



**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO
FACULTAD DE ESTUDIOS SUPERIORES
ZARAGOZA**



**Metodología de las Superficies de Respuesta en el Análisis de Mezclas.
Aplicación Práctica en la Industria Farmoquímica.**

TESINA
QUE PARA OBTENER EL TÍTULO DE
QUÍMICO FARMACÉUTICO BIÓLOGO

PRESENTA:
RAYMUNDO VEGA MARTÍNEZ

Director: M. en C. Armando Cervantes Sandoval
Asesor: Q.F.B. Jorge Luis Guevara Deloya

Ciudad de México, 2017



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

TABLA DE CONTENIDO

I.	RESUMEN.....	1
II.	INTRODUCCIÓN.....	2
III.	PLANTEAMIENTO DEL PROBLEMA.....	4
IV.	JUSTIFICACIÓN.....	5
V.	OBJETIVOS.....	6
VI.	MARCO TEÓRICO.....	7
	A. FUNDAMENTOS DE LA METODOLOGÍA DE RESPUESTA.....	11
	B. GRÁFICAS DE SUPERFICIE DE RESPUESTA.....	12
	C. MODELOS POLINOMIALES APROXIMADOS A LA SUPERFICIE REAL.....	15
	D. MÉTODOS PARA DETERMINAR PUNTOS ÓPTIMOS EN UNA SUPERFICIE DE RESPUESTA.....	16
	1. Naturaleza del problema.....	17
	2. Método del ascenso más pronunciado.....	22
	3. Localización del punto estacionario.....	24
	4. Caracterización de la superficie de respuesta mediante el análisis canónico.....	27
	5. Respuestas múltiples.....	28
	6. Análisis Ridge.....	30

E. DISEÑOS EXPERIMENTALES PARA AJUSTAR SUPERFICIES DE	
RESPUESTA.....	31
1. Diseños experimentales de mezclas.....	32
2. Consideraciones.....	33
3. Diseños Símplex de redes para mezclas.....	36
4. Diseño de centroide símplex.....	42
5. Modelo del diseño de mezclas.....	44
VII. EJEMPLOS DE APLICACIÓN.....	48
A. ESTUDIO DE ESTABILIDAD OXIDATIVA DEL BIODIESEL B100 EN	
UNA MEZCLA CON ANTIOXIDANTES.....	48
1. Planteamiento del problema.....	48
2. Resultados en Statgraphics®	49
3. Interpretación de los resultados en Statgraphics®.....	51
B. OPTIMIZACIÓN DE LOS COMPONENTES DEL MEDIO DE	
CULTIVO PARA LA PRODUCCIÓN DE POLIHIDROXIALKANOATO	
(PHA) DE ÁCIDOS ORGÁNICOS POR <i>Ralstonia eutropha</i>.....	53
1. Planteamiento del problema.....	53
2. Resultados en Statgraphics®.....	54
3. Interpretación de los resultados en Statgraphics®.....	57

C. DISEÑO DE MEZCLAS COMO PRIMER PASO PARA MEJORAR LA PRODUCCIÓN DE GLUTAMINASA EN LA FERMENTACIÓN EN ESTADO SÓLIDO POR <i>Bacillus sp.</i> RSP-GLU.....	59
1. Planteamiento del problema.....	59
2. Resultados en Statgraphics®	60
3. Interpretación de los resultados en Statgraphics®	62
VIII. CONCLUSIONES.....	64
IX. REFERENCIAS.....	68
X. APÉNDICE.....	71

ÍNDICE DE FIGURAS

Figura 1a.	Superficie de respuesta cuadrática para dos factores.....	13
Figura 1b.	Curvas de nivel de respuesta cuadrática.....	13
Figura 2.	Gráfica de superficie de respuesta tridimensional.....	14
Figura 3.	Gráfica de contornos combinada con una superficie de respuesta.....	14
Figura 4.	Búsqueda del punto óptimo en una gráfica de contornos.....	19
Figura 5.	Superficie de respuesta de primer orden y trayectoria del ascenso más pronunciado.....	23
Figura 6.	Determinación de las coordenadas del punto estacionario de un sistema de respuesta.....	25
Figura 7.	Punto máximo, punto mínimo y punto silla para dos factores.....	26
Figura 8.	Forma canónica del modelo de segundo orden.....	28
Figura 9.	Círculos concéntricos del análisis Ridge en los cuales se localizan valores máximos de la respuesta.....	30
Figura 10.	Área experimental para una mezcla binaria que corresponde a los puntos de la línea $x_1 + x_2 = 1$	35
Figura 11a.	Área experimental para una mezcla de tres componentes (triángulo equilátero).....	35
Figura 11b.	Área experimental para una mezcla de cuatro componentes (tetraedro).....	35
Figura 12.	Diseño Símpex de grado 2 para 3 componentes $\{3, 2\}$	39

Figura 13.	Diseño del Centroide Simplex $(x_1, x_2, x_3)x_i = (1/q), i = 1, 2, 3....$	42
Figura 14.	Combinaciones dentro del área de respuesta, variante de los diseños centrales.....	43
Figura 15.	Superficie de respuesta estimada para la mezcla de antioxidantes.....	49
Figura 16.	Gráfica de residuales para el periodo de incubación.....	50
Figura 17.	Superficie de respuesta estimada para la mezcla de ácidos volátiles.....	55
Figura 18.	Gráfica de residuales para el contenido de PHA.....	56
Figura 19.	Superficie de respuesta estimada para la mezcla de sustratos...	60
Figura 20.	Gráfica de residuales para la Actividad de Glutaminasa.....	61

ÍNDICE DE TABLAS

Tabla 1.	Número de experimentos para un arreglo $\{q, m\}$	38
Tabla 2.	Mezclas específicas de un diseño Láttice $\{3, 2\}$	38
Tabla 3.	Diseño de mezclas ara el periodo de inducción (h).....	48
Tabla 4.	Estimaciones de los efectos del modelo completo para Periodo de inducción (horas).....	49
Tabla 5.	Resultados completos de los modelos para el Periodo de inducción (horas).....	49
Tabla 6.	ANOVA del periodo de inducción-Antioxidantes.....	49
Tabla 7.	Resultados del modelo cuadrático ajustado del Periodo de inducción.....	50
Tabla 8.	Valores óptimos de antioxidantes para el Periodo de inducción.....	50
Tabla 9.	Diseño de mezclas para la producción de PHA.....	54
Tabla 10.	Estimaciones de los efectos del modelo completo para PHA (%)...	54
Tabla 11.	Resultados completos de los modelos para la producción de PHA.....	55
Tabla 12.	ANOVA del PHA-Ácidos orgánicos.....	55
Tabla 13.	Resultados del modelo cúbico especial ajustado para PHA.....	55
Tabla 14.	Valores óptimos de ácidos orgánicos para la producción de PHA.....	56
Tabla 15.	Diseño de mezclas para la producción de Glutaminasa.....	59
Tabla 16.	Estimaciones de los efectos del modelo completo para la Actividad	

	glutaminasa (U/gds).....	60
Tabla 17.	Resultados completos de los modelos para la Actividad glutaminasa.....	60
Tabla 18.	ANOVA de la Actividad glutaminasa-Sustratos.....	60
Tabla 19.	Resultados del modelo cúbico especial ajustado para la Actividad Glutaminasa.....	61
Tabla 20.	Valores óptimos de sustratos para la producción de Glutaminasa.....	61

I. RESUMEN

El diseño de mezclas es una herramienta estadística basada en la metodología de la superficie de respuesta utilizada para la optimización de las proporciones de los componentes de una mezcla con el objetivo de obtener un producto con atributos de calidad deseados. Facilita la optimización de un proceso utilizando un número pequeño de experimentos, la aplicación de esta técnica es limitada en las industrias dedicadas a la producción de especies químicas (industrias farmoquímicas) debido a la escasa difusión de sus fundamentos, aplicaciones y ventajas.

El presente trabajo contiene información teórica necesaria que abarca aspectos como: fundamentos de la metodología de superficie de respuesta, fundamentos de los diseños de mezclas, así como ejemplos experimentales para optimizar la producción de especies químicas mediante la proporción de los componentes de la mezcla para producirlos; en donde se visualiza la aplicación de estos diseños y la manera de trabajarlos en el software estadístico Statgraphics® Centurion y cómo interpretar los resultados obtenidos.

En este trabajo se presenta una guía de usuario que permite realizar la captura y análisis de datos con Statgraphics®, software estadístico auxiliar para la rápida obtención de resultados, análisis e interpretación de los mismos por el usuario.

II. INTRODUCCIÓN

La actividad profesional es, en todo momento, un proceso de aprendizaje y aunque es posible realizar cualquier actividad profesional sin planearla, bajo este esquema de trabajo, el proceso de aprendizaje se hace difícil y es poco probable obtener resultados satisfactorios.

Bajo esta misma premisa, es posible hacer investigación sin diseñar un experimento, pero la inducción de los resultados suele ser difícil, obteniendo información ambigua.

La estadística y el diseño de experimentos son más efectivos cuando se combinan con el apropiado conocimiento del tema o del fenómeno. Son un gran aliado, no un sustituto del profesional.

Hoy en día es de suma importancia contar con procesos óptimos, con el fin de minimizar tiempos y recursos, obteniendo productos que satisfagan la calidad que exige actualmente la legislación. Investigadores de prácticamente todos los campos de estudio llevan a cabo experimentos, por lo general para descubrir algo acerca de un proceso.

En un sentido literal, un experimento es una prueba. En una perspectiva más formal, un experimento puede definirse como una prueba o serie de pruebas en las que se hacen cambios deliberados en las variables de entrada de un proceso o sistema para medir e identificar las razones de los cambios que pudieran medirse en la respuesta de salida.

En todas las aplicaciones del diseño de experimentos, es necesario definir claramente los objetivos del estudio, las hipótesis de trabajo y proponer entonces el mejor diseño experimental que permita estimar el efecto de las variables, controlar el error experimental y evaluar la correlación y causalidad cuando existe complejidad de los efectos estudiados.

Una vez que los factores importantes que afectan a un proceso han sido identificados, usualmente son necesarias corridas adicionales para desarrollar la localización de las condiciones óptimas de los factores. Puesto que “óptimo” implica a menudo un mínimo o máximo, los diseños para optimización implican por lo menos tres niveles (bajo, medio y alto) por cada factor, estos son necesarios para poder estimar la curvatura. Existen diferentes tipos de diseños para la optimización de la respuesta, éstos se han agrupado como Diseños de Superficie de Respuesta: Diseño Central Compuesto, Diseño Factorial a Tres-Niveles, Diseño Box-Behnken y Diseños DraperLin.

Cuando los factores experimentales a ser estudiados son ingredientes o componentes de una mezcla, la función de la respuesta típicamente depende sobre las proporciones relativas de cada componente, no de la cantidad absoluta. Puesto que las proporciones deben sumar una cantidad fija, generalmente un 100%, los factores no pueden variar independientemente sobre algún otro. Consecuentemente, los diseños normalmente usados para investigación y optimización no se pueden aplicar directamente, ni el espacio de coordenadas es rectangular, es la representación más deseable de la región experimental.

III. PLANTEAMIENTO DEL PROBLEMA

La industria química farmacéutica puede dividirse en dos grandes rubros: los establecimientos dedicados a la fabricación de medicamentos y los dedicados a la fabricación de principios activos. En este último, los principios activos son obtenidos a partir síntesis química y/o síntesis microbiológica donde se llevan a cabo mezclas de reactivos para obtener el producto deseado, involucrando una serie de variables que pueden afectar el rendimiento del proceso si no se conocen las proporciones adecuadas de las especies químicas, fuentes de nutrientes, entre otros.

Actualmente, la aplicación de las herramientas estadísticas que utilizan superficies de respuesta, tales como los diseños de mezclas, una herramienta compleja, pero no por ello difícil, se usa con escasa frecuencia en el sector industrial, debido a que el análisis de los datos sin un paquete estadístico es muy complejo. Por lo tanto, es de suma importancia conocer dicha herramienta estadística, que apoye a la localización de la mejor respuesta a las variables estudiadas, a fin de contar con procesos óptimos minimizando tiempos y recursos con el menor número de experimentos y obteniendo productos que cumplan con los lineamientos establecidos en la normatividad vigente.

IV. JUSTIFICACIÓN

En el presente trabajo se desarrollará una propuesta metodológica para aplicar el análisis de mezclas en las industrias dedicadas a la obtención de principios activos, mostrando a los profesionales una manera práctica de obtener un proceso óptimo, enfatizando en la interpretación de resultados para obtener conclusiones válidas y objetivas.

V. OBJETIVOS

OBJETIVO GENERAL:

Desarrollar una propuesta metodológica para mostrar la aplicación del análisis de mezclas en la optimización de procesos dentro de las industrias farmoquímicas.

OBJETIVOS PARTICULARES:

- Describir los elementos teóricos de la metodología de la superficie de respuesta.
- Describir las ventajas y limitaciones en el análisis de mezclas.
- Mostrar la aplicación del análisis de mezclas mediante ejemplos.
- Elaborar una guía de usuario para realizar la captura, análisis e interpretación de los datos utilizando el software estadístico Statgraphics®.

VI. MARCO TEÓRICO

La selección del diseño de experimentos depende de la etapa de desarrollo, de los recursos disponibles y del objetivo del experimento. El tipo y el tamaño del diseño para un experimento depende de qué preguntas necesitan ser respondidas por el estudio. En general, existen tres tipos de experimentos: experimentos de selección, para seleccionar los factores que afectan otros experimentos o para demostrar robustez; experimentos de interacción, para estudiar a fondo las interacciones entre factores de interés; y experimentos de optimización, para trazar un mapa más cuidadoso de una región de interés⁽¹⁾.

Un experimento de selección proporciona información amplia pero no profunda, y generalmente requiere menos corridas experimentales que un experimento de interacción u optimización, en general un DOE (*Design Of Experiments*) de selección es más eficiente que cambiar un factor a la vez. Puede haber interés en estudiar las interacciones entre factores aunque esto no es habitualmente el punto central. Típicamente, se recomienda llevar a cabo una evaluación de riesgo para identificar factores potenciales para un estudio de selección⁽¹⁾.

Los experimentos de selección se utilizan con frecuencia para identificar cuales factores tienen el mayor efecto sobre las respuestas. Estos pocos factores significativos pueden entonces ser estudiados en un DOE posterior, más extenso⁽²⁾. Los diseños en estos experimentos generalmente incluyen experimentos factoriales fraccionados o

altamente fraccionados con cada factor en dos niveles (p. ej., Diseño de Plackett-Burman). Los puntos centrales (es decir, cada factor se ajusta en el centro de su rango) se agregan frecuentemente para medir la variabilidad de la replicación y para realizar una prueba general de la curva sobre la región experimental⁽³⁾.

Conforme el fenómeno científico subyacente se vuelve más complejo, así también sucede con el modelo matemático subyacente. El objetivo en un experimento de interacción es estudiar las interacciones entre factores junto con los efectos principales. Una interacción entre dos factores significa que el efecto que un factor tiene sobre la respuesta depende del nivel del otro factor⁽²⁾. Por ejemplo, el efecto del tiempo de mezclado sobre una respuesta puede depender de la velocidad de mezclado. Los diseños de interacción generalmente consisten en diseños experimentales factoriales o fracción-factorial para evaluar los factores en dos niveles junto con un punto central replicado. Existen otros diseños que pueden evaluar las interacciones (p. ej., diseños desbalanceados producidos utilizando algoritmos D-óptimos)⁽⁴⁾.

La ventaja de la familia de diseños factoriales, incluyendo los diseños de selección e interacción, es que pueden agregarse puntos adicionales al diseño para obtener más información acerca de las interacciones de manera secuencial⁽⁵⁾. El análisis es bastante fácil y puede ser realizado por la mayoría de los programas estadísticos. Algunos diseños (p. ej. Plackett-Burman) parecen similares a los experimentos factoriales pero

no tienen las mismas propiedades y no pueden ser fácilmente sumados de manera secuencial⁽³⁾.

Los experimentos de optimización son usados para trazar un mapa de superficie de respuesta para comprender las complejidades de los factores y sus interacciones, así como las no linealidades y su efecto combinado sobre las superficies de interés⁽⁶⁾. Estos experimentos pueden ser usados para encontrar la combinación de factores que pronostiquen la respuesta óptima (un máximo o un mínimo), para encontrar una región de posibles combinaciones de factores que pronostiquen resultados aceptables, o para pronosticar el desempeño del proceso en la región estudiada⁽⁷⁾. Los diseños de optimización son generalmente más grandes, requiriendo tres o más niveles para estimar términos cuadráticos o de mayor orden. Adicionalmente, pueden requerirse más niveles de factores para capturar la respuesta con más exactitud, especialmente en los casos en donde los rangos del factor son grandes. Los diseños comunes para los experimentos de optimización son el diseño central combinado, el de Box-Behnken, el de factoriales de tres niveles, o el de mezcla. El análisis de los diseños utilizados en los experimentos de optimización puede ser más complicado que los usados en los experimentos de selección⁽⁸⁾.

Los experimentos de selección deben realizarse antes de correr un experimento de optimización. Es posible que no sea necesaria una superficie de respuesta si el objetivo es mostrar que pueden obtenerse resultados aceptables sobre una región experimental, entonces puede ser que un experimento de selección o interacción sea todo lo que se necesita. Un error común es correr un experimento de optimización muy

prematuramente en el proceso y encontrar que los rangos no fueron bien seleccionados, resultando así en recursos desperdiciados. Una estrategia económica es correr un diseño de selección o interacción, y si los resultados son razonables, agregar entonces puntos al diseño para permitir el ajuste a una superficie de respuesta⁽⁹⁾.

Existen ventajas y desventajas para llevar a cabo estudios de manera secuencial. La realización de corridas pre-experimentales puede proporcionar un buen arranque para los rangos potenciales a considerar para los estudios. La realización de puntos centrales durante la selección podría proporcionar información no sólo sobre el error de replicación, sino también sobre la presencia de curvaturas significativas en la región experimental. Aunque los puntos centrales pueden identificar la curvatura, por sí mismos no pueden identificar cuales factores contribuyen a la curvatura. Si los puntos centrales no proporcionan evidencia de curvatura, es improbable que el riesgo/beneficio sea lo suficientemente grande para avanzar a un experimento de optimización. Sin embargo, si existe curvatura, entonces el riesgo/beneficio basado en un conocimiento *a priori* y la magnitud/dirección de la curvatura deberían ser utilizados para determinar si se desarrolla un modelo de superficie de respuesta. En general, si existe curvatura prácticamente importante puede usarse ya sea el conocimiento científico para ayudar a modelar la curvatura, o debe realizarse un experimento de optimización⁽¹⁰⁾.

El objetivo de todos los experimentos incluye describir la respuesta a los factores de tratamiento. Cuando los factores tienen niveles cuantitativos, se determina la respuesta (y) con respecto al factor (x) mediante ecuaciones de regresión. La ecuación de regresión estimada, en el caso de ser cuadrática, se grafica como una curva y se puede visualizar la respuesta (y) con respecto al factor (x). Una de las mayores ventajas de la curva de respuesta es que con ella se puede visualizar las respuestas para todos los niveles de factores en el experimento⁽¹¹⁾.

La ecuación de respuesta se puede mostrar como una superficie cuando los experimentos investigan el efecto de dos factores cuantitativos, como por ejemplo, los efectos de temperatura y presión en la tasa de una reacción química⁽¹²⁾.

A. Fundamentos de la metodología de superficie de respuesta.

La metodología de la superficie de respuesta es la alternativa más adecuada en el modelo y análisis de problemas en la que una respuesta de interés recibe la influencia de diversas variables y donde el objetivo es optimizar esta respuesta^(13, 14). Esta metodología está conformada por un conjunto de herramientas estadísticas y matemáticas que se aplican de acuerdo con las siguientes etapas:

1. Diseño y recopilación de datos experimentales, los cuales se usan en la determinación de una ecuación que se emplea para hacer predicciones.
2. Aplicación de las técnicas de regresión lineal múltiple para obtener la mejor ecuación que represente el comportamiento de los datos.

3. Análisis de la superficie ajustada mediante gráficas de contorno y otras técnicas matemáticas y numéricas⁽¹⁵⁾.

Dentro de estas herramientas, los diseños experimentales tienen la ventaja de explorar la relación de los factores y la variable dependiente en toda la región experimental, y no solamente en las fronteras.

B. Gráficas de superficie de respuesta

La superficie de respuesta permite que el investigador inspeccione de manera visual, la respuesta para cierta zona de los niveles de los factores de interés y evaluar su sensibilidad a los factores de tratamiento⁽¹⁶⁾. En ciertas aplicaciones industriales, las superficies de respuesta se exploran para determinar la combinación de niveles de los factores que proporcionan una condición operativa óptima, como en la combinación de temperatura-tiempo que maximizan la producción química. En otras aplicaciones se exploran para encontrar la combinación de niveles de los factores que proporcionan mejoras económicas a las respuestas de las condiciones operativas urgentes, si no es demasiado costoso lograr esas condiciones óptimas.^(16, 17)

Las superficies de respuesta también se pueden utilizar en estudios analíticos de procesos fundamentales⁽¹⁸⁾; por ejemplo, suelen usarse en las ciencias biológicas para investigar la influencia mutua de los factores sobre la variable de respuesta, como la interacción entre el nitrógeno y el fósforo en el crecimiento de las plantas por citar un ejemplo⁽¹⁹⁾.

La ecuación de respuesta cuadrática se representa como una superficie en tres dimensiones (figura 1a), en la figura 1b la ecuación aparece como curvas de nivel donde las líneas con valores de respuesta iguales son similares a las elevaciones mostradas en los mapas topográficos⁽¹³⁾.

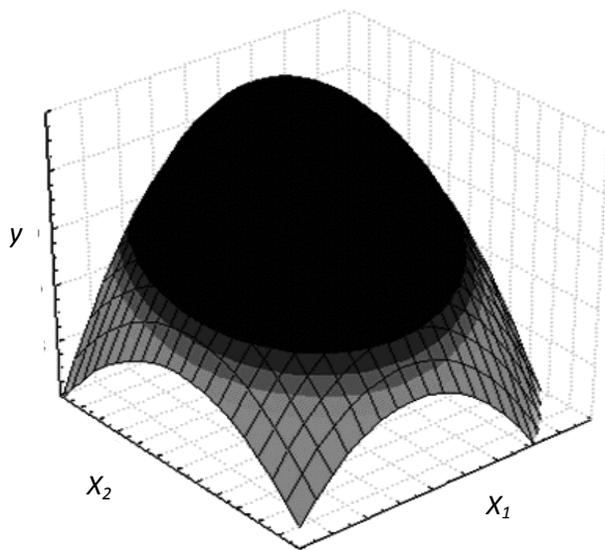


Figura 1a. Superficie de respuesta cuadrática para dos factores

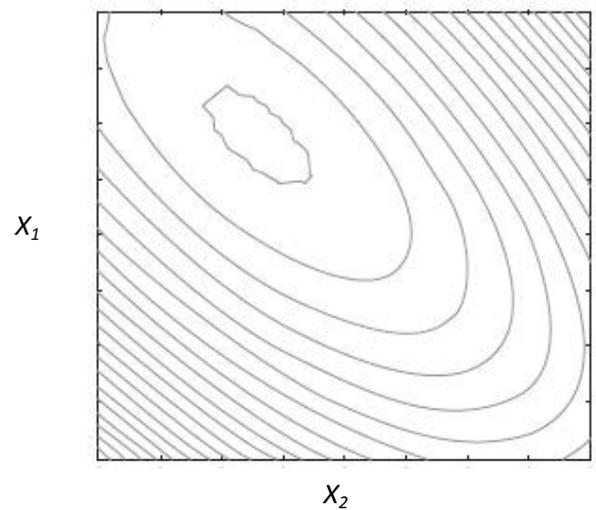


Figura 1b. Curvas de nivel de respuesta cuadrática

Por lo general se evalúa a Y como una función de las variables independientes, es decir, $Y = f(x_1, x_2, \dots, x_k) + \varepsilon$, donde ε representa el error observado en la respuesta. Si la respuesta esperada se denota por $E(Y) = f(x_1, x_2, \dots, x_k)$ entonces la superficie representada por $\eta = f(x_1, x_2, \dots, x_k)$ se denomina superficie de respuesta. La representación gráfica de la respuesta es posible, por ejemplo, para un problema en el cual se desea maximizar el rendimiento Y de un proceso que está determinado por dos factores x_1 (temperatura) y x_2 (presión); obtenemos la figura 2 donde la respuesta Y

que es el rendimiento esperado (η) se representa como una superficie sólida en un espacio tridimensional.

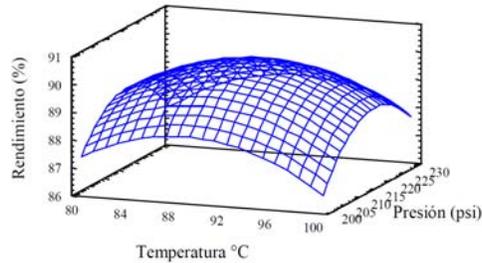


Figura 2. Gráfica de superficie de respuesta tridimensional

Con el propósito de visualizar mejor la forma de una superficie de respuesta con frecuencia se grafican los contornos de dicha superficie en forma bidimensional: en la cual se trazan líneas de respuesta en el plano x_1, x_2 . Cada contorno corresponde a una altura específica de la superficie de respuesta. La gráfica de contornos es útil para estudiar los niveles x_1 y x_2 que dan por resultado cambios en la forma o altura de la superficie, y por lo tanto con este tipo de gráficas se facilita la interpretación de las relaciones cuando existen dos o más factores, figura 3. ^(13, 19)

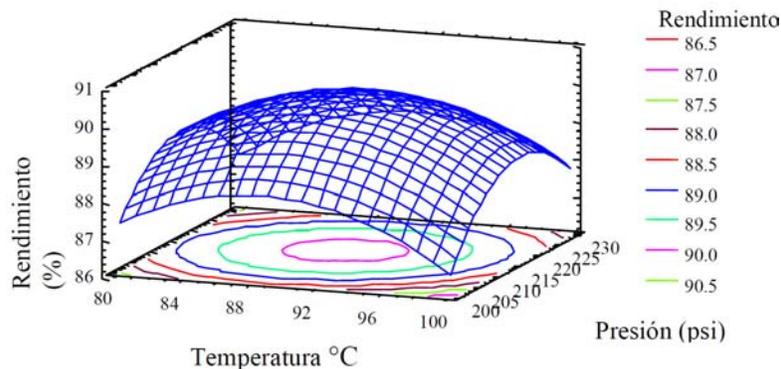


Figura 3. Gráfica de contornos combinada con una superficie de respuesta

C. Modelos polinomiales aproximados a la respuesta real

En la mayoría de los problemas de la metodología de superficies de respuesta, la forma de la relación entre la respuesta y las variables independientes se desconoce. Por lo tanto, el primer paso de la metodología es encontrar una aproximación adecuada de la verdadera relación funcional entre Y y el conjunto de variables independientes. Si la respuesta está bien modelada por una función lineal de las variables independientes, entonces la función de aproximación es el modelo de primer orden.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Si hay una curvatura en el sistema, entonces debe usarse un polinomio de orden superior, tal como el modelo de segundo orden.

$$Y = \sum_{i=1}^k \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i^2 + \sum \sum \beta_{ij} x_i x_j + \varepsilon$$

En casi todos los problemas de la metodología de superficies de Respuesta, se usa uno de estos modelos o ambos. Desde luego, es probable que un modelo polinomial solo sea una aproximación razonable de la verdadera relación funcional en el espacio completo de las variables independientes, pero para una región relativamente pequeña suelen funcionar bastante bien.

La metodología de superficies de respuesta es un procedimiento secuencial. Muchas veces, cuando se está en un punto de la superficie de respuesta alejado del óptimo, el

sistema representa una curvatura moderada y el modelo de primer orden será apropiado. El objetivo en este caso es llevar al experimentador de manera rápida y eficiente por la trayectoria del mejoramiento hasta la vecinidad general del óptimo. Una vez que ha encontrado la región del óptimo, puede emplearse un modelo más elaborado, como el de segundo orden, y llevar a cabo un análisis para localizar el óptimo.

Para estimar los parámetros de los polinomios de aproximación se utiliza el método de mínimos cuadrados. Después se realiza el análisis del sistema real. Los parámetros del modelo pueden estimarse de manera más eficiente cuando se emplean los diseños experimentales apropiados para recolectar los datos. Los diseños para ajustar superficies de respuesta se denominan diseños de superficie de respuesta.

El objetivo último de la metodología de superficies de respuesta es identificar las condiciones de operación óptimas del sistema o identificar una región del espacio de los factores en la que se satisfagan los requerimientos de operación^(13, 19)

D. Métodos para determinar puntos óptimos en una superficie de respuesta.

Existen varios procedimientos para determinar las condiciones bajo las cuales obtendremos el mejor valor de respuesta; esta respuesta podría ser un rendimiento máximo o un nivel de calidad alto, también es posible desear un costo mínimo o el menor número de defectos en un proceso. En cualquiera de estos casos se estarán buscando las condiciones óptimas

1. Naturaleza del problema.

Se supone que una respuesta Y está afectada por un número de factores cuantitativos x_1, x_2, \dots, x_k . Se lleva a cabo un programa experimental para descubrir el nivel al cual cada uno de estos factores debe ser establecido para llevar al óptimo la respuesta. Este problema es común en trabajos de investigación y desarrollo en la industria donde el investigador está buscando los medios de manufacturar un producto nuevo con ciertas características deseables u obtener un producto de manera más económica.

La respuesta puede ser la cantidad real de producto o una medida de la cantidad del mismo. En ocasiones, el objetivo es minimizar Y , como cuando Y representa la cantidad de un subproducto no deseado o el costo de manufactura por unidad de producción. Localizar el mínimo no requiere explicación separada, ya que un cambio en el signo de la respuesta lo transforma en uno de maximización.

Además de localizar el máximo de Y , es valioso aprender algo de cómo varía Y en la vecinidad del máximo, cuando los niveles de los factores se cambian de sus valores óptimos. Esto es por varias razones:

1. En aplicación a gran escala de los resultados puede no ser factible establecer cada factor exactamente a su nivel óptimo. Alguna combinación de niveles diferente del óptimo puede ser más económica de mantener.

2. Frecuentemente, más de una característica del producto es importante. Un cambio en la x_i , alejándolo del óptimo, puede ser preferible a causa de su efecto sobre una de estas otras características.
3. La forma de la superficie de respuesta cerca del óptimo puede dar indicios de la naturaleza del proceso.
4. La superficie puede carecer de un máximo verdadero en la región experimental (o en la región en la cual el proceso se está llevando a cabo en la práctica). En este caso, se desea conocer la naturaleza de la superficie en zonas de respuesta relativamente altas.

Existen varios procedimientos para determinar las condiciones bajo las cuales obtendremos el mejor valor de respuesta, es decir, las condiciones óptimas. En aplicaciones prácticas, cada método requiere de un considerable criterio por parte del investigador y los métodos no pueden presentarse como un conjunto de reglas inflexibles.

Al final, el experimentador debe decidir sobre los factores que se deben incluir en el programa. Los métodos por describir se han aplicado, en su mayor parte, para un número de factores entre 2 y 6, y no son aplicables para mayor número de ellos. Se aconseja una depuración preliminar de factores que probablemente sean de menor importancia. Debe seleccionarse también el intervalo dentro del cual varía el nivel de cada factor.^(13, 20)

Los pasos a seguir en la localización de las condiciones óptimas son:

1. Identificar a las variables que presentan mayor influencia en la respuesta, generalmente son pocas y se identifican con facilidad (análisis de regresión).
2. Expresar la respuesta de interés como una función de las variables, para la cual usualmente se utiliza un modelo polinomial de primer orden, este modelo dará las bases para realizar las pruebas iniciales, y también hará posible modelar los valores del área de respuesta estimada sobre la región experimental, dicha representación tomará la forma de una gráfica de contorno del área estimada (figura 4), dentro de esta gráfica las líneas de contorno son establecidas mediante la conexión de dos ejes (x_1 y x_2) en la región experimental que produce los mismos valores de Y .

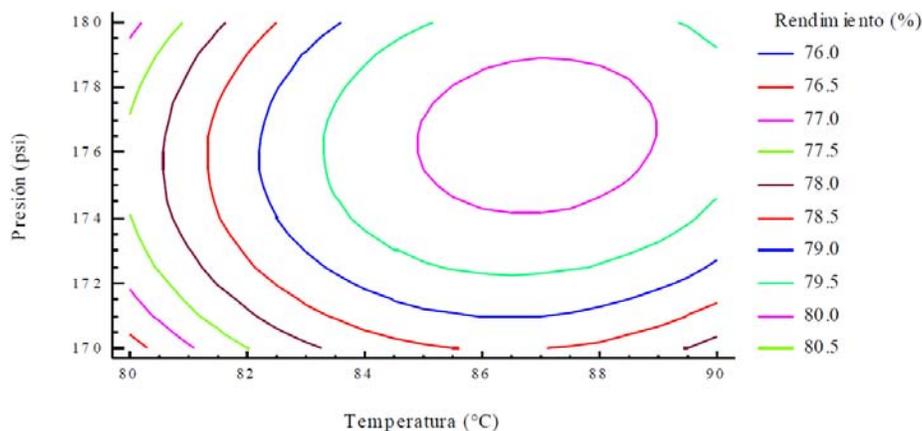


Figura 4. Búsqueda del punto óptimo en una gráfica de contornos.

En este método el experimentador hace primero una estimación preliminar de la combinación óptima de niveles del factor, la cual se identifica por $x_{11}, x_{21}, \dots, x_{k1}$ ya que cada experimento trata con un único factor, los factores deben arreglarse por el

orden en el cual van a ser probados. En general, un buen plan es principiar con el factor que se espera que dé la mayor respuesta.

En el primer experimento, todos los factores excepto el primero, se mantienen constantes a sus niveles iniciales $x_{21}, x_{31}, \dots, x_{k1}$. El propósito de este experimento es encontrar el nivel de x_1 , el cual maximiza Y para esta combinación particular de niveles de los otros factores. Hay varias formas como puede llevarse a cabo el experimento. Para establecer un máximo, deben compararse cuando menos tres niveles de x_1 . Son aconsejables 4 o 5 niveles, si el rango de x_1 es amplio y la posición de su óptimo casi no se conoce o quizá un experimento inicial con 5 niveles, ampliamente espaciados, seguidos de un segundo experimento con tres niveles a un espaciamiento estrecho. Se ha demostrado que el igual espaciamiento de los niveles, no es en general, el procedimiento más eficiente para localizar el óptimo. Para experimentos con tres niveles se ha llevado a cabo el espaciamiento más eficiente en situaciones donde la posición del óptimo x_1 , por anticipado está bien determinada.

Cuando está claro que los puntos experimentales no definen un máximo de Y , tal posición puede ser estimada ajustando una parábola a las respuestas observadas.

Si la ecuación de la parábola ajustada es:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$$

Con β_2 negativo, se sabe por el cálculo que el óptimo x_1 es:

$$x_{12} = \beta_0/2\beta_2$$

En ocasiones, el valor de x_1 que da la mayor Y observada está suficientemente cercano al óptimo estimado. El óptimo estimado se identifica por x_{12} .

En el segundo experimento, x_1 se fija al nivel x_{12} y los valores de $x_3 \dots x_k$ permanecen fijos a sus niveles iniciales. Varios niveles del segundo factor se comparan para encontrar el nivel óptimo de x_2 . El tercer experimento trata a x_3 de la misma manera; los otros factores se mantienen a los niveles $x_{12}, x_{22}, x_{41}, x_{k1}$, respectivamente. El proceso se continúa hasta que se ha estimado un óptimo para los k factores.

Esto finaliza la primera etapa de experimentos: la combinación de factores óptima estimada hasta este momento es el conjunto de niveles $x_{12}, x_{22}, \dots, x_{k2}$. Si este nuevo conjunto de niveles es parecido al conjunto inicial $x_{11}, x_{21}, \dots, x_{k1}$, y si los valores de Y han mejorado poco durante la primera etapa, el investigador puede tomar la decisión de terminar los experimentos, dando por cierto que no es posible una ganancia apreciable en la estimación inicial del óptimo. Si han ocurrido cambios más considerables, entra en una segunda etapa el proceso integro, principiándose con niveles fijos $x_{22}, x_{32}, \dots, x_{k2}$ y probando varios niveles de x_1 para determinar si la posición del óptimo ha cambiado del nivel x_{12} . Al final de esta etapa, se ha obtenido un nuevo conjunto de estimaciones $x_{13}, x_{23}, \dots, x_{k3}$, entonces el experimentador estudia la situación otra vez y decide si la lleva a término o continúa con una tercera etapa.

Al principio de la tercera etapa, todas las variables x se cambian simultáneamente: el cambio en x_i es proporcional a $(x_{i3} - x_{i2})$. Esta sugerencia se asemeja al método del ascenso más pronunciado que se describe a continuación. Puede tomarse la decisión de omitir algunos de los factores en etapas posteriores sobre la base de que no han demostrado efectos en las primeras etapas.

Este método no da una estimación de la forma de la curva de respuesta. Cuando el proceso ha terminado, será aconsejable hacer experimentos adicionales para este propósito⁽²⁰⁾.

2. Método del ascenso más pronunciado

El método del ascenso más pronunciado es un procedimiento para moverse secuencialmente sobre la trayectoria del ascenso más pronunciado, es decir, en la dirección del incremento máximo de la respuesta. Desde luego, si lo que se pretende es una minimización, entonces esta técnica se llama método del descenso más pronunciado. El modelo ajustado de primer orden es:

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X_i$$

Y la superficie de respuesta de primer orden, es decir, los contornos de \hat{Y} , es una serie de líneas paralelas como las que se muestran en la figura 5.

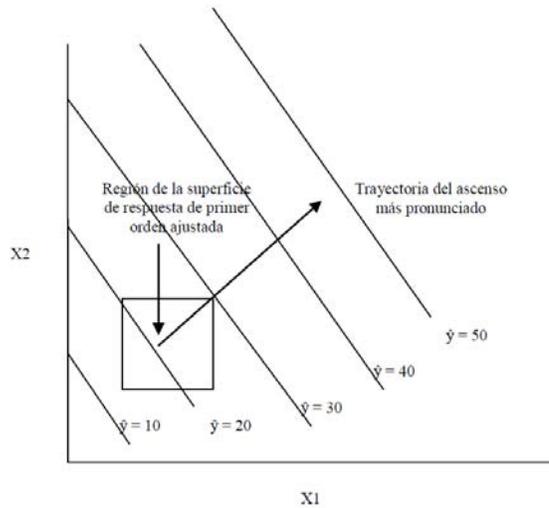


Figura 5. Superficie de respuesta de primer orden y trayectoria del ascenso más pronunciado

La dirección del ascenso más pronunciado es aquella en la que \hat{Y} se incrementa con mayor rapidez. Esta dirección es paralela a la normal de la superficie de respuesta ajustada. Por lo general se toma como la trayectoria del ascenso más pronunciado a la recta que pasa por el centro de la región de interés y que es normal a la superficie ajustada. Por lo tanto, los pasos sobre la trayectoria son proporcionales a los coeficientes de regresión $\{\beta\}$.

El tamaño real del paso lo determina el experimentador con base en el conocimiento del proceso o de otras consideraciones prácticas.

Se conducen experimentos sobre la trayectoria del ascenso más pronunciado hasta que deja de observarse un incremento adicional en la respuesta. Entonces se puede

ajustar un nuevo modelo de primer orden, determinar una nueva trayectoria del ascenso más pronunciado y el procedimiento continúa. En última instancia, el experimentador llegará a la vecindad del óptimo, en general, la falta de ajuste del modelo de primer orden indica que se ha llegado a ella, en este momento se realizan experimentos adicionales para obtener una estimación más precisa del óptimo.^(13, 19, 20)

3. Localización del punto estacionario

El primer paso en el proceso de encontrar las condiciones óptimas es identificar las variables que presentan la mayor influencia en la respuesta, las cuales generalmente son pocas y se identifican con facilidad (análisis de regresión), el segundo paso consiste en expresar la respuesta de interés como una función de las variables, para lo cual usualmente se utiliza un modelo polinomial de primer orden, este modelo nos dará las bases para realizar las pruebas iniciales, y también hará posible modelar los valores del área de respuesta estimada sobre la región experimental, dicha representación tomará la forma de una gráfica del contorno del área estimada, dentro de esta gráfica las líneas de contorno son establecidas mediante la conexión de dos ejes (x_1 y x_2) en la región experimental que produce los mismos valores de Y , lo anterior se representa en la figura 6 donde dos factores (tiempo de reacción y temperatura de reacción) modifican el rendimiento de una reacción química, es posible observar en esta figura bajo qué condiciones de los factores se obtiene el mejor rendimiento.

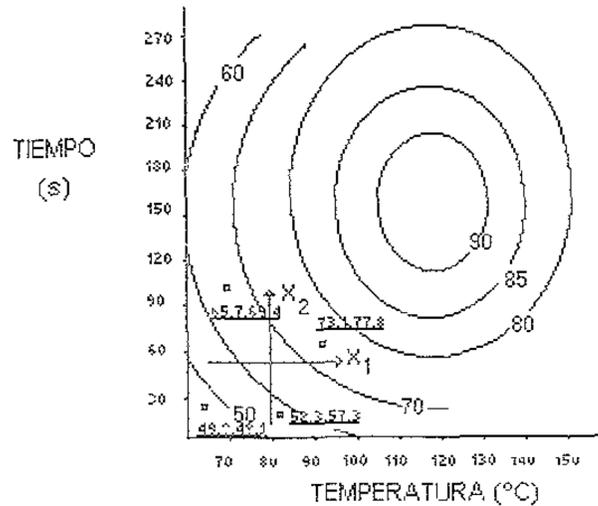


Figura 6. Determinación de las coordenadas del punto estacionario de un sistema de respuesta.

Una región estacionaria es definida donde una región como la inclinación del área, o los gradientes a lo largo de los ejes de las variables, es menor comparada con el estimado del error experimental. El punto estacionario de una región estacionaria es el punto en el cual la inclinación del área de respuesta es cero cuando es tomada en todas direcciones.

Suponiendo que se desea determinar los niveles de x_1, x_2, \dots, x_k que optimizan la respuesta predicha. Éste óptimo, si existe, será el conjunto de x_1, x_2, \dots, x_k tal que las derivadas parciales $\frac{\partial Y}{\partial X_1} = \frac{\partial Y}{\partial X_2} = \dots, \frac{\partial Y}{\partial X_k} = 0$. Dicho punto, se denomina punto estacionario.

El punto estacionario podría representar:

- a) Una respuesta máxima.
- b) Una respuesta mínima.
- c) Un punto silla⁽¹³⁾

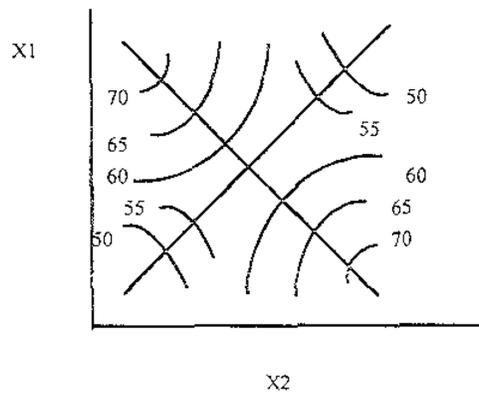
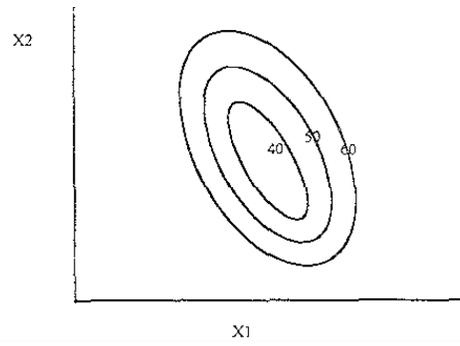
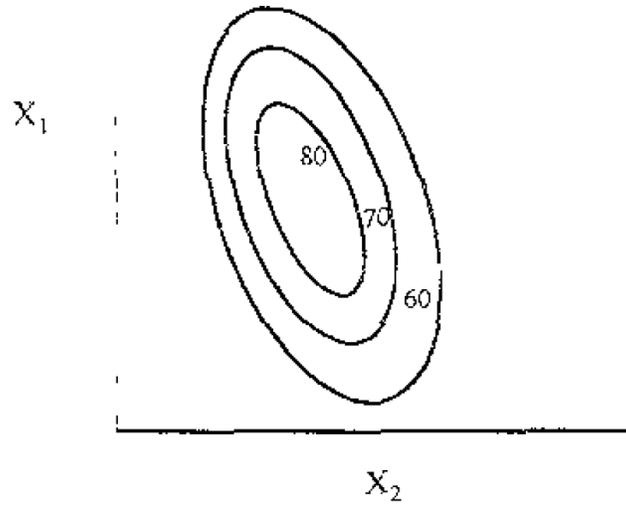


Figura 7. Punto máximo, punto mínimo y punto silla para dos factores.

4. Caracterización de la superficie de respuesta mediante el análisis canónico

Una vez que se ha encontrado el punto estacionario generalmente es necesario caracterizar la superficie de respuesta en la vecindad de este punto, es decir, determinar si el punto estacionario es el punto de una respuesta máxima, mínima o un punto silla. Por lo general también se desea estudiar la sensibilidad relativa de la respuesta a las variables x_1, x_2, \dots, x_k .

La forma más directa de hacer esto es examinando una gráfica de contorno del modelo ajustado, si solo hay dos o tres variables en el proceso (las x 's), la construcción e interpretación de esta gráfica de contorno es relativamente sencilla, ya que se trata de curvas correspondientes a valores constantes de la respuesta sobre el plano X_1X_2 (plano cuyos ejes coordenados vienen dados por los niveles X_1 y X_2 de los factores). Geométricamente, cada línea de contorno es una proyección sobre el plano X_1X_2 de una sección de la superficie de respuesta al intersectar con un plano paralelo al X_1X_2 . Sin embargo, incluso cuando hay un número relativamente reducido de variables, un análisis más formal, llamado análisis canónico es utilizado.

Es conveniente transformar primero el modelo en un nuevo sistema de coordenadas con el origen en el punto estacionario x , y después hacer la rotación de los ejes de este sistema hasta que sean paralelos a los ejes principales de la superficie de respuesta ajustada. Esta transformación se ilustra en la figura 8.

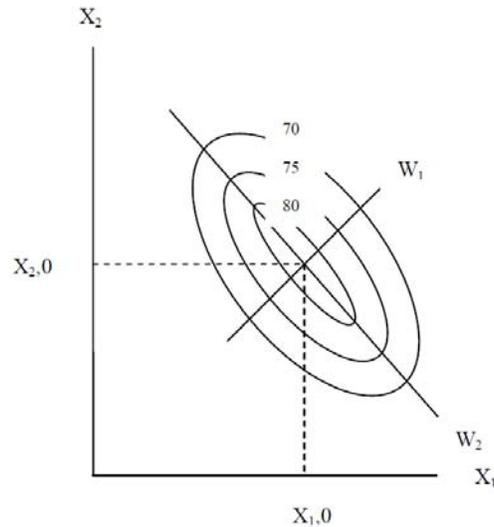


Figura 8. Forma canónica del modelo de segundo orden.

Puede demostrarse que se obtiene así el modelo ajustado.

$$\hat{Y} = \hat{Y}_0 + \lambda_1 W_1^2 + \lambda_2 W_2^2 + \dots + \lambda_k W_k^2$$

Donde las (W_i) son las variables independientes transformadas y las (λ_i) son constantes. La naturaleza de la superficie de respuesta se puede determinar a partir del punto estacionario, del signo y de la magnitud de (λ_i) . Si todas las λ son positivas, entonces x_0 es un punto de respuesta mínima y si las λ son negativas, entonces x_0 es un punto de respuesta máxima y si las λ tienen distintos signos, x_0 corresponde a un punto silla.^(13, 19)

5. Respuestas múltiples

Existen problemas de superficie de respuesta que incluyen el análisis de varias respuestas. La evaluación simultánea de respuestas múltiples requiere construir

primero un modelo de superficie de respuesta apropiado para cada respuesta y después intentar encontrar un conjunto de condiciones de operación que optimice en cierto sentido todas las respuestas o que al menos las mantenga en los rangos deseados.

Un enfoque relativamente directo para optimizar varias respuestas que funciona bien cuando sólo hay pocas variables en el proceso es la superposición de las gráficas de contorno de cada respuesta. El experimentador puede hacer el examen visual de la gráfica de contorno para determinar las condiciones de operación apropiadas.

Cuando hay más de tres variables en el diseño, se hace muy complicada la superposición de las gráficas de contorno, ya que la gráfica de contorno es bidimensional y $k - 2$ de las variables del diseño deben mantenerse constante para construir la gráfica. Con frecuencia se necesita una gran cantidad de ensayo y error para determinar cuáles son los factores que deben mantenerse constantes y qué niveles seleccionar para obtener la mejor vista de la superficie. Por lo tanto existe interés práctico en métodos de optimización más formales para las respuestas múltiples.⁽¹³⁾

6. Análisis Ridge

Durante el análisis del área de respuesta ajustada se podrá descubrir que el punto estacionario no está dentro de la región experimental, pero si se sospecha que el valor óptimo de la respuesta se encuentra en los bordes de la región experimental, es posible utilizar el análisis Ridge para localizar ese punto.

En general, el análisis Ridge se utiliza para encontrar un valor máximo absoluto o un mínimo de la respuesta estimada de \hat{Y} en esferas concéntricas de radio variante, RL ($L = 1, 2, \dots, 3$) los cuales son centrados en $(x_1, x_2, \dots, x_k) = (0, 0, \dots, 0)$ y contenidos dentro de la región experimental. Lo anterior se muestra en la figura 9, donde tres círculos se trazan con diferente radio.

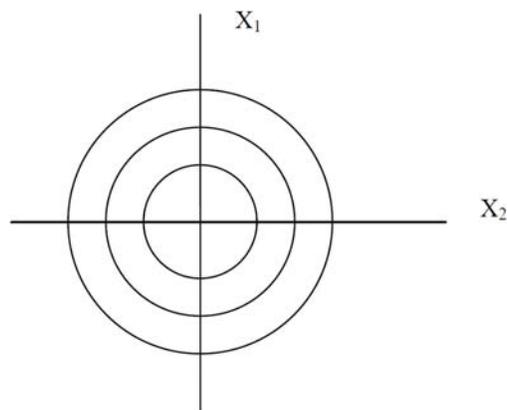


Figura 9. Círculos concéntricos del análisis Ridge en los cuales se localizan valores máximos de la respuesta.

Asumiendo que el modelo ajustado sobre la región de las k variables codificadas x_1, x_2, \dots, x_k es de segundo orden y suponiendo además que el punto estacionario está

en los bordes de una esfera de radio R , se restringirá el examen a encontrar las coordenadas de las variables que maximicen y sujetas a la condición:

$$\sum_{i=1}^k x_i^2 = R^2$$

Una vez encontradas esas coordenadas para un valor particular de R , se podría cambiar el valor de R y repetir el procedimiento para varios valores elegidos de R y hacer una gráfica con las apropiadas coordenadas x_1, x_2, \dots, x_k contra \hat{Y} ; esto generaría gráficas de valores de \hat{Y} para varias distancias del punto central.^(13, 21)

E. Diseños experimentales para ajustar superficies de respuesta.

El ajuste y análisis de superficie de respuesta se facilita en gran medida con la elección apropiada del diseño experimental, algunas de las características deseables en el diseño son las siguientes:

1. Proporcionar una distribución razonable de los puntos de los datos (y en consecuencia información) en toda la región de interés.
2. Permitir que se investigue la adecuación del modelo, incluyendo la falta de ajuste.
3. Permitir que los experimentos se realicen en bloques.
4. Permitir que los diseños de orden superior se construyan secuencialmente.
5. Proporcionar una estimación interna del error.
6. Proporcionar estimaciones precisas de los coeficientes del modelo.

7. Proporcionar un buen perfil de la varianza de predicción en toda la región experimental.
8. Proporcionar una robustez razonable contra los puntos atípicos o los valores faltantes.
9. No requerir de gran número de corridas.
10. No requerir de demasiados niveles de las variables independientes.
11. Asegurar la simplicidad en el cálculo de los parámetros del modelo.

1. Diseños experimentales de mezclas

Hasta ahora se han presentado diseños de superficie de respuesta donde los factores involucrados son independientes; sin embargo, existen otros tipos de diseños de superficie de respuesta en los cuales los factores involucrados no son independientes, este es el caso de los diseños de mezclas, en donde al modificar algún componente de la mezcla se tendrá que modificar otro u otros componentes de la mezcla, en estos diseños se considera que la variable dependiente sólo es resultado de la cantidad de los elementos de la mezcla. Retomando lo anterior, al evaluar la capacidad de una ecuación matemática para representar una superficie de respuesta en un diseño de mezclas, se supone que existe alguna correlación de una función $Y = f(x_1, x_2, \dots, x_k)$ la cual define la dependencia de Y en las proporciones x_1, x_2, \dots, x_k de los componentes. Esta función es continua en x_1 , y es representada por un polinomio de primer o segundo orden.

Definición.

Los diseños de superficies de respuesta para mezclas son una técnica estadística donde se busca conseguir una respuesta óptima (máximos o mínimos) en función de los elementos constitutivos de la mezcla. En un programa experimental de ese tipo se desea modelar el área experimental de combinaciones de la mezcla para:

- Predecir una respuesta debida a algún o algunos de los ingredientes de la mezcla.
- Obtener alguna medida de la influencia en la respuesta de cada componente o de sus combinaciones.⁽²²⁾

2. Consideraciones.

En un diseño experimental de mezclas, dos o más componentes se mezclan o combinan juntos para formar un producto final. Las mediciones de las propiedades físicas de varias de las combinaciones de los ingredientes se realizan con el fin de encontrar la combinación que arroje el “mejor” resultado.

Existen características que hacen al diseño experimental de mezclas diferente de los demás tipos de diseños experimentales:

- La característica de los productos finales (respuesta) tales como la calidad del producto, depende únicamente de las proporciones relativas de los ingredientes (componentes) presentes en la mezcla y no de la cantidad total de la misma.
- Los cambios en la proporción de uno de los componentes no se puede variar de manera independiente, con respecto a la proporción de los otros componentes presentes en la mezcla.⁽²³⁾

Las variables controlables en el diseño de mezclas representan cantidades proporcionales de la mezcla, y estas proporciones se pueden expresar en fracción de volumen, fracción peso o por fracción mol o porcentaje, y están comprendidas entre el cero y la unidad. Cuando son expresadas como fracción de la mezcla, la suma de las proporciones debe ser igual a la unidad. Esto se puede expresar de la siguiente manera:⁽¹⁹⁾

$$0.0 \leq x_i \leq 1.0$$

$$i = 1, 2, \dots, q$$

$$\sum_{i=1}^q x_i = x_1 + x_2 + \dots + x_q = 1, 0$$

Donde:

q = número de componentes (ingredientes)

x_1 = fracción del i -ésimo componente (ingrediente)

Las ecuaciones representan las principales restricciones de las proporciones de los componentes en un diseño de mezclas. Como resultado de estas restricciones, sobre los valores de x_1 , la región experimental o área de interés es un Símplex regular (región donde se puede trazar una línea recta y todos los puntos de la recta quedan dentro de la misma región con $q - 1$ dimensiones. Por ejemplo: para $q - 2$ componentes, el área de interés es una línea recta (figura 10).

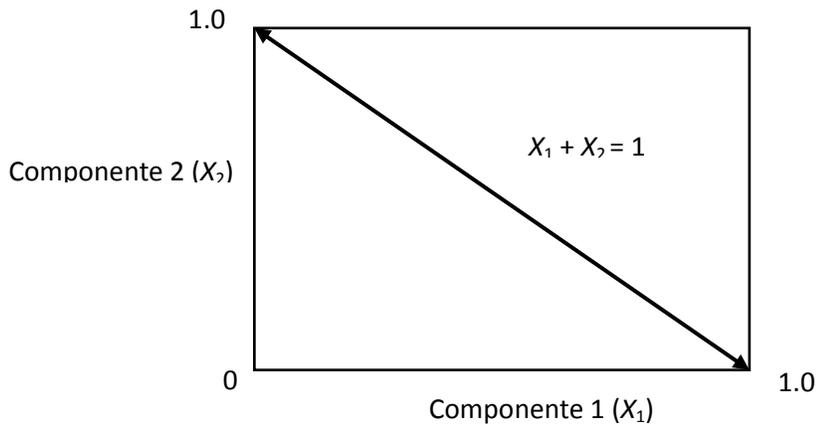


Figura 10. Área experimental para una mezcla binaria que corresponde a los puntos de la línea $x_1 + x_2 = 1$

Para $q - 3$ el área de interés es un triángulo equilátero, y para $q - 4$ el área de interés es un tetraedro (Figura 11).

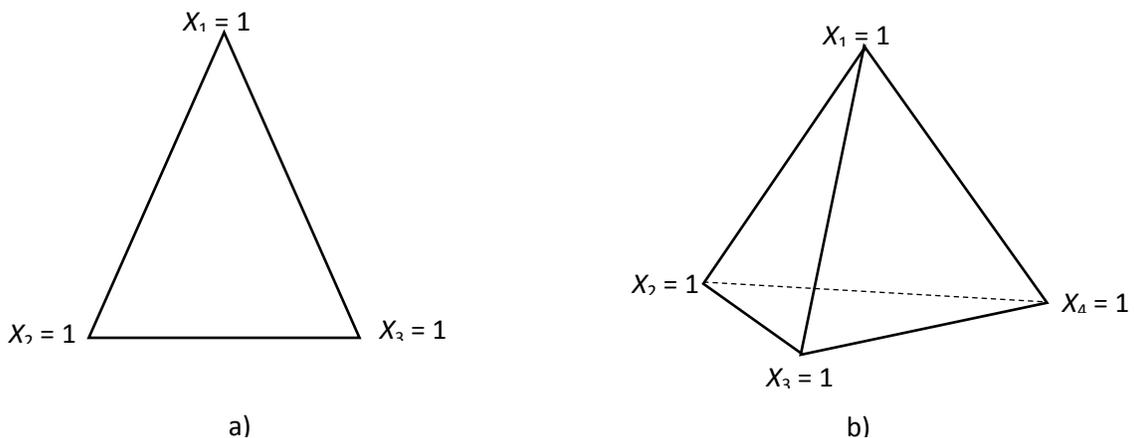


Figura 11. a) Área experimental para una mezcla de tres componentes (triángulo equilátero).

b) área experimental para una mezcla de cuatro componentes (tetraedro).

Se debe notar que la suma de las proporciones de los componentes es igual a la unidad, por lo que al variar la proporción de alguno de los componentes se altera la proporción de los demás componentes de la mezcla dentro de la región experimental.⁽²⁴⁾

3. Diseño Símplex de redes para mezclas

El planteamiento del diseño experimental debe incluir puntos en todo el espacio experimental del Símplex, para que así sea más sencillo el ajuste de un modelo a estos datos. Cuando el arreglo espacial de estos puntos sigue un orden de puntos igualmente espaciados, se dice que se tiene un diseño de tipo Ládice.

En los diseños de mezclas con Símplex, se utilizan las ecuaciones polinomiales canónicas para el ajuste de los puntos a un modelo. El ládice tiene un arreglo espacial para dar soporte al modelo y cubrir sus requerimientos.

La cantidad y la distribución de los puntos en el espacio experimental se definen esencialmente por dos valores:

- a) El número de componentes en la mezcla: q
- b) El grado de polinomio que se desea ajustar: m

Entonces se designa que se tiene un modelo Símplex Látice $\{q, m\}$, el cual tiene $m + 1$ valores igualmente espaciados para las proporciones, es decir, la proporción de cada componente es:

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \frac{3}{m}, \dots, 1$$

Por lo que cada componente, presentará un valor de proporción dado por la ecuación anterior, y el modelo Símplex Látice constará de todas las posibles mezclas entre los componentes con estos valores de proporción.

El número de puntos en el diseño Símplex $\{q, m\}$, está dado por la siguiente fórmula:

$$\binom{q + m - 1}{m} = \frac{(q + m - 1)!}{m! (q - 1)!}$$

La siguiente tabla resume el número de puntos en el diseño Símplex $\{q, m\}$ para los valores de q (de 3 a 10 componentes) y m (de 1 a 4 proporciones).

Grado del modelo <i>m</i>	Número de componentes							
	<i>q</i>							
	3	4	5	6	7	8	9	10
2 (cuadrático)	6	10	15	21	28	36	45	55
2 (cúbico especial)	7	14	25	41	63	92	129	172
3 (cúbico completo)	10	20	35	56	84	120	165	220
4 (cuártico)	15	35	70	126	210	330	495	715

Tabla 1. Número de experimentos para un arreglo $\{q, m\}$

Por ejemplo, para un diseño, en el cual se evaluarán tres componentes ($q = 3$), y se desea ajustar un modelo polinomial canónico cuadrático ($m = 2$), se tiene un modelo Simplex Látime $\{3, 2\}$, donde se tienen 3 proporciones para cada componente ($x_i = 0, \frac{1}{2}, 1$, para $y = 1, 2, 3$), y el número de puntos que se deben tener para las mezclas son:

$$\binom{3 + 2 - 1}{2} = \frac{4!}{2! * 2!} = \frac{24}{4} = 6$$

La siguiente tabla muestra las mezclas específicas de un diseño Látime $\{3, 2\}$

Número	X_1	X_2	X_3	Y
1	1	0	0	Y_1
2	0	1	0	Y_2
3	0	0	1	Y_3
4	$\frac{1}{2}$	$\frac{1}{2}$	0	Y_{12}
5	$\frac{1}{2}$	0	$\frac{1}{2}$	Y_{13}
6	0	$\frac{1}{2}$	$\frac{1}{2}$	Y_{23}

Tabla 2. Mezclas específicas de un diseño Látime $\{3, 2\}$

Se debe observar que la tabla ha respetado las proporciones de las mezclas $(0, \frac{1}{2}, 1)$, y no se ha hecho otra cosa que desplegar las combinaciones posibles de los componentes. La última columna muestra el orden de las respuestas (y) con el único propósito de mostrar su ubicación en el espacio experimental.

Debe observarse también que se cumplen las dos condiciones generales del diseño:

$$0.0 \leq X_i \leq 1.0 \quad \text{y} \quad \sum X_i = 1$$

Cuando se hace la gráfica del Símplex, estos puntos representan un arreglo simétrico de los componentes de la combinación con respecto a la orientación del Símplex (es decir, es simétrico con respecto a los vértices y a los lados del triángulo), como lo muestra la figura 12.

$$(x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), \left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right)$$

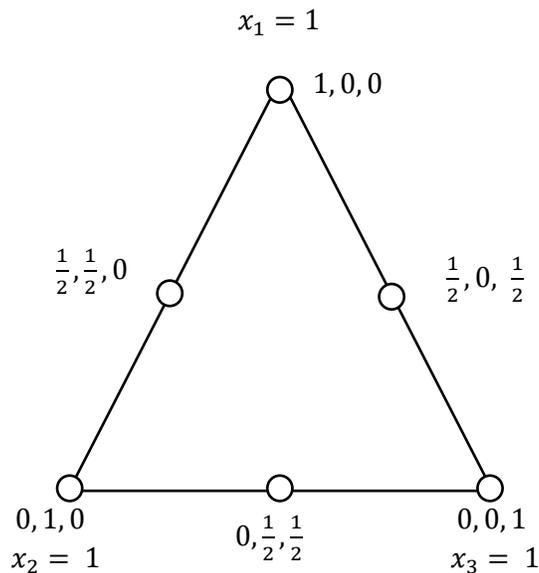


Figura 12. Diseño Símplex de grado 2 para 3 componentes $\{3, 2\}$

Los puntos experimentales se indican con un círculo y se definen por las mezclas determinadas en la tabla 2; cada mezcla corresponde a cada una de las corridas descritas.

En la tabla 1 se muestran los modelos cuadrático ($m = 2$), cúbico especial ($m = 2$), y cúbico completo ($m = 3$), para q componentes.

Una vez seleccionado el modelo que se empleará, la pregunta obligada es con respecto al número de experimentos que se deben realizar para el diseño particular. El número de puntos o experimentos a realizar se calcular con la siguiente ecuación para todos los modelos, excepto para el cúbico especial:

$$p = \binom{q + m - 1}{m} = \frac{(q + m - 1)!}{m!(q - 1)!}$$

Donde q es el número de componentes y m el orden del modelo.

Para el caso del modelo cúbico especial se emplea la ecuación:

$$p = \frac{q(q + 1)}{2} + \frac{q(q - 1)(q - 2)}{6}$$

Es altamente recomendable añadir a estos puntos experimentales, los denominados “puntos de prueba”, los cuales son algunas mezclas adicionales que serán utilizadas para el cálculo del error y de la bondad de ajuste del modelo. La composición de los puntos de prueba se estima con el arreglo:

$$\left\{ \frac{(q + 1)}{2q}, \frac{1}{2q} \right\}$$

El siguiente paso es construir la tabla de mezclas que deberán ser elaboradas en cada experimento. Hasta este punto se tendrán los siguientes elementos: componentes (q), orden del modelo (m), modelo experimental, fracciones de la mezcla, tabla de experimentos con mezclas del modelo y mezclas de puntos pruebas. El siguiente paso será realizar las mezclas experimentalmente y obtener las mediciones de las variables de respuesta seleccionadas para todas las mezclas.

En este punto, un aspecto importante para evitar el sesgo y asegurar la correcta distribución del error en toda el área experimental es la aleatorización del orden en que las mezclas se llevan a cabo. Es necesario establecer un mecanismo que permita reducir los posibles errores (aleatorios y sistemáticos), adicionalmente de los cuidados y precauciones que la naturaleza de cada mezcla impone. Para esto, el total de las mezclas se debe aleatorizar. Esto se puede lograr con una tabla de números aleatorios, hasta encontrar el total de las mezclas en la tabla.⁽⁶⁾

El diseño Símplex es un diseño con puntos frontera, es decir, que todos los puntos se encuentran sobre los límites del Símplex. Cuando nos interesa una predicción en el interior del Símplex, es deseable aumentar el tipo de diseño Símplex con un *Diseño Centroide Símplex*.⁽¹³⁾

4. Diseño de Centroide Símplex

Este diseño contiene los mismos puntos sobre la red que el diseño Símplex, pero además cuenta con un punto central en el triángulo en donde los componentes tienen un valor equivalente a la unidad dividida por el número de componentes en estudio. En este tipo de diseños de q componentes hay $2^q - 1$ puntos, que corresponden a q permutaciones de $(1, 0, 0, \dots, 0)$, $(q/2)$ permutaciones de $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$, $(q/3)$ permutaciones de $(\frac{1}{3}, \frac{1}{3}, 0, \dots, 0)$ y el centroide global $(\frac{1}{q}, \frac{1}{q}, 0, \dots, \frac{1}{q})$. En otras palabras, el diseño de centroide Símplex es un diseño en el sistema de coordenadas que se componen de mezclas que contienen 1, 2, 3, ..., q componentes en proporciones iguales. En consecuencia, existen q mezclas de un componente, todas las mezclas posibles de dos componentes con proporción $\frac{1}{2}$ para cada una, todas las mezclas posibles de tres componentes $\frac{1}{3}$ para cada una, así sucesivamente hasta una mezcla de q componentes con proporción de $\frac{1}{q}$.⁽²²⁾

El punto central se localiza mediante la determinación del nivel medio de todos los niveles de la mezcla; es decir, en este punto los componentes tienen un valor equivalente a la unidad dividida por el número de componentes en el estudio.

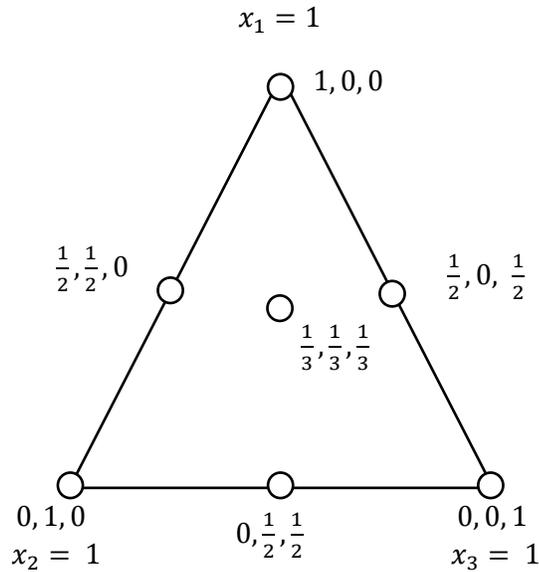


Figura 13. Diseño del centroide Símples. (x_1, x_2, x_3) $x_i = (1/q), i = 1, 2, 3$.

El diseño centroide Símples puede presentar variantes que incluyen combinaciones lineales de los componentes en estudio que caen interiormente dentro del área que forma o el triángulo o región de estudio.

En el ejemplo de la figura 14 corresponde a un centroide agregado con tres factores que presenta 3 puntos en el interior del triángulo; correspondientes a los puntos $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ y $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$.

Con estos diseños se puede dilucidar con mayor precisión la forma del área de respuesta dentro del área del triángulo.

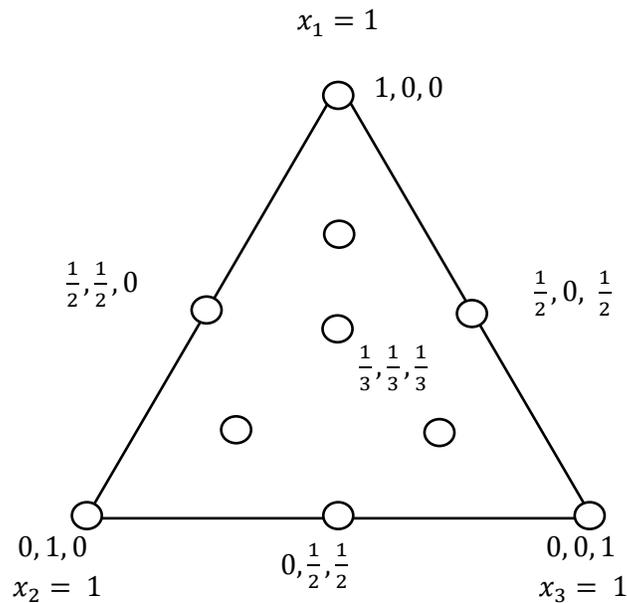


Figura 14. Combinaciones dentro del área respuesta, variante de los diseños centrales.

Las combinaciones de la mezcla para los diseños Símplex y diseño centroide Símplex se encuentra en las orillas del espacio de factores Símplex con la excepción de un punto centroide que contiene la mezcla de todos los componentes. Es posible obtener mezclas más completas si se aumentan los diseños de centroide Símplex con mezclas sobre los ejes del espacio de factores.⁽²⁵⁾

5. Modelos del Diseño de Mezclas

La forma canónica del polinomio de mezcla es derivada por la aplicación de la restricción $x_1 + x_2 + \dots + x_q = 1$ para simplificar los términos del polinomio estándar; es decir, se elimina el término constante de las ecuaciones polinomiales. Por ejemplo, para dos componentes x_1 y x_2 ; el polinomio estándar de primer grado es:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Sin embargo, como $x_1 + x_2 = 1$, se puede reemplazar β_0 por $\beta_0(x_1 + x_2)$ en Y para obtener:

$$Y = (\beta_0 + \beta_1)x_1 + (\beta_0 + \beta_2)x_2$$

$$Y = \beta_1'X_1 + \beta_2'X_2$$

De esta manera se remueve el término β_0 . En el caso del polinomio de segundo grado, los términos cuadráticos $\beta_{11}x_1$ y $\beta_{22}x_2$ son removidos del modelo, así como el término constante β_0 . Por consiguiente, los modelos de mezclas tienen menos términos que los polinomios estándar; a este tipo de formas canónicas de polinomios se les denomina modelos tipo Scheffé.⁽²⁶⁾

Este trabajo introduce a los diseños de mezclas, definida como una técnica estadística que tiene por objetivo obtener una respuesta óptima (máxima o mínima) en función de los componentes de una mezcla, consecuentemente los niveles de dichos elementos o ingredientes no son independientes, generándose con ello dos restricciones, una que indica que la cantidad máxima de un componente no es posible que sea mayor a la unidad y que la suma de los componentes de la mezcla no debe exceder la unidad (o el 100%). Por consiguiente, el espacio donde se realizarán los experimentos corresponderá a un Símplex regular, el cual en el caso de tres componentes corresponde a un triángulo equilátero. Al explorar una superficie de respuesta que corresponde a un Símplex se deben utilizar los diseños Símplex en red, los centroides Símplex o alguna de sus variantes, generándose con ello una matriz de experiencias, con ayuda de los resultados de estos experimentos se busca un polinomio de

respuesta que permita predecir respuestas para ciertas proporciones de los componentes.

En general, los modelos canónicos más importantes de mezclas son:

A) Modelo lineal

$$Y = \sum_{i=1}^k \beta_i x_i \dots (6)$$

El modelo lineal se usa en los casos en que la mezcla de los componentes es aditiva y la calidad del producto se define como una combinación lineal de sus proporciones.

B) Modelo cuadrático

$$Y = \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^j \sum_{i < j}^k \beta_{ij} x_i x_j + \varepsilon$$

El modelo cuadrático se emplea si existe interacción (antagonismo o sinergismo) entre los constituyentes de la mezcla, y por consiguiente la calidad fuese superior o inferior a la que se hubiera obtenido con la combinación lineal de sus proporciones.

C) Modelo cúbico especial

$$Y = \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^j \sum_{i < j}^k \beta_{ij} x_i x_j + \sum_{i=1}^j \sum_{i < j}^{k-j} \sum_{i < j < k}^k \beta_{ijk} x_i x_j x_k + \varepsilon$$

D) Modelo cúbico completo

$$Y = \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^j \sum_{i < j}^k \beta_{ij} x_i x_j + \sum_{i=1}^j \sum_{i < j}^k \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i=1}^j \sum_{i < j}^{k-j} \sum_{i < j < k}^k \beta_{ijk} x_i x_j x_k + \varepsilon$$

El término cúbico β_{123} identifica la diferencia en los valores de la variable dependiente con la mezcla $x_1 = \frac{1}{3}, x_2 = \frac{1}{3}, x_3 = \frac{1}{3}$ y el valor pronosticado en dicho punto con el modelo cuadrático.⁽²⁷⁾

Los modelos lineal, cuadrático y cúbico completo son generalmente asociados con $\{q, 1\}$, $\{q, 2\}$ y $\{q, 3\}$ del diseño de red Simplex, porque establecido el número de componentes $\{q\}$ sólo se requerirá una combinación binaria para generar el modelo lineal, dos para el modelo cuadrático y tres para el modelo cúbico. Con respecto a la ecuación cúbica especial se observa que se reduce a la forma de un polinomio de tercer grado que posee las combinaciones ternarias de los componentes i, j y k , por consiguiente representa un polinomio al menos de grado 2, donde es posible calcular el siguiente número de términos.

$$N = \frac{q (q^2 + 5)}{6}$$

Mientras que la fórmula para calcular el número de términos para el modelo cúbico completo es:⁽²²⁾

$$N = \frac{q (q + 1)(q + 2)}{6}$$

VII. Ejemplos de aplicación

A. Estudio de la Estabilidad Oxidativa del Biodiesel B100 en una mezcla con antioxidantes.⁽²⁸⁾

1. Planteamiento del problema.

El Biodiesel es una mezcla de ésteres de ácidos grasos, renovable y biodegradable producido por la reacción de aceites vegetales o grasa animal con metanol, en presencia de catalizadores. Los materiales usados como los aceites vegetales y las grasas usadas en el proceso utilizado en la producción de biodiesel sufren degradación hidrolítica a través de reacciones oxidativas. Los antioxidantes son una alternativa para prevenir o retardar la degradación del biocombustible. El objetivo es evaluar la estabilidad oxidativa del biodiesel B100 en la presencia de antioxidantes sintéticos, BHA (butilhidroxianisol), BHT (butilhidroxitolueno) y TBHQ (terbutilhidroxiquinona), puros o en una mezcla.

Se utilizó un diseño Símplex centroide con 7 réplicas en el punto central, con $2^q - 1$ combinaciones, donde q es el número de componentes con suma igual a 1 o 100%.

Se evaluó el periodo de inducción en horas como variable de respuesta. Tabla 3.

Tratamiento	Mezcla*	Periodo de inducción (h) 100 °C
1	(1;0;0)	12.38
2	(0; 1; 0)	12.23
3	(0; 0; 1)	11.17
4	(½; ½; 0)	11.83
5	(½; 0; ½)	14.42
6	(0; ½; ½;)	6.09
7	(1/3; 1/3; 1/3)	12.12

Tabla 3. Diseño de mezclas para el periodo de inducción (h)

* (%_{BHA}, %_{BHT}, %_{TBHQ})

**Se realizaron 7 réplicas con los mismos resultados para el cálculo del error y de la bondad de ajuste del modelo
 Los datos de la tabla 3 se trabajaron con el software estadístico Statgraphics® Centurion XVI Versión 16.1.11. La forma de crear el modelo y la introducción de los datos se muestra en el Apéndice I.

2. Resultados en Statgraphics®.

Tabla 4. Estimaciones de los efectos del modelo completo para Periodo de incubación (horas)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Mean	1839.56	1	1839.56		
Linear	18.8378	2	9.4189	1.71	0.2254
Quadratic	56.9107	3	18.9702	41.69	0.0000
Special Cubic	3.64014	1	3.64014		
Error	-6.08402E-14	7	-8.69146E-15		
Total	1918.95	14			

Tabla 5. Resultados completos de los modelos para el Periodo de inducción (horas)

Model	SE	R-Squared	Adj. R-Squared
Linear	2.34619	23.73	9.86
Quadratic	0.67455	95.41	92.55
Special Cubic	0.0	100.00	100.00

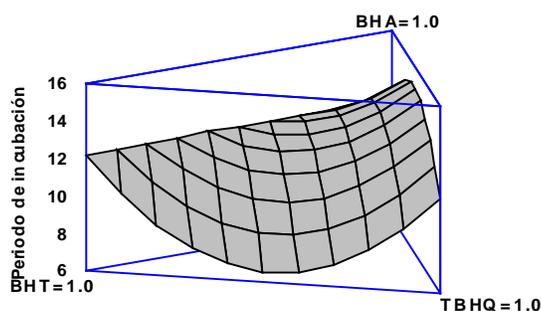


Figura 15. Superficie de respuesta estimada para la mezcla de antioxidantes.

Tabla 6. ANOVA del Periodo de inducción - Antioxidantes

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Quadratic Model	75.7485	5	15.1497	33.29	0.0000
Total error	3.6402	8	0.455024		
Total (corr.)	79.3887	13			

R-squared = 95.4147 percent

R-squared (adjusted for d.f.) = 92.5489 percent

Standard Error of Est. = 0.674555

Mean absolute error = 0.402599

Durbin-Watson statistic = 1.71212 (P=0.3044)

Lag 1 residual autocorrelation = -0.0113642

Tabla 7. Resultados del modelo cuadrático ajustado del Periodo de inducción.

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
A:BHA	12.2626	0.475172		
B:BHT	12.1126	0.475172		
C:TBHQ	11.0526	0.475172		
AB	0.448495	2.18423	0.205333	0.8424
AC	12.9285	2.18423	5.91901	0.0004
BC	-20.0915	2.18423	-9.19844	0.0000

R-squared = 95.4147 percent

R-squared (adjusted for d.f.) = 92.5489 percent

Standard Error of Est. = 0.674555

Mean absolute error = 0.402599

Durbin-Watson statistic = 1.71212 (P=0.3044)

Lag 1 residual autocorrelation = -0.0113642

The equation of the fitted model is

$$\text{Periodo de incubación} = 12.2626 \cdot \text{BHA} + 12.1126 \cdot \text{BHT} + 11.0526 \cdot \text{TBHQ} + 0.448495 \cdot \text{BHA} \cdot \text{BHT} + 12.9285 \cdot \text{BHA} \cdot \text{TBHQ} - 20.0915 \cdot \text{BHT} \cdot \text{TBHQ}$$

Respuesta óptima

Objetivo: Maximizar el Periodo de incubación

Optimum value = 14.918

Tabla 8. Valores óptimos de antioxidantes para el Periodo de Inducción.

<i>Factor</i>	<i>Low</i>	<i>High</i>	<i>Optimum</i>
BHA	0.0	1.0	0.546807
BHT	0.0	1.0	2.97049E-10
TBHQ	0.0	1.0	0.453193

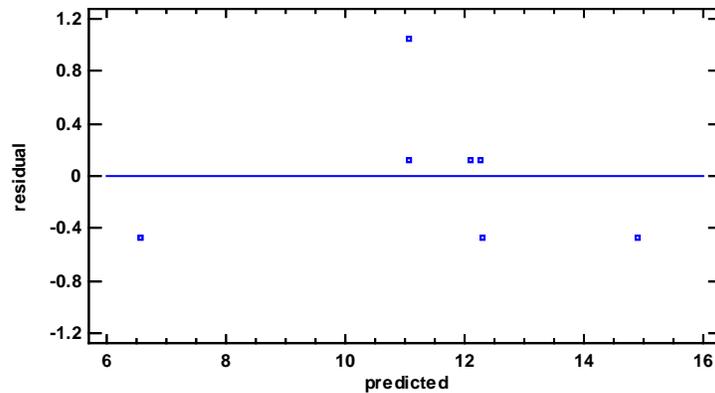


Figura 16. Gráfica de residuales para el periodo de incubación.

3. Interpretación de los resultados de Statgraphics®

La tabla 4 muestra cada uno de los modelos que explican la variación entre las variables involucradas. Normalmente, se selecciona el modelo más complicado con un *p-value* menor a 0.05, suponiendo que se está trabajando en el nivel de confianza del 95.0%. De acuerdo con este criterio, el modelo seleccionado es el modelo cuadrático. La R-cuadrada ajustada para el modelo cuadrático es de 92.55%.

El estadístico R-cuadrada indica que el modelo ajustado explica el 95,4147% de la variabilidad en el Periodo de incubación.

El análisis de varianza (tabla 6) indica que el modelo seleccionado es el adecuado para describir la variación observada en los datos, dado que el *p-value* para el modelo cuadrático es inferior a 0.05, existe una relación estadísticamente significativa entre el Periodo de incubación y los antioxidantes al 95.0% de confianza.

La prueba de Durbin-Watson (DW) para los residuos determina si existe alguna correlación significativa basada en el orden en el que ocurren las mezclas. Dado que el *p-value* es mayor que 0.05 (0.3044), no hay ninguna indicación de autocorrelación en los residuos en el nivel de significación 5,0%.

En la figura 16 se puede apreciar que no hay tendencia en los datos, los residuos no presentan estructura definida respecto de los valores predichos por el modelo, por lo que no debemos rechazar la hipótesis de homocedasticidad.

Este mismo gráfico se puede observar un comportamiento aleatorio, esto es indicio de que el modelo propuesto describe adecuadamente los datos.

La ecuación que describe el modelo es:

$$\text{Periodo de incubación} = 12.2626*BHA + 12.1126*BHT + 11.0526*TBHQ + 0.448495*BHA*BHT + 12.9285*BHA*TBHQ - 20.0915*BHT*TBHQ$$

Todos los términos positivos en la ecuación del modelo cuadrático indican que hay una influencia positiva en el periodo de inducción. Esto indica que el BHA, BHT y TBHQ cuando son utilizados aisladamente incrementan el periodo de incubación, posponiendo la oxidación del Biodiesel. Por otro lado, la combinación de los antioxidantes BHA y TBQH tienen una mayor influencia positiva a diferencia de la combinación BHA-BHT. En cambio, el empleo de la combinación de los antioxidantes BHT y TBHQ tiene un efecto negativo en la oxidación. Esto indica que el BHT no presenta sinergismo cuando se mezcla con BHA y TBHQ.

En la tabla 7 se puede observar la mezcla óptima para obtener el mayor periodo de incubación, es decir, empleando 0.546807 de BHA, 0.453193 de TBHQ y 2.9704E-10 de BHT (mol/litro) se obtendrá un tiempo de estabilidad oxidativa de 14.918 horas. Siendo éste el óptimo para el Biodiesel. Se puede observar que efectivamente el BHT es el antioxidante que utilizado en una mezcla de antioxidantes puede ocasionar efectos negativos en la oxidación.

B. Optimización de los componentes del medio de cultivo para la producción de Polihidroxiálcanoato (PHA) de Ácidos Orgánicos por *Ralstonia eutropha*.⁽²⁹⁾

1. Planteamiento del problema

Los bioplásticos actualmente están siendo desarrollados para reemplazar los productos a base de petróleo. Los PHAs tienen una mejor biocompatibilidad que los plásticos a base de petróleo, están siendo usados para desarrollar dispositivos médicos, órganos artificiales y compuestos terapéuticos. POME es una corriente acuosa que se genera en grandes cantidades durante la molienda de frutos del aceite de palma; es una fuente de carbono y puede ser fermentado para la producción de PHA por *Ralstonia eutropha*. El objetivo es encontrar mediante un análisis de mezclas cuáles son las concentraciones óptimas de acetato, propionato y butirato en el medio de cultivo para obtener el mayor contenido en por ciento peso-peso de PHA.

Para el diseño de análisis de mezclas se utilizó un modelo Símples lattice aumentado. El diseño de experimentos contiene una serie de 13 combinaciones. La variable de respuesta es el contenido de PHA en por ciento de peso.

La serie de mezclas se muestra en la tabla 9.

Corrida	Acetato	Propionato	Butirato	PHA (%)
1	1.00	0.00	0.00	66.5
2	0.67	0.33	0.00	71.1
3	0.67	0.00	0.33	76.8
4	0.33	0.67	0.00	72.9
5	0.33	0.33	0.33	77.6
6	0.33	0.00	0.67	83.0
7	0.00	1.00	0.00	70.0
8	0.00	0.67	0.33	76.8
9	0.00	0.33	0.67	83.7
10	0.00	0.00	1.00	82.2
11	0.67	0.17	0.17	77.7
12	0.17	0.67	0.17	74.6
13	0.17	0.17	0.67	78.1

Tabla 9. Diseño de mezclas para la producción de PHA

Los datos de la tabla 9 se trabajaron con el software estadístico Statgraphics® Centurion XVI Versión 16.1.11. La forma de crear el modelo y la introducción de los datos se muestra en el Apéndice I.

2. Resultados en Statgraphics®.

Tabla 10. Estimaciones de los efectos del modelo completo para PHA (%)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Mean	75544.7	1	75544.7		
Linear	242.538	2	121.269	15.86	0.0008
Quadratic	41.7218	3	13.9073	2.80	0.1182
Special Cubic	4.06593	1	4.06593	0.80	0.4069
Cubic	6.90255	3	2.30085	0.29	0.8316
Error	23.7795	3	7.9265		
Total	75863.7	13			

Tabla 11. Resultados completos de los modelos para la producción de PHA.

Model	SE	R-Squared	Adj. R-Squared
Linear	2.76532	76.03	71.23
Quadratic	2.228	89.11	81.33
Special Cubic	2.26134	90.38	80.76
Cubic	2.8154	92.55	70.18

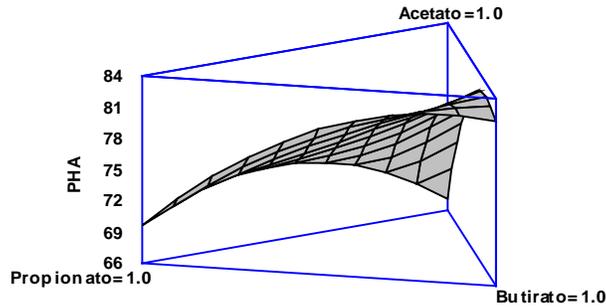


Figura 17. Superficie de respuesta estimada para la mezcla de ácidos volátiles.

Tabla 12. ANOVA del PHA - Ácidos orgánicos

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Special Cubic Model	285.548	6	47.5914	8.53	0.0098
Total error	33.4594	6	5.57657		
Total (corr.)	319.008	12			

R-squared = 89.5114 percent

R-squared (adjusted for d.f.) = 79.0228 percent

Standard Error of Est. = 2.36148

Mean absolute error = 1.17753

Durbin-Watson statistic = 1.27529 (P=0.1015)

Lag 1 residual autocorrelation = 0.117684

Tabla 13. Resultados del modelo Cúbico especial ajustado para PHA

Parameter	Estimate	Standard Error	T Statistic	P-Value
A:Acetato	67.1243	2.18947		
B:Propionato	69.5545	2.18947		
C:Butirato	81.7823	2.18947		
AB	18.8241	10.532	1.78733	0.1241
AC	24.2304	10.532	2.30066	0.0610
BC	17.0936	10.532	1.62303	0.1557
ABC	-61.066	69.1524	-0.883064	0.4112

R-squared = 89.5114 percent

R-squared (adjusted for d.f.) = 79.0228 percent

Standard Error of Est. = 2.36148

Mean absolute error = 1.17753

Durbin-Watson statistic = 1.27529 (P=0.1015)

Lag 1 residual autocorrelation = 0.117684

The equation of the fitted model is

$$\text{PHA} = 67.1243 \cdot \text{Acetato} + 69.5545 \cdot \text{Propionato} + 81.7823 \cdot \text{Butirato} + 18.8241 \cdot \text{Acetato} \cdot \text{Propionato} + 24.2304 \cdot \text{Acetato} \cdot \text{Butirato} + 17.0936 \cdot \text{Propionato} \cdot \text{Butirato} - 61.066 \cdot \text{Acetato} \cdot \text{Propionato} \cdot \text{Butirato}$$

Respuesta óptima

Objetivo: Maximizar PHA

Valor óptimo = 82.7277

Tabla14. Valores óptimos de ácidos orgánicos para la producción de PHA.

<i>Factor</i>	<i>Low</i>	<i>High</i>	<i>Optimum</i>
Acetato	0.0	1.0	0.197505
Propionato	0.0	1.0	2.23033E-9
Butirato	0.0	1.0	0.802495

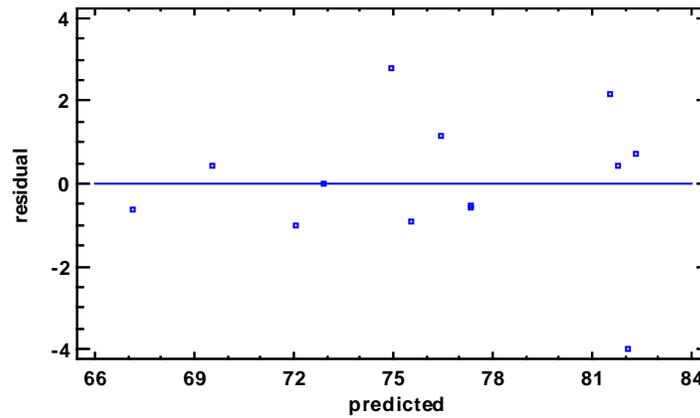


Figura 18. Gráfica de residuales para el contenido de PHA.

3. Interpretación de los resultados de Statgraphics®

La tabla 10 muestra cada uno de los modelos que explican la variación entre las variables involucradas, El modelo seleccionado es el cúbico especial. La R-cuadrada ajustada para el modelo es de 80.76%. Se selecciona el modelo cúbico especial para poder interpretar el efecto de los 3 ácidos en el medio de cultivo.

El estadístico R-cuadrada indica que el modelo ajustado explica el 80.76% de la variabilidad en el proceso de producción de PHA.

El análisis de varianza (tabla 12) indica que el modelo seleccionado es el adecuado para describir la variación observada en los datos, dado que el *p-value* para el modelo cúbico especial es inferior a 0.05, existe una relación estadísticamente significativa entre los ácidos orgánicos y la producción de PHA al 95.0% de confianza.

La prueba de Durbin-Watson (DW) para los residuos determina si existe alguna correlación significativa basada en el orden en el que ocurren las mezclas. Dado que el *p-value* es mayor que 0.05 (0.3044), no hay ninguna indicación de autocorrelación en los residuos en el nivel de significación 5,0%.

En la figura 18 se puede se puede apreciar que no hay tendencia en los datos, los residuos no presentan estructura definida respecto de los valores predichos por el modelo por lo que no debemos rechazar la hipótesis de homocedasticidad.

Este mismo gráfico se puede observar un comportamiento aleatorio, esto es indicio de que el modelo propuesto describe adecuadamente los datos.

La ecuación que describe el modelo es:

$$\begin{aligned}
 \text{PHA} = & 67.1243 \cdot \text{Acetato} + 69.5545 \cdot \text{Propionato} + 81.7823 \cdot \text{Butirato} + \\
 & 18.8241 \cdot \text{Acetato} \cdot \text{Propionato} + 24.2304 \cdot \text{Acetato} \cdot \text{Butirato} + 17.0936 \cdot \text{Propionato} \cdot \text{Butirato} \\
 & - 61.066 \cdot \text{Acetato} \cdot \text{Propionato} \cdot \text{Butirato}
 \end{aligned}$$

Todos los términos positivos en la ecuación del modelo cúbico especial indican que hay una influencia positiva en la producción de PHA. Esto indica que el acetato, propionato y butirato cuando son utilizados aisladamente incrementan la producción de PHA. Por otro lado, la combinación de acetato-butirato y propionato-butirato tiene mayor influencia en la producción de PHA, a diferencia de la combinación acetato-propionato, esto se debe a que el butirato es el ácido que va a producir la mayor cantidad de PHA en el medio de cultivo, por lo tanto, si se utiliza en combinación con cualquiera de los otros ácidos, la producción de PHA se verá beneficiada.

En la tabla 14 se puede observar la mezcla óptima para obtenerla mayor cantidad de PHA en el medio de cultivo, es decir, empleando 0.197505 de acetato, 0.802495 de butirato y 2.23033E-9 de propionato (por ciento p/v) se obtendrá 82.7277 % (p/p) de Polihidroxiálcanoato (PHA). Se puede observar que el propionato es el ácido orgánico que utilizado en una mezcla de ácidos puede ocasionar efectos negativos en la producción de PHA, por lo tanto, a mayor cantidad de butirato, el cual es consumido por *R. eutropha* más rápido que los otros ácidos, mayor será la cantidad producida de PHA.

C. Diseño de Mezclas como Primer Paso para Mejorar la Producción de Glutaminasa en la Fermentación en Estado Sólido por *Bacillus sp.* RSP-GLU⁽³⁰⁾

1. Planteamiento del problema

La glutaminasa tiene importancia en el sector industrial y farmacéutico como un agente efectivo en el tratamiento de la leucemia aguda, como un agente analítico, agente biosensor, como potenciador de sabor y en la producción de especialidades químicas como la treonina por reacciones de transferencia de glutamil gama. La comunidad científica está involucrada en desarrollar diferentes estrategias de fermentación para mejorar la productividad. La producción de glutaminasa se ha reportado bajo fermentación en estado sólido utilizando salvado de trigo, aceite de sésamo y perlas de poliestireno. El objetivo es seleccionar la mejor combinación de sustratos sólidos y evaluar el impacto de la mezcla en la producción de glutaminasa. Se utilizó un diseño Simplex centroide aumentado que consiste en $7(2^3 - 1)$ corridas, además se adicionaron 3 corridas de acuerdo a la siguiente tabla:

Corrida	Salvado de trigo	Garbanzo	Semilla de palma	Actividad Glutaminasa (U/gds)
1	1	0	0	1059.63
2	0	1	0	2789.34
3	0	0	1	1964.31
4	½	½	0	2927.06
5	0	½	½	2671.32
6	½	0	½	2132.0
7	1/3	1/3	1/3	2946.59
8	2/3	1/6	1/6	2698.61
9	1/6	2/3	1/6	3165.58
10	1/6	1/6	2/3	2605.82

Tabla 15. Diseño de mezclas para la producción de Glutaminasa.

*U/gds = Unidades por gramo de sustrato seco

Los datos de la tabla 15 se trabajaron con el software estadístico Statgraphics® Centurion XVI Versión 16.1.11. La forma de crear el modelo y la introducción de los datos se muestra en el Apéndice I.

2. Resultados en Statgraphics®.

Tabla 16. Estimaciones de los efectos del modelo completo para la Actividad Glutaminasa (U/gds)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Mean	6.23015E7	1	6.23015E7		
Linear	1.02225E6	2	511123.	1.45	0.2967
Quadratic	2.35521E6	3	785069.	29.08	0.0035
Special Cubic	46312.2	1	46312.2	2.25	0.2303
Error	61663.8	3	20554.6		
Total	6.57869E7	10			

Tabla 17. Resultados completos de los modelos prara la Actividad glutaminasa

Model	SE	R-Squared	Adj. R-Squared
Linear	593.197	29.33	9.14
Quadratic	164.298	96.90	93.03
Special Cubic	143.369	98.23	94.69

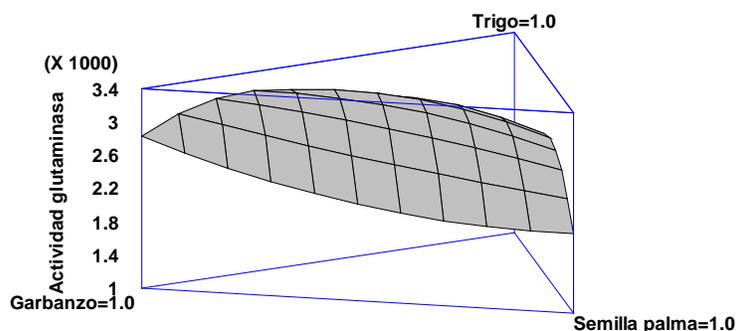


Figura 19. Superficie de respuesta estimada para la mezcla de sustratos.

Tabla 18. ANOVA de la Actividad Glutaminasa - Sustratos

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Special Cubic Model	3.42376E6	6	570627.	27.76	0.0101
Total error	61662.9	3	20554.3		
Total (corr.)	3.48543E6	9			

R-squared = **98.2308** percent

R-squared (adjusted for d.f.) = **94.6925** percent

Standard Error of Est. = **143.368**

Mean absolute error = **56.5548**

Durbin-Watson statistic = 1.8011 (P=0.3856)

Lag 1 residual autocorrelation = 0.096132

Tabla 19. Resultados del modelo Cúbico especial ajustado para la Actividad Glutaminasa.

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
A:Trigo	1079.84	138.592		
B:Garbanzo	2828.69	138.592		
C:Semilla palma	1949.25	138.592		
AB	4129.43	697.645	5.9191	0.0096
AC	4647.71	697.645	6.662	0.0069
BC	-930.711	697.645	-1.33408	0.2744
ABC	6903.88	4599.35	1.50105	0.2303

R-squared = 98.2308 percent

R-squared (adjusted for d.f.) = 94.6925 percent

Standard Error of Est. = 143.368

Mean absolute error = 56.5548

Durbin-Watson statistic = 1.8011 (P=0.3856)

Lag 1 residual autocorrelation = 0.096132

Actividad glutaminasa = 1079.84*Trigo + 2828.69*Garbanzo + 1949.25*Semilla palma + 4129.43*Trigo*Garbanzo + 4647.71*Trigo*Semilla palma - 930.711*Garbanzo*Semilla palma + 6903.88*Trigo*Garbanzo*Semilla palma

Respuesta óptima

Objetivo: maximizar la producción de Glutaminasa

Optimum value = 3172.61

Tabla 20. Valores óptimos de sustratos para la producción de glutaminasa.

<i>Factor</i>	<i>Low</i>	<i>High</i>	<i>Optimum</i>
Trigo	0.0	1.0	0.309885
Garbanzo	0.0	1.0	0.642345
Semilla palma	0.0	1.0	0.0477707

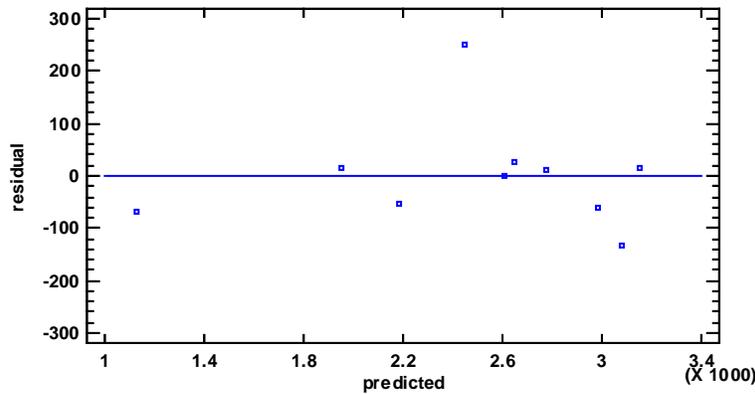


Figura 20. Gráfica de residuales para la Actividad de Glutaminasa

3. Interpretación de los resultados de Statgraphics®.

La tabla 16 muestra cada uno de los modelos que explican la variación entre las variables involucradas, El modelo seleccionado es el cúbico especial. La R-cuadrada ajustada para el modelo es de 94.69%. Se selecciona el modelo cúbico especial para poder interpretar el efecto de los 3 sustratos en el medio de cultivo.

El estadístico R-cuadrada indica que el modelo ajustado explica el 94.69% de la variabilidad en el proceso de producción de la enzima glutaminasa.

El análisis de varianza (tabla 18) indica que el modelo seleccionado es el adecuado para describir la variación observada en los datos, dado que el *p-value* para el modelo cúbico especial es inferior a 0.05, existe una relación estadísticamente significativa entre los sustratos y la producción de glutaminasa al 95.0% de confianza.

La prueba de Durbin-Watson (DW) para los residuos determina si existe alguna correlación significativa basada en el orden en el que ocurren las mezclas. Dado que el *p-value* es mayor que 0.05 (0.7556), no hay ninguna indicación de autocorrelación en los residuos en el nivel de significación 5,0%.

En la figura 20 se puede apreciar que no hay tendencia en los datos, los residuos no presentan estructura definida respecto de los valores predichos por el modelo por lo que no debemos rechazar la hipótesis de homocedasticidad. En este mismo gráfico se puede observar un comportamiento aleatorio, esto es indicio de que el modelo propuesto describe adecuadamente los datos.

La ecuación que describe el modelo es:

$$\text{Actividad glutaminasa} = 1079.84 * \text{Trigo} + 2828.69 * \text{Garbanzo} + 1949.25 * \text{Semilla palma} + 4129.43 * \text{Trigo} * \text{Garbanzo} + 4647.71 * \text{Trigo} * \text{Semilla palma} - 930.711 * \text{Garbanzo} * \text{Semilla palma} + 6903.88 * \text{Trigo} * \text{Garbanzo} * \text{Semilla palma}$$

Todos los términos positivos en la ecuación del modelo cúbico especial indican que hay una influencia positiva en la producción de la enzima glutaminasa. Esto indica que el trigo, garbanzo y la semilla de palma cuando son utilizados aisladamente incrementan la producción de glutaminasa; principalmente el garbanzo. Por otro lado, la combinación trigo-garbanzo y trigo-semilla de palma, favorecen la producción de glutaminasa; a diferencia de la combinación garbanzo-semilla de palma, teniendo un efecto negativo en la producción de la enzima. La combinación de sustratos que dará la mayor producción de enzima, es cuando se combinan los 3 sustratos.

En la tabla 20 se puede observar la mezcla óptima para obtenerla mayor cantidad de glutaminasa por fermentación de sustratos como el trigo, garbanzo y semilla de palma, es empleando 0.309885g de trigo, 0.642345g de garbanzo y 0.0477707g de semilla de trigo; se obtendrá 3,172.61 U/gds de glutaminasa.

VIII. Conclusiones

Los diseños de mezclas son una herramienta estadística de gran utilidad en el desarrollo y optimización de procesos dentro de las industrias dedicadas a la producción de precursores y/o sustancias químicas, tal es el caso de las farmoquímicas. Mediante este tipo de diseños, el experimentador tiene una visión general de las posibles respuestas generadas a partir de una cantidad mínima de datos experimentales independientemente del diseño de mezclas que se utilice.

Algunos diseños de tratamientos tienen dos o más factores que son ingredientes de una mezcla, en la que los porcentajes de cada uno deben sumar el 100% de la mezcla. En los experimentos factoriales, los niveles de cada factor son independientes de los niveles de otros factores; por lo tanto, una de las ventajas de los diseños de mezclas es que los niveles de un factor no son independientes de los niveles de otros. En los experimentos con mezclas, los factores son los componentes o ingredientes de la mezcla y por consecuencia, sus niveles no son independientes, de esta manera, estas variables controlables representan cantidades proporcionales a la mezcla en vez de cantidades sin restricción.

La variación en las proporciones de los ingredientes de las mezclas puede afectar las propiedades del producto final. Más que en las cantidades totales de los factores, las investigaciones de experimentos con mezclas se concentran en la relación de la

variable de respuesta medida con las proporciones relativas de cada ingrediente presente en el producto.

El espacio de experimentación de los diseños de mezclas corresponde a un símplex regular y para explorarlo se deben utilizar los diseños símplex en red, los símplex centroide o alguna de sus variantes; el primero de estos diseños se emplea cuando se desea analizar un problema en donde es posible tener cada uno de los componentes en una cantidad del 100%, sin embargo, se recomienda añadir puntos centrales (símplex centroide) los cuales son unas mezclas adicionales para el cálculo del error y la bondad de ajuste. En muchos experimentos con mezclas existen restricciones en las proporciones de los componentes que no permiten explorar toda la región del símplex. Regularmente estas restricciones toman formas de límites inferiores y superiores para cada una de las proporciones de los componentes.

Los modelos que se generan con los diseños de mezclas tienen un menor número de términos que los polinomios estándar, los modelos usuales en el diseño de mezclas son: el lineal, cuadrático, cúbico especial y cúbico completo; el criterio para seleccionar alguno de estos modelos se fundamenta en la existencia de interacciones en los componentes de la mezcla. Por consiguiente, el primero de los modelos mencionados; el lineal, se empleará cuando la mezcla de los componentes presente un comportamiento aditivo en la respuesta; el cuadrático, si existe alguna interacción en la combinación binaria de los componentes de la mezcla; el cúbico especial y cúbico

completo evalúan el comportamiento de la combinación de los componentes. El siguiente paso después de generar cualesquiera de los modelos anteriores es, a través de una prueba de hipótesis, determinar si la respuesta depende de los componentes de la mezcla y cuando ello resulta ser afirmativo, se realiza la determinación del coeficiente de ajuste para decidir si el modelo explica la suficiente cantidad de variación en los valores de la respuesta.

Una de las ventajas del diseño de mezclas es que son capaces de evaluar más de una respuesta mediante la superposición de las zonas de mejor respuesta, consiguiendo con ello, obtener las proporciones de los componentes en las cuales el producto cumplirá con las características de calidad deseadas, siempre considerando que se deberá corroborar dicha información mediante la experimentación.

El uso del diseño de mezclas se presenta en este trabajo como una herramienta sencilla de utilizar, siempre y cuando se cuente con un paquete estadístico que facilite la elaboración de los cálculos. Statgraphics® es un software “fácil” de utilizar ya que presenta una interfaz amigable para el usuario, similar a cualquier hoja de cálculo electrónica, proporciona la simplicidad suficiente para permitir a un investigador con poca experiencia realizar procedimientos complejos.

Los ejemplos de aplicación expuestos en este trabajo proporcionan una visión detallada de los resultados que se obtienen con el software estadístico Statgraphics®, los cuales

se pueden manejar de acuerdo a las necesidades de cada experimentador y el objetivo que se quiera lograr: minimizar o maximizar una variable de respuesta.

Con el presente trabajo se pretende impulsar el uso de herramientas estadísticas para la resolución de problemas en la investigación. El diseño de mezclas no solo tiene aplicación en la industria para la síntesis de especies químicas (farmoquímica), también tiene aplicación en la industria farmacéutica, en el desarrollo de nuevas formulaciones o mejorar las existentes. El alcance del diseño de mezclas es para cualquier industria que elabore un producto a partir de componentes de una mezcla, desde la industria alimentaria, biotecnológica hasta la industria petroquímica, etcétera.

IX. REFERENCIAS

1. Stan Altan JB, Lori Pfahler, Edith Senderak, Shanthi Sethuraman, y Kim Erland Vukovinsky. Consideraciones Estadísticas en el Desarrollo del Espacio de Diseño Parte I de III. Pharm Technol. 2010;8(4).
2. Kalariya PD, Namdev D, Srinivas R, Gananadhamu S. Application of experimental design and response surface technique for selecting the optimum RP-HPLC conditions for the determination of moxifloxacin HCl and ketorolac tromethamine in eye drops. Journal of Saudi Chemical Society. 2014;18(2).
3. Gutierrez Pulido H. Análisis y Diseño de Experimentos 3ed. México: McGraw-Hill Interamericana; 2012.
4. Castillo P, Magninb J-P, Velasquez M, Willisonc J. Modeling and Optimization of Hydrogen Production By The Photosynthetic Bacterium Rhodobacter capsulatus By The Methodology Of Design Of Experiments (DOE): Interaction Between Lactate Concentration And Light Luminosity. Energy Procedia. 2012;29:357-66.
5. Kavanaugh CF. Los Diseños de Taguchi contra los Diseños Clásicos de Experimentos. Concinecia Tecnológica. 2002(19).
6. Ramirez MLV. Diseño de Experimentos con Mezclas. México: Asociación Farmacéutica Mexicana; 2003.
7. Salazar. JC, Zapata AB. Análisis y Diseño de Experimentos Aplicados a Estudios de Simulación. Dyna. 2008;76(159):249-57.
8. Encinia FB, Garza ÁM. Eficiencia de Algunos Diseños Experimentales en la Estimación de una Superficie de Respuesta. Agrociencia. 2002;36(2):201-10.

9. Stan Altan J, Bergum, Loru Pfhaler, Edith Senderak, Shanthi, Kim Erland Vukovinsky. Consideraciones Estadísticas en el Desarrollo del Espacio Diseño parte II de III. Pharm Technol. 2010;8(5):22-30, 3.
10. Sangshetti JN, Mrinmayee Deshpande , Arote R, Zaheer Z, Shinde DB. Quality by Design Approach: Regulatory Need. Arabian Journal of Chemistry. 2014;7(1).
11. Sánchez Ruiz JF. Aplicaciones del Diseño de Experimentos a la Farmacia y Química Clínica. México: UNAM; 1995.
12. Tostado EC. Diseño de Experimentos: Estrategias y Análisis en Ciencia y Tecnología. Querétaro: Centro de Investigación en Matemáticas; 2010.
13. Montgomery DC. Diseño y Analisis de Experimentos. 2 ed. México: Limusa; 2002.
14. Box GEP, Draper NR. Empirical Model-Building and Response Surfaces. USA: John Wiley and Sons; 1987.
15. Erjavec J, Madrigal JL. Estrategias Experimentales para el Mejoramiento de la Calidad en la Industria. México: Iberoamericana; 1992.
16. Box GEP. Estadística para Investigadores : Introduccion al Diseño de Experimentos, Analisis de Datos y Construccion de Modelos. México: Reverté; 1999.
17. Box GEP. Estadística para Investigadores : Diseño, Innovación y Descubrimiento. Barcelona: Reverté; 2008.
18. Box GEP, Draper NR. Response Surfaces, Mixtures, and Ridge Analyses. New Jersey: Wiley InterScience; 2007.
19. Kuehl RO. Diseño de Experimentos : Principios Estadísticos de Diseño y Análisis de Investigación. 2^a ed. México: Thomson Learning; 2001. 666 p.
20. Cochran WG, Cox GM. Diseños Experimentales. México: Trillas; 1983. 373-408 p.

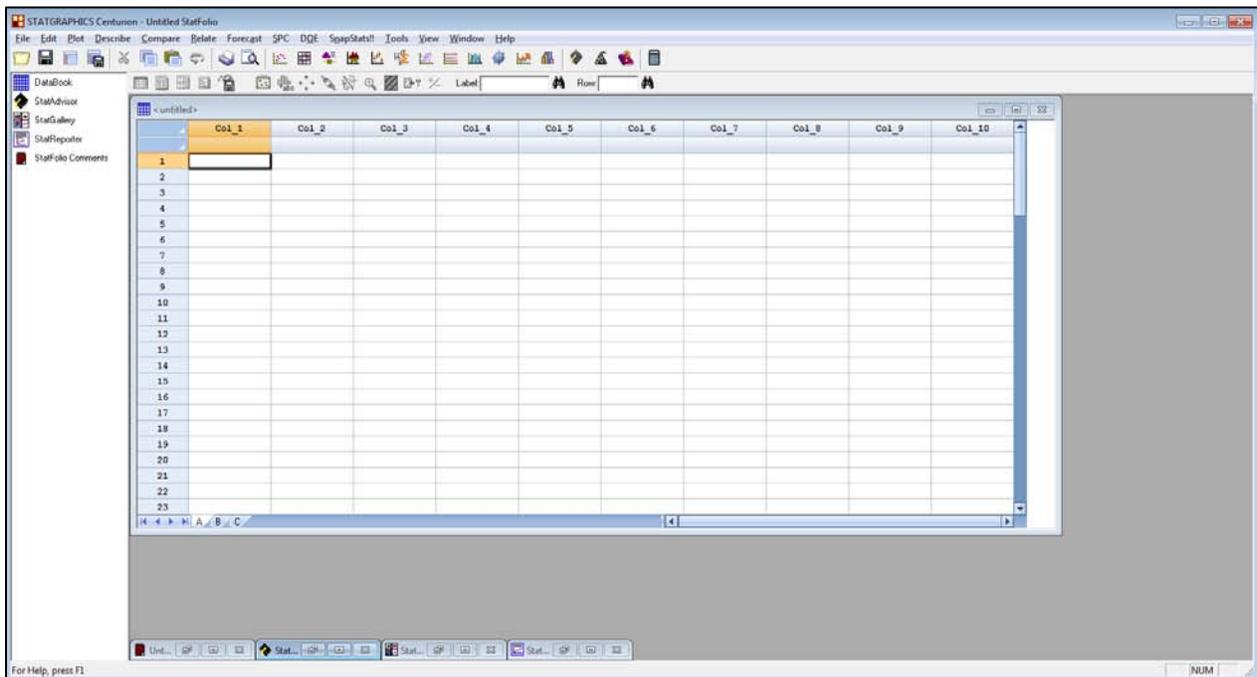
21. Wu J, Hamada MS. Experiments: Planning, Analysis, and Optimization. USA: John Wiley and Sons, Inc.; 2000.
22. Cornell JA. Experiments With Mixtures: Design, Models, and The Analysis of Mixture Data. 2 ed 2011. 1-52 p.
23. Piepel GF, Cornell JA. Mixtures Experiments Approaches: Examples, Discussion, and Recommendations. Journal of Quality Technology. 1994;26(3):177-96.
24. Mason RL, Gunst RF, Hess JL. Statistical Design and Analysis of Experiments: With applications to Engineering and Science. USA 1989.
25. Cornell JA, Deng JC. Combining Process Variables and Ingredient Components in Mixing Experiments. Journal of Food Science. 1992; 47(3): 836-43.
26. Atkinson AC, Donev AN. Optimum Experimental Designs. New York: Oxford University; 1992.
27. Zamora Ramirez R. Diseño de Mezclas: Una Herramienta Estadística en la Industria Farmacéutica. México: Facultad de Estudios Superiores Zaragoza UNAM; 2000.
28. Maia EC, Borsato D, Moreira I, Spacino KR, Rodrigues PR, Gallina AL. Study of the Biodiesel B100 Oxidative Stability in Mixture with Antioxidants. Fuel Processing Technology. 2011;92(9):1750-5.
29. Yang Y-H, Brigham CJ, Budde CF, Boccuzzi P, Willis LB, Hassan MA, et al. Optimization of Growth Media Components for Polyhydroxyalkanoate (PHA) Production from Organic Acids By *Ralstonia eutropha*. Biotechnological Products And Process Engineering. 2010;87(6):2037-45.
30. Satnish T, Lakshmi GS, Rao C, Brahmaiah P, Prakasham RS. Mixture Design as First Step for Improved Glutaminase Production in Solid-State Fermentation by Isolated *Bacillus* sp. RSP-GLU. Letters in Applied Microbiology. 2008;47(4):256-62.

X. APÉNDICE

GUÍA DE USUARIO DEL SOFTWARE ESTADÍSTICO STATGRAPHICS® CENTURION XVI VERSIÓN 16.1.11 PARA ANÁLISIS DE MEZCLAS

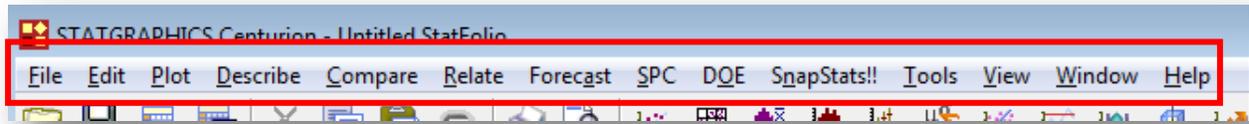
1. Introducción.

El programa Statgraphics® es un software que está diseñado para facilitar el análisis estadístico de datos. Mediante su aplicación es posible realizar un análisis descriptivo de una o varias variables, utilizando gráficos que expliquen su distribución o calculando sus medidas características. El programa trabaja en un entorno WINDOWS® y su pantalla principal (a la que se ingresa ejecutando el programa SGWIN.EXE o directamente dando clic sobre el icono correspondiente) es la siguiente:



En la pantalla principal de Statgraphics® podemos distinguir los siguientes elementos:

1. Barra de menú



3. Barra de herramientas



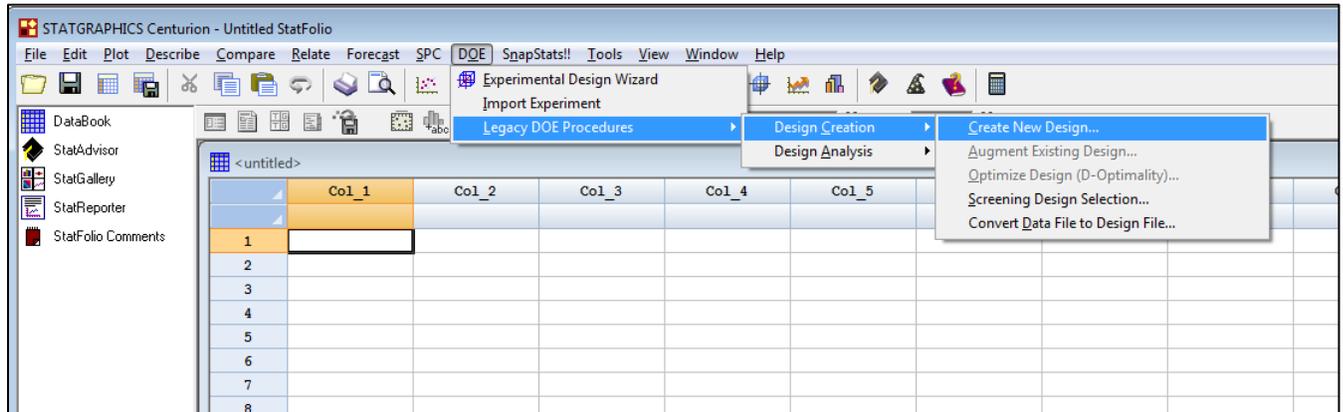
3. Barra de tareas



2. Creación del diseño.

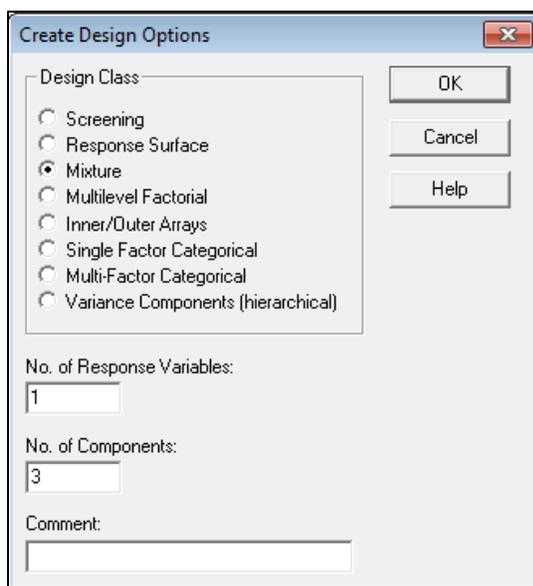
Los experimentos de mezclas son creados seleccionando en la barra de herramientas

DOE → Legacy DOE Procedures → Design Creation → Create New Design...



Cuadro de Dialogo #1 – Tipo de Diseño

El primer cuadro de diálogo que se despliega durante la creación del diseño se muestra abajo:

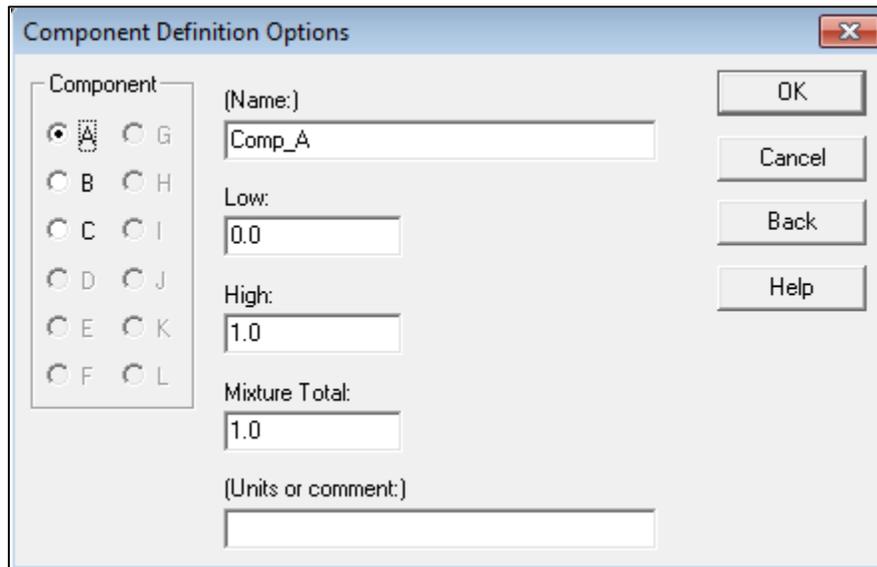


- **Design Class** (Tipo de Diseño): Tipo de diseño a ser creado.
- **No. of Response Variables** (No. de variables respuesta): El número de variables de respuestas *Y* que deberán medirse durante cada corrida experimental. Este número está en un rango de 1 hasta 16.
- **No. of Components** (Número de componentes): el número de componentes *X* que se pueden variar de corrida a corrida. Para diseños de mezclas, el número de componentes está en un rango de 2 hasta 12.
- **Comments** (Comentarios): Un comentario que aparecerá sobre las salidas de los procedimientos de análisis.

Seleccionar en *Design Class, Mixture*; en el campo *No. of Response Variables* seleccionar el número de variables de respuesta *Y* que desean medirse y en el campo *No. of Components*, el número de componentes de la mezcla. Dar clic en el botón *OK* y aparecerá el siguiente cuadro de diálogo.

Cuadro de Dialogo #2 – Componentes experimentales

El segundo cuadro de diálogo requiere información acerca de cada uno de los componentes de la mezcla



Dar clic sobre las letras A, B, C,..., solo una vez e ingrese la siguiente información para cada componente del diseño:

- **Name** (Nombre): El nombre del componente es similar a un nombre de columna en una base de datos estándar de Statgraphics®, cada nombre debe tener hasta 32 caracteres y no contener cualquier carácter especial u operadores algebraicos.
- **Low** (Bajo): El límite inferior L_j para el componente.
- **High** (Alto): El límite superior U_j para el componente.
- **Mixture Total** (Mezcla Total): El total o suma de T de q componentes. Los niveles de los componentes en todas las corridas deberán sumar este total.

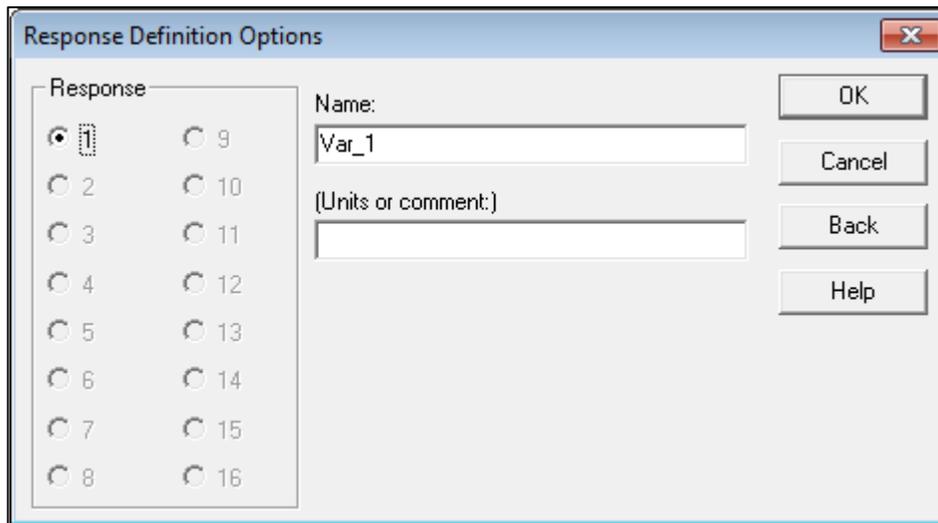
- ***Units or Comment*** (Unidades o Comentario): Una etiqueta opcional o un comentario de hasta 64 caracteres que se incluyen sobre la hoja de trabajo experimental.

Cuando el botón *OK* es presionado para continuar, algunos de estos diversos eventos pueden ocurrir:

1. Si cualquier límite superior es menor que cualquier límite inferior, se presenta un mensaje de error.
2. Si la suma del campo *Low* es mayor que *T*, se presenta un mensaje de error.
3. Si las restricciones son inconsistentes, por ejemplo, un valor alto nunca será alcanzado, el valor es automáticamente ajustado para hacer las restricciones consistentes, después de presionar el botón *OK* debe presionarlo otra vez.

Si las restricciones no plantean ningún problema, se presenta el siguiente cuadro de dialogo:

Cuadro de Diálogo #3 – Variables de Respuesta

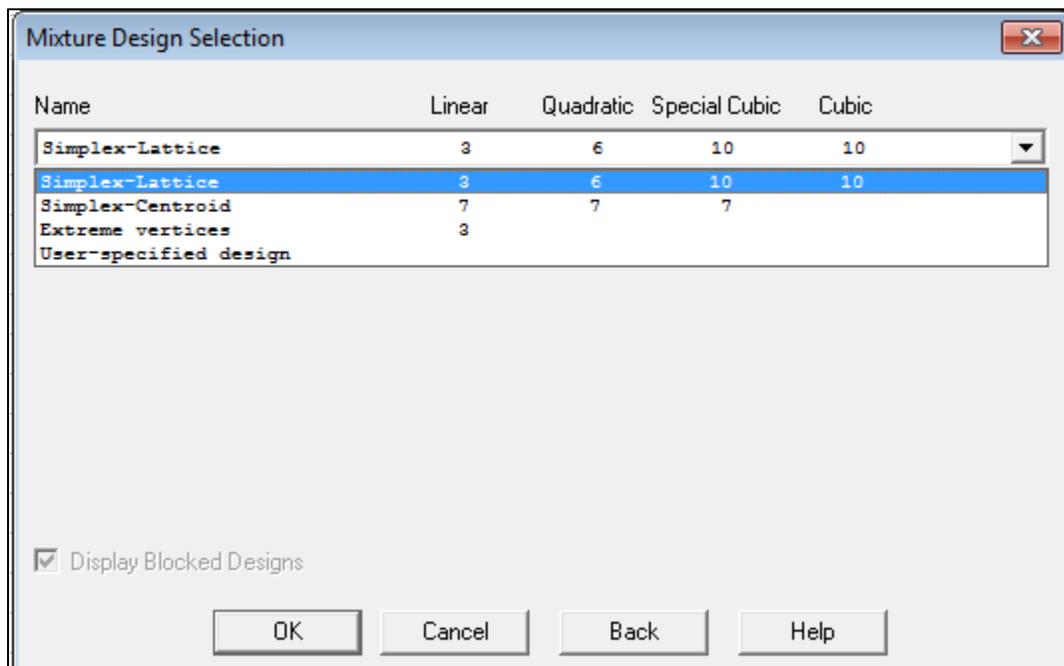


Dar clic sobre los números 1, 2, 3,..., solo una vez e ingrese la siguiente información para cada variable respuesta del experimento:

- **Name** (Nombre): Un nombre para cada respuesta conteniendo hasta 32 caracteres.
- **Units or Comment** (Unidades o Comentario): Una etiqueta opcional o comentario hasta 64 caracteres que se incluyen sobre la hoja de trabajo experimental.

Dar clic en el botón *OK* para pasar al siguiente cuadro de diálogo.

Cuadro de Diálogo #4 – Catálogo de Diseños.



El cuarto cuadro de diálogo muestra una lista de todos los diseños de mezclas del catálogo en Statgraphics® para un número seleccionado de componentes:

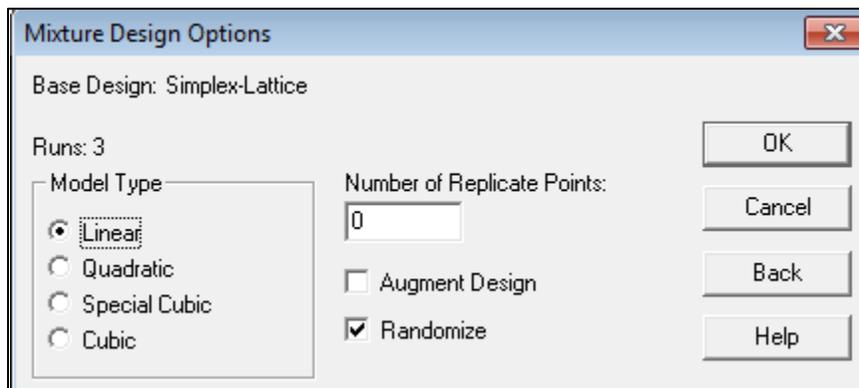
- **Name** (Nombre): El nombre del diseño. Para diseños de mezclas, los siguientes tipos aparecen en la lista, dependiendo del número de componentes y la naturaleza de cualquier restricción impuesta:
 1. Simplex-Lattice: Un conjunto de puntos uniformemente espaciados sobre un simplex, donde el simplex es regularmente el lado de un figura con q vértices en $q - 1$ dimensiones (Ejemplo: un triángulo en un espacio bi-dimensional). Dependiendo de las restricciones, estos diseños pueden no estar disponibles.
 2. Simplex-Centroide: Un diseño que consiste de $2(q - 1)$ corridas en todas las mezclas primarias, mezclas binarias, mezclas terciarias, etc., hasta el diseño

centroide. Dependiendo de las restricciones, estos diseños pueden no estar disponibles.

3. Vértices Extremos: Un diseño que consiste de una corrida por cada vértice de la región factible. Este diseño siempre está disponible.
4. Especificado-Usuario: Las columnas son creadas en la base de datos para cada componente y la variable respuesta, pero el usuario debe definir las corridas a ser desarrolladas.

Una vez seleccionado el tipo de diseño que se requiere realizar dar clic en el botón OK para continuar y aparecerá el último cuadro de diálogo:

Cuadro de Diálogo #5 – Opciones del Diseño.



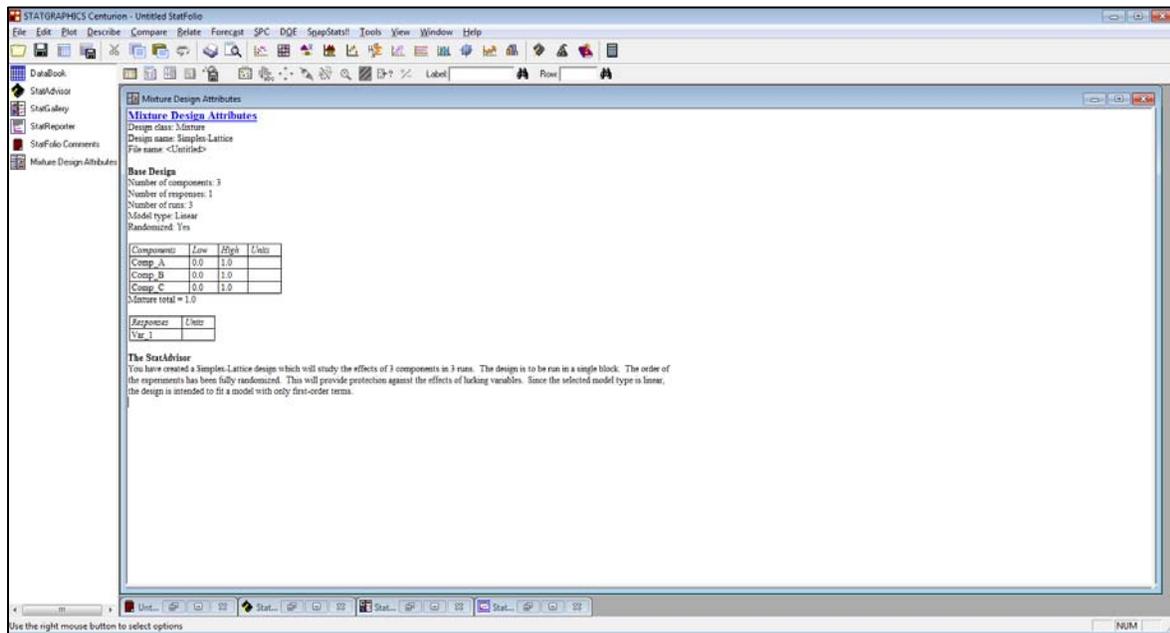
El cuadro de dialogo final permite al analista agregar corridas adicionales para el diseño y especificar el orden en el cual las corridas serán desarrolladas.

- **Model Type** (Tipo de Modelo): La forma del modelo más complicado que debe ajustarse. Suficientes corridas deben agregarse para estimar el tipo de modelo. Cuando se analizan los resultados, el modelo por supuesto se simplifica.

- **Number of Replicant Points** (Número de Puntos Replicas): Si no es cero, el número de puntos especificados podrá seleccionarse al azar sobre los puntos del diseño original y replicarse. Esto puede usarse para proporcionar grados de libertad adicionales para estimar el error experimental. Alternativamente, el analista puede cambiar manualmente las corridas a replicar más que permitir que el programa lo haga, también muy simple, puede hacerlo agregando corridas al presionar el botón en la base de datos después de la creación del diseño.
- **Augment Design** (Aumentar Diseño): Si es seleccionado, corridas adicionales deberán colocarse en mezclas especiales, por ejemplo, combinaciones de los componentes que pueden ser utilizados para revisar la estimación del modelo seleccionado. Para los diseños simplex-lattice y simplex-centroide, estas corridas están localizadas a lo largo de una línea axial ejecutando las mezclas puras a través del centroide. Para los diseños de vértices extremos, el programa agrega puntos sobre la línea axial ejecutándose del centroide a los vértices, en los centros de los bordes conectando los vértices, en los centroide restringidos del plano, y en los centroide totales. Nota: para los diseños de vértices extremos con simples restricciones, puede ser necesario seleccionar esta opción en orden para tener suficientes corridas estimando todos los tipos de modelos simples.
- **Randomize** (Aleatorizar); Si esta opción se selecciona, las corridas podrán ordenarse en un orden al azar.

Dar clic en el botón *OK* para concluir con la construcción del diseño.

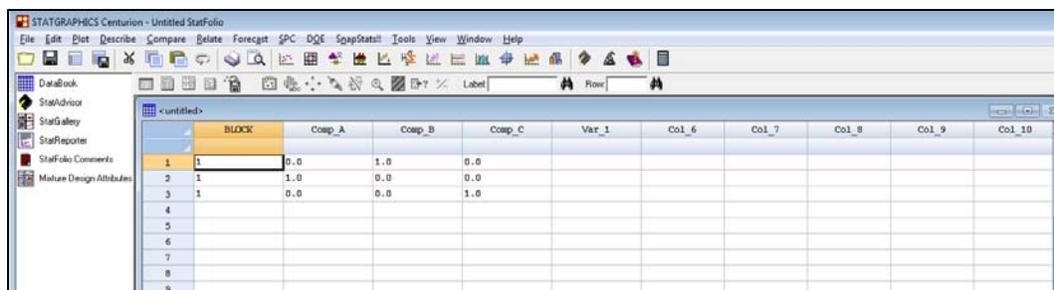
Enseguida, el software nos muestra un resumen del experimento creado en una ventana llamada; *Mixture Design Attributes*.



La secuencia de corridas mostradas sobre la hoja de trabajo (*Databook*) es exactamente la misma que en la base de datos de Statgraphics®. Después de desarrollar el experimento, es más fácil para el analista la transferencia de datos de la hoja de trabajo escrita al archivo de datos experimentales.

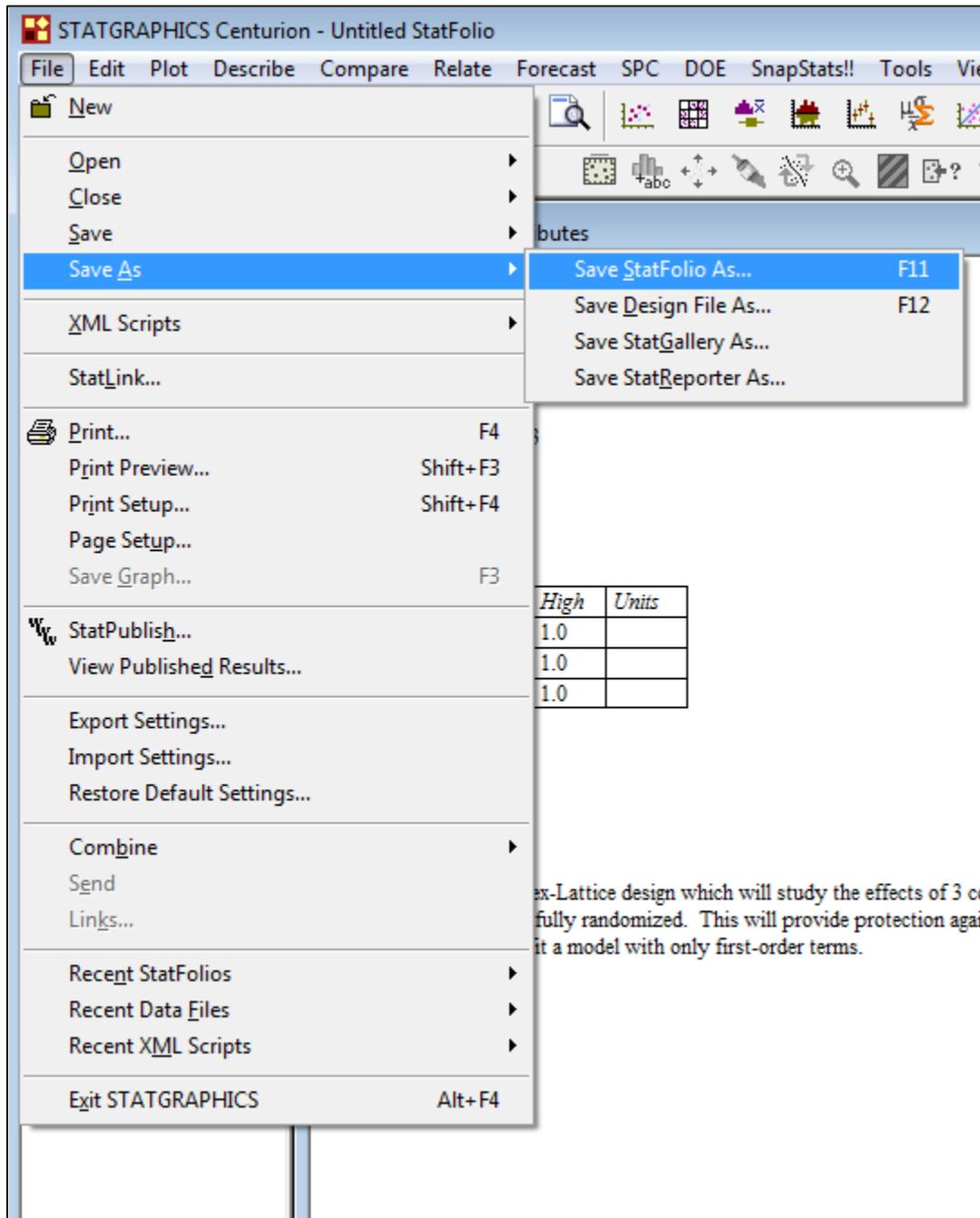
2.1. Agregando Corridas Adicionales

En algunos casos, es deseable agregar corridas adicionales al diseño. Esto puede realizarse manualmente agregando las corridas al final de la base de datos:



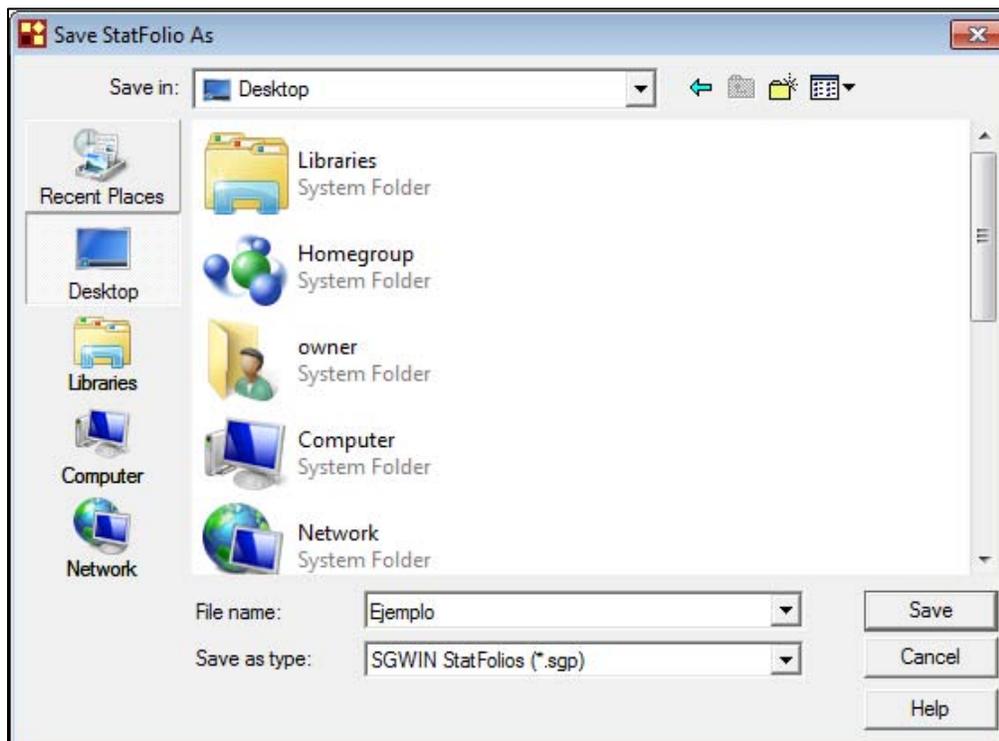
3. Guardando el archivo del Diseño.

Una vez que el diseño fue creado, deberá grabarse en una ruta especificada por el analista seleccionando *File* → *Save As* → *Save StatFolio As...*



Aparecerá el cuadro de diálogo número seis en el que el analista debe indicar la ruta en la que el diseño será almacenado.

Cuadro de Diálogo #6 – Guardar Diseño.

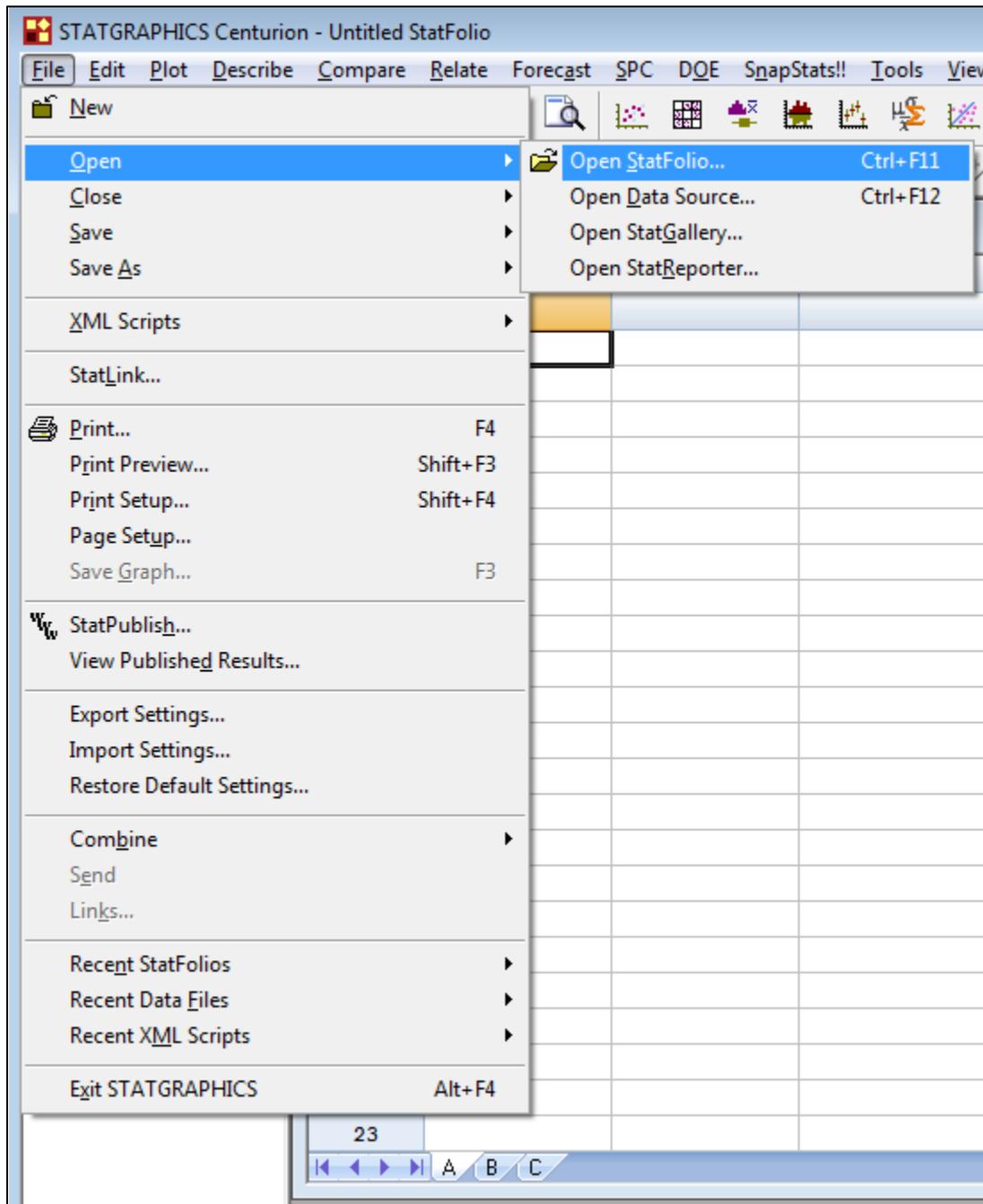


Dar clic en el botón *Save* para guardar el diseño. Los archivos de diseño son como los archivos de datos y tienen la extensión *.sgp*.

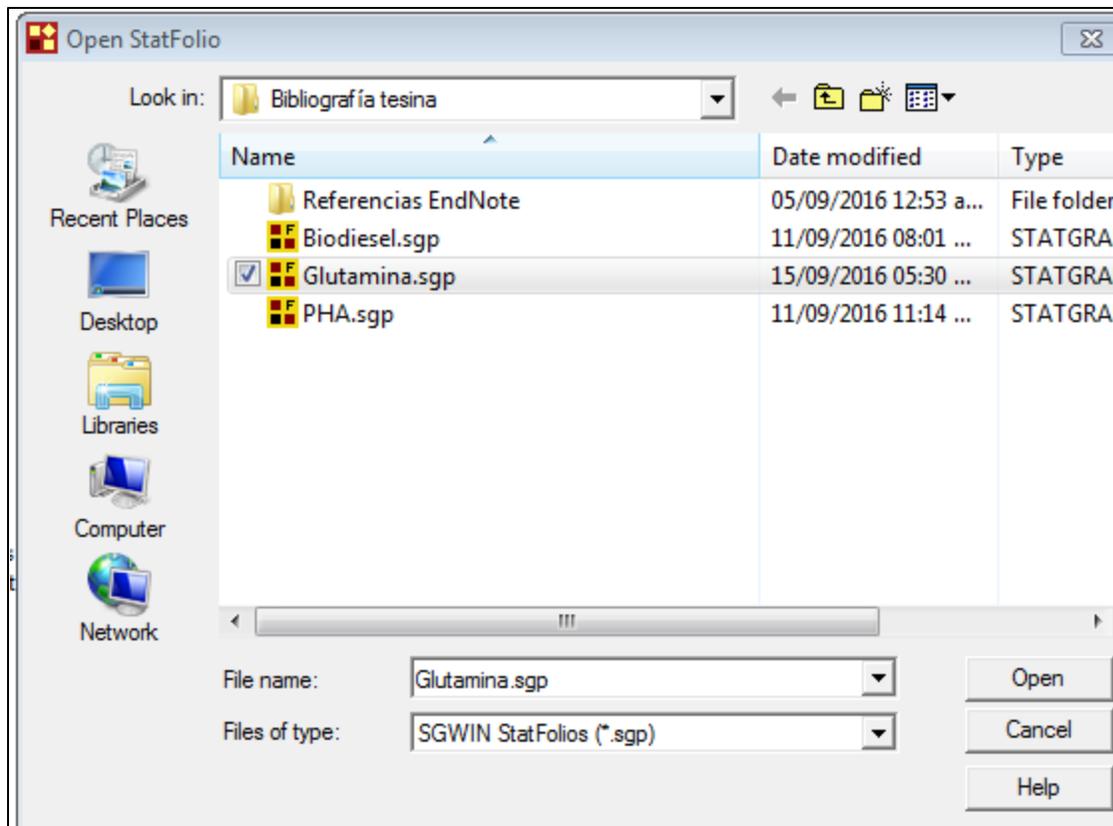
4. Analizando los Resultados.

Después que los archivos de diseño fueron creados y guardados, los experimentos pueden desarrollarse. En una fecha más adelante, una vez que los resultados fueron recolectados, el experimentador deberá regresar a Statgraphics® y re-abrir el archivo de diseño guardado.

Para re-abrir un archivo de experimento, seleccione *File* → *Open* → *Open Stat folio...*



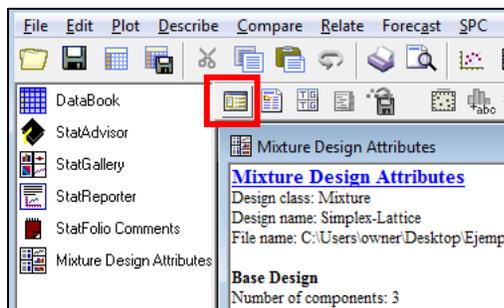
Seleccionar el archivo de diseño guardado y dar clic en el botón *Open*.



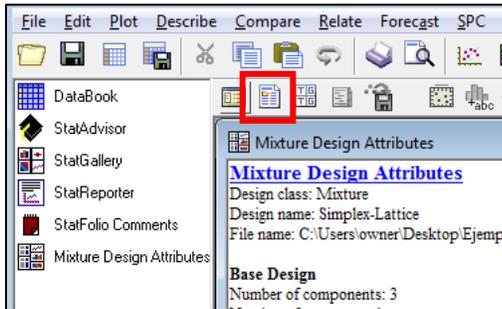
Los datos podrán recargarse dentro de la base de datos, y la ventana de *Mixture Design Attributes* aparecerá como cuando el diseño fue creado originalmente.

4.1. Para cambiar el diseño una vez que fue creado, despliegue la ventana *Mixture Design Attributes* y realice:

- Presionar el botón de *Input dialog* (Dialogo Entrada) siga la secuencia a través de los cuadros de dialogo y cambie algunas sentencias si es necesario.



- Seleccionar *Analysis Options* (Opciones del Análisis) para desplegar la última entrada del cuadro de dialogo (*Mixture Design Options*). Use el botón *Back* (Atrás) para regresar a través de la secuencia.



Notas Importantes:

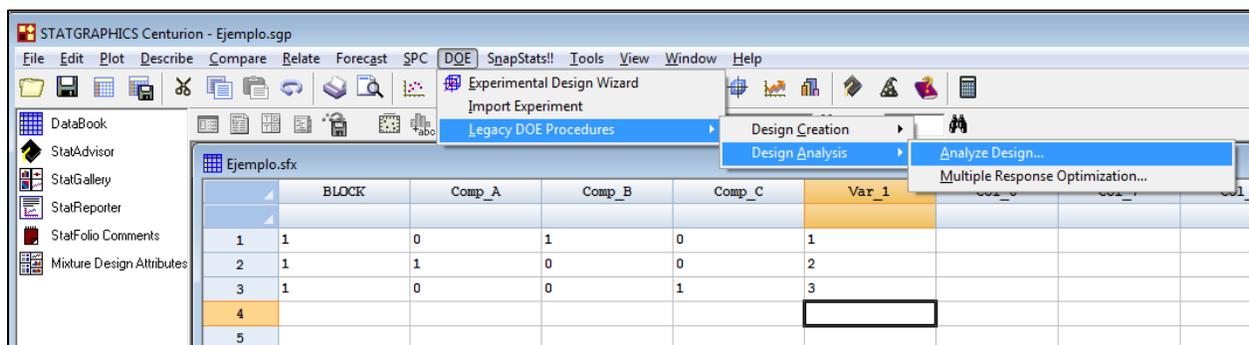
- Si más de una muestra fue tomada de cada conjunto de condiciones experimentales, ingrese el valor promedio de Y sobre las muestras dentro de la columna respuesta. También podemos incorporar el logaritmo de la desviación estándar como una respuesta si desea estudiar el efecto de los factores sobre la variabilidad dentro de muestras. Las estadísticas dentro de grupos pueden calcularse por el procedimiento Estadísticas por Filas, dependiendo de la estructura de los datos. No trate a las muestras como replicas a menos que se reajuste el proceso entre cada muestra.
- Si algún experimento no pudo ser desarrollado, deje el espacio correspondiente en blanco. El programa reconocerá el desbalance del diseño y lo manejará.
- Si alguna corrida experimental fue realizada en diferentes condiciones a las planeadas originalmente, cambie las entradas en la columna del componente correspondiente a los valores que fueron utilizados realmente.

- Si fueron desarrolladas corridas adicionales, estas se pueden agregar al final de la base de datos. Estas serán incluidas en la estimación.

Al abrir el experimento guardado se despliega la hoja de cálculo donde aparecen las columnas *Block*, junto con las columnas con el nombre para los componentes (*Comp_A*, *Comp_B*, *Comp_C*, etc.) de la mezcla con sus proporciones, y además la columna(s) correspondiente(s) a la(s) variable(s) de respuesta (*Var_1*). Ingresar los datos de la variable de respuesta.

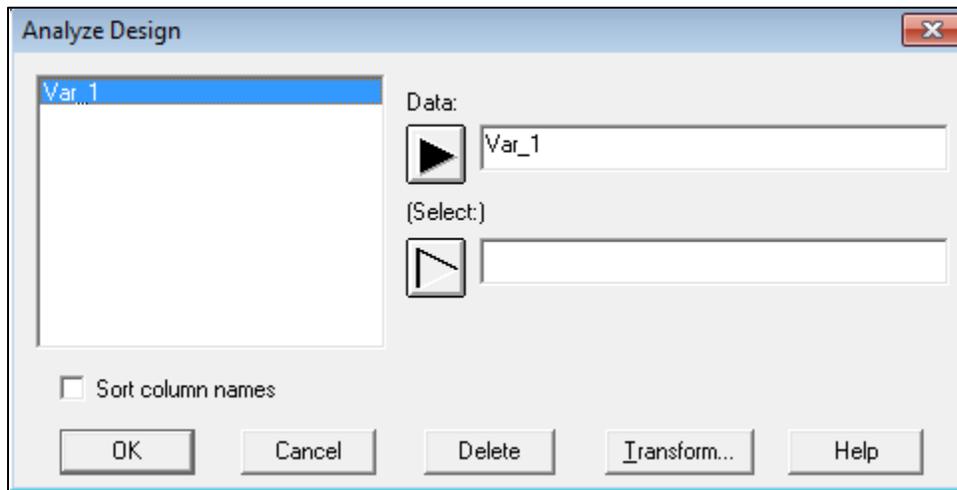
	BLOCK	Comp_A	Comp_B	Comp_C	Var_1
1	1	0	1	0	
2	1	1	0	0	
3	1	0	0	1	
4					
5					

Una vez que se ingresaron los datos de la variable de respuesta ir a la barra de menús:
DOE → *Legacy DOE Procedures* → *Design Analysis* → *Analyze Design...*



Se abrirá el siguiente cuadro de diálogo;

Cuadro de diálogo #7 – Análisis del Diseño.

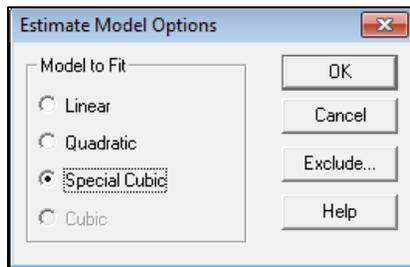


- **Data** (Datos): Columna que contiene los valores de la variable respuesta a ser analizada.
- **Select** (Selección): Selección de un subconjunto de los datos.

Si más de una respuesta fue medida, se puede realizar un análisis por separado para cada una.

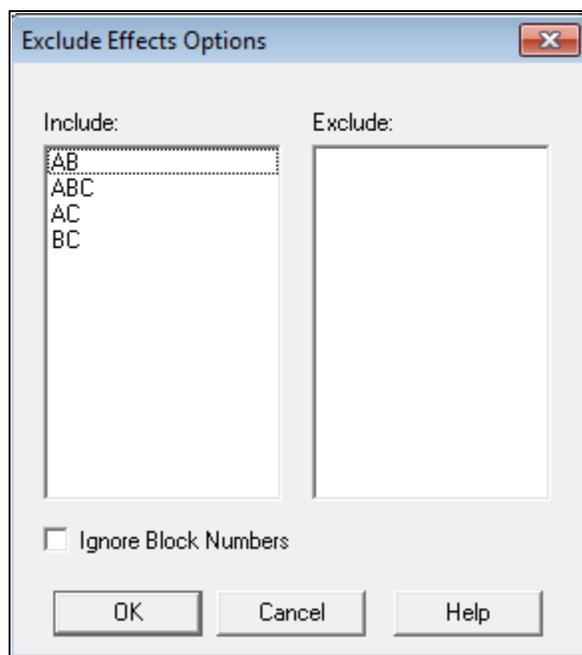
Seleccionar la variable de respuesta a analizar y dar clic en el botón *OK*. Aparecerá el siguiente cuadro de diálogo en donde se debe seleccionar el modelo seleccionado para el diseño de mezclas

Cuadro de diálogo #8 – Opciones del Modelo.



- **Modelo Fit** (Modelo a Estimar): Modelo para el cual los resultados se presentaran en las tablas y gráficas.
- Botón **Exclude** (Excluir): Presiónelo para excluir un término del modelo estimado. El siguiente cuadro de dialogo será desplegado:

Cuadro de diálogo #8 – Opciones de excluir efectos.

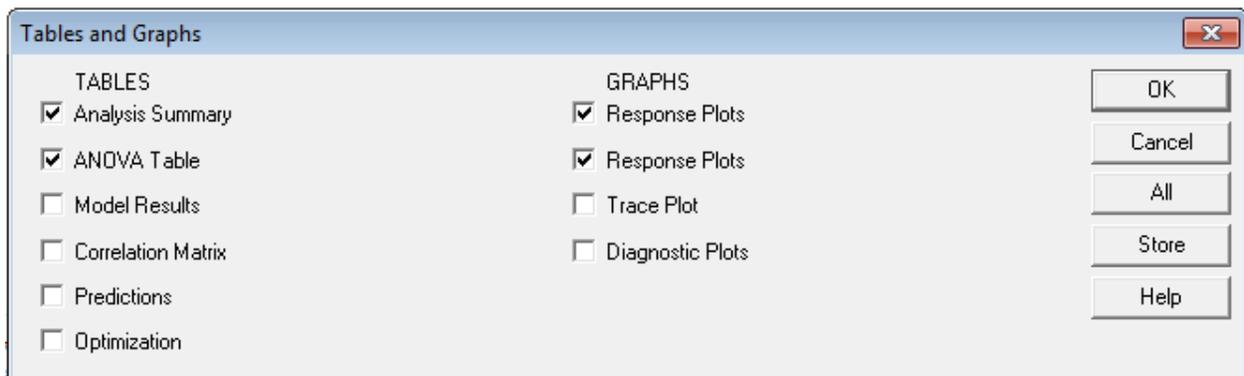


Los efectos pueden excluirse del modelo con doble-clic solo una vez sobre ellos. Dar doble-clic sobre un efecto en cualquiera de las dos columnas lo mueve a la otra columna.

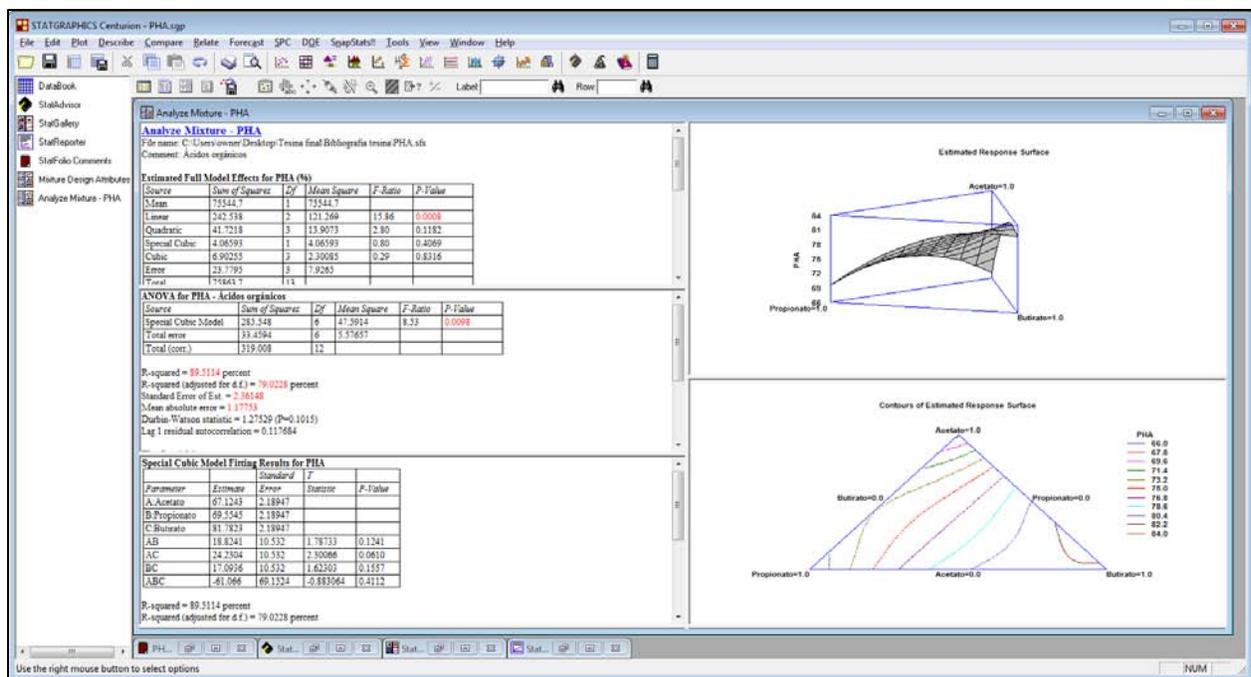
Dar clic en el botón *OK* para regresar al cuadro de diálogo número 7 – Opciones del modelo y dar clic en el botón *OK* para continuar con el análisis.

Aparecerá el siguiente cuadro de diálogo en donde se puede elegir cuáles serán las salidas del software Statgraphics®, el analista deberá seleccionar las salidas de interés.

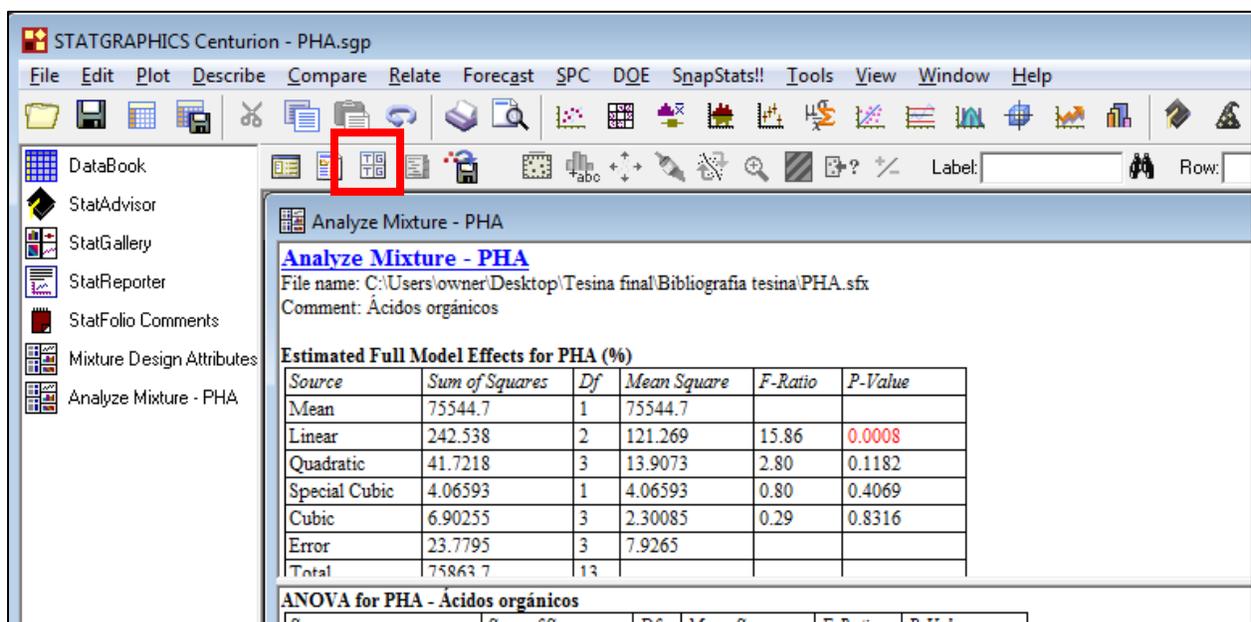
Cuadro de diálogo #9 – Opciones de salidas de Statgraphics® (Tablas y gráficas).

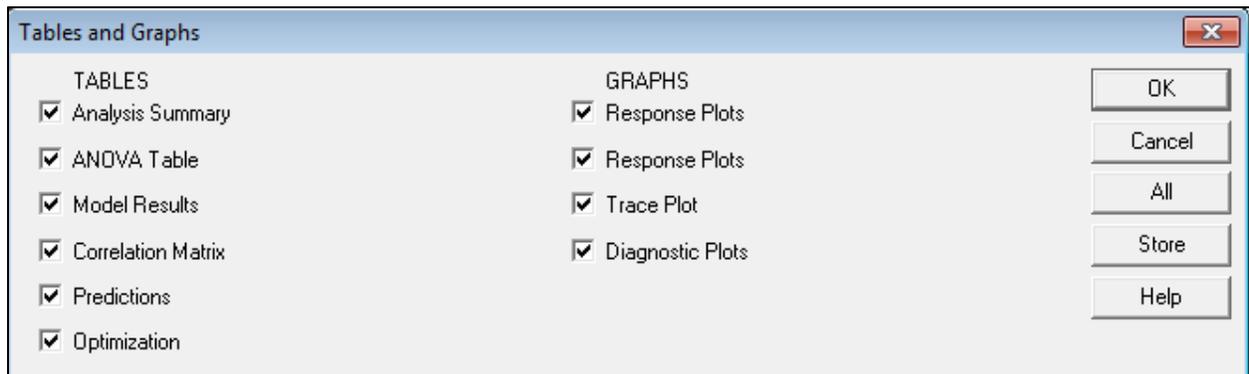


Enseguida se muestran los resultados para el experimento



Para obtener más resultados como la tabla de optimización, la matriz de correlación, etcétera, seleccionar la opción *Tables and graphs* para que se abra el cuadro de diálogo número 9 y se puedan seleccionar más tablas o gráficas de acuerdo a lo que el analista desee obtener.





5. Guardar los resultados del experimento

Para guardar los resultados en *Statfolio* con la extensión *.sgp*, ir a menú *File* → *Save as* → *Save StatFolio as...*

STATGRAPHICS Centurion - PHA.sgp

File Edit Plot Describe Compare Relate Forecast SPC DOE SnapStats!! Tools Vi

New

Open

Close

Save

Save As

XML Scripts

StatLink...

Print... F4

Print Preview... Shift+F3

Print Setup... Shift+F4

Page Setup...

Save Graph... F3

StatPublish...

View Published Results...

Export Settings...

Import Settings...

Restore Default Settings...

Combine

Send

Links...

Recent StatFolios

Recent Data Files

Recent XML Scripts

Exit STATGRAPHICS Alt+F4

Save StatFolio As... F11

Save Design File As... F12

Save StatGallery As...

Save StatReporter As...

Squares	Df	Mean Square	F-Ratio	P-Value
7	1	75544.7		
8	2	121.269	15.86	0.00
8	3	13.9073	2.80	0.11
3	1	4.06593	0.80	0.40
5	3	2.30085	0.29	0.83
5	3	7.9265		
7	12			

os orgánicos

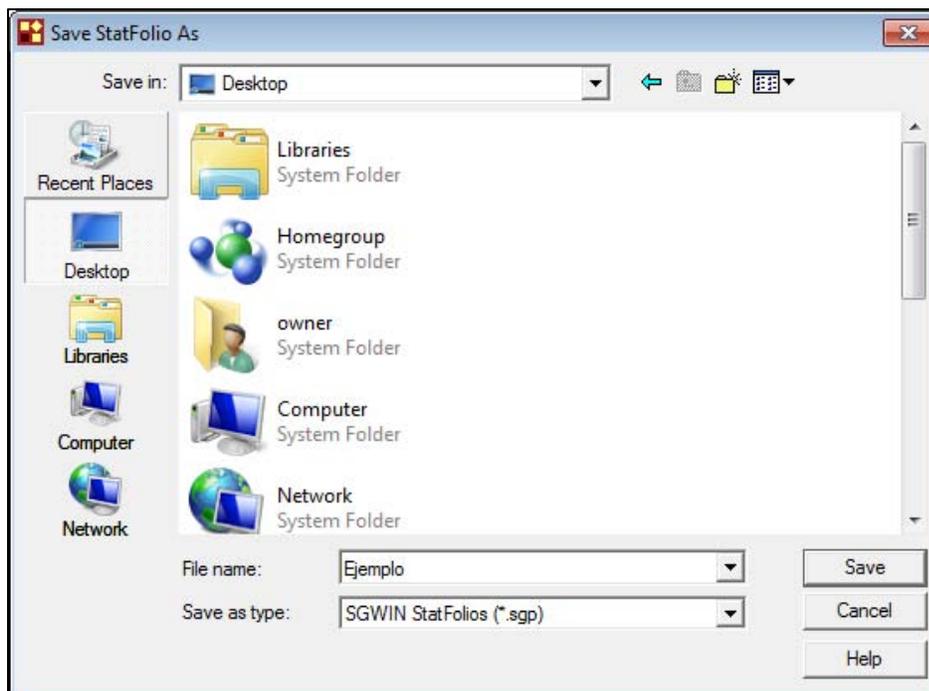
Sum of Squares	Df	Mean Square	F-Ratio
285.548	6	47.5914	8.53
33.4594	6	5.57657	
319.008	12		

cent
f.) = 79.0228 percent
36148
7753
= 1.27529 (P=0.1015)
tion = 0.117684

Setting Results for PHA

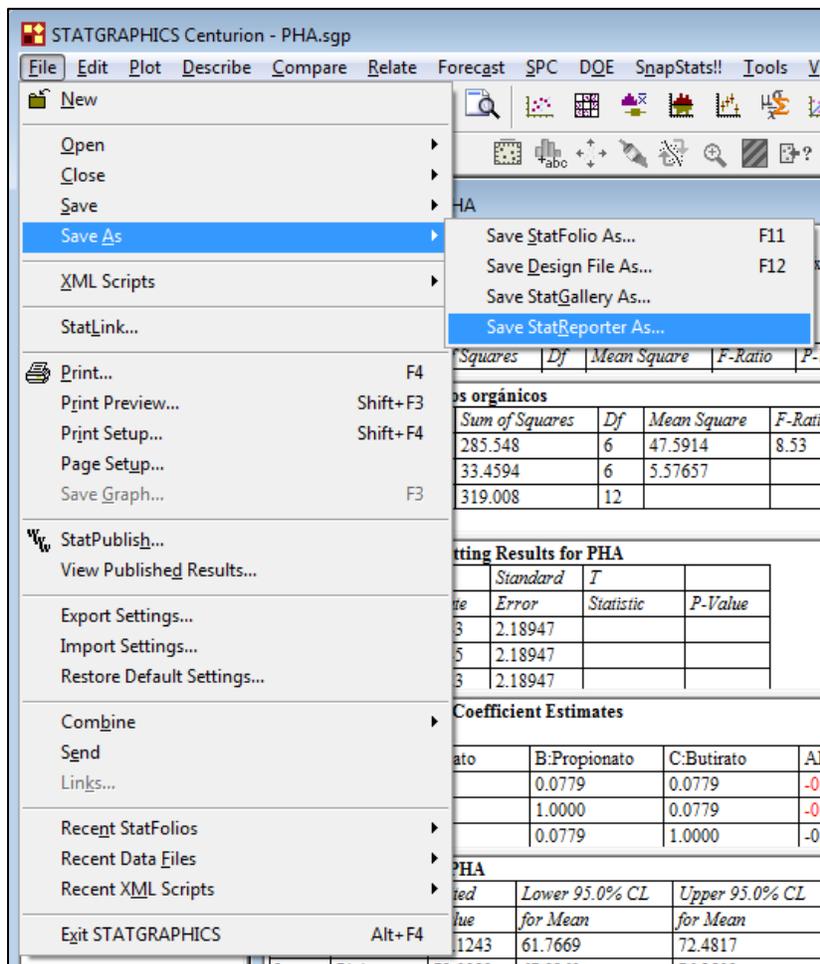
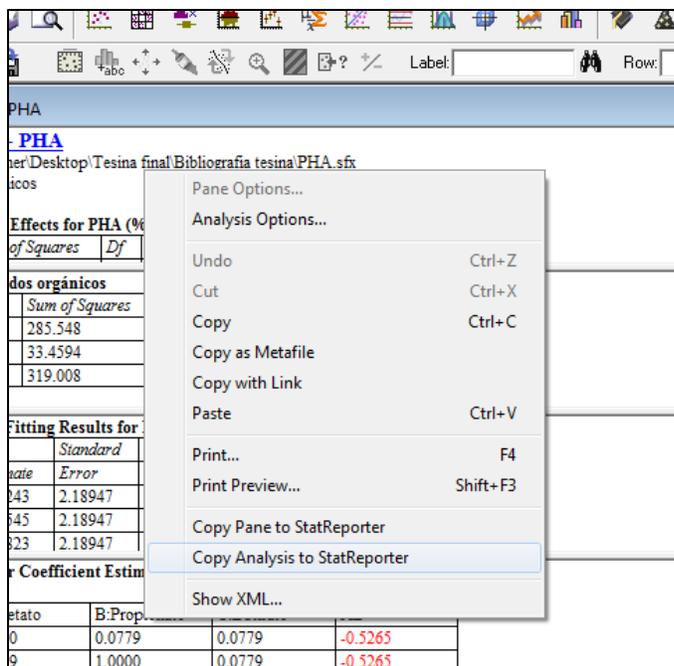
	Standard	T	
ne	Error	Statistic	P-Value
A:Acetato	67.1243	2.18947	

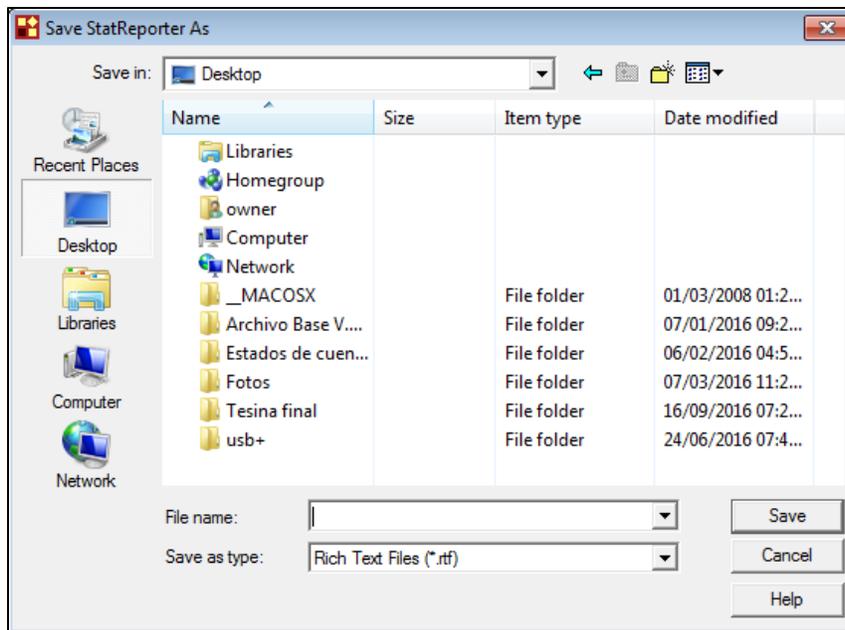
Cuadro de Diálogo #10 – Guardar Resultados del Diseño.



Dar clic en el botón *Save* para guardar los resultados del diseño. Los archivos de diseño son como los archivos de datos y tienen la extensión *.sgp*.

Los resultados se pueden salvar en un formato de texto enriquecido (que se puede leer en cualquier procesador de texto, como Word de Microsoft®) dando un clic derecho sobre los resultados en la ventana *Analyze Mixtures* elegir la opción *Copy to StatReporter*, ir al menú *File* → *Save* → *StatReporter* y nombrar el archivo, este se guardará con la extensión *.rtf*.





6. Análisis de resultados

La interpretación de los resultados dependerá del tipo de datos ingresados, del tipo de modelo elegido, la cantidad de factores, las variables de respuesta medidas y con el criterio del propio analista.