la hipótesis de que los efectos de los dos tipos de biofiltros son iguales. Esto permite al investigador usar cualquiera de ellos para trabajos posteriores. Finalmente, la pareja de hipótesis sobre la interacción entre suministro de metanol y tipo de biofiltro se plantea como:

H₀: (MS*TB)₁₁ = (MS*TB)_{1,2}, (MS*TB)₂₁ = (MS*TB)_{2,2}, ..., (MS*TB)_{5,1} = (MS*TB)_{5,2},

H_A: al menos una de las igualdades de H₀ no se cumple.

No hay evidencia para rechazar la hipótesis nula de igualdad de efectos de interacción, puesto que p = 0.4968. Al no haber interacción, los cambios que produce la velocidad de carga de metanol suministrado para el biofiltro de corriente ascendente, son iguales a los cambios que produce la velocidad de carga de metanol suministrado para el biofiltro de corriente descendente.

Tabla 12.2. Anova para velocidad de carga de metanol suministrado y tipo de biofiltro utilizado.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Bloque	1	12.8	12.8	1.4922	0.2529
Suministro metanol	4	27417.7	6854.4	799.0910	1.807E ⁻¹¹
Tipo de biofiltro	1	0.2	0.2	0.0233	0.8820
Interacción SM*TB	4	31.3	7.8	0.9122	0.4968
Error	9	77.2	8.6		

Las medias de eliminación de metanol para los niveles de metanol suministrado están en la Tabla 12.3.

Tabla 12.3. Velocidades de carga de metanol suministrado y eliminación de metanol.

Velocidad de carga de metanol suministrado (MS)	0	50	100	150	175
Media de eliminación de metanol (EM)	0.00	49.75	94.25	96.50	88.50

Debido a que no hubo efecto de bloque, de tipo de biofiltro ni de interacción entre suministro de metanol y tipo de biofiltro, se puede hacer un análisis de regresión con los datos de velocidad de carga de metanol suministrado y eliminación de metanol sin considerar los otros factores.

El diagrama de dispersión de EM (eliminación de metanol) sobre MS (velocidad de carga de metanol suministrado) muestra una curva en la parte superior derecha (Figura 12.1), esta curva permite determinar el valor de velocidad de carga de metanol suministrado que produce el máximo de eliminación. Se plantea el modelo de segundo orden o cuadrático para la media de EM.

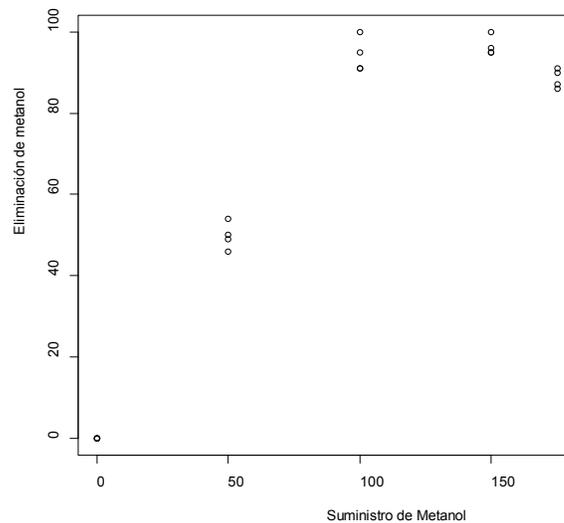


Figura 12.1. Diagrama de dispersión de eliminación de metanol sobre velocidad de carga de metanol suministrado.

Se escribe el modelo que relaciona a la media de EM con la variable explicativa MS, luego está el modelo para la variable aleatoria EM y finalmente se tiene el modelo para los estimadores (variables aleatorias).

Modelo cuadrático para la media de EM es: $\mu_{EM/MS,MS^2} = \beta_0 + \beta_{MS}MS + \beta_{MS^2}MS^2$,

modelo cuadrático para la variable aleatoria EM es: $EM = \mu_{EM/MS,MS^2} + \varepsilon = \beta_0 + \beta_{MS}MS + \beta_{MS^2}MS^2 + \varepsilon$, con $\varepsilon \sim N(0, \sigma_{EM/MS,MS^2}^2)$ y el valor de la varianza $\sigma_{EM/MS,MS^2}^2$ se desconoce.

El modelo para los estimadores (variables aleatorias) es: $\hat{\mu}_{EM/MS,MS^2} = \hat{\beta}_0 + \hat{\beta}_{MS}MS + \hat{\beta}_{MS^2}MS^2$.

Una vez que se tienen los datos y se calculan los valores para los estimadores se obtiene el modelo estimado (Tabla 12.4):

$$\mu'_{EM/MS,MS^2} = -2.3440722 + 1.4291753MS - 0.0051526MS^2,$$

recuerde que se usa $\mu'_{EM/MS,MS^2}$ para enfatizar que no es variable aleatoria, sino la estimación de la media usando los datos.

El resultado del análisis de regresión está en la Tabla 12.4. La t calculada y el valor p de cada renglón en la Tabla 12.4 prueba si el coeficiente es cero o no. Para el coeficiente de MS (lineal) se tiene $H_0: \beta_{MS} = 0$, $H_A: \beta_{MS} \neq 0$ con valor $p = 9.06E^{-14}$, por lo que la hipótesis nula se rechaza, β_{MS} es diferente de cero y el término con MS debe permanecer en la ecuación. Para el coeficiente de MS^2 (cuadrático) se plantea $H_0: \beta_{MS^2} = 0$, $H_A: \beta_{MS^2} \neq 0$, el valor p para la prueba es $p = 7.21E^{-11}$ y se rechaza la hipótesis nula, el coeficiente β_{MS^2} es diferente de cero y se comprueba que la tendencia es cuadrática. Esto se afirma al revisar el valor de $R^2 = 0.9844$, que indica que la regresión explica el 98.44% de la variación de la variable respuesta EM.

Tabla 12.4. Pruebas parciales de regresión cuadrática para Eliminación-suministro de metanol.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante β_0	-2.3440722	2.4108306	-0.972	0.345
Lineal β_{MS}	1.4291753	0.0664341	21.513	9.06E-14
Cuadrático β_{MS^2}	-0.0051526	.0003624	-14.22	7.21E-11
R cuadrada	0.9844		F con 2 y 17 g.l.	537.0

El diagrama de dispersión con los valores estimados para la media se muestra en la Figura 12.2, donde los círculos son los datos observados y los triángulos son los valores estimados para la media en la regresión cuadrática.

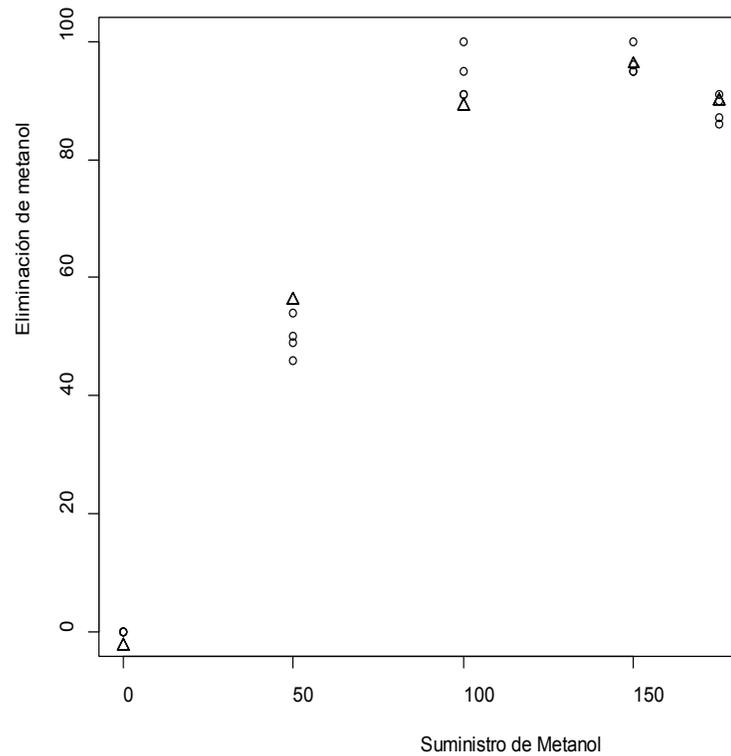


Figura 12.2. Diagrama de dispersión de los datos de eliminación de metanol observados -○- y los estimados -△- por el modelo cuadrático.

Para obtener el máximo de la estimación de la media de eliminación de metanol, se deriva la ecuación estimada con respecto a MS, se iguala a cero la derivada y se resuelve para MS.

$$\frac{d(-2.3440722 + 1.4291753MS - 0.0051526MS^2)}{d(MS)} = 1.4291753 - 2 * 0.0051526MS = 0$$

produce $MS = 138.68$ que da un máximo de eliminación de metanol de 96.758.

El intervalo con 95% de confianza para la media de metanol eliminado en el óptimo se muestra en la Tabla 12.5.

Tabla 12.5. Intervalo de confianza para la media estimada de eliminación de metanol.

MS	Media estimada	Límite inferior del I.C. 95% para la media	Límite superior del I.C. 95% para la media
138.6849	96.75885	93.63489	99.8828

El análisis de la variable conteo de bacterias y la relación entre eliminación de metanol y conteo de bacterias se deja como ejercicio a resolver por el estudiante (Ejercicio 12.1).

El análisis de eliminación de metanol se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares.

Guía para resolver con R

```

datosTyM<-read.table(file = "tipoYmetanol.csv",header = T,sep = ",")

# Se leen los datos y se crea la tabla. Datos en Excel con tipo csv

datosTyM

Blo<-datosTyM$bloque           # Se definen los vectores de observaciones ,
MS<-datosTyM$metSum           # se extraen de la tabla usando $ y el nombre de la variable

TB<-datosTyM$tipo

eli<-datosTyM$elimin

bac<-datosTyM$bact

# Blo # MS # TB # eli # bac

baseTyM<-data.frame(Bloque = factor(Blo),MSmn = factor(MS),TBft = factor(TB),ElimMet = eli,CtBac = bac)

# Se crea el marco para el AdeV usando aov. Los factores se definen usando factor

baseTyM

attach(baseTyM)

emAdeV<-aov(ElimMet~Bloque+MSmn*TBft,data = baseTyM)

# El término MSmn*TBft genera los factores MSmn y TBft y su interacción

replications(ElimMet~Bloque+MSmn*TBft,data = baseTyM,na.fail)

# Revise que los niveles de cada factor e interacción tengan igual número de repeticiones. Lo requiere aov

summary(emAdeV)

medias<-model.tables(emAdeV,type = "means",se = FALSE)

# Genera las medias de los niveles de cada factor e interacción

medias

# TukeyHSD(emAdeV, which = "MSmn",orderer = TRUE, conf.level = 0.95)

# No se pidió la prueba de Tukey, pero puede verla

# library(gplots) # Este paquete es necesario para usar plotmeans. No se pidieron

# plotmeans(ElimMet~MSmn,data = baseTyM,n.label = FALSE,bars = FALSE,connect = FALSE,

# main = "Medias de Eliminación de Metanol-Suministro de Metanol")

```

```

# boxplot(ElimMet~MSmn,data = baseTyM,outline = TRUE,beside = TRUE,boxwex = .5,
# xlab = "Suministro de metano",ylab = "Metanol Eliminado",main = "GRAFICA DE CAJAS PARA Metanol")

par(mfrow = c(1,1))

plot(eli~MS,xlab = "Suministro de Metanol",ylab = "Eliminación de metanol",main = " ")

# Diagrama de dispersión de eli sobre MS

baseTM<-data.frame(MS = MS,ME = eli,MS2 = MS*MS) # Se crea la base con MS al cuadrado

attach(baseTM)

EmSm2<- lm(ME~MS+MS2,data = baseTM) # Regresión cuadrática (tiene Ms y MS2)

EmSm2

summary(EmSm2)

anova(EmSm2)

EmSm2pred<- predict(EmSm2) # Se obtienen las medias estimads para hacer la gráfica que las incluye

# EmSm2pred

plot(eli~MS,xlab = "Suministro de Metanol",ylab = "Eliminación de metanol",
main = " ")

points(EmSm2pred~MS,pch = 2) # Agrega los puntos de medias estimadas al diagrama de dispersión

optimoMS<- 1.4291753/(2*0.0051526) # Cálculo delo optimo para MS

optimoMS

optimoEM<-2.3440722+1.4291753*optimoMS-0.0051526*optimoMS*optimoMS

# Cálculo para el óptimo de EM

optimoEM

optimoMS2<-optimoMS^2

pOp<-data.frame(MS = optimoMS,MS2 = optimoMS2)

icOptimo<-predict(EmSm2,pOp,interval = "confidence")

icOptimo

```

12.1. Ejercicio

Haga el análisis de conteo de bacterias del problema anterior resuelto y encuentre la relación de regresión entre eliminación de metanol y conteo de bacterias. Son los mismos datos.

Bloque	Metanol suministrado (MS) g/ (m ³ h)	Eliminación de metanol (EM)		Conteo de Bacterias (CB)	
		Corriente ascendente	Corriente descendente	Corriente ascendente	Corriente descendente
1	0	0	0	8	1.9
1	50	49	50	5	2
1	100	100	91	10	3.5
1	150	100	95	13	11
1	175	91	90	16	14
2	0	0	0	8.5	1.8
2	50	46	54	6	2.5
2	100	91	95	12	4
2	150	95	96	13.5	12
2	175	87	86	15	15

Bibliografía

-  Krailas, S., Pham, Q.T., Amal, R., Jiang, J.K. and Heitz, M. (2000) Effect of inlet mass loading, water and total bacteria count on methanol elimination using upward flow and downward flow biofilters. *Journal of Chemical Technology and Biotechnology*. 75: 299-305.

Nota: La referencia se utilizó para plantear un problema similar pero llevado al nivel y el interés del curso.

13. Optimización del medio de cultivo para producción de estreptomina

La producción de antibióticos como la estreptomina sigue teniendo una gran demanda en nuestros días. Con la finalidad de obtener una mejora en la producción del antibiótico, se evaluó la composición de un medio de cultivo. Los factores evaluados fueron las concentraciones de glucosa: 12.93, 15, 20, 25 y 27.07 g/l (x_1), de extracto de levadura de cerveza: 10.86, 15, 25, 35 y 39.14 g/l (x_2), de NaCl: 2.93, 5, 10, 15 y 17.07 g/l (x_3) y de K_2HPO_4 : 0.293, .5, 1, 1.5 y 1.707 g/l (x_4). La variable de respuesta (y), fue la producción de estreptomina obtenida a los tres días de cultivo, determinada mediante una titulación y expresada como concentración relativa. Para el estudio se usó un diseño compuesto central ortogonal para cuatro factores con cuatro repeticiones en el punto central. Con esto se puede obtener y analizar el error puro y luego, utilizando el residual de las regresiones, la falta de ajuste. La codificación del arreglo experimental y los datos obtenidos en el experimento se muestran en la Tabla 13.1.

Tabla 13.1. Arreglo experimental y datos obtenidos por tratamiento.

ensayo	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_1	-1	-1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	1
x_2	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1
x_3	-1	-1	1	-1	-1	1	1	1	-1	-1	-1	1	1	-1
x_4	-1	1	-1	-1	-1	1	-1	-1	1	1	-1	1	1	1
y	0.639	0.624	0.521	0.879	1.114	0.637	0.969	0.971	0.809	1.022	0.949	1.028	0.959	1.132
ensayo	15	16	17	18	19	20	21	22	23	24	25	26	27	28
x_1	1	1	0	0	0	0	1.414	-1.414	0	0	0	0	0	0
x_2	1	1	0	0	0	0	0	0	1.414	-1.414	0	0	0	0
x_3	1	1	0	0	0	0	0	0	0	0	1.414	-1.414	0	0
x_4	-1	1	0	0	0	0	0	0	0	0	0	0	1.414	-1.414
y	1.09	1.195	0.988	1.039	1.156	1.322	1.231	0.709	0.782	0.699	1.115	1.01	1.372	1.007

En los diseños de superficie de respuesta, como este diseño compuesto central ortogonal, interesa obtener la regresión de segundo orden para obtener los valores de las variables explicativas que maximizan el valor de la media de la variable de respuesta. Debido a esto, es deseable que cada uno de los coeficientes de la regresión cuadrática resulten diferentes de cero.

La Figura 13.1 muestra las posibles relaciones de cada factor evaluado con la producción de estreptomina. Como se observa, la relación de la variable de respuesta sobre las explicativas no es muy clara porque los datos presentan mucha variación, pero se puede ver que la producción de estreptomina tiene una respuesta de tipo cuadrático sobre glucosa y extracto de levadura. En el caso de NaCl y K_2HPO_4 también parece haber una respuesta de tipo cuadrático, pero hay varias observaciones fuera de esa tendencia.

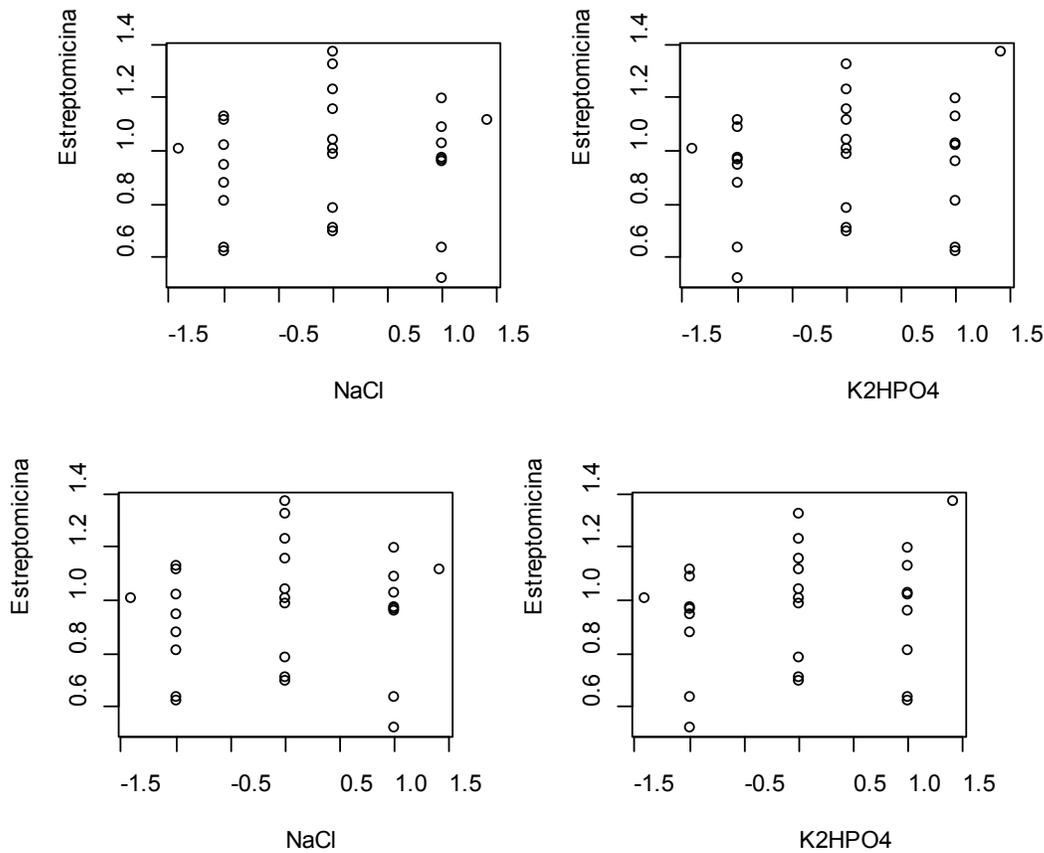


Figura 13.1. Diagrama de dispersión de estreptomicina sobre glucosa (a); estreptomicina sobre extracto de levadura (b); estreptomicina sobre NaCl (c) y estreptomicina sobre K_2HPO_4 (d).

Las cuatro repeticiones del punto central permiten estimar el error puro, que es su varianza muestral y que vale $S^2_{ep} = 0.021976$ con tres grados de libertad.

Las variables explicativas están codificadas de acuerdo con el diseño compuesto central y como tales se usan en el análisis. Se plantea el modelo cuadrático en los cuatro factores:

$$y \sim N(\mu_{y/x_1, x_2, x_3, x_4, \text{cuadrático}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{1,1} x_1^2 + \beta_{2,2} x_2^2 + \beta_{3,3} x_3^2 + \beta_{4,4} x_4^2 + \beta_{1,2} x_1 x_2 + \beta_{1,3} x_1 x_3 + \beta_{1,4} x_1 x_4 + \beta_{2,3} x_2 x_3 + \beta_{2,4} x_2 x_4 + \beta_{3,4} x_3 x_4, \sigma^2_{y/x_1, x_2, x_3, x_4, \text{cuadrático}})$$

que se acostumbra escribirlo como sigue:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{1,1} x_1^2 + \beta_{2,2} x_2^2 + \beta_{3,3} x_3^2 + \beta_{4,4} x_4^2 + \beta_{1,2} x_1 x_2 + \beta_{1,3} x_1 x_3 + \beta_{1,4} x_1 x_4 + \beta_{2,3} x_2 x_3 + \beta_{2,4} x_2 x_4 + \beta_{3,4} x_3 x_4 + \varepsilon$$

con $\varepsilon \sim N(0, \sigma^2_{y/x_1, x_2, x_3, x_4, \text{cuadrático}})$.

Con los datos se obtienen las estimaciones y se hacen las pruebas para cada una de las 14 pruebas parciales sobre los coeficientes de regresión $H_0: \beta_i = 0$ con $H_A: \beta_i \neq 0$ para $i = 1,2,3,4$, $H_0: \beta_{ij} = 0$ con $H_A: \beta_{ij} \neq 0$ para $i = 1,2,3,4$ y $H_0: \beta_{ij} = 0$ con $H_A: \beta_{ij} \neq 0$ para $i = 1,2,3$, $j = i+1, i+2,4$. Para cada prueba se reporta el valor calculado de t y el valor p . Una revisión de la Tabla 13.2 indica que sólo son significativas las pruebas parciales de los coeficientes de glucosa, extracto de levadura, extracto de levadura cuadrática y la interacción glucosa por extracto de levadura. Esto sugiere que conviene tener un modelo cuadrático en glucosa y extracto de levadura.

Tabla 13.2. Pruebas de hipótesis parciales de la regresión cuadrática de producción de estreptomicina

Coeficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	1.109727	0.045353	24.468	2.94E ⁻¹²
Glucosa (β_1)	0.153215	0.024380	6.284	2.81E ⁻⁰⁵
Levadura (β_2)	0.084073	0.024380	3.448	0.004321
NaCl (β_3)	0.017525	0.024380	0.719	0.484979
K ₂ HPO ₄ (β_4)	0.039508	0.024380	1.621	0.129116
Glucosa cuadrada ($\beta_{1,1}$)	-0.061623	0.036264	-1.699	0.113050
Levadura cuadrada ($\beta_{2,2}$)	-0.176408	0.036264	-4.865	0.000309
NaCl cuadrado ($\beta_{3,3}$)	-0.015359	0.036264	-0.424	0.678824
K ₂ HPO ₄ ($\beta_{4,4}$)	0.048160	0.036264	1.328	0.207010
Glucosa*Levadura ($\beta_{1,2}$)	-0.060250	0.027257	-2.210	0.045613
Glucosa*NaCl ($\beta_{1,3}$)	-0.012875	0.027257	-0.472	0.644502
Glucosa* K ₂ HPO ₄ ($\beta_{1,4}$)	0.005875	0.027257	0.216	0.832690
Levadura*NaCl ($\beta_{2,3}$)	0.051500	0.027257	1.889	0.081342
Levadura* K ₂ HPO ₄ ($\beta_{2,4}$)	0.017500	0.027257	0.642	0.532006
NaCl* K ₂ HPO ₄ ($\beta_{3,4}$)	0.016375	0.027257	0.601	0.558325

En el análisis de varianza que se muestra en la Tabla 13.3, con las pruebas secuenciales (usando el residual de todo el modelo) muestra significación para glucosa cuadrada. Debido a que se tienen 13 grados de libertad para el residual de la regresión, se obtiene una estimación de $S^2_{glu*lev*NC*kp, cuadrático} = 0.01189$, que es menor que el error puro $S^2_{ep} = 0.021976$. Esto ocurre cuando hay mucha variación en los datos, la variación de las cuatro observaciones en el punto central (0,0,0) es mayor que la variación alrededor de la regresión estimada.

Tabla 13.3. Análisis de varianza de la regresión cuadrática para producción de estreptomicina.

Estreptomicina-4 factores					
Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
glu	1	0.46947	0.46947	39.4942	2.813E ⁻⁰⁵
lev	1	0.14136	0.14136	11.8918	0.0043210
NC	1	0.00614	0.00614	0.5167	0.4849792
KP	1	0.03122	0.03122	2.6260	0.1291160

glu2	1	0.07350	0.07350	6.1835	0.0272659
lev2	1	0.27665	0.27665	23.2735	0.0003321
NC2	1	0.00076	0.00076	0.0637	0.8047486
KP2	1	0.02096	0.02096	1.7637	0.2070105
glu:lev	1	0.05808	0.05808	4.8861	0.0456128
glu:NC	1	0.00265	0.00265	0.2231	0.6445022
glu:KP	1	0.00055	0.00055	0.0465	0.8326895
lev:NC	1	0.04244	0.04244	3.5700	0.0813418
lev:KP	1	0.00490	0.00490	0.4122	0.5320058
NC:KP	1	0.00429	0.00429	0.3609	0.5583249
Residuals	13	0.15453	0.01189		

Quitando las variables explicativas cuyos coeficientes no fueron significativos, se plantea el modelo cuadrático en glucosa y levadura que corresponde a la distribución siguiente:

$$y \sim N(\mu_{y/x1,x2,\text{cuadrático}} = \beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_{1,1} x1^2 + \beta_{2,2} x2^2 + \beta_{1,2} x1x2, \sigma_{y/x1,x2,\text{cuadrático}}^2),$$

que se escribe: $y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_{1,1} x1^2 + \beta_{2,2} x2^2 + \beta_{1,2} x1x2 + \varepsilon$, con $\varepsilon \sim N(0, \sigma_{y/x1,x2,\text{cuadrático}}^2)$.

Los estimadores del modelo son $\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x1 + \hat{\beta}_2 x2 + \hat{\beta}_{1,1} x1^2 + \hat{\beta}_{2,2} x2^2 + \hat{\beta}_{1,2} x1x2$,

evaluado con los datos da la ecuación estimada (Tabla 13.4):

$$\mu'_{y/x1,x2,\text{cuadrático}} = 1.12612 + 0.15321x1 + 0.08407 x2 - 0.05670x1^2 - 0.17148x2^2 - 0.06025x1x2$$

con las siguientes pruebas parciales:

Para glucosa en el modelo cuadrático de glucosa*levadura $H_0: \beta_1 = 0$, $H_A: \beta_1 \neq 0$, el valor $p = 3.04E^{-06}$ lleva a rechazar la hipótesis nula y el coeficiente de glucosa no es cero.

Para levadura en el modelo cuadrático de glucosa*levadura $H_0: \beta_2 = 0$, $H_A: \beta_2 \neq 0$, el valor $p = 0.00255$ lleva a rechazar la hipótesis nula y el coeficiente de levadura no es cero.

Para glucosa cuadrada en el modelo cuadrático de glucosa*levadura $H_0: \beta_{1,1} = 0$, $H_A: \beta_{1,1} \neq 0$, el valor $p = 0.12968$ indica que no hay evidencia para rechazar la hipótesis nula y el coeficiente de glucosa al cuadrado puede ser cero.

Para levadura cuadrada en el modelo cuadrático de glucosa*levadura $H_0: \beta_{2,2} = 0$, $H_A: \beta_{2,2} \neq 0$, con el valor $p = 9.4E^{-05}$ se rechaza la hipótesis nula y el coeficiente de levadura al cuadrado no es cero.

Para la interacción glucosa*levadura en el modelo cuadrático de glucosa*levadura $H_0: \beta_{1,2} = 0$, $H_A: \beta_{1,2} \neq 0$, con el valor $p = 0.04009$ se rechaza la hipótesis nula y el coeficiente del producto de glucosa por levadura no es cero. Todos estos valores están en la Tabla 13.4.

Tabla 13.4. Pruebas de hipótesis parciales de la regresión cuadrática considerando glucosa y extracto de levadura en la producción de estreptomicina.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	1.12612	0.03905	28.835	$< 2E^{-16}$
Glucosa (β_1)	0.15321	0.02470	6.203	$3.04E^{-06}$
Levadura (β_2)	0.08407	0.02470	3.404	0.00255
Glucosa cuadrada ($\beta_{1,1}$)	-0.05670	0.03601	-1.574	0.12968
Levadura cuadrada ($\beta_{2,2}$)	-0.17148	0.03601	-4.762	$9.40E^{-05}$
Glucosa*Levadura ($\beta_{1,2}$)	-0.06025	0.02762	-2.182	0.04009

El término glucosa al cuadrado no resultó significativo en la prueba parcial, como se mostró en la tabla anterior, pero en la prueba secuencial sí lo es, con valor p de 0.022494, como se muestran en la Tabla 13.5 del análisis de varianza. Se usará el modelo con los cinco términos para buscar el óptimo de producción de estreptomicina.

Tabla 13.5. Análisis de varianza de la regresión cuadrática considerando glucosa y extracto de levadura en la producción de estreptomicina.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Glucosa	1	0.46947	0.46947	38.4750	$3.044E^{-06}$
Levadura	1	0.14136	0.14136	11.5849	0.002549
Glucosa cuadrada	1	0.07350	0.07350	6.0239	0.022494
Levadura cuadrada	1	0.27665	0.27665	22.6729	$9.398E^{-05}$
Glucosa*Levadura	1	0.05808	0.05808	4.7600	0.040095
Residuales	22	0.26844	0.01220		

Para obtener los valores de glucosa y levadura que dan el máximo valor de producción de estreptomicina, en la ecuación estimada:

$$\mu'_{y/x_1, x_2, \text{cuadrático}} = 1.12612 + 0.15321x_1 + 0.08407x_2 - 0.05670x_1^2 - 0.17148x_2^2 - 0.06025x_1x_2,$$

se deriva con respecto a x_1 y x_2 , las derivadas se igualan a cero y se resuelve el sistema de ecuaciones resultantes.

Los valores codificados para el óptimo de glucosa y levadura son 1.346498 y 0.00855826, respectivamente. El valor óptimo de la media estimada de producción de estreptomicina vale 1.2296 y se obtiene para 26.7324 de glucosa y 25.0858 de levadura. El intervalo con 95% de confianza para la media estimada en los valores de glucosa y levadura para el óptimo es (1.096373, 1.362904). Note que para transformar los valores codificados de las variables explicativas a los valores originales se multiplica el valor codificado por el incremento entre los valores centrales sin codificar y este valor se suma al valor central sin codificar, así:

$$(20-15)*1.3464+20=26.7325 \text{ y } (25-15)*0.0085+25=25.0858.$$

Conviene mencionar que debido a que los coeficientes de regresión que correspondían a las variables explicativas NaCl y K_2HPO_4 no resultaron significativos, y que para fines prácticos valen cero, para la formulación del medio de cultivo, se puede usar cualquiera de los valores analizados para esos dos factores.

El análisis de producción de estreptomina se hace con las instrucciones del paquete R que siguen.

Guía para resolver con R

```

datosEstr<-read.table(file = "estreptomina.csv",header = T,sep = ",")

datosEstr

glu<-datosEstr$Glucosa

lev<-datosEstr$Levadura

NC<-datosEstr$NaCl

KP<-datosEstr$K2HPO4

y<-datosEstr$y

glu2<-glu*glu

lev2<-lev*lev

NC2<-NC*NC

KP2<-KP*KP

gluLev<-glu*lev

# glu # lev # NC # KP # y

par(mfrow = c(2,2))

plot(y~glu,main = " ",ylab = "Estreptomina", xlab = "Glucosa")

plot(y~lev,main = " ",ylab = "Estreptomina", xlab = "Levadura")

plot(y~NC,main = " ",ylab = "Estreptomina", xlab = "NaCl")

plot(y~KP,main = " ",ylab = "Estreptomina", xlab = "K2HPO4")

trata<-c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,17,17,21,22,23,24,25,26,27,28)

baseTrat<-data.frame(tratam = factor(trata),y = y)

baseTrat

tratamAdev<-lm(y~tratam,data = baseTrat)

summary(tratamAdev)

```

```

anova(tratamAdev)

EstrR2<-lm(y~glu+lev+NC+KP+glu2+lev2+NC2+KP2+glu:lev+glu:NC+glu:KP+lev:NC+lev:KP+NC:KP)

summary(EstrR2)

anova(EstrR2)

#EstrR2a<-lm(y~glu+lev+NC+KP+glu:lev+lev:NC)

#summary(EstrR2a)

#anova(EstrR2a)

EstrR2b<-lm(y~glu+lev+glu2+lev2+gluLev)

summary(EstrR2b)

anova(EstrR2b)

b0<-c(1.12612)

b1<-c(0.15321)

b2<-c(0.08407)

b11<-(-0.05670)

b22<-(-0.17148)

b12<-(-0.06025)

x1op<-(-2*b1*b22-b2*b12)/(b12^2-4*b11*b22)

x2op<-(-2*b2*b11-b1*b12)/(b12^2-4*b11*b22)

x1op

x2op

x1orig<-c(12.93,15,20,25,27.07)

x2orig<-c(10.86,15,25,35,39.14)

x1orgOp<-(x1orig[3]-x1orig[2])*x1op+x1orig[3]

x2orgOp<-(x2orig[3]-x2orig[2])*x2op+x2orig[3]

x1orgOp

x2orgOp

EsR2bp<-data.frame(glu = x1op,lev = x2op,glu2 = x1op^2,lev2 = x2op^2,gluLev = x1op*x2op)

```

```
icEstrR2b<-predict(EstrR2b,EsR2bp,interval = "confidence")
```

```
icEstrR2b
```

13.1 Ejercicio

Se hizo un experimento como el del problema 13 pero con los dos factores que resultaron significativos: con glucosa (x1) y extracto de levadura de cerveza (x2), centrando el experimento con 5 g/l más en ambas variables queda Glucosa: 17.93, 20, 25, 30 y 32.07 g/l (x1) y extracto de levadura de cerveza: 15.86, 20, 30, 40 y 44.14 g/l (x2). El diseño es compuesto central ortogonal para dos factores, con cuatro repeticiones en el punto central. Haga el análisis.

ensayo	1	2	3	4	5	6
x1	-1	-1	1	1	-1.414	1.414
x2	-1	1	-1	1	0	0
y	0.623	0.832	1.102	1.426	1.119	1.192
ensayo	7	8	9	10	11	12
x1	0	0	0	0	0	0
x2	-1.414	1.4140	0	0	0	0
y	0.974	1.020	1.183	1.772	1.156	1.214

Bibliografía

- Saval, S., Pablos, L. and Sanchez, S. (1993) Optimization of a culture medium for streptomycin production using response-surface methodology. *Bioresource Technology*. 43:19-25.

Nota: La referencia se utilizó, para plantear un problema similar pero llevado al nivel y el interés del curso.

14. Proceso de cocido por extrusión en amaranto

El amaranto es ampliamente utilizado en la industria de alimentos para obtener harinas de buena calidad nutricional. Uno de los puntos importantes en su obtención es el cocimiento y extrusión de la harina, de forma que pueda ser utilizada en diversas formulaciones y aplicaciones.

En el presente ensayo, se acondicionó harina de amaranto variedad Oscar Blanco hasta obtener la humedad adecuada para la extrusión, se guardó en bolsas de polietileno y se almacenó a 5°C por 48 horas. Para la extrusión se usó un aparato con barril de 20 mm de diámetro y relación 20:1 alimentando 70 g de harina/minuto. Los factores que se estudiaron fueron temperatura (T) del barril (parte central) y contenido de humedad (H) de la masa. Las variables de respuesta fueron la tasa de expansión (TE, diámetro del producto extruido sobre diámetro del hoyo), y densidad mg/mm³ (D, en ambas variables se tomó el promedio de 32 mediciones). El objetivo del trabajo es ajustar la regresión de segundo orden para obtener los valores de las variables explicativas que maximizan el valor de la variable de respuesta. Se usó un diseño central compuesto, los niveles de temperatura evaluados fueron 128.79, 135, 150, 165 y 171.21 y los niveles de humedad fueron 12.172, 13, 15, 17 y 17.828. Los datos obtenidos se muestran en la Tabla 14.1.

Tabla 14.1. Arreglo experimental codificado y datos obtenidos por tratamiento.

Ensayo	Humedad codificada	Temperatura codificada	Tasa de Expansión (TE)	Densidad (DE) mg/mm ³
1	-1	-1	2.45	.239
2	-1	1	2.48	.228
3	1	-1	2.47	.254
4	1	1	2.32	.236
5	1.414	0	2.42	.248
6	-1.414	0	2.48	.243
7	0	1.414	2.53	.216
8	0	-1.414	2.48	.239
9	0	0	2.84	.227
10	0	0	2.76	.229
11	0	0	2.75	.220
12	0	0	2.70	.224
13	0	0	2.67	.219
14	0	0	2.65	.236

Una visión preliminar de la relación de las variables de respuesta con las explicativas se presenta en la Figura 14.1. Las gráficas de tasa de expansión y la de densidad con humedad muestran una tendencia cuadrática, la gráfica de densidad con temperatura pudiera ser cuadrática, pero no es claro. En las cuatro gráficas se ve mucha variación en los puntos con valor cero en el eje horizontal, pero se puede esperar que los dos puntos que se alejan de los otros correspondan a los puntos que no son centrales, esto es, que tienen valores diferentes de cero para la variable que no está en la gráfica.

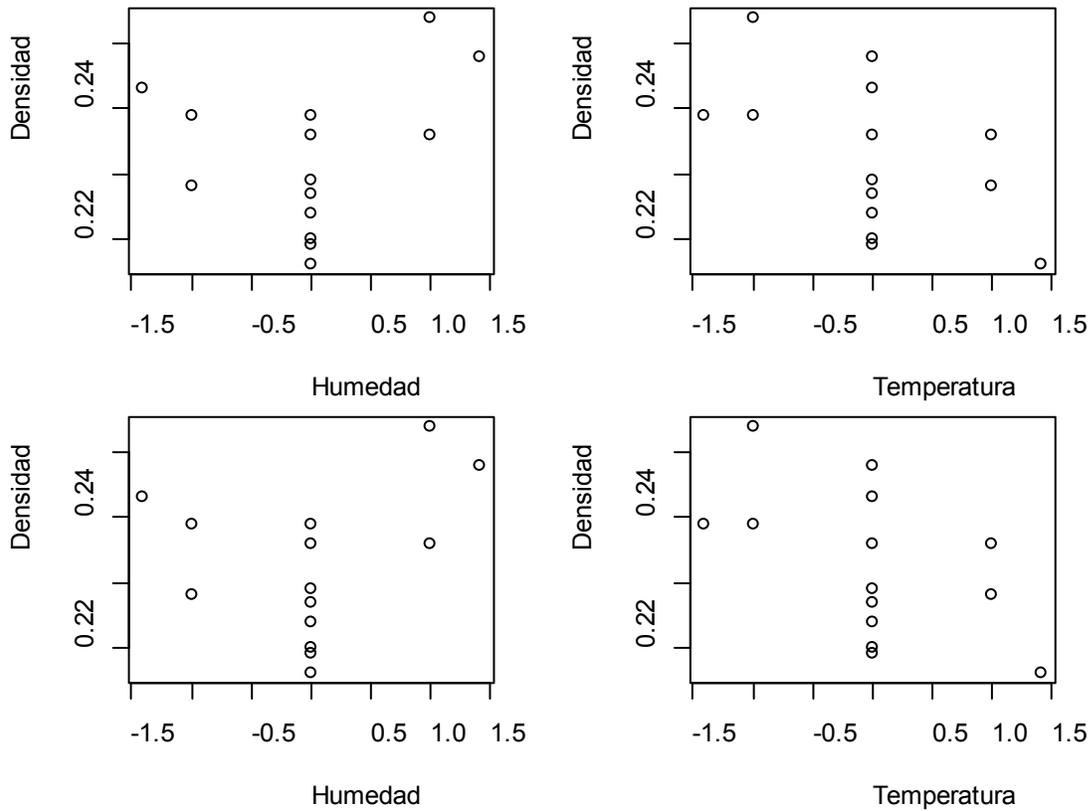


Figura 14.1. Diagrama de dispersión de la tasa de expansión sobre humedad (a); tasa de expansión sobre temperatura (b); densidad sobre humedad (c) y densidad sobre temperatura (d).

Los puntos centrales del diseño se repiten para tener una estimación del error puro, que es la variación de las observaciones centrales que tienen el mismo tratamiento. El análisis de varianza del diseño da la estimación de la variación o error puro, la que se obtiene a partir de los seis puntos centrales. Para tasa de expansión $S^2TE = 0.004857$ y para densidad $S^2D = 0.000039767$ con cinco grados de libertad, estos valores, junto con sus respectivas sumas de cuadrados 0.024283 y 0.00019883 servirán para conocer la falta de ajuste de las regresiones.

Se estudia por separado a la tasa de expansión, después, como ejercicio se hará un estudio similar para la densidad.

Se plantea el modelo cuadrático, donde la distribución de TE es:

$$TE \sim N(\mu_{TE/H,T, \text{cuadrático}} = \beta_0 + \beta_1 H + \beta_2 T + \beta_{1,1} H^2 + \beta_{2,2} T^2 + \beta_{1,2} H * T, \sigma_{TE/H,T, \text{cuadrático}}^2),$$

que (recuerde), se acostumbra escribir como sigue:

$$TE = \beta_0 + \beta_1 H + \beta_2 T + \beta_{1,1} H^2 + \beta_{2,2} T^2 + \beta_{1,2} H * T + \varepsilon \text{ con } \varepsilon \sim N(0, \sigma_{TE/H,T, \text{cuadrático}}^2),$$

este modelo da lugar al modelo de estimadores:

$$\hat{\mu}_{TE/H,T, \text{cuadrático}} = \hat{\beta}_0 + \hat{\beta}_1 H + \hat{\beta}_2 T + \hat{\beta}_{1,1} H^2 + \hat{\beta}_{2,2} T^2 + \hat{\beta}_{1,2} H * T .$$

el cual, al evaluarse en los valores de los datos produce el modelo estimado:

$$\mu'_{TE/H,T, \text{cuadrático}} = 2.728338 - 0.028109H - 0.006163T - 0.151071H^2 - 0.123562T^2 - 0.045000H*T$$

En la Tabla 14.2 se presentan los coeficientes de regresión estimados con los valores de t y valores p para las pruebas parciales. Para cada coeficiente β_i ; $i = 1, 2, 1,1, 2,2, 1,2$ se plantea la hipótesis de que el coeficiente es cero $H_0: \beta_i = 0$; $H_A: \beta_i \neq 0$. En los valores p se ve que sólo los coeficientes de las variables explicativas al cuadrado resultaron significativos con valores p de 0.000227 y 0.000851.

Tabla 14.2. Pruebas de hipótesis parciales de la regresión cuadrática para tasa de extrusión.

Coeficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	2.728338	0.026499	102.959	8.85E ⁻¹⁴
Humedad (β_1)	-0.028109	0.022951	-1.225	0.255502
Temperatura (β_2)	-0.006163	0.022951	-0.269	0.795071
Humedad cuadrada ($\beta_{1,1}$)	-0.151071	0.023891	-6.323	0.000227
Temperatura cuadrada ($\beta_{2,2}$)	-0.123562	0.023891	-5.172	0.000851
Humedad*Temperatura ($\beta_{1,2}$)	-0.045000	0.032455	-1.387	0.202995

En la Tabla 14.3 del análisis de varianza se presentan las pruebas secuenciales, que se pueden usar para comprobar o modificar los resultados de las pruebas parciales; son las mismas parejas de hipótesis nula y alternativa, pero aplicadas en secuencia a modelos en los que cada vez se incluye la variable explicativa que identifica al renglón. En este caso se obtiene el mismo tipo de resultado, sólo los coeficientes de humedad al cuadrado y temperatura al cuadrado resultan significativos (valores p de 0.0003444 y 0.0008512). El coeficiente de determinación (R^2) vale 0.8913, indicando que la regresión explica el 89.13% de la variación de la variable de respuesta tasa de expansión.

La prueba de falta de ajuste, se refiere a las hipótesis H_0 : No hay falta de ajuste y H_A : Hay falta de ajuste; entendiendo por falta de ajuste que la diferencia de la variación residual en el modelo ajustado y la del error puro, que se llama variación de falta de ajuste, es mayor que la variación del error puro. Se puede escribir $H_0: \sigma^2_{\text{Falta de ajuste}} = \sigma^2_{\text{TE igual Tratamiento}}$; $H_A: \sigma^2_{\text{Falta de ajuste}} > \sigma^2_{\text{TE igual Tratamiento}}$. Esta hipótesis se prueba con una F, que se obtiene usando la ecuación:

$$F_c = \frac{(\text{scResReg} - \text{scResEPuro}) / (\text{glResReg} - \text{glResEPuro})}{S^2_{\text{EPuro}}}$$

donde scRes se refiere a suma de cuadrados residual scResReg es la suma de cuadrados residual de la regresión y scResEPuro es la suma de cuadrados del error puro. Del mismo modo se hace referencia a los grados de libertad usando gl. La ecuación se evalúa como sigue:

$$F_c = \frac{(0.033706 - 0.024283) / (8 - 5)}{0.004857} = 0.6466$$

El residual de $S^2_{TE/H,T, \text{cuadrático}} = 0.004213$ es muy similar al del error puro $S^2_{TE} = 0.004857$, y la prueba de falta de ajuste da una F_c de 0.6466 con 3 y 5 grados de libertad, que no da evidencia para rechazar la hipótesis nula de que no hay falta de ajuste, ya que el valor p vale = 0.6178, lo que indica que el modelo se ajusta de manera adecuada.

Tabla 14.3. Análisis de varianza de la regresión cuadrática para la tasa de extrusión.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Humedad	1	0.006320	0.006320	1.5000	0.2555019
Temperatura	1	0.000304	0.000304	0.0721	0.7950713
Humedad cuadrada	1	0.148845	0.148845	35.3277	0.0003444
Temperatura cuadrada	1	0.112696	0.112696	26.7479	0.0008512
Humedad*Temperatura	1	0.008100	0.008100	1.9225	0.2029952
Residual	8	0.033706	0.004213		
R cuadrada = 0.8913			R cuadrada ajustada = 0.8233		

Aunque los coeficientes de H, T y H*T no sean significativos, es decisión del investigador dejar el modelo cuadrático completo o plantear un modelo reducido para buscar el óptimo. Como ejercicio, conviene plantear el modelo reducido.

$TE \sim N(\mu_{TE/H^2, T^2} = \beta_0 + \beta_{1,1}H^2 + \beta_{2,2}T^2, \sigma_{TE/H^2, T^2}^2)$, o escrito de otra manera:

$$TE = \beta_0 + \beta_{1,1}H^2 + \beta_{2,2}T^2 + \varepsilon \text{ con } \varepsilon \sim N(0, \sigma_{TE/H^2, T^2}^2)$$

El modelo estimado es:

$$\mu'_{TE/H^2, T^2} = 2.72834 - 0.15107 H^2 - 0.12356 T^2$$

En la Tabla 14.4 se ve que los coeficientes de las variables explicativas al cuadrado resultan significativos con valores p de 0.0000685 y 0.000367.

Tabla 14.4. Pruebas de hipótesis parciales de la regresión cuadrática con el modelo reducido.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	2.72834	0.02709	100.719	$< 2E^{-16}$
Humedad cuadrada ($\beta_{1,1}$)	-0.15107	0.02442	-6.186	0.0000685
Temperatura cuadrada ($\beta_{2,2}$)	-0.12356	0.02442	-5.059	0.000367

Para comprobar o modificar los resultados de las pruebas parciales, en la Tabla 14.5 del análisis de varianza se presentan las pruebas secuenciales. En este caso se obtiene el mismo tipo de resultado, los coeficientes de humedad al cuadrado y temperatura al cuadrado resultan significativas (valores p de 0.0001168 y 0.0003667). El valor del coeficiente de determinación (0.8438) está indicando que la regresión explica el 84.38% de la variación de la variable respuesta tasa de expansión.

Tabla 14.5. Análisis de varianza de la regresión cuadrática con el modelo reducido.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Humedad cuadrada	1	0.14885	0.148845	33.807	0.0001168
Temperatura cuadrada	1	0.11270	0.112696	25.597	0.0003667
Residual	11	0.04843	0.004403		
R cuadrada = 0.84376			R cuadrada ajustada = 0.81535		

Como pudo apreciarse, el valor del coeficiente de determinación (0.8438) es menor que el del modelo completo (0.8913), por lo que el investigador puede optar por el modelo completo. Los modelos cuadrático completo y el modelo reducido con solo los términos cuadráticos son jerárquicos, ya que el segundo modelo se obtuvo eliminando del primer modelo variables explicativas y sus respectivos coeficientes. Debido a esto, es mejor comparar los valores de las R cuadradas ajustadas, que son 0.8233 y 0.81535 (Tablas 14.5 y 14.3). Se ve que la diferencia de las R cuadradas ajustadas es menor que la que se obtuvo con los valores de R cuadrada (sin ajustar), esto quiere decir que al quitar las variables explicativas H y T el modelo reducido explica casi la misma variación que el modelo que las incluye.

Para completar el ejercicio, siguiendo con el modelo reducido, la prueba de falta de ajuste, que se refiere a las hipótesis H0: No hay falta de ajuste y HA: Hay falta de ajuste, se prueba con:

$$F_c = \frac{(\text{scResReg} - \text{scResEPuro}) / (\text{glResReg} - \text{glResEPuro})}{S_{\text{EPuro}}^2}$$

que se evalúa con los valores ya dados de:

$$F_c = \frac{(0.04843 - 0.024283) / (11 - 5)}{0.004858} = 0.8286$$

con 6 y 5 grados de libertad, se obtiene un valor p de 0.5935 y no hay evidencia para rechazar la hipótesis de que no hay falta de ajuste, el modelo es adecuado desde este punto de vista.

El óptimo se obtiene como en el modelo completo. En la ecuación estimada se deriva con respecto a las variables explicativas H y T, las derivadas se igualan a cero y se resuelve el sistema de ecuaciones. En este caso, al igualar a cero se obtiene cero para los valores codificados de Hc = 0 y Tc = 0, que corresponde a los valores de H = 150 y T = 15. El óptimo estimado para la media de tasa de expansión vale 2.72834.

El intervalo con 95% de confianza para la media estimada en los valores de humedad y temperatura para el óptimo es (2.668717, 2.78796). El hecho de que el modelo con sólo las variables al cuadrado (modelo cuadrático reducido) produzca siempre el óptimo en los valores codificados de cero, hace que los investigadores prefieran el modelo cuadrático completo.

Los óptimos en el modelo cuadrático completo para Hc y Tc se obtienen con las ecuaciones que resultan de derivar la ecuación de regresión estimada con respecto a Hc y TC, luego igualar las derivadas a cero y resolver. Denotando bi al coeficiente i de la regresión se tienen las ecuaciones:

$$Hc_{\text{Opt}} = (2 * b_1 * b_{22} - b_2 * b_{12}) / (b_{12}^2 - 4 * b_{11} * b_{22}) \quad Tc_{\text{Opt}} = (2 * b_2 * b_{11} - b_1 * b_{12}) / (b_{12}^2 - 4 * b_{11} * b_{22})$$

El modelo cuadrático completo produce el óptimo en los valores codificados de Hc = -0.09181 y Tc = -0.0082211, que corresponden a los valores de H = 148.62289 y T = 14.98356. El óptimo vale 2.72965. El intervalo con 95% de confianza para la media estimada en los valores de humedad y temperatura para el óptimo es (2.66861, 2.790698).

La relación entre densidad y tasa de expansión se ve en la Figura 14.2, donde se ve una tendencia cuadrática con mucha variación.

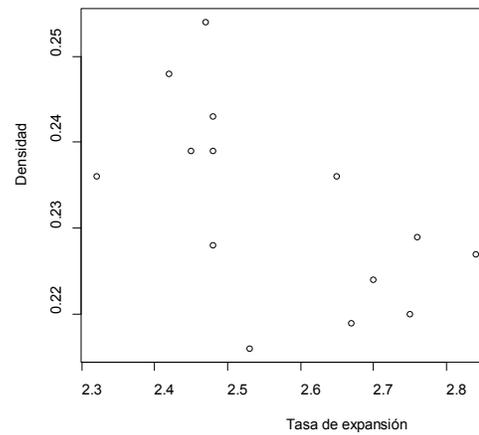


Figura 14.2. Diagrama de dispersión de densidad sobre tasa de expansión.

Para la regresión se plantea el modelo siguiente:

$$D \sim N(\mu_{D/TE, TE^2} = \beta_0 + \beta_1 TE + \beta_{1,1} TE^2, \sigma_{D/TE, TE^2}^2)$$

$$D = \beta_0 + \beta_1 TE + \beta_{1,1} TE^2 + \varepsilon \text{ con } \varepsilon \sim N(0, \sigma_{D/TE, TE^2}^2) . \text{ El modelo de estimadores es}$$

$$\hat{\mu}_{D/TE, TE^2} = \hat{\beta}_0 + \hat{\beta}_1 TE + \hat{\beta}_{1,1} TE^2, \text{ que al evaluarlo en los datos produce:}$$

$$\mu'_{D/TE, TE^2} = 0.56611 - 0.21622 TE + 0.03355 TE^2$$

Se obtiene la Tabla 14.6 con las pruebas parciales y se ve que ninguna de ellas sobre los coeficientes de TE o TE² son significativos, todos los valores p son mayores que 0.05.

Tabla 14.6. Pruebas de hipótesis parciales de la regresión cuadrática para densidad.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	0.56611	0.88534	0.639	0.536
T. de Expansión (β_1)	-0.21622	0.68557	-0.315	0.758
T. de Expansión al cuadrado ($\beta_{1,1}$)	0.03355	0.13236	0.253	0.805

En la Tabla 14.7 del análisis de varianza se tienen las pruebas secuenciales y sólo es significativo el coeficiente de TE con un valor p de 0.03743. El modelo explica el 33.98% de la variación y es claro que no conviene tratar de explicar la media de densidad con la variable explicativa tasa de expansión.

Tabla 14.7. Análisis de varianza de la regresión cuadrática para densidad. Elaboración propia del taller de diseño experimental

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
T. de Expansión	1	0.00055993	0.00055993	5.5967	0.03743
T. de Expansión alcuadrado	1	0.00000643	0.00000643	0.0643	0.80457
Residual	11	0.00110050	0.00010005		

El análisis de tasa de expansión se hace con las instrucciones del paquete R que siguen.

Guía para resolver con R

```

datosAm<-read.table(file = "amaranto.csv",header = T,sep = ",")

datosAm

TE<-datosAm$TE

D<-datosAm$D

Hc<-datosAm$Hc

Tc<-datosAm$Tc

Hc2<-Hc*Hc

Tc2<-Tc*Tc

HcTc<-Hc*Tc

# TE # D # Hc # Tc # Hc2 # Tc2 # HcTc

par(mfrow = c(2,2))

plot(TE~Hc,main = "",ylab = "Tasa de expansión", xlab = "Humedad")

plot(TE~Tc,main = "",ylab = "Tasa de expansión", xlab = "Temperatura")

plot(D~Hc,main = "",ylab = "Densidad", xlab = "Humedad")

plot(D~Tc,main = "",ylab = "Densidad", xlab = "Temperatura")

tr<-c(1,2,3,4,5,6,7,8,9,9,9,9,9,9)

baseTrat<-data.frame(trat = factor(tr),Te = TE,D = D)

baseTrat

tTEAdev<-aov(TE~trat,data = baseTrat)

options(digits = 8)

anova(tTEAdev)

```

```
tDAdev<-aov(D~trat,data = baseTrat)
anova(tDAdev)
TER2<-lm(TE~Hc+Tc+Hc2+Tc2+HcTc)
summary(TER2)
anova(TER2)
Fcalcu <-((0.033706-0.024283)/(8-5))/0.004857
Fcalcu
valorp<-1-pf(Fcalcu,3,5)
valorp
TER2r<-lm(TE~Hc2+Tc2)
summary(TER2r)
anova(TER2r)
Fcalcu <-((0.04843-0.024283)/(11-5))/0.004857
Fcalcu
valorp<-1-pf(Fcalcu,6,5)
valorp
Hc2opt<-0
Tc2opt<-0
TER2rp<-data.frame(Hc2 = Hc2opt,Tc2 = Tc2opt)
icTER2r<-predict(TER2r,TER2rp,interval = "confidence")
icTER2r
TER2<-lm(TE~Hc+Tc+Hc2+Tc2+HcTc)
summary(TER2)
b0<-c(2.728338)
b1<-c(-0.028109)
b2<-c(-0.006163)
b11<-c(-0.151071)
```

```

b22<-(-0.123562)

b12<-(-0.045000)

x1op<-(2*b1*b22-b2*b12)/(b12^2-4*b11*b22)

x2op<-(2*b2*b11-b1*b12)/(b12^2-4*b11*b22)

x1op

x2op

x1orig<-c(128.79,135,150,165,171.21)

x2orig<-c(12.172,13,15,17,17.828)

x1orgOp<-(x1orig[3]-x1orig[2])*x1op+x1orig[3]

x2orgOp<-(x2orig[3]-x2orig[2])*x2op+x2orig[3]

x1orgOp

x2orgOp

TER2p<-data.frame(Hc = x1op,Tc = x2op,Hc2 = x1op^2,Tc2 = x2op^2,HcTc = x1op*x2op)

icTER2<-predict(TER2,TER2p,interval = "confidence")

icTER2

par(mfrow = c(1,1))

plot(D~TE,main = " ",xlab = "Tasa de expansión",ylab = "Densidad")

TE2 = TE*TE

DTER2<-lm(D~TE+TE2)

summary(DTER2)

anova(DTER2)

```

14.1 Ejercicio

Haga el análisis de los datos de densidad del problema anterior.

15. Optimización de producción de dextrina

Con el objetivo de optimizar el medio de cultivo de *Metarhizium anisopliae* para la producción de dextrina A y B, se hizo un experimento usando un diseño central compuesto con $\alpha = 2$ en cuatro dimensiones. Los factores evaluados fueron maltosa, pectina, β -alanina y glucosa, las variables de respuesta fueron la producción de dextrina A (DA) y de dextrina B (DB). La unidad experimental fue un frasco Erlenmeyer de 500 ml con 100 ml de medio líquido y 5% de inóculo. Se dejó crecer el hongo durante catorce días en una incubadora rotatoria a 28°C.

Los niveles de los factores están en la Tabla 15.1.

Tabla 15.1. Factores y niveles codificados estudiados.

Factor	Notación	-2	-1	0	1	2
Maltosa	M	2.1	2.3	2.5	2.7	2.9
Pectina	P	0.3	.05	0.7	0.9	1.1
β -alanina	bA	0	0.01	0.02	0.03	0.04
Glucosa	G	0.04	0.20	0.36	0.52	0.68

Se tuvieron 26 puntos experimentales incluyendo el punto central con dos repeticiones y los resultados están en la Tabla 15.2.

Tabla 15.2. Arreglo experimental codificado y resultados obtenidos en cada tratamiento.

Ensayo	M	P	Ba	G	DA	DB
1	-1	-1	-1	-1	5.7	92
2	-1	-1	-1	1	6.2	148
3	-1	-1	1	-1	10.9	145
4	-1	-1	1	1	8.6	125
5	-1	1	-1	-1	5.0	85
6	-1	1	-1	1	10.5	184
7	-1	1	1	-1	3.6	185
8	-1	1	1	1	12.5	218
9	1	-1	-1	-1	6.9	99
10	1	-1	-1	1	14.6	203
11	1	-1	1	-1	6.0	146
12	1	-1	1	1	12.3	94
13	1	1	-1	-1	12.9	155
14	1	1	-1	1	8.2	165

15	1	1	1	-1	5.1	141
16	1	1	1	1	16.4	150
17	-2	0	0	0	1.9	56
18	2	0	0	0	11.2	144
19	0	-2	0	0	5.4	75
20	0	2	0	0	6.4	125
21	0	0	-2	0	8.7	121
22	0	0	2	0	13.4	148
23	0	0	0	-2	11.6	141
24	0	0	0	2	14.2	163
25	0	0	0	0	17.0	223
26	0	0	0	0	17.1	233

Se hará el análisis para dextrina A (DA). Las gráficas de dispersión de producción de dextrina A sobre las variables explicativas muestran mucha variación (Figura 15.1).

El diseño tiene sólo dos repeticiones del punto central y la estimación de la varianza residual $S^2_{DA} = 0.005$ tiene un grado de libertad, por lo que no es muy apropiado como estimador de la varianza ni para pruebas de falta de ajuste.

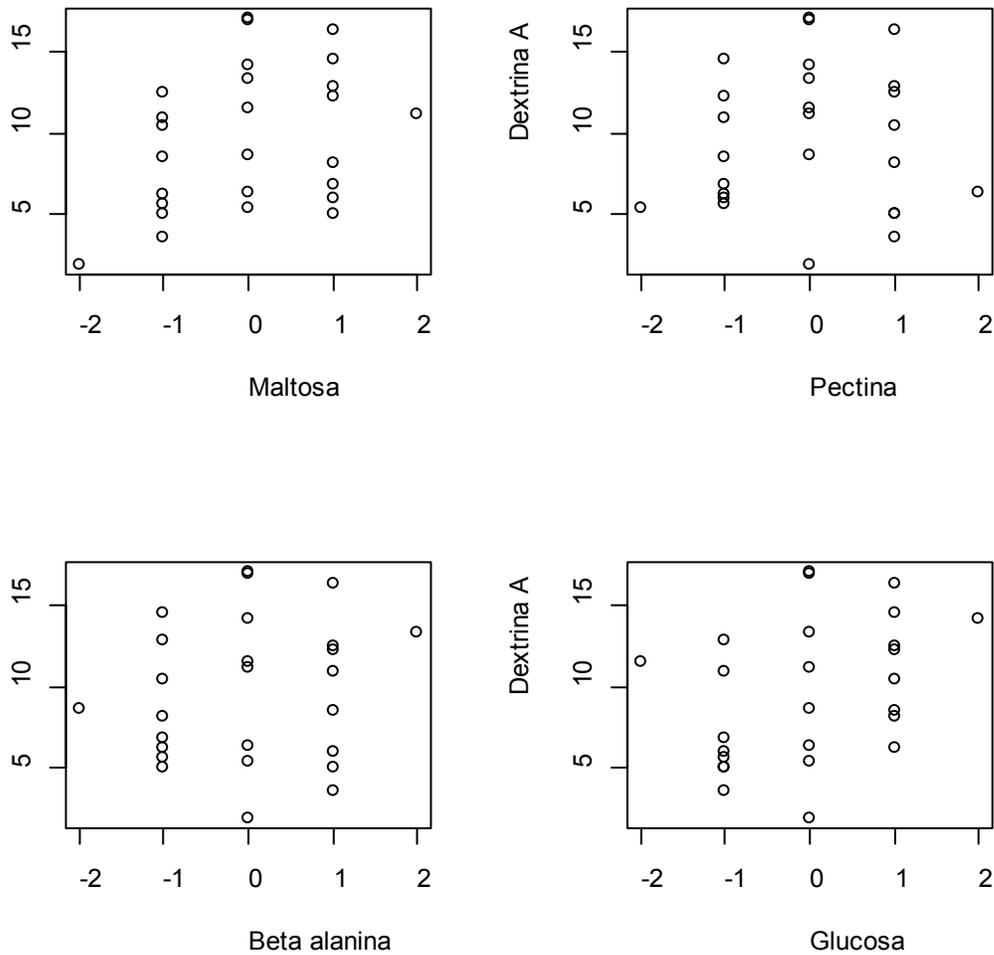


Figura 15.1. Diagrama de dispersión de producción de dextrina A sobre: maltosa (a); pectina (b); β -alanina (c) y glucosa (d).

La regresión cuadrática de las cuatro variables explicativas sobre la producción de dextrina tipo A se hace con la distribución:

$$DA \sim N(\mu_{DA/x1,x2,x3,x4,\text{cuadrático}} = \beta_0 + \beta_1x1 + \beta_2x2 + \beta_3x3 + \beta_4x4 + \beta_{1,1}X1^2 + \beta_{2,2}X2^2 + \beta_{3,3}X3^2 + \beta_{1,2}x1x2 + \beta_{1,3}x1x3 + \beta_{1,4}x1x4 + \beta_{2,3}x2x3 + \beta_{2,4}x2x4 + \beta_{3,4}x3x4, \sigma_{DA/x1,x2,x3,x4,\text{cuadrático}}^2)$$

En el modelo se usan las etiquetas x_1 , x_2 , x_3 y x_4 para representar a las variables: maltosa, pectina, β -alanina y glucosa, respectivamente.

Las hipótesis parciales de que cada coeficiente del modelo es cero, se prueban con el valor p que le corresponde en la fila de la Tabla 15.3. Los valores de p de los coeficientes de maltosa, glucosa, maltosa cuadrada y pectina cuadrada son menores que 0.05, por tanto, con 5% de significación se rechaza la hipótesis nula de que los respectivos coeficientes de regresión valen cero. Con 10% de significación se rechaza la hipótesis nula de que el coeficiente de regresión de β -alanina cuadrada es cero. Las pruebas para los demás coeficientes no dan evidencia para rechazar la hipótesis de que valen cero.

Tabla 15.3. Pruebas de hipótesis parciales de la regresión cuadrática para producción de dextrina A.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	17.0500	2.3153	7.364	1.42E ⁻⁰⁵
Maltosa (β_1)	1.5833	0.6684	2.369	0.03722
Pectina (β_2)	0.2083	0.6684	0.312	0.76109
β -alanina (β_3)	0.6167	0.6684	0.923	0.37598
Glucosa (β_4)	1.6000	0.6684	2.394	0.03561
Maltosa cuadrada ($\beta_{1,1}$)	-2.6271	0.7837	-3.352	0.00645
Pectina cuadrada ($\beta_{2,2}$)	-2.7896	0.7837	-3.559	0.00448
β -alanina cuadrada ($\beta_{3,3}$)	-1.5021	0.7837	-1.917	0.08162
Glucosa cuadrada ($\beta_{4,4}$)	-1.0396	0.7837	-1.326	0.21157
Maltosa*Pectina ($\beta_{1,2}$)	0.1625	0.8186	0.199	0.84626
Maltosa* β -alanina ($\beta_{1,3}$)	-0.6875	0.8186	-0.840	0.41886
Maltosa*Glucosa ($\beta_{1,4}$)	0.5000	0.8186	0.611	0.55373
Pectina* β -alanina ($\beta_{2,3}$)	-0.2125	0.8186	-0.260	0.79997
Pectina*Glucosa ($\beta_{2,4}$)	0.5500	0.8186	0.672	0.51551
β -alanina*Glucosa ($\beta_{3,4}$)	0.9500	0.8186	1.161	0.27039

La Tabla 15.4 del análisis de varianza con las pruebas secuenciales (usando el residual de todo el modelo) muestra significación al 5% para los coeficientes de las mismas variables, pero no presenta evidencia para rechazar con 10% de significación la hipótesis nula de que el coeficiente de β -alanina cuadrada vale cero.

Tabla 15.4. Análisis de varianza de la regresión cuadrática para dextrina.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F calculada	Valor p
Maltosa	1	60.167	60.167	5.6120	0.03722
Pectina	1	1.042	1.042	0.0972	0.76109
β -alanina	1	9.127	9.127	0.8513	0.37598
Glucosa	1	61.440	61.440	5.7308	0.03561
Maltosa cuadrada	1	47.646	47.646	4.4442	0.05877

Pectina cuadrada	1	97.067	97.067	9.0539	0.01189
β -alanina cuadrada	1	23.319	23.319	2.1751	0.16830
Glucosa cuadrada	1	18.864	18.864	1.7595	0.21157
Maltosa*Pectina	1	0.422	0.422	0.0394	0.84626
Maltosa* β -alanina	1	7.563	7.563	0.7054	0.41886
Maltosa*Glucosa	1	4.000	4.000	0.3731	0.55373
Pectina* β -alanina	1	0.722	0.722	0.0674	0.79997
Pectina*Glucosa	1	4.840	4.840	0.4515	0.51551
β -alanina*Glucosa	1	14.440	14.440	1.3469	0.27039
Residual	11	117.931	10.721		

Se decide reducir el modelo y dejar sólo las variables significativas al 5%: maltosa, glucosa, maltosa cuadrada y pectina cuadrada. La distribución de DA es:

$$DA \sim N(\mu_{DA/x1,x4,x1^2,x2^2} = \beta_0 + \beta_1x1 + \beta_4x4 + \beta_{1,1}x1^2 + \beta_{2,2}x2^2, \sigma_{DA/x1,x4,x1^2,x2^2}^2)$$

con modelo de estimadores:

$$\hat{\mu}_{DA/x1,x4,x1^2,x2^2} = \hat{\beta}_0 + \hat{\beta}_1x1 + \hat{\beta}_4x4 + \hat{\beta}_{1,1}x1^2 + \hat{\beta}_{2,2}x2^2$$

y modelo estimado:

$$\mu'_{DA/x1,x4,x1^2,x2^2} = 13.2375 + 1.5833x1 + 1.6000x4 - 1.8328x1^2 - 1.9953x2^2$$

Las pruebas parciales en este modelo con x_1 , x_4 , x_1^2 y x_2^2 se refieren a las parejas de hipótesis:

$$H_0: \beta_{Mc} = 0, H_A: \beta_{Mc} \neq 0, \text{ que da un valor p de } 0.02081,$$

$$H_0: \beta_{Gc} = 0, H_A: \beta_{Gc} \neq 0, \text{ que da un valor p de } 0.01966,$$

$$H_0: \beta_{Mc^2} = 0, H_A: \beta_{Mc^2} \neq 0, \text{ que da un valor p de } 0.00826$$

y

$$H_0: \beta_{pc^2} = 0, H_A: \beta_{pc^2} \neq 0, \text{ que da un valor p de } 0.00457,$$

Las cuatro hipótesis nulas se rechazan, los coeficientes de maltosa glucosa, maltosa cuadrada y pectina cuadrada son diferentes de cero como se observa en la Tabla 15.5.

Tabla 15.5. Pruebas de hipótesis parciales de la regresión cuadrática para dextrina A con el modelo reducido.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	13.2375	1.0973	12.064	$6.6E^{-11}$
Maltosa (β_1)	1.5833	0.6335	2.499	0.02081
Glucosa (β_4)	1.6000	0.6335	2.526	0.01966
Maltosa cuadrada ($\beta_{1,1}$)	-1.8328	0.6285	-2.916	0.00826
Pectina cuadrada ($\beta_{2,2}$)	-1.9953	0.6285	-3.175	0.00457

Las pruebas secuenciales del análisis de varianza producen valores p que llevan al rechazo de las hipótesis de que los coeficientes valen cero.

En la gráfica de residuales se ven residuales muy grandes, mientras que la gráfica de probabilidad normal presenta una curvatura acentuada en los dos extremos (Figura 15.2), esto indica que hay discrepancias para la suposición de distribución normal.

La prueba de normalidad de Shapiro-Wilk para los residuales plantea la pareja de hipótesis:

H0: Los residuales se distribuyen de acuerdo con la distribución normal, con

HA: Los residuales no se distribuyen de acuerdo con la distribución normal.

Haciendo la prueba se obtiene un valor p de 0.2433, por lo que la muestra no da evidencia para rechazar la suposición de que los residuales se distribuyen normalmente.

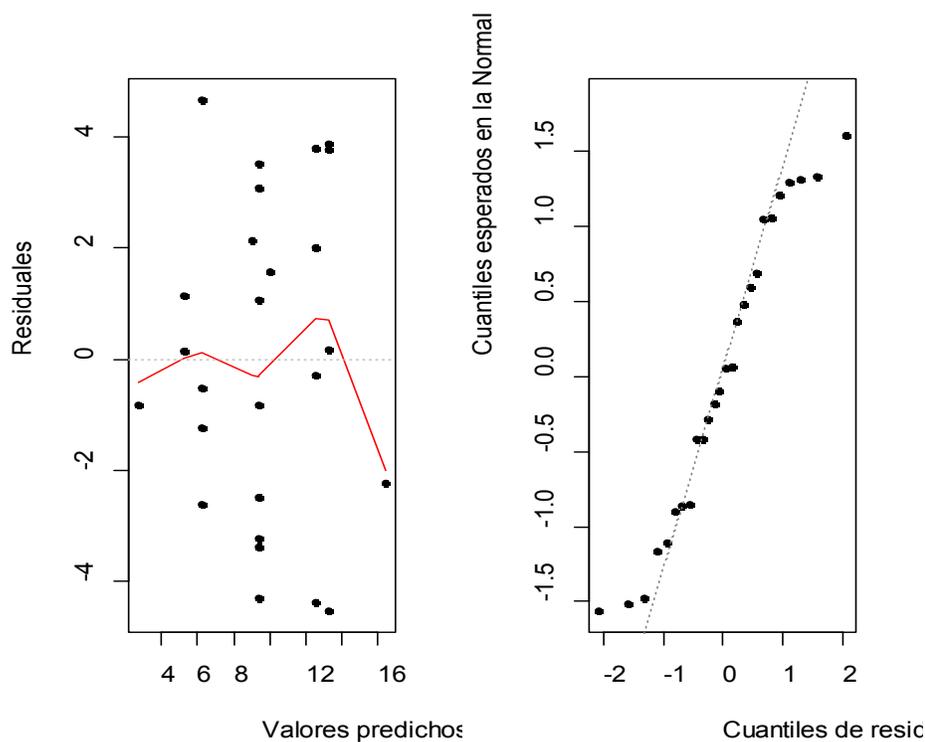


Figura 15.2. Residuales contra valores predichos (a) y gráfica de probabilidad normal (b).

Para obtener un valor máximo de producción de dextrina, se deriva la ecuación estimada con respecto a las variables explicativas maltosa, glucosa y pectina. El modelo tiene lineal y cuadrático de maltosa (x_1), lineal de glucosa (x_4) y cuadrático de pectina (x_2). Para pectina, al derivar se obtiene que el máximo ocurre en el valor codificado como 0. Para glucosa, lineal positivo, el máximo ocurre en la frontera del área de exploración, el valor codificado 2, mientras que para maltosa se obtiene una ecuación que da el máximo cuando x_1 vale 0.4319 (codificado).

La relación entre dextrina tipo B y dextrina tipo A se ve en la Figura 15.3, donde se ve una tendencia lineal con mucha variación. Para la regresión se plantea el modelo siguiente:

$$DB \sim N(\mu_{DB/DA} = \beta_0 + \beta_1 DA, \sigma_{DB/DA}^2)$$

la evaluación del modelo de estimadores:

$$\hat{\mu}_{DB/DA} = \hat{\beta}_0 + \hat{\beta}_1 DA \text{ da el modelo de estimación}$$

$$\mu'_{DB/DA} = 78.159464 + 6.864265 DA$$

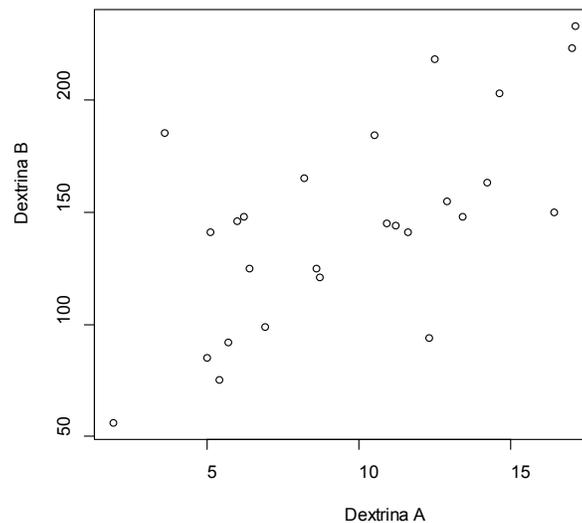


Figura 15.3. Diagrama de dispersión de producción de dextrina B sobre producción de dextrina A.

Se obtiene la Tabla 15.5 con las pruebas parciales y se ve que el coeficiente de dextrina A, para la pareja de hipótesis $H_0: \beta_{DA} = 0$ con la hipótesis alternativa $H_A: \beta_{DA} \neq 0$, tiene valor $p = 0.000303$. Se rechaza la hipótesis nula y el coeficiente de DA en la regresión es diferente de cero.

Tabla 15.5. Pruebas de hipótesis parciales de la regresión cuadrática para la producción de dextrina B.

Coefficiente	Estimación	Error Estándar	t calculada	Valor p (bilateral)
Constante (β_0)	78.159	17.237	4.5344.534	0.000136
Dextrina A (β_1)	6.864	1.627	4.218	0.000303
R cuadrada	0.3523			

El modelo explica el 35.23% de la variación, dado que es un valor menor que 50%, no conviene tratar de explicar la media de dextrina B usando dextrina A como variable explicativa.

El análisis de dextrina A se hace con las instrucciones del paquete R que siguen. Se escriben algunos comentarios que ayudan a modificar las instrucciones para resolver problemas similares.

Guía para resolver con R

```

datosDxt<-read.table(file = "dextrina.csv",header = T,sep = ",")

datosDxt

Mc<-datosDxt$M
Pc<-datosDxt$P
Bac<-datosDxt$Ba
Gc<-datosDxt$G
DA<-datosDxt$DA
DB<-datosDxt$DB

# Mc # Pc # Bac # Gc # DA #DB

par(mfrow = c(2,2))

plot(DA~Mc,main = " ",ylab = "Dextrina A",xlab = "Maltosa")
plot(DA~Pc,main = " ",ylab = "Dextrina A",xlab = "Pectina")
plot(DA~Bac,main = " ",ylab = "Dextrina A",xlab = "Beta alanina")
plot(DA~Gc,main = " ",ylab = "Dextrina A", xlab = "Glucosa")

trata<-c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,25)

baseTrat<-data.frame(tratam = factor(trata),DA = DA)

baseTrat

tratDAadev<-aov(DA~tratam,data = baseTrat)

anova(tratDAadev)

Mc2<-Mc^2
Pc2<-Pc^2
Bac2<-Bac^2
Gc2<-Gc^2

DAxR2<-lm(DA~Mc+Pc+Bac+Gc+Mc2+Pc2+Bac2+Gc2+Mc:Pc+Mc:Bac+Mc:Gc+Pc:Bac+Pc:Gc+Bac:Gc)

```

```

summary(DAxR2)

anova(DAxR2)

DAxR2r<-lm(DA~Mc+Gc+Mc2+Pc2)

summary(DAxR2r)

anova(DAxR2r)

# par(mfrow = c(2,2))

# plot(DAxR2r)          # Son las gráficas estándar para revisar residuales

par(mfrow = c(1,2))

plot(DAxR2r,which = 1,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Valores predichos",ylab = "Residuales",col.lab = 1)

plot(DAxR2r,which = 2,caption = "",col.lab = 0,id.n = NULL,pch = 20)

title(main = expression(" "),xlab = "Cuantiles de residuales estándar",
ylab = "Cuantiles esperados en la Normal",col.lab = 1)

residuales<-DAxR2r$resid

shapiro.test(residuales)

par(mfrow = c(1,1))

plot(DB~DA,main = " ",ylab = "Dextrina B",xlab = "Dextrina A")

DADBR<-lm(DB~DA)

summary(DADBR)

anova(DADBR)

```

15.1 Ejercicio

Haga el análisis para la producción de dextrina B utilizando los datos del problema anterior.

Bibliografía

-  Liu, B-L., Chen, J-W. and Tzeng, Y-M. (2000) Production of Cyclodepsipeptides Destrin A and B from *Metarhizium anisopliae*. *Biotechnology Progress*. 16 (6): 993-999.

Nota: La referencia se utilizó para plantear un problema similar pero llevado al nivel y el interés del curso. Se utilizaron los datos en la referencia.

Bibliografía General

-  Castillo M. A. (2013) *Estadística aplicada*. Trillas.
-  Daniel, W. W. (2006) *Bioestadística. Base para el análisis de las ciencias de la salud*, México: Limusa-Wiley.
-  Devore, J.L. (2001) *Probabilidad y estadística para ingeniería y ciencias*. México: Thompson Learning.
-  Escarela, G. P. (2014) *R para todos*. Un enfoque aplicado al análisis estadístico básico. UAM-I.
-  Gutiérrez, P. H., De la Vara, S.R., Cano, C.A. y Osorio, S.M. (2008) *Análisis y diseño de experimentos*. México: Mc Graw-Hill.
-  Kuehl, R.O. (2001) *Diseño de experimentos, Principios estadísticos de diseño y análisis de investigación*. 2ª ed. México: International Thompson Editores.
-  Márquez de Cantú, M.J. (1991) *Probabilidad y Estadística para Ciencias Químico-Biológicas*, México: Mc Graw Hill – Interamericana.
-  Ostle, B. (1988) *Estadística Aplicada México*. Limusa.
-  Quevedo Urias, H. y Perez Salvador B.R. (2008) *Estadística para ingeniería y ciencias*, México: Grupo Editorial Patria.
-  Steel, R.G.D. y Torrie J.H. (1985) *Bioestadística: principios y procedimientos*, México: Mc Graw Hill–Interamericana.

Ejercicios para el taller de diseño experimental
Se terminó de imprimir en diciembre de 2018,
con un tiraje de 200 ejemplares, mas sobrantes para reposición.



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

UNIDAD IZTAPALAPA

División de Ciencias Biológicas y de la Salud

Av. San Rafael Atlixco No.186, Col. Vicentina
C.P. 09340, Del. Iztapalapa, México D.F.
Tel.: (01) 58044600