

## ***Diseño de Experimentos – Diseños de Cribado***

### **Resumen**

La sección de *Diseño Experimental* en STATGRAPHICS contiene un conjunto de procedimientos que soportan el diseño y análisis de muchos tipos de experimentos diferentes. Estos procedimientos habilitan el análisis para construir un conjunto de corridas experimentales que pueden calcular la cantidad de rendimiento máximo de información acerca de un proceso con el número mínimo de corridas. En contraste para experimentos con altos riesgos, el diseño de experimentos caracteriza la manipulación sistemática de un proceso para determinar los efectos atribuibles de diferentes factores.

En las etapas iniciales de una investigación, el analista tiene una faceta con una lista grande de factores que pueden afectar el proceso. Por ejemplo, en un proceso químico típico, fácilmente podemos detectar los factores que tienen un alto impacto sobre el rendimiento de un proceso, así como también la temperatura en la cual deberá correrse, la cantidad de catalizador adicional, la velocidad máxima, y otras más. Debido a que es difícil estudiar muchos factores a detalle simultáneamente, los diseños de investigación pueden emplearse para determinar rápidamente cuales factores tienen el mayor impacto sobre un proceso.

Este documento describe la construcción y análisis de diseños conducidos para identificar los factores más importantes. Posteriormente que los factores críticos son determinados, un diseño experimental más complicado se ejecuta de un gran conjunto de niveles para un factor, construidos para encontrar la configuración óptima para estos factores.

### **Ejemplo**

Como un ejemplo, un experimento de investigación típico puede construirse sobre 5 factores y 1 respuesta. El ejemplo, presenta una reacción química involucrada, es discutida en el Capítulo 12 del conocido libro de Box, Hunter y Hunter (1976). Los factores que pueden ser controlados son:

- X1: Tasa de alimentación
- X2: Cantidad de catalizador
- X3: Tasa de agitación
- X4: Temperatura
- X5: Concentración

Hay una sola variable respuesta:

- Y: Porcentaje de reacción

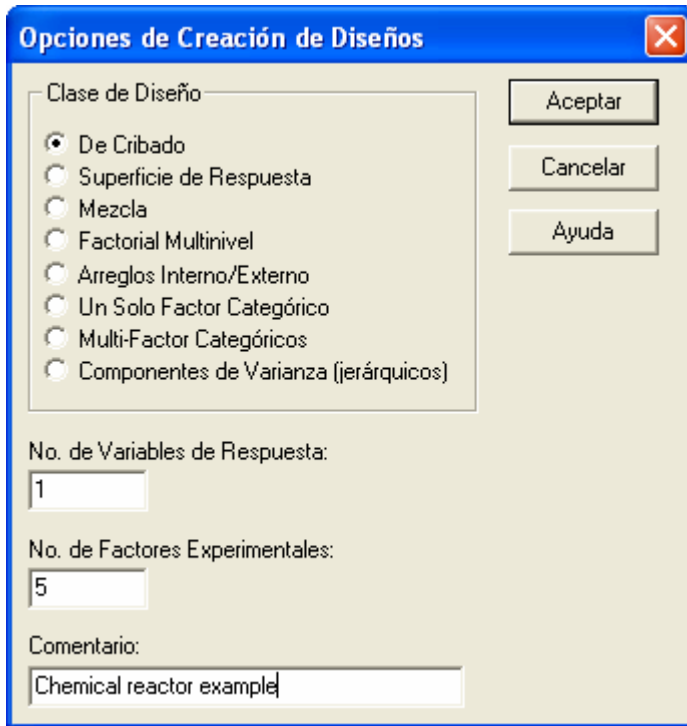
**Ejemplo StatFolio:** doe screening.sgp

## Creación del Diseño

Los experimentos de investigación son creados seleccionando *Crear Diseño* del menú de diseño de experimentos y al completar diferentes cajas de diálogos.

### Caja de Dialogo #1 – Tipo de Diseño

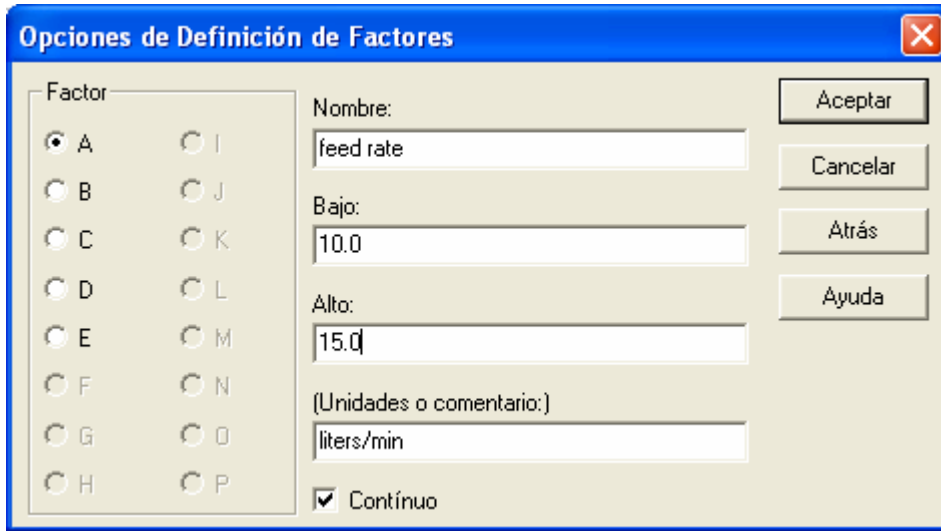
La primera caja de dialogo se despliega durante la creación del tipo de diseño especificado para ser construido:



- **Clase de Diseño:** Tipo de diseño a ser creado.
- **No. de Variables Respuestas:** El número de variables respuestas  $Y$  que deberán medirse durante cada corrida experimental. Este numero esta en un rango de 1 hasta 16.
- **No. de Factores Experimentales:** el número de factores experimentales  $X$  que pueden variar de corrida a corrida. Para diseños de investigación, el número de factores experimentales con un rango de 2 hasta 16.
- **Comentario:** Un comentario que aparecerá sobre las salidas de los procedimientos de análisis.

### Caja de Dialogo #2 – Factores Experimentales

La segunda caja de dialogo requiere información acerca de cada uno de los factores experimentales:



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

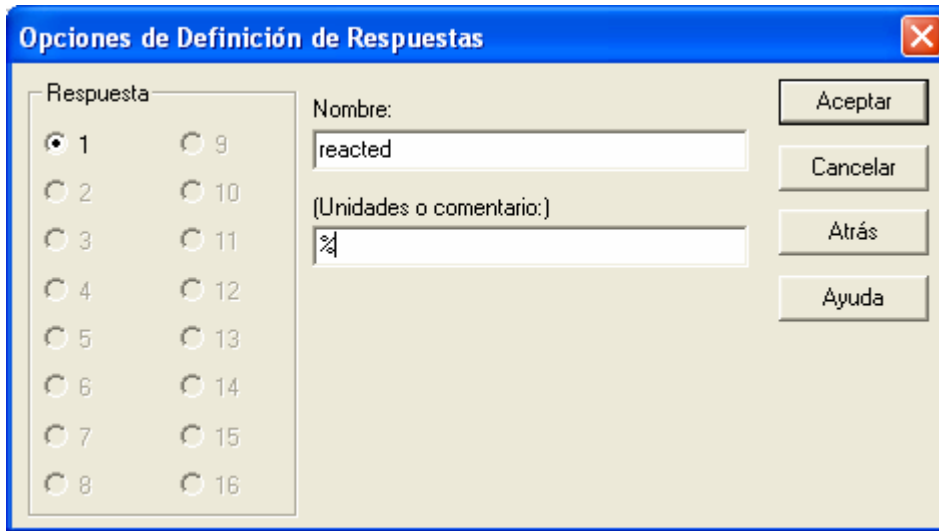
- **Nombre** – El nombre del factor es similar a un nombre de columna en una base de datos estándar de STATGRAPHICS, cada nombre debe tener hasta 32 caracteres y no contener cualquier carácter especial u operadores algebraicos.
- **Bajo** – El nivel bajo del factor. Para un diseño de investigación, ingresar un valor numérico o etiqueta. Si se ingresa una etiqueta no numérica, la caja etiquetada *Continua* debe desactivarse.
- **Alto** – El nivel alto del factor. Para un diseño de investigación, ingresar un valor numérico o etiqueta. Si la caja *Continua* esta activada, el nivel alto será mayor que el nivel bajo.
- **Continuo** – Si esta habilitada, el procedimiento localiza puntos al centro en un valor intermedio entre el nivel bajo y alto si lo requiere. Esta caja no puede activarse si etiquetas no numéricas son ingresadas en el nivel bajo y alto.
- **Unidades o Comentario** - Una etiqueta opcional o un comentario hasta 64 caracteres que se incluyen sobre la hoja de trabajo experimental.

En un diseño de investigación a dos-niveles, las corridas experimentales pueden desarrollarse en los niveles bajo y alto de cada factor, y algunas en un nivel central entre bajo y alto. Para el ejemplo actual, 5 factores cuantitativos se definieron a continuación:

Nombre	Bajo	Alto	Unidades	Continuo
<i>feed rate</i>	10	15	liters/min	Si
<i>catalyst</i>	1	2	%	Si
<i>agitation</i>	100	120	rpm	Si
<i>temperature</i>	140	180	degrees	Si
<i>concentration</i>	3	6	%	Si

Caja de Dialogo #3 – Variable Respuesta

La tercera caja de dialogo requiere información acerca de cada variable respuesta:

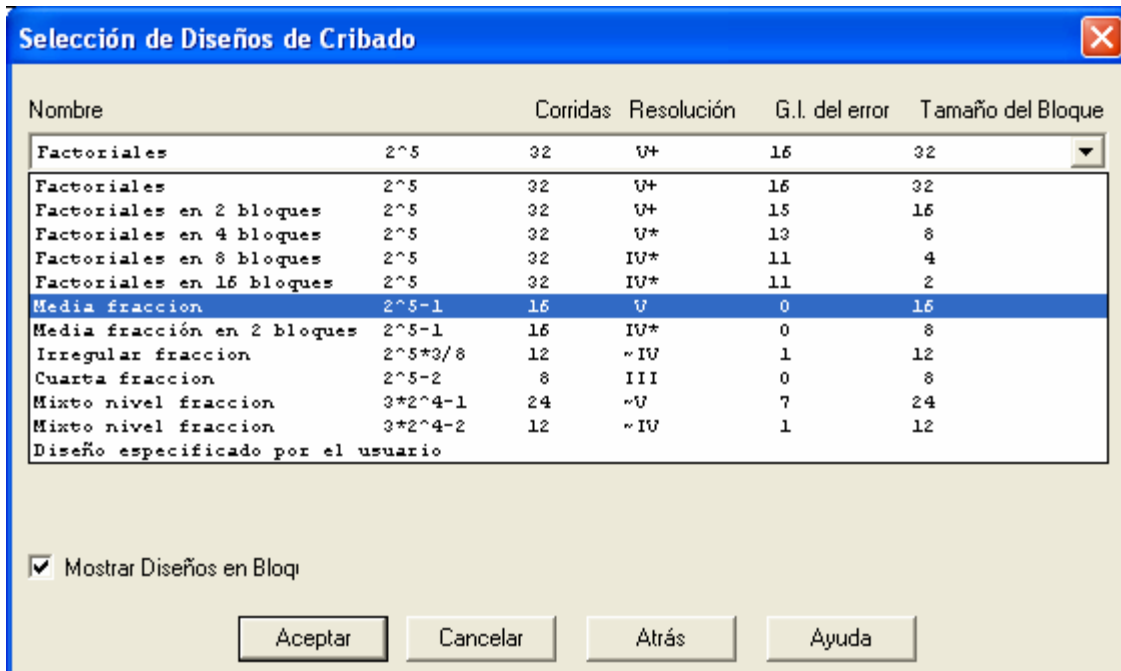


Clic sobre los números 1, 2, 3,..., solo una vez e ingrese la siguiente información para cada variable respuesta del experimento:

- **Nombre** – Un nombre para cada respuesta conteniendo hasta 32 caracteres.
- **Unidades o Comentario** – Una etiqueta opcional o comentario hasta 64 caracteres que se incluyen sobre la hoja de trabajo experimental.

Caja de Dialogo #4 – Catalogo de Diseño

La cuarta caja de dialogo muestra una lista de todos lo diseños de investigación del catálogo en STATGRAPHICS para un número seleccionado de factores experimentales:



- **Nombre** – El nombre del diseño, incluyendo una abreviación como  $2^5$  si es relevante. Para diseños de investigación, los siguientes tipos pueden aparecer en la lista, dependiendo sobre el número de factores experimentales:
  1. *Factorial* – Incluye corridas en todas las combinaciones de los niveles bajo y alto de cada factor, para un total  $2^k$  corridas. Muchos diseños con capacidad de estimar los efectos principales de todos los factores y todas las interacciones entre los factores.
  2. *Factorial en m bloques* – Incluye las mismas corridas que un diseño factorial completo. Sin embargo, las corridas son divididas dentro de bloques, las cuales son grupos de corridas que se mezclaron (en el mismo día, o por el mismo operador, o del mismo lote o tipo de material) para eliminar el efecto de uno o más factores de ruido. Como el número de bloques se incrementa, la habilidad para estimar ciertos factores se pierde. La tabla de *Estructura Alias* despliega después de haber creado el diseño cuales interacciones están confundidas con efectos de bloques.
  3. *Media fracción (o cuarta, octava, ...)* – Un subconjunto de las corridas de un diseño factorial completo, cualquiera de las dos mitades de un  $2^k$  completo, un-cuarto de las corridas, un-octavo de las corridas, o alguna otra fracción regular. El número de corridas en el diseño es igual a  $2^{k-p}$ , donde  $p = 1$  para una media fracción,  $p = 2$  para un cuarto de fracción,  $p = 3$  para un octavo de fracción, etc. Para muchos diseños, el campo *Resolución* indica información importante acerca del orden de las interacciones que pueden ser estimadas por tal diseño, como se describe abajo. Como factoriales con bloques, la tabla de *Estructura Alias* muestra el patrón de confusión del diseño.
  4. *Fracción Irregular* - Es un diseño factorial completo en el cual el número de corridas no está en una potencia de 2. Ciertas fracciones irregulares, no son completamente ortogonales, tienen patrones atractivos de confusión. Los diseños que se incluyen aquí fueron descritos por Haaland (1989).
  5. *Fracción de Nivel Mixto* - En contraste con todos los diseños de investigación, estos diseños asignan un factor (factor A) para correrse en 3 niveles. Esto asigna un efecto cuadrático a estimarse para ese factor, el cual puede ser cuantitativo. Para los demás factores, forman las corridas de un diseño factorial estándar. Los diseños que se incluyen aquí fueron descritos por Haaland (1989).
  6. *Plackett-Burman* – Diseños con dos niveles creados para investigar un gran número de factores con un número de corridas pequeñas, donde el número de corridas no está en una potencia de 2. Por ejemplo, un diseño está disponible para estudiar 11 factores en 12 corridas. Los efectos principales están confundidos con interacciones de 2 factores, también estos diseños pueden usarse cuando las interacciones no están presentes o se sabe que son pequeñas.
  7. *Plackett-Burman Cruzado* – Similar a los diseños Plackett-Burman, excepto que las interacciones de dos factores no están confundidas con el efecto principales. Sin embargo, las interacciones de dos factores están gravemente confundidas entre ellas mismas y no pueden resolverse.

8. *Especificar Diseño* - Asigna una base de datos experimental vacía para que sea construida por el analista al ingresar sus propias corridas. Esto asigna al análisis procedimientos que se ejecutan usando un diseño creado por uno mismo. El usuario debe ser cuidadoso, cuando crea este diseño, al ingresar los valores reales bajo y alto de cada factor sobre la caja de dialogo, será la manera en la cual los efectos se calculan durante los análisis, dependiendo de la configuración de los valores bajo y alto.
- **Corridas** – El número de corridas en el diseño base, antes de sumar replicas o puntos al centro adicionales.
  - **Resolución** – Una indicación sobre el patrón de confusión en el diseño. Los diseños se clasifican teniendo alguna de las siguientes resoluciones:

*Resolución III:* Un diseño en el cual se confunden las estimaciones de los efectos principales con interacciones de dos factores. Tales diseños pueden interpretarse con seguridad si solamente todas las interacciones de dos factores son pequeñas o no existentes.

*Resolución IV:* Un diseño que es capaz de obtener estimaciones limpiamente de todos los efectos principales. Sin embargo, algunas o todas las interacciones de dos factores son confundidas con otras interacciones de dos factores o efectos de bloques. La tabla de *Estructura Alias* descrita abajo indica como ocurre la confusión.

*Resolución V:* Un diseño que es capaz de obtener limpiamente todas las estimaciones de los efectos principales e interacciones de dos factores. Sin embargo las interacciones de alto orden se confunden con estos efectos. En muchos casos, esto no es un problema puesto que efectos de tercer o alto orden son generalmente asumidos como pequeños o no existentes. Los diseños con Resolución V típicamente son excelentes selecciones.

*Resolución V+:* El diseño con resolución mayor que 5, asigna las estimaciones de interacciones de tres factores o mayor orden si lo desea.

Para diseños por bloques, un asterisco se muestra para la siguiente resolución del diseño que indica el estado de resolución asumido para los factores de bloque que no interactúan con los factores experimentales, esta asunción es estándar cuando se desarrollan los análisis.

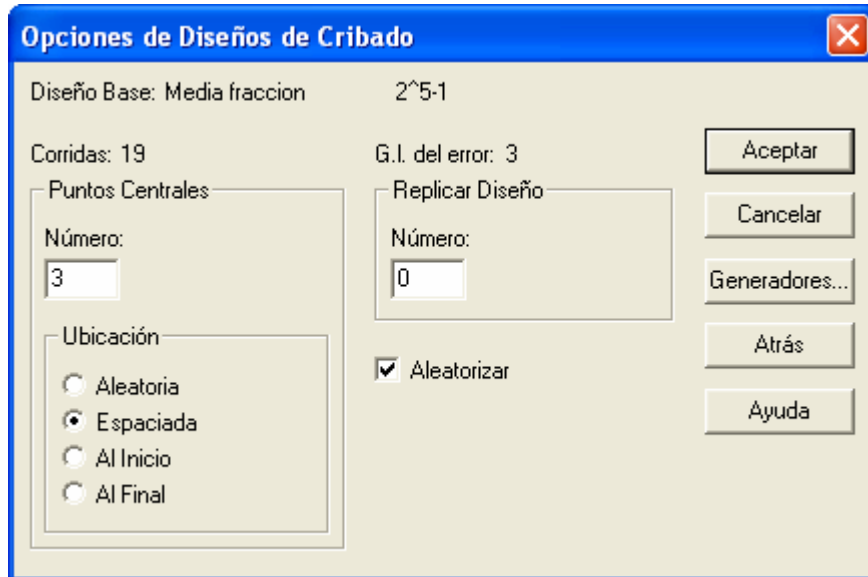
- **g.l. del Error** – El número de grados de libertad para que el error experimental pueda estimarse después de calcular todos los efectos principales, interacciones de segundo orden, y efectos cuadráticos (si son relevantes). Esto es previo para cualquier replicación o puntos al centro adicionales. En general, al menos 3 grados de libertad deben estar disponibles para que la prueba estadística a ser desarrollada durante un análisis tenga una razonable potencia estadística.
- **Tamaño de Bloque** - Para un diseño que tiene corridas en más de un bloque, el número de corridas dentro de los bloques.

Para el ejemplo actual, 16 corridas de una media fracción fueron. Este diseño tiene resolución V, lo cual significa que es capaz de estimar todos los efectos principales e interacciones dobles. Sin

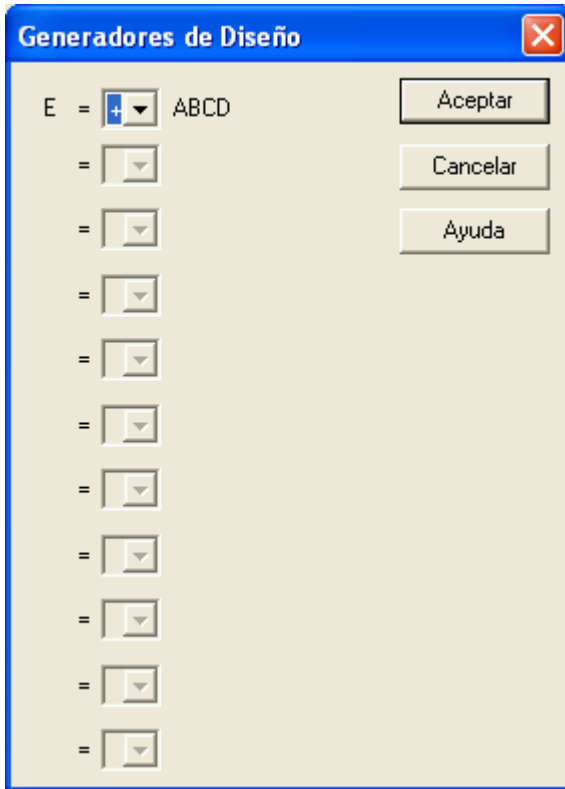
embargo, hay o grados de libertad restantes para estimar el error experimental. En orden para hacer una prueba estadística formal, deberán adicionarse corridas a la base del diseño.

### Caja de Dialogo #5 –Opciones del Diseño

La caja de dialogo final permite al analista incluir corridas adicionales al diseño y especificar el orden en el cual las corridas serán desarrolladas:



- **Puntos al Centro (Numero)** – El número de puntos al centro que serán adicionados en el diseño base, los cuales son corridas experimentales adicionales localizadas en un punto medio entre el nivel bajo y alto de todos los factores. Cada punto central adicional agrega un grado de libertad para estimar el error experimental. Si el diseño resuelve un solo factor categórico, los puntos al centro pueden localizarse en el nivel medio de los factores cuantitativos y se dividen igualmente entre los dos niveles del factor categórico.
- **Puntos al Centro (Localización)** – La posición de los puntos al centro con respecto a las corridas en el diseño base. Estas pueden ponerse *al azar* a través de las otras corridas experimentales, *espaciados* regularmente a través de las corridas, o localizarse al *inicio* o *final* del experimento. Las primeras dos opciones son generalmente las preferidas.
- **Replica del Diseño** - Si otro número distinto de 0 es ingresado, el diseño entero deberá repetirse el número de veces indicado.
- **Aleatorización** - Activar esta caja para aleatorizar el orden de las corridas en el experimento. Aleatorizar generalmente es una buena idea, puesto que reduce el efecto de variables ocultas como tendencias sobre el tiempo. Sin embargo, cuando replicamos los ejemplos en esta documentación, **no se aleatorizan** las corridas en el diseño.
- **Botón de Generadores** – Este botón despliega una caja de dialogo que permite al analista experimentado cambiar los generadores del diseño para diseños factoriales fraccionados:



En orden para generar un diseño factorial fraccionado  $2^{k-p}$ , primero se genera un diseño factorial completo para  $k - p$  factores. Las columnas adicionales para  $p$  factores son creadas por la multiplicación de varias combinaciones de columnas del diseño inicial. En el ejemplo actual, la columna para el factor E es creada por la multiplicación de la columna A por la columna B por la columna C por la columna D. Esto se abrevia como:

$$E = ABCD \tag{1}$$

Alternativamente, la columna E puede crearse por la multiplicación de las mismas cuatro columnas cambiando el signo, ej.

$$E = -ABCD \tag{2}$$

esto da como resultado un conjunto de 16 corridas diferentes. Para más detalles de estos procedimientos, ver Box, Hunter y Hunter (1976).

Note que tenemos 3 puntos al centro adicionados en el diseño base del ejemplo actual, resultando un total de 19 corridas y proporcionando 3 grados de libertad al error para estimar el error experimental. Puesto que la opción *Espaciados* fue seleccionada, los tres puntos al centro fueron localizados en el inicio, mitad y final del experimento.

### Atributos del Diseño

Una vez que la caja de dialogo final fue completada, una ventana de análisis será presentada mostrando los atributos del diseño seleccionado. El panel de *Atributos del Diseño* se muestra para el ejemplo actual:



**Screening Design Attributes**

Clase de diseño: De Cribado  
 Nombre del Diseño: Media fraccion 2^5-1  
 Nombre del archivo: C:\DocData\chemical reaction.sfx  
 Comentario: Chemical reactor example

**Diseño Base**

Número de factores experimentales: 5  
 Número de bloques: 1  
 Número de respuestas: 1  
 Número de corridas: 19, incluyendo 3 puntos centrales por bloque  
 Grados de libertad para el error: 3  
 Aleatorizar: No

<i>Factores</i>	<i>Bajo</i>	<i>Alto</i>	<i>Unidades</i>	<i>Continuo</i>
feed rate	10.0	15.0	liters/min	Sí
catalyst	1.0	2.0	%	Sí
agitation	100.0	120.0	rpm	Sí
temperature	140.0	180.0	degrees	Sí
concentration	3.0	6.0	%	Sí

<i>Respuestas</i>	<i>Unidades</i>
reacted	%

El resumen incluye:

- **Resumen del Diseño** – El tipo del diseño creado, los comentarios especificados, el nombre del archivo con el cual el diseño fue almacenado. Los experimentos no se graban hasta que la opción *Archivo – Guardar Diseño* es seleccionado del menú principal de STATGRAPHICS. Los archivos de experimentos son guardados como archivo de datos con la extensión .SFX. Los archivos se pueden abrir de una manera similar a los archivos de datos estándar. Sin embargo, los nombres y tipos de columna no pueden modificarse excepto regresando a la ventana de análisis de *Atributos del Diseño* actual y accediendo a *Opciones del Análisis*.
- **Diseño Base** – El número de factores experimentales, el número de variables respuestas, el número total de corridas incluyendo puntos al centro adicionales, el número de grados de libertad disponibles para la estimación del error experimental, y cuando el orden de las corridas fue aleatorizado.
- **Factores** – Información acerca de cada uno de los k factores experimentales.
- **Respuestas** – Información acerca de cada una de las variables respuestas.

## Hoja de Trabajo

El botón de *Opciones Tabulares* sobre la barra de herramientas del análisis en la ventana de *Atributos del Diseño* contiene una opción para desplegar una *Hoja de Trabajo*, tiene una lista de las corridas experimentales en el orden en el cual fueron desarrolladas:

Hoja de trabajo para C:\DocData\chemical reaction.sfx - Chemical reactor example						
<i>corrida</i>	<i>feed rate</i>	<i>catalyst</i>	<i>agitation</i>	<i>temperature</i>	<i>concentration</i>	<i>reacted</i>
	(liters/min)	(%)	(rpm)	(degrees)	(%)	(%)
1	12.5	1.5	110.0	160.0	4.5	65.0
2	10.0	1.0	100.0	140.0	6.0	56.0
3	15.0	1.0	100.0	140.0	3.0	53.0
4	10.0	2.0	100.0	140.0	3.0	63.0
5	15.0	2.0	100.0	140.0	6.0	65.0
6	10.0	1.0	120.0	140.0	3.0	53.0
7	15.0	1.0	120.0	140.0	6.0	55.0
8	10.0	2.0	120.0	140.0	6.0	67.0
9	15.0	2.0	120.0	140.0	3.0	61.0
10	12.5	1.5	110.0	160.0	4.5	67.0
11	10.0	1.0	100.0	180.0	3.0	69.0
12	15.0	1.0	100.0	180.0	6.0	45.0
13	10.0	2.0	100.0	180.0	6.0	78.0
14	15.0	2.0	100.0	180.0	3.0	93.0
15	10.0	1.0	120.0	180.0	6.0	49.0
16	15.0	1.0	120.0	180.0	3.0	60.0
17	10.0	2.0	120.0	180.0	3.0	95.0
18	15.0	2.0	120.0	180.0	6.0	82.0
19	12.5	1.5	110.0	160.0	4.5	63.0

La hoja de trabajo esta diseñada para ser enviada a una impresora y contiene espacios en blanco para capturar cada respuesta.

### *Opciones del Panel*

**Opciones de Hoja de Trabajo**

Espaciado de Línea

Espaciado Sencillo

Espaciado Doble

Espaciado Triple

Aceptar

Cancelar

Ayuda

Fila Inicial: 1 Fila Final: 256

Respuestas a Incluir:

reacted

- **Espaciado de Línea** – El espaciado de las corridas cuando se envían a la impresora. Doble o triple espaciado permiten más espacio para escribir los resultados a mano.
- **Respuestas a Incluir** – Las respuestas que se incluyen sobre la hoja de trabajo a imprimir. El analista puede pedir imprimir por separada la hoja de trabajo para cada respuesta si el número de factores experimentales es grande.

La secuencia de corridas mostradas sobre la hoja de trabajo 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.

## Estructura Alias

La tabla de *Estructura Alias* despliega una lista de todas las estimaciones que serán obtenidas cuando se accedan a los procedimientos de análisis, suponiendo que el diseño fue corrido completo:

Estructura de Alias	
Contraste	Estimados
1	A
2	B
3	C
4	D
5	E
6	AB
7	AC
8	AD
9	AE
10	BC
11	BD
12	BE
13	CD
14	CE
15	DE

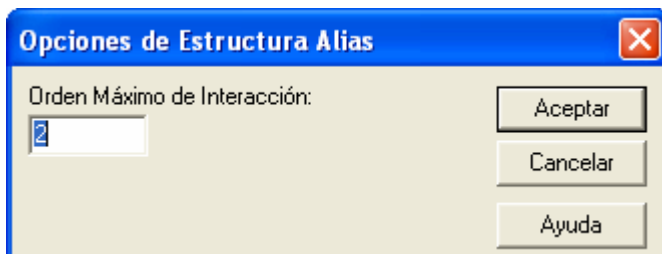
En la tabla, las letras son usadas para representar varios:

- Una simple letra como una A representa el efecto principal del factor A.
- Un par de letras como AB representan la interacción de dos factores entre el factor A y el factor B.
- Un par de letras como AA representa el efecto cuadrático del factor A.
- Un conjunto de 3 letras como ABC representa una interacción de tres factores, etcétera.

Por defecto, se asumen los efectos que serán estimados hasta de segundo orden. Para el ejemplo actual, cada uno de los efectos principales e interacciones de dos factores están listados sobre una línea separada, indicando que pueden estimarse separadamente (no existe confusión).

### Opciones del Panel

La caja de dialogo de *Opciones del Panel* permite al analista determinar el alcance para el máximo orden de confusión de las interacciones:



- **Máximo Orden de Interacción:** La interacción más grande para desplegar en la tabla.

Solicitando un análisis de interacciones con un orden de 5 los resultados se muestran en la siguiente tabla:

<b>Estructura de Alias</b>	
<i>Contraste</i>	<i>Estimados</i>
1	A+BCDE
2	B+ACDE
3	C+ABDE
4	D+ABCE
5	E+ABCD
6	AB+CDE
7	AC+BDE
8	AD+BCE
9	AE+BCD
10	BC+ADE
11	BD+ACE
12	BE+ACD
13	CD+ABE
14	CE+ABD
15	DE+ABC
16	ABCDE

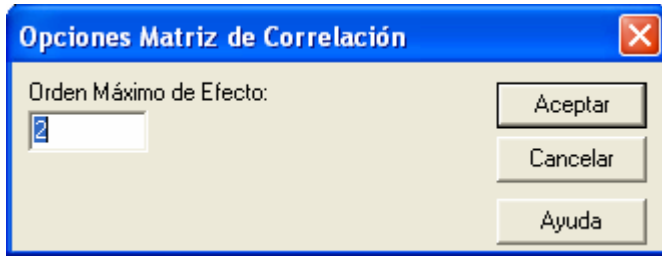
Una entrada como A+BCDE indican que el efecto principal del factor A esta confundido con una interacción de 4 factores con los factores B, C, D, y E. Típicamente, la confusión de 3 factores con interacciones de alto orden son consideradas como aceptables.

### Matriz de Correlación

Un segundo método para examinar el patrón de confusión del diseño seleccionado es desplegar la Matriz de Correlación. Una porción de la salida se muestra abajo:

<b>Matriz de Correlación</b>													
	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD
A	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
B	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
D	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
E	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AB	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AC	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
BC	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
BD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
BE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
CD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
CE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

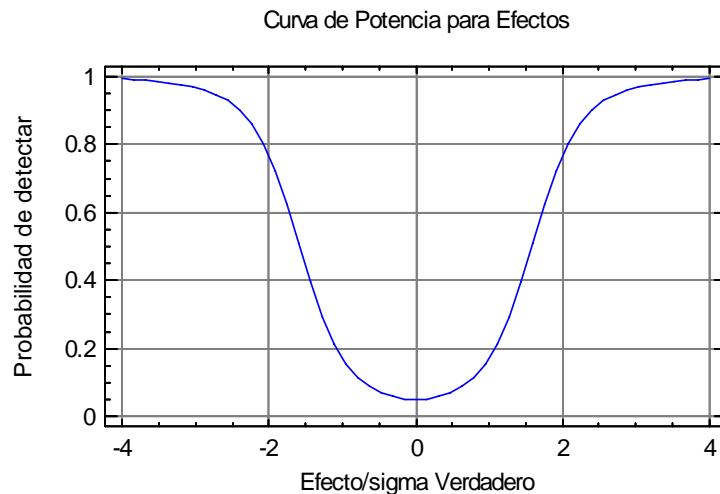
Esta matriz de correlación despliega la correlación lineal entre las columnas de la matriz X que puede ser usada para estimar un modelo estadístico para los resultados del experimento. Los coeficientes de correlación están entre -1 y 1, con 0 indica no correlación. La diagonal de la matriz muestra una indicación que el diseño es perfectamente ortogonal. En otros casos, las columnas fuera de la diagonal pueden indicar correlaciones no cero entre un par de efectos, lo cual nos indica que esos efectos están confundidos.

*Opciones del Panel*

- **Máximo Orden del Efecto:** El máximo orden de la interacción o el efecto para desplegar en la tabla.

**Curva de Potencia**

Puesto que todos los experimentos están sujetos a cierta cantidad de ruido o error experimental, las estimaciones de los efectos de los factores experimentales también están sujetas al error. Cuando probamos la significancia estadística de los efectos principales y las interacciones, los efectos que son demasiados pequeños es probable que se pierdan con el ruido. La *Curva de Potencia* puede ser usada para determinar cual es la magnitud de los efectos que son probables detectar por el diseño actual:



El eje vertical equis, etiquetado “Probabilidad de Detección”, muestra la probabilidad de un efecto de cierta magnitud pueda detectarse por una prueba estadística la cuál deberá ejecutarse después que los datos son recolectados. El eje horizontal muestra el cociente  $\Delta/\sigma$ , donde

$\Delta$  = verdadero efecto en unidades de la variable respuesta

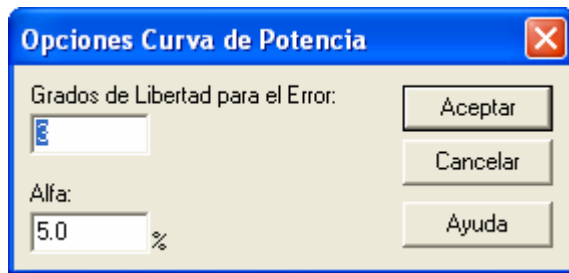
$\sigma$  = desviación estándar del error experimental

El cociente  $\Delta/\sigma$  es un tipo de cociente señal-a-ruido. En este caso, la curva indica una probabilidad relativamente alta de detectar efectos que son 5 veces o más que la sigma.

*Opciones del Panel*

La curva de poder es sensible a diversos factores, incluyendo el número de corridas experimentales que fueron desarrolladas, la configuración del factor, y los grados de libertad

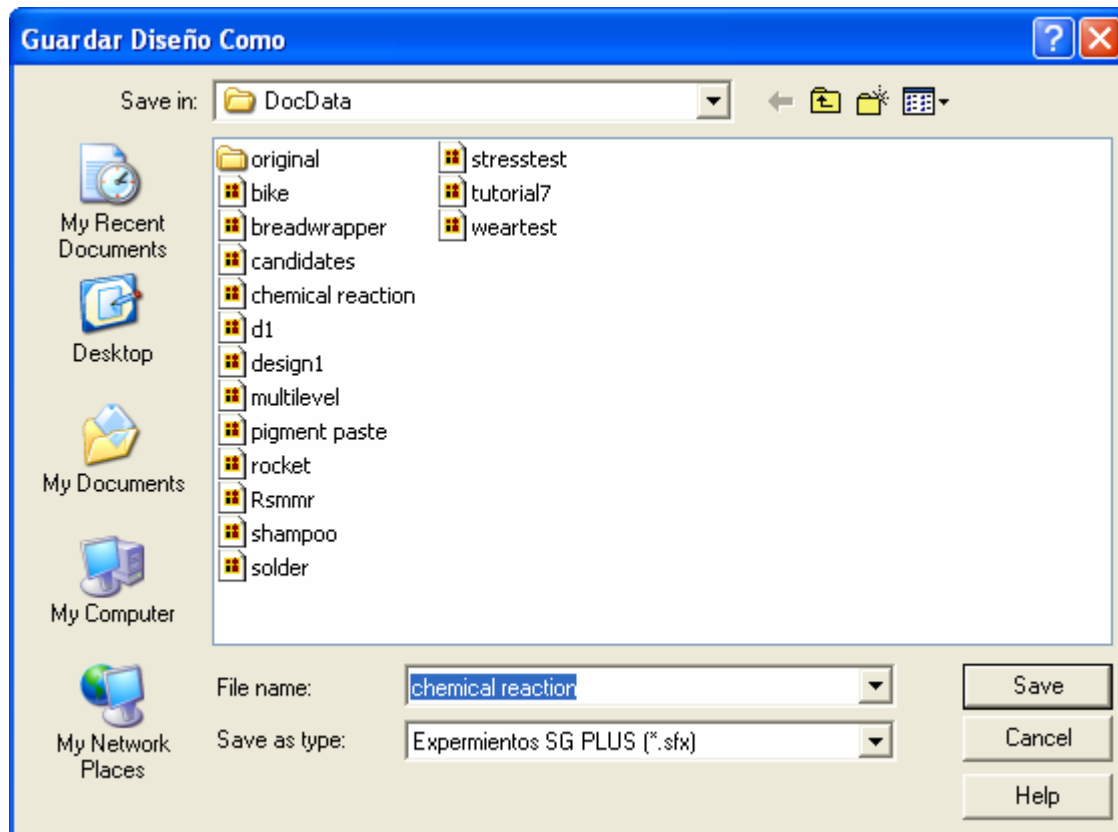
disponibles para estimar el error experimental. La caja de dialogo de las Opciones del Panel permite al usuario examinar el último efecto:



- **Grados de Libertad del Error** – Los grados de libertad pueden ser utilizados para estimar el error experimental. Los grados de libertad del error pueden cambiarse durante la estimación, si ciertos efectos pueden eliminarse o adicionarse al modelo. Por defecto se asume que todos los efectos principales y las interacciones de dos factores pueden estimarse.
- **Alfa** – El riesgo alfa de la prueba estadística que fue desarrollada. Esto especifica el nivel de significancia requerido para determinar que un efecto es *estadísticamente significativo*. Los valores comunes son 1%, 5%, o 10%.

## Grabando el Archivo del Diseño

Una vez que el experimento fue creado, deberá grabarse en un disco seleccionando *Grabar Archivo de Diseño* del menú *Archivo*.



Los archivos de diseño son como los archivos de datos y tienen la extensión *.sfx*. Ellos incluyen los datos mezclándolos con otra información que fue ingresada sobre la entrada de la caja de dialogo.

Para re-abrir un archivo de experimento, seleccione *Abrir Archivo de Datos* del menú *Archivo*. Los datos podrán recargarse dentro de la base de datos, y la ventana de *Atributos del Diseño* aparecerá como cuando el diseño fue creado originalmente.

Nota: Para cambiar el diseño una vez que fue creado, despliegue la ventana de *Atributos del Diseño* y realice:

1. Presionar el botón de dialogo *Entrada* siga la secuencia a través de las cajas de dialogo, cambie algunas sentencias si es necesario.
2. Seleccionar *Opciones del Análisis* para desplegar la ultima entrada de la caja de dialogo. Use el botón *Atrás* para regresar a través de la secuencia.

### 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 usando la selección *Abrir Datos Fuente* sobre el menú principal *Archivo*. Los resultados pueden capturarse en las columnas respuesta. Los resultados para el ejemplo se despliegan abajo:

<i>run</i>	<i>feed rate</i> (liters/min)	<i>catalyst</i> (%)	<i>agitation</i> (rpm)	<i>temperature</i> (degrees)	<i>concentration</i> (%)	<i>reacted</i> (%)
1	12.5	1.5	110.0	160.0	4.5	65.0
2	10.0	1.0	100.0	140.0	6.0	56.0
3	15.0	1.0	100.0	140.0	3.0	53.0
4	10.0	2.0	100.0	140.0	3.0	63.0
5	15.0	2.0	100.0	140.0	6.0	65.0
6	10.0	1.0	120.0	140.0	3.0	53.0
7	15.0	1.0	120.0	140.0	6.0	55.0
8	10.0	2.0	120.0	140.0	6.0	67.0
9	15.0	2.0	120.0	140.0	3.0	61.0
10	12.5	1.5	110.0	160.0	4.5	67.0
11	10.0	1.0	100.0	180.0	3.0	69.0
12	15.0	1.0	100.0	180.0	6.0	45.0
13	10.0	2.0	100.0	180.0	6.0	78.0
14	15.0	2.0	100.0	180.0	3.0	93.0
15	10.0	1.0	120.0	180.0	6.0	49.0
16	15.0	1.0	120.0	180.0	3.0	60.0
17	10.0	2.0	120.0	180.0	3.0	95.0
18	15.0	2.0	120.0	180.0	6.0	82.0
19	12.5	1.5	110.0	160.0	4.5	63.0

### Notas Importantes:

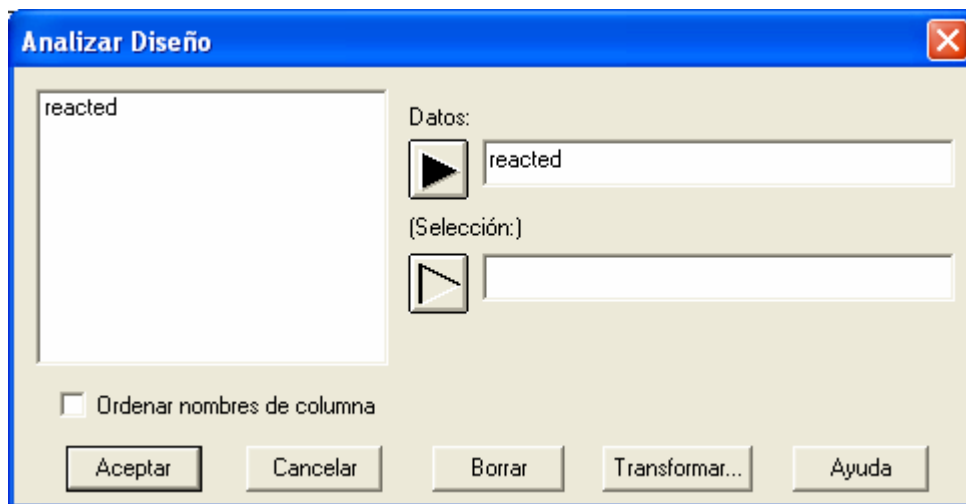
1. Si más de una muestra fue tomada de cada conjunto de condiciones experimentales, ingrese el valor promedio de Y sobre las muestras dentro de la columna respuesta. También podemos incorporar el logaritmo de la desviación estándar como una respuesta si desea estudiar el efecto de los factores sobre 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.

2. Si algún experimento no pudo ser desarrollado, deje el espacio correspondiente en blanco. El programa reconocerá el desbalance del diseño y lo manejará.
3. Si alguna corrida experimental fue realizada en diferentes condiciones a las planeadas originalmente, cambie las entradas en la columna del factor experimental correspondiente a los valores que fueron utilizados realmente.
4. Si fueron desarrolladas corridas adicionales, estas se pueden agregar al final de la base de datos. Estas serán incluidas en la estimación.

Para analizar los resultados, el menú de DDE contiene una selección llamada *Analizar Diseño*. Esto brinca a una caja de dialogo que despliega todas las columnas de respuestas en el archivo de datos:



- **Datos:** Columna que contiene los valores de la variable respuesta a ser analizada.
- **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. Nota: Si intenta utilizar el procedimiento *Optimización de Múltiples Respuestas*, abra una ventana por separado para cada una de las variables respuesta y seleccione el modelo estadístico deseado.

## Modelo Estadístico

El modelo estadístico en el que se basa el análisis de diseño de investigación expresan la variable respuesta  $Y$  como una función lineal de los factores experimentales, interacciones entre los factores, y términos del error. Existen dos tipos de modelos que generalmente se estiman, se ilustran abajo para 5 factores experimentales:

1. *Modelos de Primer Orden* – Contiene términos que representan solamente efectos principales.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (3)$$

2. *Modelos de Segundo Orden* – Contiene términos que representan efectos principales e interacciones de segundo orden.

$$\begin{aligned} Y = & \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \\ & + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{14} X_1 X_4 + \beta_{15} X_1 X_5 + \beta_{23} X_2 X_3 \\ & + \beta_{24} X_2 X_4 + \beta_{25} X_2 X_5 + \beta_{34} X_3 X_4 + \beta_{35} X_3 X_5 + \beta_{45} X_4 X_5 + \varepsilon \end{aligned} \quad (4)$$

El error experimental  $\varepsilon$  se asume que fue aleatorio y que proviene de una distribución normal con media 0 y desviación estándar igual a  $\sigma$ . En casos raros, las interacciones de 3 o más factores pueden incluirse como términos adicionales consistentes en el producto de más de dos factores. Ocasionalmente para los diseños de investigación que tienen 3 niveles en un factor, como el término  $\beta_{11} X_1^2$  también pueden incluirse en el modelo de segundo orden.

Para las variables cuantitativas, STATGRAPHICS representa el factor experimental  $X_j$  usando los valores originales dentro de la base de datos. Para los factores categóricos, las variables indicadoras son usadas de la forma

$$X_j = \begin{cases} -1 & \text{en el nivel } \textit{bajo} \text{ del factor } j \\ +1 & \text{en el nivel } \textit{alto} \text{ del factor } j \end{cases} \quad (5)$$

donde los niveles “bajo” y “alto” fueron definidos cuando el diseño fue construido.

Efectos

En orden para simplificar la interpretación de los diseños de investigación, es común re-expresar el modelo en términos de los “efectos”. El *efecto principal* del factor  $j$  es definido como el cambio en la variable respuesta  $Y$  cuando  $X_j$  cambia de su nivel bajo al alto, con el resto de los factores fijados como constantes, en un punto central, nivel bajo o alto. En un diseño factorial a dos niveles balanceado, la estimación del efecto del factor  $j$  es igual a la diferencia entre la respuesta promedio en el nivel alto menos la respuesta promedio del nivel bajo del factor:

$$\hat{\Delta}_j = \bar{y}_+ - \bar{y}_- \quad (6)$$

donde

$\bar{y}_+$  = respuesta promedio en el nivel alto del factor  $j$

$\bar{y}_-$  = respuesta promedio en el nivel bajo del factor  $j$

En un diseño des-balanceado, el efecto es una función de los coeficientes más complicado, pero la interpretación básica sigue siendo la misma.

Las interacciones de dos-factores también pueden definirse. En general, una interacción de dos factores se puede pensar como el efecto adicional de un factor sobre el efecto principal cuando el segundo factor se lleva a cabo en el nivel alto. En un diseño factorial a dos niveles balanceado, este efecto de interacción es igual a:

$$\hat{\Delta}_{jk} = \bar{y}_{++} - (\bar{y} + \hat{\Delta}_j + \hat{\Delta}_k) \tag{7}$$

donde

$\bar{y}_{++}$  = respuesta promedio en los niveles altos de ambos factores  $j$  y  $k$ .

y  $\bar{y}$  es el gran promedio:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \tag{8}$$

Una característica adicional importante de los efectos es que todos ellos se expresan en unidades de la variable respuesta, también es que los efectos de los factores pueden ser comparados directamente, sin importar las unidades en las cuales se expresan los factores.

### Resumen del Análisis

El análisis de un diseño de investigación se resuelve estimando el promedio o los efectos principales de cada factor experimental y las interacciones entre los factores. El *Resumen del Análisis* despliega información acerca de la estimación sobre los efectos:

<b><u>Analizar Experimento - reacted</u></b>			
Nombre del archivo: C:\DocData\chemical reaction.sfx			
Comentario: Chemical reactor example			
<b>Efectos estimados para reacted (%)</b>			
<i>Efecto</i>	<i>Estimado</i>	<i>Error Estd.</i>	<i>V.I.F.</i>
promedio	65.2105	0.378313	
A:feed rate	-2.0	0.824515	1.0
B:catalyst	20.5	0.824515	1.0
C:agitation	0.0	0.824515	1.0
D:temperature	12.25	0.824515	1.0
E:concentration	-6.25	0.824515	1.0
AB	1.5	0.824515	1.0
AC	0.5	0.824515	1.0
AD	-0.75	0.824515	1.0
AE	1.25	0.824515	1.0
BC	1.5	0.824515	1.0
BD	10.75	0.824515	1.0
BE	1.25	0.824515	1.0
CD	0.25	0.824515	1.0
CE	2.25	0.824515	1.0
DE	-9.5	0.824515	1.0

Errores estándar basados en el error total con 3 g.l.

La tabla muestra:

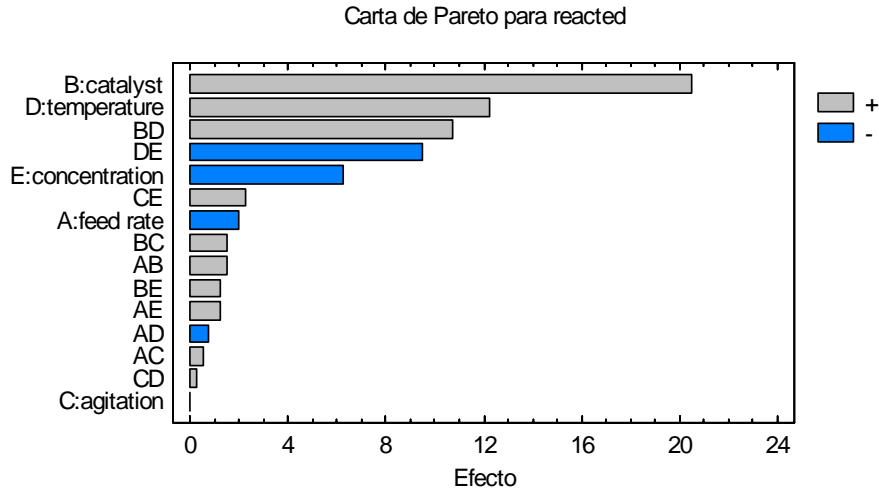
- **Promedio** - La estimación promedio en el centro de la región del diseño. Para datos completos de muchos diseños ortogonales, este es igual al gran promedio de todos los valores de los datos  $\bar{y}$ .
- **Estimación Efectos Principales** - La diferencia entre la respuesta del nivel alto y bajo de un factor cuando el resto de los factores se fijan en los valores centrales.
- **Estimación Interacciones de 2-Factores** - El efecto adicional de un factor cuando un Segundo se fija en su nivel alto. Las interacciones ocurren cuando el efecto de un factor es distinto en los diferentes niveles de cualquier otro factor.
- **Otros Efectos** – Definido como dos veces el coeficiente asociado con el correspondiente termino en el modelo de regresión cuando todas las variables fueron estandarizados de acuerdo a:

$$X'_j = \frac{X_j - low_j}{high_j - low_j} \quad (9)$$

- **Errores Estándar** - Para cada efecto se muestra por defecto cual es la estimación del error estándar. Los errores estándar son mediciones de las estimaciones del error asociado con cada efecto. La salida puede cambiarse al mostrar los intervalos de confianza para cada efecto usando la caja de dialogo de *Opciones del Análisis*.
- **V.I.F.** – Varianza de Inflación de los Factores mide el grado en el cual cualquier des-balance en el experimento puede inflar la varianza de la estimación de los efectos. Par aun diseño ortogonal perfecto, los factores son iguales a 1.0. Cualquier valor de 10 o mayor es generalmente tomado como una señal de serios problemas de correlación existentes entre la estimación de los efectos, causaría que las estimaciones fueran mucho más variables que las que serían en un buen diseño experimental.
- **Errores Estándar basados sobre...** - Una indicación de como el error experimental fue estimado, este es determinado por la caja de dialogo *Opciones del Análisis*.

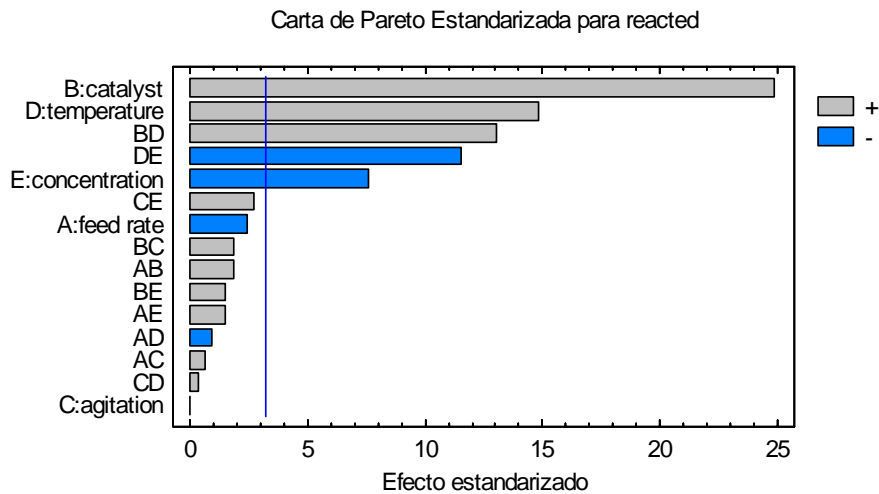
## Grafico de Pareto

El gráfico de Pareto presenta una pintura gráfica de cada uno de los efectos en la tabla de abajo. Hay dos tipos de grafico de Pareto, una forma estandarizada y otra sin estandarizar. El gráfico sin estandarizar muestra los valores absolutos de los efectos en orden decreciente:



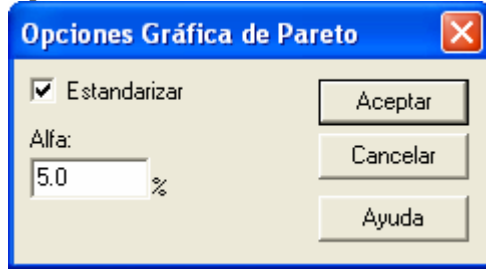
El color de las barras muestra si un efecto es positivo o negativo. Del gráfico anterior, es fácil ver que los tres efectos más importantes en el ejemplo son *catalyst*, *temperature*, y *concentration*.

Para crear un gráfico de Pareto estandarizado, cada efecto es transformado a un estadístico-t dividiéndolo entre su error estándar. Esta estandarización de efectos son graficados en orden decreciente de la magnitud absoluta:



En adición, una línea es dibujada sobre el gráfico para determinar si un efecto es estadísticamente significativo en un nivel de significancia especificado, generalmente del 5%. En el gráfico anterior los efectos de los factores B, D, y E son significativos, así como las interacciones BD y DE. Perceptiblemente se abstiene de la lista cualquier efecto que implique los factores A y C.

Opciones del Panel



- **Estandarizar:** Activa al grafico los efecto estandarizados en lugar de los efectos absolutos.
- **Alfa:** El nivel de significancia  $\alpha$  correspondiente a la línea vertical sobre el gráfico. Las barras que se extienden más allá de la línea son estadísticamente significativas en el nivel de significancia  $\alpha$ .

Tabla ANOVA

Para determinar el nivel de significancia para cada efecto, se puede utilizar la *Tabla ANOVA*:

Análisis de Varianza para reacted - Chemical reactor example					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
A:feed rate	16.0	1	16.0	5.88	0.0937
B:catalyst	1681.0	1	1681.0	618.17	0.0001
C:agitation	0.0	1	0.0	0.00	1.0000
D:temperature	600.25	1	600.25	220.74	0.0007
E:concentration	156.25	1	156.25	57.46	0.0048
AB	9.0	1	9.0	3.31	0.1664
AC	1.0	1	1.0	0.37	0.5870
AD	2.25	1	2.25	0.83	0.4301
AE	6.25	1	6.25	2.30	0.2268
BC	9.0	1	9.0	3.31	0.1664
BD	462.25	1	462.25	169.99	0.0010
BE	6.25	1	6.25	2.30	0.2268
CD	0.25	1	0.25	0.09	0.7815
CE	20.25	1	20.25	7.45	0.0720
DE	361.0	1	361.0	132.75	0.0014
Error total	8.15789	3	2.7193		
Total (corr.)	3339.16	18			

El ANOVA particiona la varianza de la respuesta en diversos componentes: uno para cada efecto principal, uno para cada interacción, y uno para el error experimental. La tabla del ANOVA muestra:

- **Suma de Cuadrados** – La suma de cuadrados Tipo III atribuible para cada término del modelo. Esta mide el incremento sobre la varianza del error experimental si cada término fuera eliminado por separado del modelo. La suma de cuadrados totales para el error también se incluye, donde

$$SC_{error} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{10}$$

$e_i$  es el i-ésimo residual, que mide la diferencia entre la respuesta observada para la corrida  $i$  y el valor predicho por la estimación del modelo.

- **g.l.** - Los grados de libertad asociados con cada término.
- **Cuadrados Medios** - El cuadrado medio asociado con cada término, obtenido por dividir la suma de cuadrados asociadas entre sus grados de libertad. El cuadrado medio del error (CME) estima la varianza del error experimental:

$$\hat{\sigma}^2 = CME = \frac{SC_{error}}{gl_{error}} \quad (11)$$

- **Cociente-F** – Un cociente F el cual divide el cuadrado medio de un efecto por el cuadrado medio del error:

$$F = \frac{CM_{efecto}}{CME} \quad (12)$$

El cociente F es utilizado para determinar la significancia estadística de cada efecto.

- **Valor P** – El valor-p asociado al probar la hipótesis nula de que el coeficiente para un efecto seleccionado es igual a 0, implicando que el efecto no esta presente. Un valor-p mayor que un nivel critico (como un 0.05 si esta operando el nivel de significancia del 5%) indica que el correspondiente efecto es estadísticamente significativo a un nivel de significancia.
- **R-Cuadrado** – El porcentaje de la variabilidad en la variable respuesta que ha sido considerado por el modelo estimado, calculado por

$$R^2 = 100 \left( 1 - \frac{SC_{error}}{SC_{total}} \right) \% \quad (13)$$

El rango del R-cuadrado va de 0% al 100% y mide que tan bien el modelo estimado explica los datos observados de la respuesta..

- **R-Cuadrado (ajustado por g.l.)** – El R-cuadrado ajustado, ha sido calculado por el numero de grados de libertad sobre el modelo estimado. En algunas situaciones como la actual donde es relativamente grande el número de coeficientes sobre el modelo estimado para el número de corridas totales, el estadístico ordinario R-Cuadrado puede exagerar la habilidad del modelo estimado para predecir la respuesta. La compensación del R-Cuadrado ajustado por este efecto es

$$R^2_{adj} = 100 \left[ 1 - \left( \frac{n-1}{n-p} \right) \frac{SC_{error}}{SC_{total}} \right] \% \quad (14)$$

donde  $p$  es el numero de coeficientes estimados en el modelo estimado.

- **Error Estándar de la Est.** - La desviación estándar del error experimental, calculado por

$$\hat{\sigma} = \sqrt{CME} \quad (15)$$

Este valor es utilizado cuando se construyen intervalos de predicción para la respuesta.

- **Media del Error Absoluto** - El promedio de valores absolutos de los residuales, calculado por

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n} \quad (16)$$

Este valor indica el error promedio de la predicción en la respuesta observada usando el modelo estimado.

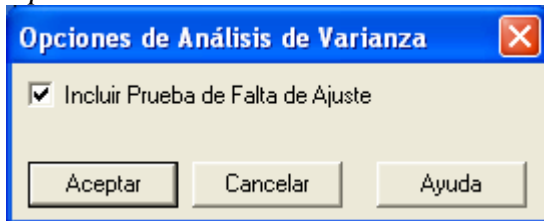
- **Estadístico Durbin-Watson** – Un estadístico calculado de los residuales de acuerdo a

$$DW = \frac{\sum_{i=1}^{n-1} (e_{i+1} - e_i)^2}{\sum_{i=1}^n e_i^2} \quad (17)$$

El estadístico Durbin-Watson mide la correlación serial en los residuales para determinar donde existe cualquier dependencia entre sucesivas observaciones. En este caso, no se detectan derivas sobre el curso del experimento. Un valor P pequeño indica que el analista debe realizar una mirada a detalle sobre los residuales para detectar cualquier tendencia, se puede hacer esto usando la opción grafica *Grafico de Diagnósticos*.

Hay cinco efectos en la tabla del ANOVA con valores p debajo de 0.05. Estos son los mismos cinco efectos identificados como significativos sobre el gráfico de Pareto estandarizado, (los dos métodos son equivalentes). En su totalidad, el modelo considera al menos un 98% de la variabilidad observada en la respuesta. Es innecesariamente complicado, sin embargo, puesto que algunos efectos no son significativos. En una sección posterior se ilustra como se remueve el efecto seleccionado usando Opciones del análisis.

#### Opciones del Panel



- **Incluir Prueba de Falta-de-Ajuste:** Si esta activada, una línea deberá adicionarse en la tabla del ANOVA para determinar si el modelo actual representa adecuadamente los datos observados. Nota: Esta opción no tiene ningún efecto a menos que se tengan replicas de corridas experimentales que son idénticas en la configuración de los factores experimentales.

El resultado de la tabla ANOVA se muestra abajo:



Análisis de Varianza para reacted - Chemical reactor example					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
A:feed rate	16.0	1	16.0	4.00	0.1835
B:catalyst	1681.0	1	1681.0	420.25	0.0024
C:agitation	0.0	1	0.0	0.00	1.0000
D:temperature	600.25	1	600.25	150.06	0.0066
E:concentration	156.25	1	156.25	39.06	0.0247
AB	9.0	1	9.0	2.25	0.2724
AC	1.0	1	1.0	0.25	0.6667
AD	2.25	1	2.25	0.56	0.5315
AE	6.25	1	6.25	1.56	0.3377
BC	9.0	1	9.0	2.25	0.2724
BD	462.25	1	462.25	115.56	0.0085
BE	6.25	1	6.25	1.56	0.3377
CD	0.25	1	0.25	0.06	0.8259
CE	20.25	1	20.25	5.06	0.1534
DE	361.0	1	361.0	90.25	0.0109
Falta de ajuste	0.157895	1	0.157895	0.04	0.8609
Error puro	8.0	2	4.0		
Total (corr.)	3339.16	18			

R-cuadrada = 99.7557 porcentaje  
R-cuadrada (ajustada por g.l.) = 98.5341 porcentaje  
Error Estándar del Est. = 2.0  
Error absoluto medio = 0.254848  
Estadístico Durbin-Watson = 1.37903 (P=0.4687)  
Autocorrelación residual de Lag 1 = 0.00827674

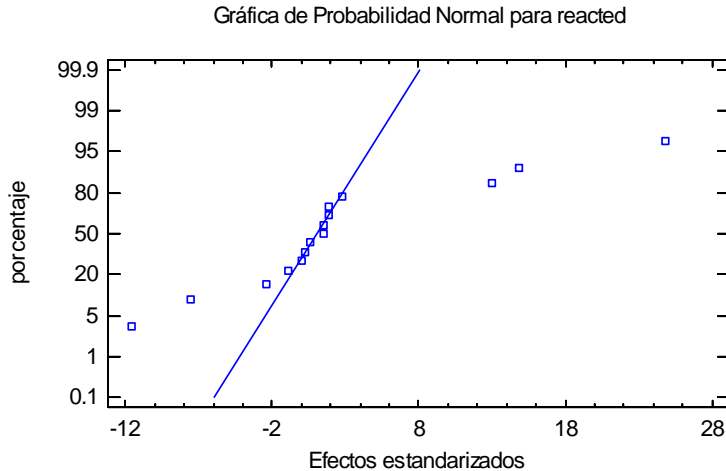
Nota las líneas etiquetadas Falta-de-Ajuste y Error Puro, proporcionan dos estimaciones por separado de la sigma del error experimental:

1. *Error Puro*: Una estimación calculada por la acumulación de la varianza dentro del conjunto de observaciones en los niveles idénticos de X. Es “puro” en el sentido que es la estimación del error experimental  $\sigma$  si no se ha seleccionado el modelo apropiado.
2. *Falta-de-Ajuste*: Una estimación calculada de la desviación entre la respuesta promedio para cada grupo de valores replicados y los valores predichos por el modelo estimado. Si el modelo es incorrecto, la estimación de  $\sigma$  adiciona una cantidad positiva que mide la falta-de-ajuste sobre el modelo seleccionado.

El valor P en la línea de Falta-de-ajuste es usado para probar la hipótesis de que el modelo actual es el adecuado. Un valor P pequeño indica un modelo inadecuado. En el ejemplo actual, el valor P es bueno arriba de 0.05, la selección del modelo parece ser la adecuada.

### Gráfico de Probabilidad Normal de Efectos

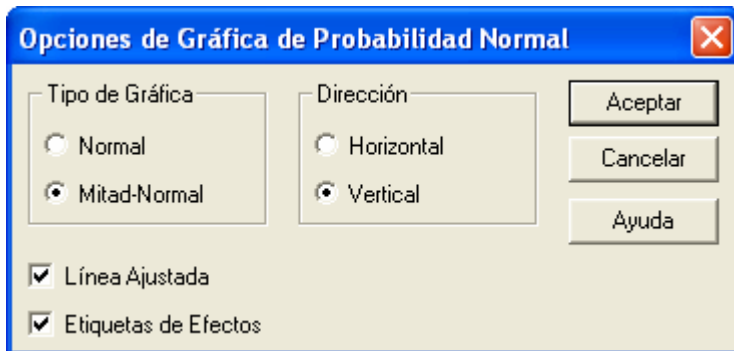
Cuando los grados de libertad disponibles para estimar el error experimental son pocos, la prueba formal F conducida en el ANOVA puede no tener mucha potencia, de modo que los efectos más pequeños no parecen ser significativos. Por otro lado, probando un número grande de efectos, cada uno a un nivel de significancia del 5%, podemos generar resultados más significativos que los que se presentan. Una manera menos rigurosa de juzgar cuales efectos son verdaderos y cuales son probablemente manifestaciones justas del ruido es a través del *Gráfico de Probabilidad Normal de Efectos*:



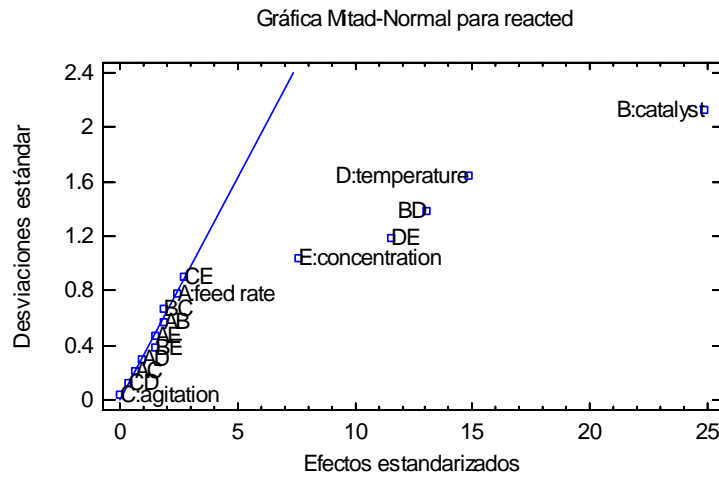
En este gráfico, los efectos estandarizados son ordenados del menor al mayor y graficados contra cuantiles de una distribución normal. Cualquier estimación que sea un ruido justo bajara aproximadamente a través de una línea recta. Cualquier estimación que corresponda a señales verdaderas mentira fuera de la línea por la izquierda o por la derecha.

Dos tipos de gráficos de probabilidad normal están disponibles, uno normal completo y Medio-normal, los cuales se cambian usando *Opciones del Panel*.

#### *Opciones del Panel*



- **Tipo de Gráfico:** Seleccione *Normal* para graficar cada efecto mientras conserva su signo positivo o negativo. Seleccione *Medio Normal* para graficar los valores absolutos de los efectos.
- **Dirección:** Seleccione *Horizontal* para dibujar porcentaje sobre las equis, o *Vertical* para graficar los mismo sobre el eje vertical.
- **Línea Estimada:** Si esta activada, una línea de referencia es agregada sobre el grafico por la estimación de mínimos cuadrados para una línea de regresión por el 50% de los efectos.
- **Etiquetar Efectos:** Si esta activada, las nombres de los efectos de adicionan al grafico.

Ejemplo: Grafico Medio-Normal con Efectos Etiquetados.

Este grafico tiene la ventaja de que todas las señales caen a la derecha de la línea del ruido. This plot has the advantage that all signals fall to the right of the noise line. El grafico anterior confirma la conclusión de que 5 efectos significativos están presentes.

### Opciones del Análisis

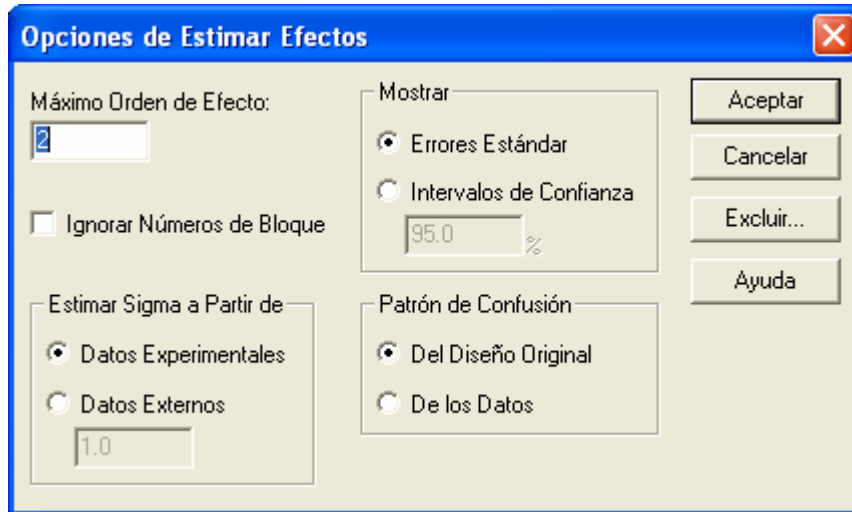
El modelo matemático utilizado para estimar los datos que contienen 5 efectos principales y 10 interacciones de los factores. Según lo visto arriba, muchos de estos términos ni tienen significancia estadística. Al construir los modelos empíricos basados solamente con los datos observados, esto también es importante para mantener los modelos tan simples como sea posible, puesto que modelos simples tienden a facilitar la interpretación y tener una mejor opción de extrapolar en otras combinaciones de los factores experimentales.

De acuerdo con el principio de parsimonia o *K.I.S.S.* (Keep It Simple Statistically- subsistencia del más simple estadístico), los efectos insignificantes deberán removerse del modelo de acuerdo a las siguientes reglas:

1. Remover cualquier interacción de dos factores insignificantes (u otros efectos de segundo orden).
2. Remover cualquier efecto principal insignificante que no este implicado en interacciones significativas.

Note que los efectos principales correspondientes a los factores que están implicados en interacciones significativas no deben ser removidos normalmente, puesto que hacen un lugar artificial de contrastes sobre el modelo polinomial subyacente.

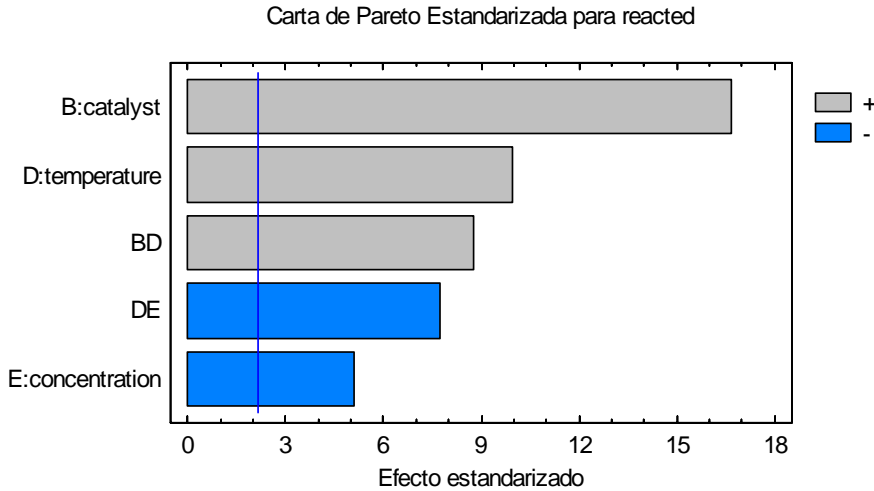
Para eliminar un efecto de un modelo, seleccione *Opciones del Análisis*:



- **Máximo Orden del Efecto** – El máximo orden del efecto a ser incluido en el modelo. Se determina por defecto en 2 para solicitar la estimación de los efectos principales e interacciones de dos factores. Si se fija en 1, solamente los efectos principales podrán estimarse.
- **Ignorar Número de Bloques** – Para diseños que contiene más de un bloque, indica cuando los efectos de bloque deben estimarse o ignorarse. Note que la columna 1 de la base de datos para cualquier archive de experimento contiene los números de bloques correspondientes a cada fila.
- **Estimar la Sigma por** – Cuando la desviación estándar del error experimental  $\sigma$  es estimada por el error experimental, o cuando el analista proporciona un valor conocido. Si un valor conocido es proporcionado, las pruebas estadísticas y los intervalos de confianza usaran este valor.
- **Desplegar** – Afecta la salida a desplegar para cada efecto sobre el *Resumen del Análisis*.
- **Patrón de Confusión** - Específica como el procedimiento determina cuales efectos serán estimados cuando se ajuste el modelo. Las opciones son:
  1. *Del Diseño Original* - Examina el patrón de confusión del diseño original para determinar cuales efectos podrán estimarse. Por ejemplo, en un diseño con resolución IV, el programa estimara las combinaciones específicas de interacciones dobles. Esta es la opción por defecto y es apropiada para todas excepto en circunstancias muy especiales.
  2. *De los Datos* – Examina la matriz X para todas las corridas desarrolladas para determinar cuales efectos podrán estimarse. Esto puede ser deseable cuando el analista agrego corridas adicionales al diseño base para estimar limpiamente ciertas interacciones que serían confundidas. Puesto que el programa procura estimar todos los efectos cuando se cambia a esta opción, la caja de dialogo *Excluir* podrá utilizarse para decirle al programa exactamente cuales efectos serán estimados.
- **Excluir** - Cuando es presionado, genera la caja de dialogo que se presenta más abajo.

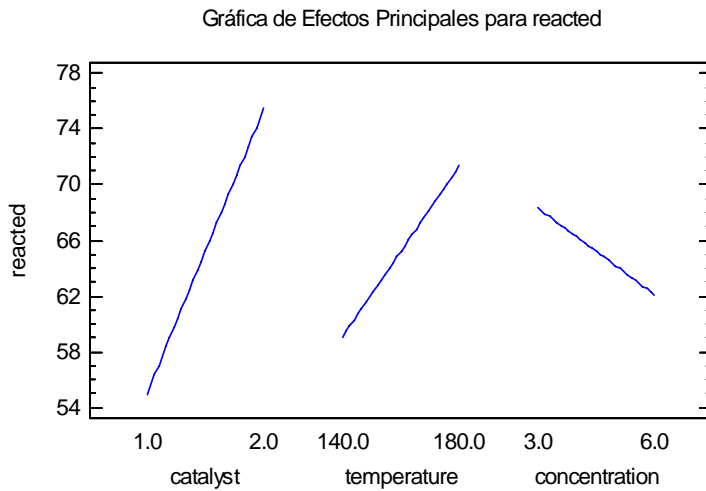
Los efectos pueden excluirse del modelo con doble-clic solo una vez sobre ellos. Dar doble-clic sobre un efecto en cualquiera de las dos columnas lo mueve a la otra columna.

En el ejemplo actual, todos los demás efectos aparte de los 5 que parecieron ser estadísticamente significativos fueron removidos. El grafico de Pareto estandarizado para el nuevo modelo se muestra con los efectos restantes:



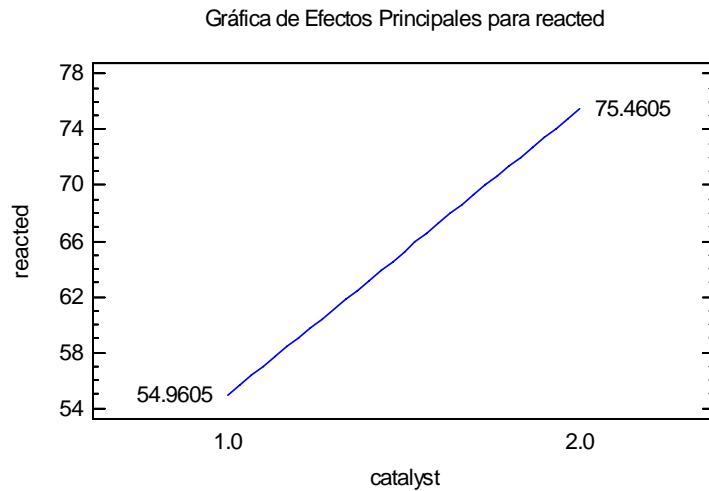
### Gráfico de Efecto Principales

Una vez que se ha estimado y comprobado un modelo, los resultados se deben presentar de una manera que sea lo más comprensible para todos. Puesto que a menudo es difícil ganar penetraciones mirando una ecuación matemática, varios gráficos se proporcionan para presentar la estimación del modelo. El gráfico de *Efectos Principales* es uno de los más importantes:



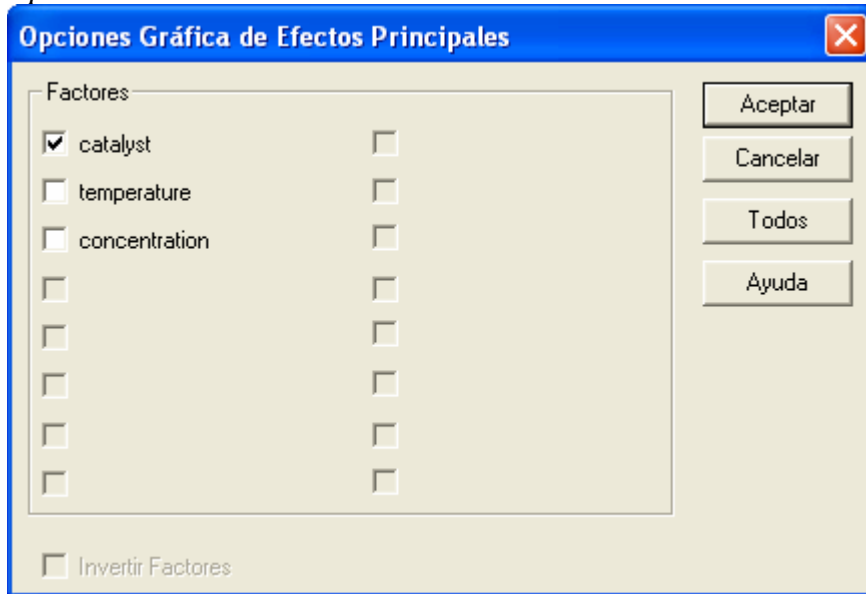
Se presenta como la respuesta predicha varia cuando cada uno de los factores del modelo cambia del nivel bajo al alto, con todos los demás factores fijados en el centro de la región experimental (un punto medio entre el nivel bajo y alto). Cuando todos los factores se dibujan en el grafico

anterior, es más fácil juzgar cuales factores tienen el más alto impacto. Cuando se dibujan individualmente, se muestra la respuesta predicha en los extremos del factor seleccionado:



Nota: En algunos casos, los valores mostrados en los puntos finales de la línea del gráfico anterior son iguales a la respuesta promedio en el nivel bajo y alto del factor graficado. Esto no es un caso general, sin embargo. Es importante notar que STATGRAPHICS grafica la respuesta predicha del modelo actual, no de los datos observados. Esto permite que el gráfico sea utilizado con muchos tipos de diseños además del factorial a dos niveles.

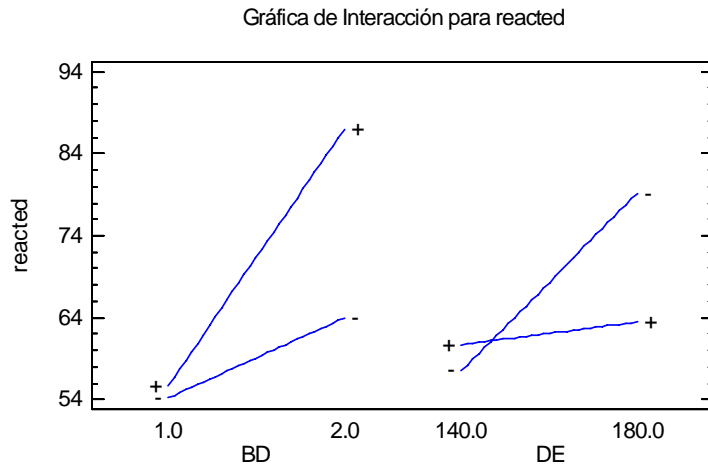
#### Opciones del Panel



- **Factores:** Factores a ser incluidos en el gráfico.

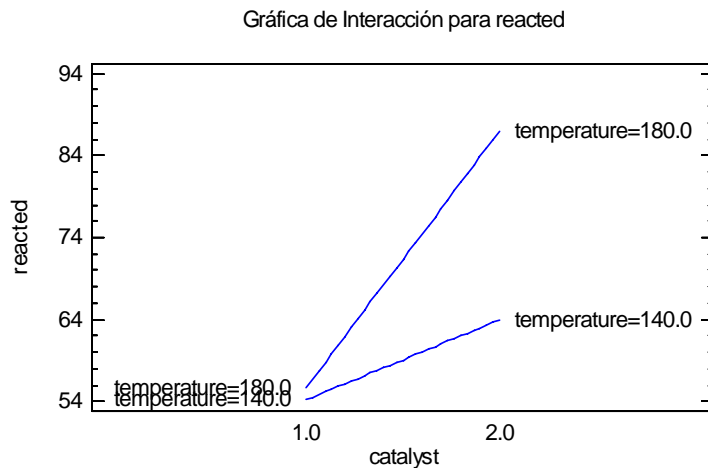
## Gráfico de Interacción

Cuando existen interacciones significativas en los factores experimentales, el gráfico de efectos principales no cuenta la toda la historia sobre los factores que interactúan e incluso pueden ser engañosos. En tales casos, un *Gráfico de Interacción* debe producirse para cada par de factores. Si más de una interacción es graficada, la presentación tomara la siguiente forma:



Un par de líneas deberán graficarse para cada interacción, correspondientes a la respuesta predicha cuando un factor se varia del valor bajo al alto, en cada nivel del otro factor. Todos los factores que no están implicados en la interacción se llevan a cabo en su valor central.

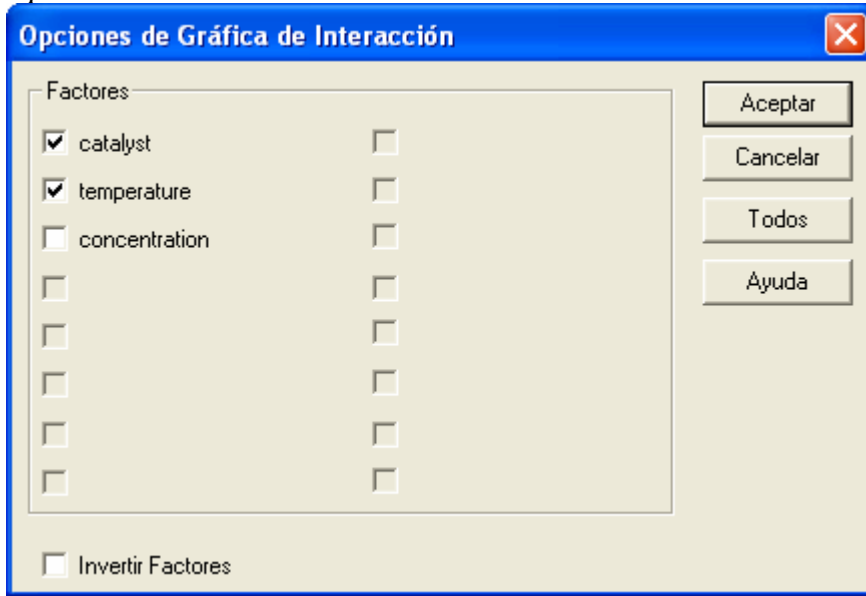
El gráfico generalmente es fácil de comprender si las *Opciones del Panel* son utilizadas para graficar cada interacción por separado:



La respuesta predicha para cada combinación de los niveles bajo y alto de dos factores es presentada al final de cada segmento de la línea. Si dos factores no interactúan, el efecto de un factor no dependerá sobre el nivel del otro y las dos líneas sobre el gráfico de interacción deberán ser aproximadamente paralelas. Si los factores interactúan, como en la figura anterior, las líneas no serán paralelas e incluso pueden cruzarse. La interpretación de los gráficos de

interacción generalmente es altamente informativa. Por ejemplo, el gráfico anterior muestra que la *temperature* tiene un efecto suave en el nivel bajo del *catalyst*. Sin embargo, tiene un efecto pesado en el nivel alto del *catalyst*.

*Opciones del Panel*



- **Factores:** Dos o más factores a incluirse sobre el gráfico. Todas las interacciones para ambos factores deben activarse para ser incluidas.
- **Factores al Revés:** Si esta activada, el primer factor en lugar del segundo factor será usado para definir las líneas sobre el gráfico.

**Coeficientes de Regresión**

El modelo de regresión subyacente puede ser presentado al seleccionar el panel de *Coeficientes de Regresión*:

Coef. de regresión para reacted - Chemical reactor example	
Coeficiente	Estimado
constante	9.83553
B:catalyst	-65.5
D:temperature	0.2125
E:concentration	23.25
BD	0.5375
DE	-0.158333

**El StatAdvisor**  
 Esta ventana despliega la ecuación de regresión que se ha ajustado a los datos. La ecuación del modelo ajustado es

$$\text{reacted} = 9.83553 - 65.5 * \text{catalyst} + 0.2125 * \text{temperature} + 23.25 * \text{concentration} + 0.5375 * \text{catalyst} * \text{temperature} - 0.158333 * \text{temperature} * \text{concentration}$$

en donde los valores de las variables están especificados en sus unidades originales. Para hacer que STATGRAPHICS evalúe esta función, seleccione Predicciones de la lista de Opciones Tabulares. Para graficar la función, seleccione Gráficas de Respuesta de la lista de Opciones Gráficas.



El StatAdvisor despliega la ecuación, la cual corresponde al modelo de regresión descrito anteriormente. Esta es la ecuación utilizada para predecir la respuesta en valores específicos de los factores experimentales.

Nota: En la ecuación de regresión, todos los factores que fueron definidos como *continuos* cuando el experimento fue inicialmente creado son expresados en sus unidades originales (Ej., la temperatura es expresada en °C). Los factores que no se definieron como *continuos* usan la codificación de -1 para el nivel bajo y +1 par el nivel alto.

## Matriz de Correlación

La matriz de correlación despliega las correlaciones estimadas entre los coeficientes sobre la estimación del modelo de regresión:

Matriz de Correlación para los Efectos Estimados							
		(1)	(2)	(3)	(4)	(5)	(6)
(1)	promedio	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
(2)	B:catalyst	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
(3)	D:temperature	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
(4)	E:concentration	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
(5)	BD	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
(6)	DE	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Las correlaciones son estimadas sobre los coeficientes de la matriz de varianza-covarianzas, calculada por

$$s^2(b) = CME(X'X)^{-1} \quad (18)$$

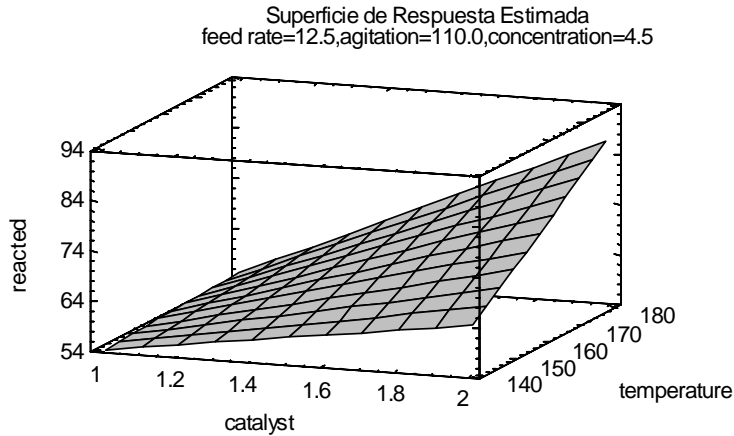
Una matriz diagonal tal como se muestra arriba indica que las estimaciones de cada uno de los efectos no están correlacionadas con las demás estimaciones, esto proviene de la ortogonalidad del diseño original. Si cualquier valor en los datos esta en blanco (indica que la correspondiente celda de la columna respuesta en la base de datos esta vacía), o corridas adicionales son agregadas por el usuario, estas deben tomar valores diferentes de cero para los términos fuera de la diagonal. Las correlaciones fuertes son probables que conduzcan a efectos mal definidos que son difíciles de interpretar.

## Gráficos de Respuestas

Una lista de los gráficos disponibles sobre el procedimiento *Analizar Datos* contiene dos selecciones etiquetadas como *Gráficos de Respuestas* que permiten predecir los valores de la respuesta por gráficos de diversas maneras. Por defecto, la primera selección crea un grafico de superficie de respuesta y la segunda un grafico de contornos. Sin embargo, cada una es controlada por la misma caja de dialogo *Opciones del Panel*, la cual permite al analista desplegar cualquiera de los cuatro tipos de gráficos:

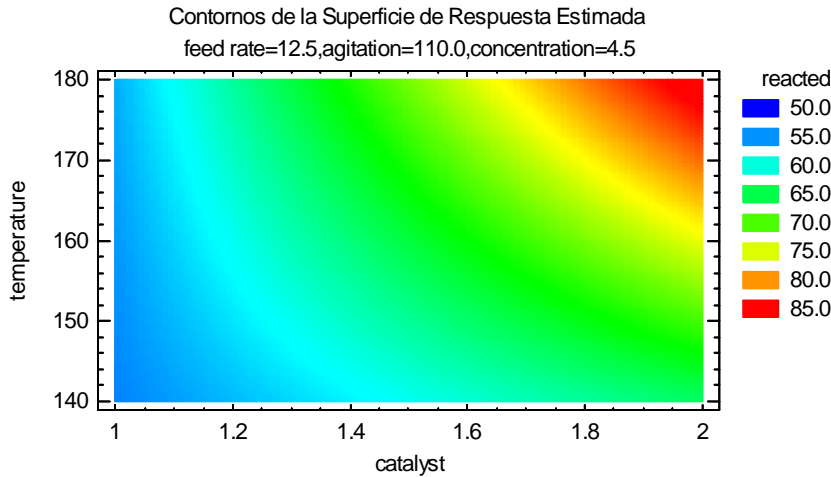
1. Un grafico de Superficie.
2. Un grafico de Contornos.
3. Un grafico de Cuadro.
4. Un grafico de Cubo.

El *Gráfico de Superficie* presenta un grafico de la respuesta predicha como una función de cualquiera par de factores experimentales, con las demás factores se debe seleccionar los valores. Por ejemplo, el grafico de abajo muestra la *reacción* como una función del *catalizador* y la *temperatura*:



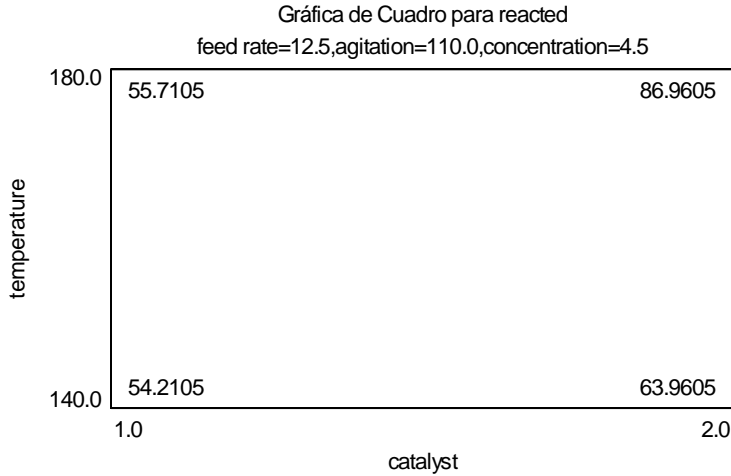
La altura de la superficie representa el valor predicho  $\hat{Y}$ , el cual fue graficado sobre el rango de los factores experimentales.

El *Grafico de Contorno* dibuja líneas o regiones coloreadas basándose sobre los valores de las respuestas predichas. Por ejemplo, el grafico de abajo presenta el rango de los valores predichos para la *reacted* usando colores que se extienden desde el azul en 50 hasta el rojo en 85:

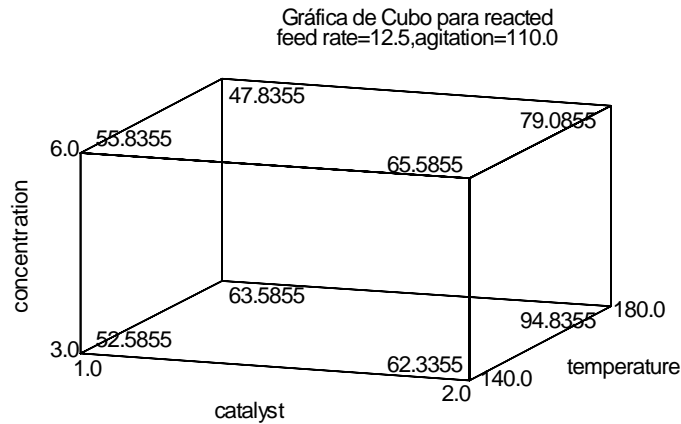


La rampa del color utilizado es controlada por la *Paleta de Colores* sobre la caja de dialogo de *Opciones Graficas*.

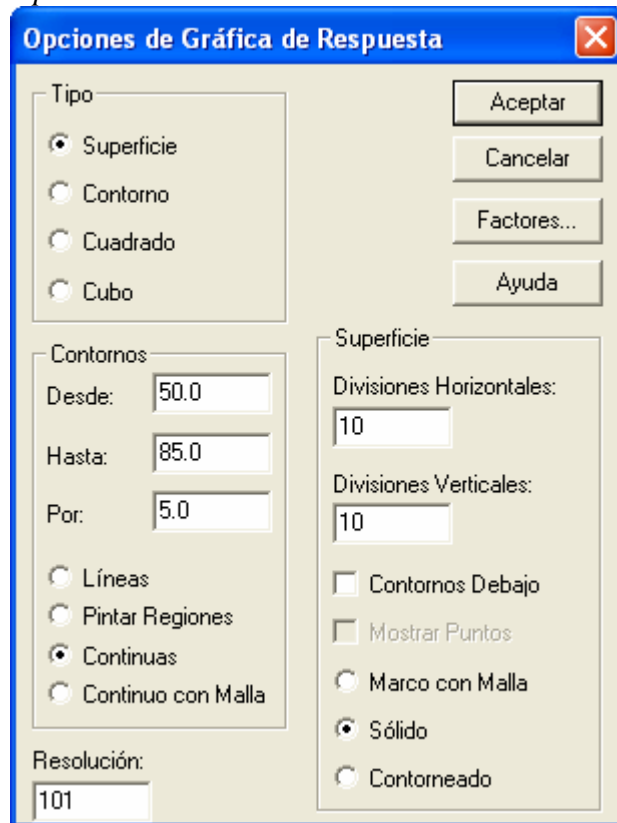
El *Grafico de Cuadro* muestra la respuesta predicha en las combinaciones de los niveles bajo y alto para cualquier par de factores:



El *Gráfico de Cubo* muestra la respuesta predicha en combinaciones de niveles bajo y alto para cualquier tercia de factores:



En el ejemplo, el valor de predicción más alto para la *reacción* es obtenido con *catalyst* = 2, *temperature* = 180, y *concentration* = 3. In the example, the highest predicted value for *reacted* is obtained at *catalyst* = 2, *temperature* = 180, y *concentration* = 3. La interacción más fuerte entre los factores da lugar a un valor predicho aproximado de 95 en esa combinación de los factores.

*Opciones del Panel*

- **Tipo:** Tipo de grafico de respuesta a ser creado.
- **De:** Localización de la primera línea de contorno a dibujarse, o el inicio de la primera región.
- **Hasta:** Localización de la ultima línea de contorno a dibujarse, o el fin de la primer región.
- **Por:** Espaciado entre las líneas de contornos o las regiones.
- **Líneas:** Si esta seleccionada, líneas de contornos secuenciales son dibujadas en los niveles seleccionados de la variable respuesta, como un mapa topográfico.
- **Pintar Regiones:** Si esta seleccionada, un conjunto de regiones serán dibujadas cubriendo varios rangos de la respuesta predicha.
- **Resolución:** Define la resolución  $m$  de una malla de  $m$  por  $m$  de valores predichos los cuales son utilizados para dibujar la superficie y líneas del contorno. Incrementar la resolución ayuda a mejorar el suavizamiento y definición de los gráficos, a expensas de tiempo y memoria de la computadora.
- **Divisiones Horizontales:** El número de divisiones a lo largo del primer eje experimental. Esto determina cuantas líneas verticales deberán dibujarse sobre el gráfico de superficie.
- **Divisiones Verticales:** El número de divisiones a lo largo del segundo eje experimental. Esto determina cuantas líneas horizontales deberán dibujarse sobre el gráfico de superficie.

- **Contornos Abajo:** Solicita un tipo específico de gráfico de contorno por abajo, será dibujado en la cara inferior del gráfico 3-D.
- **Mostrar Puntos:** Solicita que los valores observados de los datos  $Y_i$  sean agregados al gráfico, con líneas dibujadas para cada punto sobre la superficie.
- **Malla de Alambre:** Solicita que la superficie se dibuje por medio líneas cruzadas como se muestra en la figura anterior. Esta es la opción más eficaz para la presentación en blanco y negro.
- **Solidó:** Solicita que la superficie se dibuje usando un color solidó.
- **Contorneado:** Solicita que la superficie se dibuje mostrando niveles de contornos para la respuesta.
- **Continuo:** Dibuja contornos usando un rango de colores continuos.
- **Factores:** Especifica los factores a ser graficados en cada eje y los niveles en los cuales los demás factores serán fijados:

	Bajo	Alto	Mantener
<input type="checkbox"/> feed rate	10.0	15.0	12.5
<input checked="" type="checkbox"/> catalyst	1.0	2.0	1.5
<input type="checkbox"/> agitation	100.0	120.0	110.0
<input checked="" type="checkbox"/> temperature	140.0	180.0	160.0
<input checked="" type="checkbox"/> concentration	3.0	6.0	4.5
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0
<input type="checkbox"/>	0.0	1.0	0.0

Si creamos un gráfico de superficie, contorno o cuadrado deben activarse dos factores. Si creamos un gráfico de cubo, tres factores deberán activarse.

El ejemplo actual grafica los valores predichos contra *catalyst* y *temperature*, cuando *feed rate* = 12.5, *agitation* = 110, y *concentration* = 4.5.

### Predicciones

El panel de *Predicciones* puede ser utilizado para generar predicciones del modelo estimado:

Resultados Estimados para reacted						
	Observados	Ajustados		Estudentizado	Inferior 95.0% LC	Superior 95.0% LC
Fila	Valor	Valor	Residuo	Residuo	para Media	para Media
1	65.0	65.2105	-0.210526	-0.0846423	63.9933	66.4277
2	56.0	55.8355	0.164474	0.0807762	52.6295	59.0415
3	53.0	52.5855	0.414474	0.203853	49.3795	55.7915
4	63.0	62.3355	0.664474	0.327704	59.1295	65.5415
5	65.0	65.5855	-0.585526	-0.28848	62.3795	68.7915
6	53.0	52.5855	0.414474	0.203853	49.3795	55.7915
7	55.0	55.8355	-0.835526	-0.413139	52.6295	59.0415
8	67.0	65.5855	1.41447	0.708878	62.3795	68.7915
9	61.0	62.3355	-1.33553	-0.667797	59.1295	65.5415
10	67.0	65.2105	1.78947	0.735268	63.9933	66.4277
11	69.0	63.5855	5.41447	4.1464	60.3795	66.7915
12	45.0	47.8355	-2.83553	-1.52039	44.6295	51.0415
13	78.0	79.0855	-1.08553	-0.539401	75.8795	82.2915
14	93.0	94.8355	-1.83553	-0.933357	91.6295	98.0415
15	49.0	47.8355	1.16447	0.57969	44.6295	51.0415
16	60.0	63.5855	-3.58553	-2.04408	60.3795	66.7915
17	95.0	94.8355	0.164474	0.0807762	91.6295	98.0415
18	82.0	79.0855	2.91447	1.57129	75.8795	82.2915
19	63.0	65.2105	-2.21053	-0.919228	63.9933	66.4277
20		62.6605			60.9012	64.4199

Promedio de 3 puntos centrales = 65.0  
 Promedio de las predicciones del modelo al centro = 65.2105

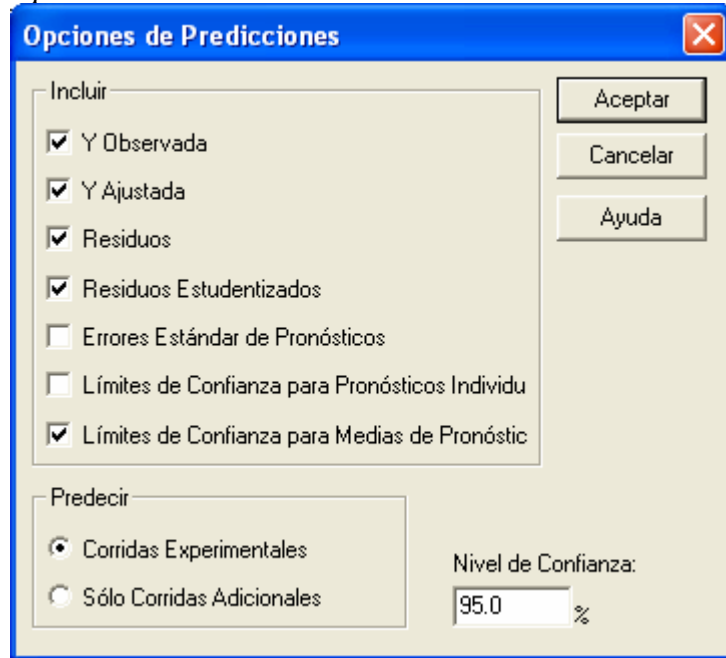
La tabla incluye todas las filas en la base de datos, o solamente filas en la cual el valor de la variable respuesta *Y* no fue incorporado. Esta última característica permite al analista hacer predicciones en las combinaciones de *X* que no fueron incluidas en el experimento. Por ejemplo, la tabla anterior muestra los resultados más una predicción en la fila 20 con *feed rate* = 12.5, *catalyst* = 1.2, *agitation* = 105, *temperature* = 165, y *concentration* = 3.5. El valor predicho de la *reacción* es 62.66. Un intervalo de confianza al 95% para el valor promedio de la *reacción* en la misma combinación sobre los rangos de los factores oscila de 60.9 hasta 64.6.

La tabla también exhibe el promedio de las corridas experimentales desarrolladas en el centro de la región experimental, junto con la respuesta predicha. Si el modelo asumido es correcto, ambos valores deben ser cercanos. Si no, puede haber una curvatura no modelada con respecto a uno o más de los factores experimentales. Para determinar la naturaleza de la curvatura se requiere desarrollar corridas adicionales en los diferentes niveles de los factores. Para agregar corridas adicionales en un experimento de investigación, podemos usar la selección *Aumentar Diseño* sobre el menú *DDE*.

Por otro lado es notable la entrada para el residual estandarizado en la tabla anterior para la fila 11. Los residuales estandarizados miden la diferencia entre la respuesta observada y la respuesta predicha, en unidades del error estándar, cuando la observación es cuestionable no es usada para la estimación del modelo. El residual estandarizado para la observación 11 es igual a 4.1. Los valores que exceden de 3.0 son inusuales y típicamente requieren una investigación a detalle. Si

el punto en cuestión no tiene un resultado deseable, volver a realizar el conjunto de condiciones experimentales de ser necesario.

*Opciones del Panel*



- **Incluir:** Cálculos a incluir en la tabla:

1. *Y Observada* – Los valores de la respuesta  $Y_i$  observada.
2. *Y Predicha* – Los valores de predicción  $\hat{Y}_i$  calculados sobre el modelo estimado.
3. *Residuales* – Los residuales  $e_i$ .
4. *Residuales Estandarizados* - Es un tipo de residual estandarizado, donde cada residual es dividido por una estimación del error estándar. STATGRAPHICS computa los *residuales estandarizados eliminados*, en donde cada observación es removida solo una vez y el modelo es re-ajustado sin estos valores. Los residuales eliminados que son igual a la respuesta observada menos el valor predicho sobre modelo estimado sin la observación, ej.

$$d_i = Y_i - \hat{Y}_{(i)} \tag{19}$$

El residual estandarizado es calculado por

$$e_i^* = \frac{d_i}{s(d_i)} \tag{20}$$

donde

$$s^2(d_i) = CME_{(i)} \left( 1 + X_i' (X_{(i)}' X_{(i)})^{-1} X_i \right) \tag{21}$$

Los residuales eliminados pueden seguir una distribución  $t$  con  $n-p-1$  grados de libertad, donde  $p$  es el número de coeficientes en el modelo estimado.

5. *Error Estándar para Pronósticos* - El error estándar para nuevas observaciones en una combinación seleccionada de los factores experimentales  $X_h$ , es calculada por

$$\sqrt{CME(1 + X'_h (X'X)^{-1} X_h)} \quad (22)$$

6. *Limites de Confianza para Pronósticos Individuales* - Limites de confianza para nuevas observaciones en una combinación seleccionada de los factores experimentales  $X_h$ , calculada por

$$\hat{Y}_h \pm t_{n-p} \sqrt{CME(1 + X'_h (X'X)^{-1} X_h)} \quad (23)$$

7. *Limites de Confianza para Pronósticos de Medias* - Limites de confianza para la respuesta promedio en una combinación seleccionada de los factores experimentales  $X_h$ , calculada por

$$\hat{Y}_h \pm t_{n-p} \sqrt{CME(X'_h (X'X)^{-1} X_h)} \quad (24)$$

- **Predicción** – Los pronósticos son desplegados para todas las corridas del archivo de datos experimentales, o solamente para las corridas que tienen un valor en blanco en la columna de la respuesta.
- **Nivel de Confianza** – El nivel de confianza para los intervalos.

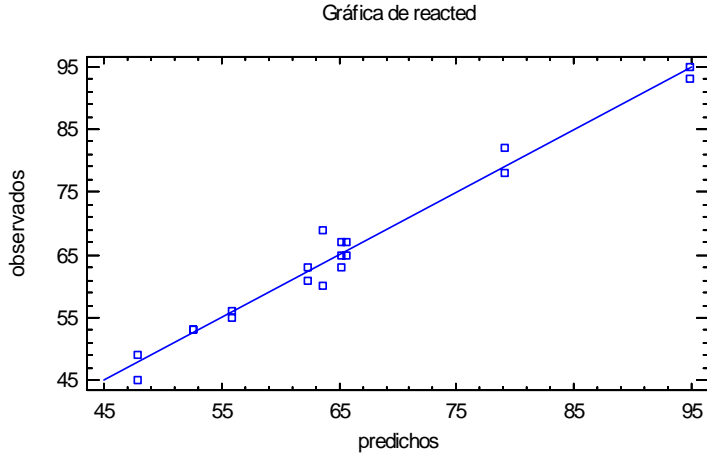
### Gráficos de Diagnósticos

Diversos gráficos también son proporcionados como *Gráficos de Diagnósticos* para examinar los residuales del modelo estimado. La caja de dialogo de *Opciones del Panel* despliega las diversas alternativas, las cuales incluyen las siguientes:

#### Observado contra Predicho

Este gráfico despliega la respuesta observada  $Y_i$  contra los valores predichos  $\hat{Y}_i$ , junto con una línea diagonal:

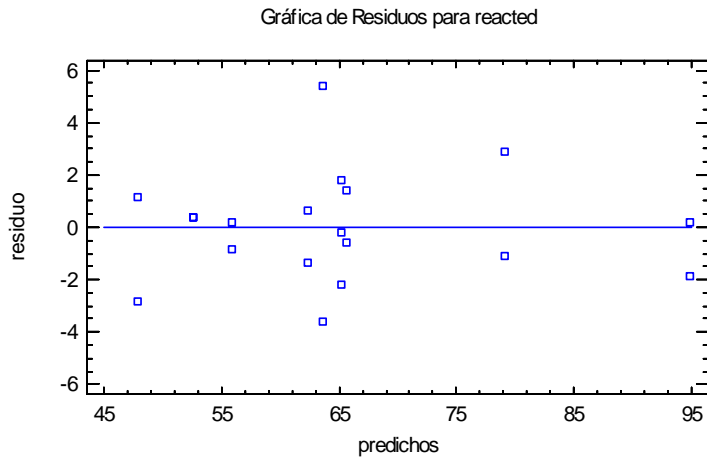




Si el modelo estimado es bueno, los valores no mienten sobre la línea, como en el ejemplo anterior. Una curvatura alrededor de la línea puede sugerir la necesidad de transformar los valores de  $Y_i$  usando un logaritmo o una función similar.

Residual contra Predicción

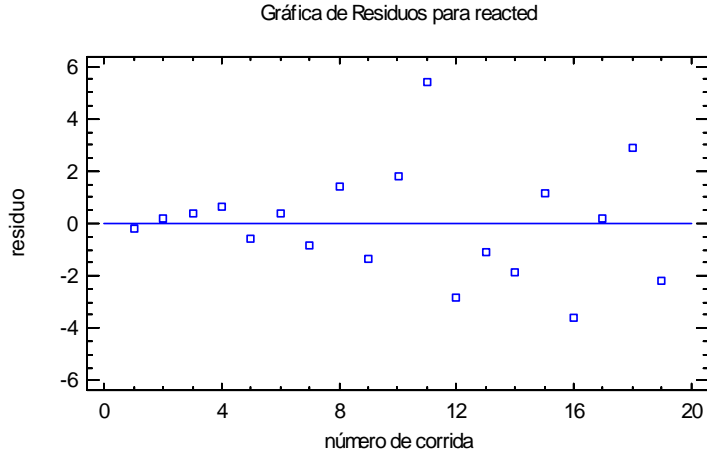
Este gráfico despliega los residuales  $e_i$  contra los valores predichos  $\hat{Y}_i$ , con una línea horizontal en cero:



Los residuales pueden variar aleatoriamente alrededor de la línea. Cambios en la magnitud de los residuales de izquierda a derecha es señal de que la varianza del error experimental varía en el nivel promedio de la respuesta. Tal heterocedasticidad puede frecuentemente ser eliminada por una transformación para estabilizar la varianza como un logaritmo o raíz cuadrada.

Residuales contra Orden de Corrida

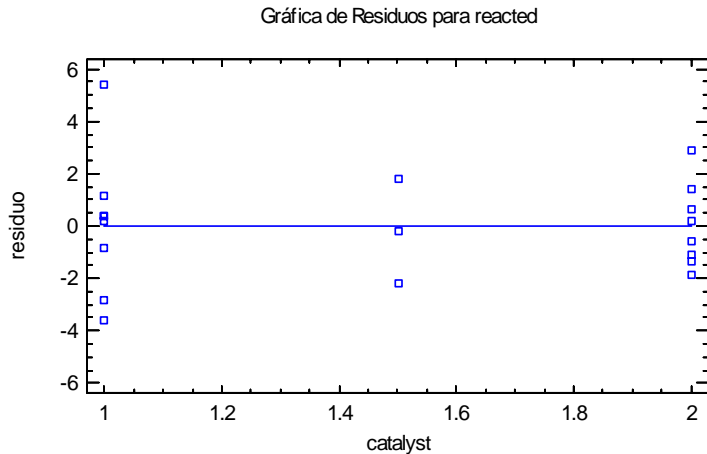
Este gráfico despliega los residuales  $e_i$  contra el número de corrida  $i$ , con una línea horizontal en cero:



Cualquier patrón no aleatorio puede indicar una tendencia en el tiempo u otro efecto. En tales casos, agregar un factor para describir los cambios que se puedan mejorar para la estimación del modelo. La gráfica de arriba sugiere un incremento en la variabilidad durante la segunda mitad del experimento, la cual puede ser causa de una investigación a detalle.

Residuales contra un Factor

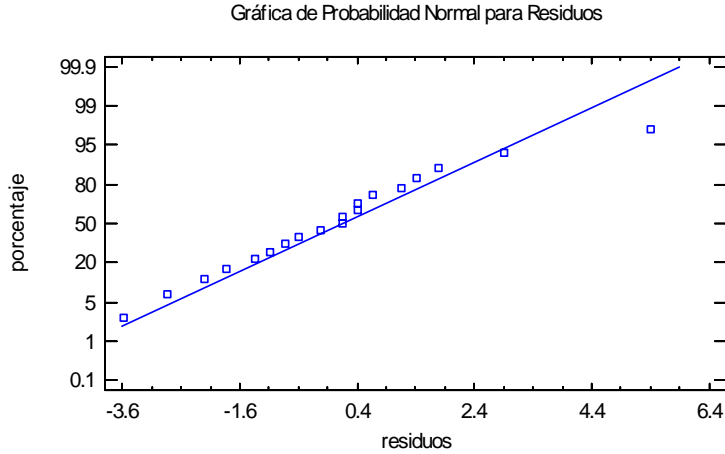
Este gráfico despliega los residuales  $e_i$  contra los valores observados de un factor experimental seleccionado:



Cualquier curvatura alrededor de la línea sugiere la necesidad de un modelo con efectos cuadráticos. El grafico anterior sugiere que la variabilidad entre los valores replicados en los puntos al centro podría ser menores que los residuales en los niveles bajo y alto del catalyst.

Gráfico de Probabilidad Normal de Residuales

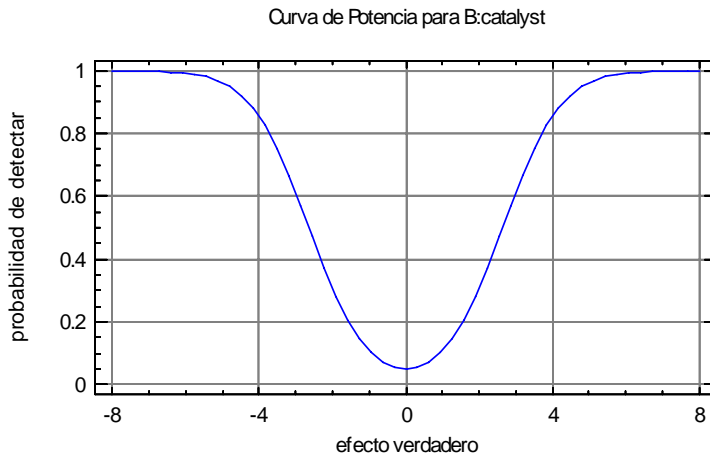
Este gráfico despliega los residuales  $e_i$  contra los cuantiles de una distribución normal, con una línea estimada opcional como referencia:



Si el error experimental sigue una distribución normal, los puntos no mienten a través de la línea recta. El gráfico anterior sugiere que el residual más grande (fila #11) sea tan alto como se espera, puesto que mentiría fuera de la línea definida por los demás. Esto indica la presencia de un valor atípico o curvatura.

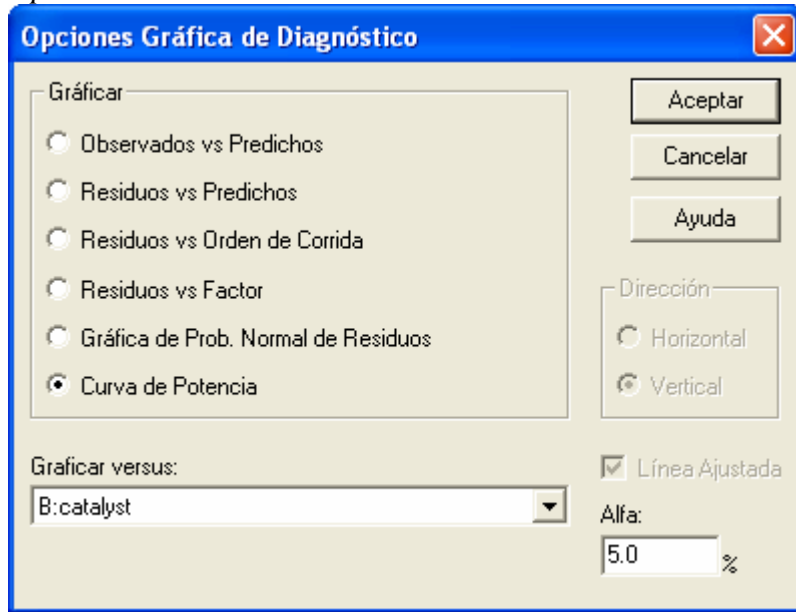
Curva de Potencia

La curva de potencia muestra la habilidad de la prueba estadística para detectar efectos de una magnitud:



Este gráfico es similar a la curva de potencia explicada a detalle anteriormente cuando el diseño fue inicialmente construido, excepto que el eje horizontal es desplegado en unidades de la respuesta sobre un cociente señal-a-ruido. El gráfico anterior muestra que el experimento actual tiene una excelente oportunidad de detectar cualquier efecto cuando el cambio en la *reacted* sea 5 o más.

Opciones del Panel



- **Gráfico:** El tipo de gráfico a ser creado.
- **Gráfico contra:** Seleccione el factor experimental a ser mostrado en el gráfico, para los gráficos donde un factor es necesario.
- **Dirección:** Defina la orientación del grafico de probabilidad normal.
- **Estimar Línea:** Especificar si una línea debe estimarse para los datos sobre un gráfico de probabilidad normal.
- **Alfa:** Especificar el riesgo- $\alpha$  asociado con la curva de potencia.

**Optimización**

Teniendo la estimación de un modelo matemático, el analista puede ahora utilizar el modelo para determinar que combinación de factores puede mejorar el rendimiento de la respuesta. La opción de *Optimización* en el procedimiento *Analizar Diseño* calcula y despliega automáticamente la configuración óptima de los factores experimentales:

<b>Optimizar Respuesta</b>			
Meta: maximizar reacted			
Valor óptimo = 94.8355			
<i>Factor</i>	<i>Bajo</i>	<i>Alto</i>	<i>Óptimo</i>
feed rate	10.0	15.0	15.0
catalyst	1.0	2.0	2.0
agitation	100.0	120.0	100.0
temperature	140.0	180.0	180.0
concentration	3.0	6.0	3.0

La tabla muestra:

- **Meta** – El tipo de optimización deseada, definida sobre la caja de dialogo *Opciones del Panel* descrita más adelante.
- **Valor Óptimo** – La respuesta predicha en la configuración optima.

- **Bajo** – El nivel bajo sobre la región en la cual la optimización es desarrollada.
- **Alto** - El nivel alto sobre la región en la cual la optimización es desarrollada.
- **Optimo** – La configuración optima de los factores experimentales.

El analista define el tipo de optimización y la región sobre la cual se desarrolla usando *Opciones del Panel*.

En el ejemplo anterior, la *reacted* tiene que ser maximizada con respecto al *catalyst*, *temperature*, y *concentration*. La *Feed rate* y *agitation* son ignoradas, puesto que ellas no tienen un efecto en la respuesta del modelo estimado. El conjunto de condiciones óptimas pueden encontrarse anticipadamente usando el *Gráfico del Cubo*.

*Opciones del Panel*

Factor	Bajo	Alto	Inicio
feed rate	10.0	15.0	12.5
catalyst	1.0	2.0	1.5
agitation	100.0	120.0	110.0
temperature	140.0	180.0	160.0
concentration	3.0	6.0	4.5

- **Tipo de Optimización** – Si se maximiza la respuesta, minimiza la respuesta o se mantiene en un valor objetivo específico.
- **Puntos de Comienzo Adicionales** – Puntos adicionales en el cual se comienza la búsqueda numérica para las condiciones óptimas. El programa siempre comienza una búsqueda empezando en el centro de la región experimental. El analista puede modificarlo para búsquedas adicionales empezando en el *mejor punto del diseño* (punto del diseño con la respuesta predicha más alta), *todos los puntos del diseño*, *mejor vértice* (la combinación del

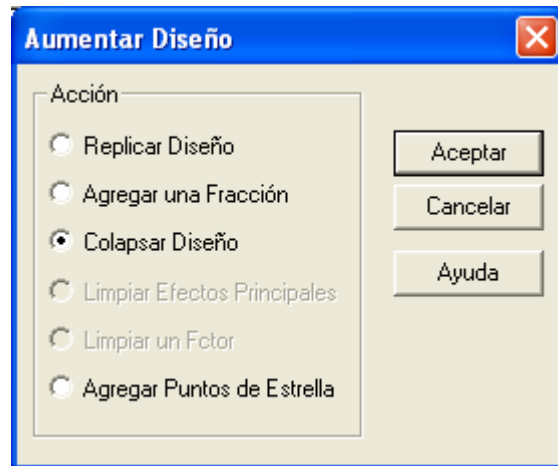
nivel bajo o alto de cada factor en la respuesta predicha más alta), o *todos los vértices*. Como con todos los procedimientos numéricos, el punto inicial tiene un efecto cuando un óptimo local o global es localizado. Esto es generalmente una buena idea para iniciar la búsqueda desde diferentes puntos de comienzo. El programa puede reportar el mejor resultado entre el conjunto de búsquedas desarrolladas.

- **Factor** – Límites inferior y superior sobre cada factor. El programa buscará sobre esta región para encontrar la configuración óptima. Configurar los valores del nivel alto y bajo igual al contraste de un factor igual al valor seleccionado.

## Ruta de Escalamiento Ascendente

Teniendo desarrollado un diseño de investigación y descubriendo 3 factores que tienen un alto impacto sobre la respuesta, el analista puede desear extrapolar el modelo estimado a través de la región experimental para determinar otras combinaciones de los factores donde podemos encontrar mejores. La Ruta de Escalamiento Ascendente genera puntos a lo largo de una curva extendiéndose desde el centro de la región experimental en la dirección sobre la cual las respuestas predichas se incrementan (o decrementan) más rápidamente para cambios pequeños en los factores experimentales. Para usar el procedimiento:

**Paso 1:** Si el modelo estimado no incluye los términos para uno a más de los factores experimentales en el diseño, estos factores pueden removerse usando la opción *Colapsar Diseño* dentro del menú *Aumentar Diseño* STATGRAPHICS. En el ejemplo actual, todos los efectos e interacciones que implican la *feed rate* y *agitation* se excluyen. Consecuentemente, el diseño puede colapsarse para remover esos dos factores. Para hacer esto, seleccione *Aumentar Diseño Existente* del menú *DDE* y seleccione *Colapsar Diseño* sobre la caja de diálogo que será desplegada:



Sobre la siguiente caja de diálogo, seleccione el factor a ser removido:



Repita el proceso para *agitation* y grabe el nuevo diseño bajo un nombre de archivo diferente.

El archivo de datos del ejemplo SCREEN2.SFX contiene el diseño colapsado, el cual implica solamente *catalyst*, *temperature*, and *concentration*. La ventana de *Atributos del Diseño* se muestra abajo:

**Atributos del Diseño de Cribado**

Clase de diseño: De Cribado  
 Nombre del Diseño: Factorial  $2^3$   
 Nombre del archivo: C:\DocData\chemical reaction.sfx  
 Comentario: Chemical reactor example

**Diseño Base**  
 Número de factores experimentales: 3  
 Número de bloques: 1  
 Número de respuestas: 1  
 Número de corridas: 19, incluyendo 3 puntos centrales por bloque  
 Grados de libertad para el error: 12  
 Aleatorizar: No

Factores	Bajo	Alto	Unidades	Continuo
catalyst	1.0	2.0	%	Sí
temperature	140.0	180.0	degrees	Sí
concentration	3.0	6.0	%	Sí

Respuestas	Unidades
reacted	%

Note que el nuevo diseño es un factorial completo  $2^3$ , con 2 replicas. Hay 12 grados de libertad disponibles para estimar el error experimental. Bajo la vista de la hoja de cálculo de los datos, el usuario notara que las columnas para la *feed rate* y *agitation* han sido eliminadas.

**Paso 2:** Un nuevo modelo debe estimarse puesto que el diseño original fue colapsado. Recuerde que para remover una interacción insignificante entre *catalyst* y *concentration*, ahora serán etiquetadas como AC.

**Paso 3:** Requiere la *Ruta de Escalamiento Ascendente*, la cual despliega la siguiente tabla similar a:

<b>Camino de Máximo Ascenso para reacted</b>			
<i>catalyst</i>	<i>temperature</i>	<i>concentration</i>	<i>Predicción para reacted</i>
(%)	(degrees)	(%)	(%)
1.5	160.0	4.5	65.2105
1.55	161.242	4.45332	66.7539
1.6	162.575	4.40465	68.4172
1.65	163.992	4.35394	70.2081
1.7	165.485	4.30116	72.1337
1.75	167.049	4.2463	74.2002
1.8	168.677	4.18938	76.4134
1.85	170.366	4.13043	78.7788
1.9	172.109	4.0695	81.3011
1.95	173.903	4.00662	83.9848
2.0	175.743	3.94187	86.834
2.05	177.626	3.8753	89.8524
2.1	179.548	3.80698	93.0437
2.15	181.507	3.73697	96.4109
2.2	183.5	3.66533	99.9571
2.25	185.525	3.59214	103.685

La tabla de la *Ruta de Escalamiento Ascendente* muestra varias combinaciones de los factores experimentales, comenzando en el centro de la región experimental. Una vez que los factores son cambiados en intervalos regulares, se especifican usando *Opciones del Panel*. Los otros factores son cambiados en la misma dirección para permanecer en la ruta que corresponde a los incrementos rápidos sobre la respuesta como un movimiento hacia otra parte del centro de la región experimental. También muestra la respuesta predicha en cada nuevo punto, extrapolando el modelo estimado.

En la tabla de arriba, *catalyst* fue cambiando con un incremento de 0.05. Como el catalizador se esta incrementando, lo ideal es que la *temperature* y *concentration* se decrementsen. La respuesta predicha constantemente se incrementa, regularmente arriba del 100%. Este es un buen ejemplo de los peligros de extrapolar demasiado afuera de la región en la cual los experimentos fueron desarrollados. Mientras el modelo estimado tiene un excelente ajuste sobre los datos observados y lo usamos para extrapolar distancias cortas fuera de la región experimental, eventualmente se romperá el flujo. Los experimentadores deben entender que los modelos estimados por este procedimiento son empíricos (basados en los datos), no modelos mecánicos basados en la teoría. Cualquier predicción hecha por el modelo debe ser verificada por experimentos confirmatorios adicionales.



## Opciones del Panel

Opciones Camino de Máximo Ascenso

Factor a Escalar

catalyst

temperature

concentration

Aceptar

Cancelar

Ayuda

Incrementar en:

Número de Incrementos:

Resolución:

- **Factor de Paso** – El factor que será cambiando con un incremento calculado por  $\Delta$  para generar los puntos a lo largo de la ruta. Como la selección del factor cambia, el programa puede determinar la configuración adecuada para cada uno de los demás factores estando sobre la *Ruta de Escalamiento Ascendente*.
- **Paso por** – La cantidad  $\Delta$  a cambiar sobre el factor seleccionado en cada paso hacia la ruta.
- **Numero de Pasos** – El número de pasos tomado hacia la ruta.
- **Resolución** – El número de veces que la pendiente es re-evaluada durante cada paso. Una resolución más alta, tiene más exactitud si el programa sigue la verdadera ruta.

## Grabar Resultados

Los siguientes resultados pueden grabarse sobre la base de datos:

1. *Valores Predichos* – El valor predicho de  $Y$  correspondiente en cada una de las  $n$  observaciones.
2. *Errores Estándar de Predicciones* – El error estándar para  $n$  valores predichos.
3. *Límites Inferiores de Predicciones* – El límite de Predicción Inferior para cada valor predicho.
4. *Límites Superiores de Predicciones* – El límite de Predicción Superior para cada valor predicho.
5. *Errores Estándar de las Medias* – El error estándar para el valor promedio de  $Y$  en cada uno de  $n$  valores de  $X$ .
6. *Límites Inferiores de Medias Pronosticadas* – El límite de confianza inferior para el valor promedio de  $Y$  en cada uno de  $n$  valores de  $X$ .
7. *Límites Superiores de Medias Pronosticadas* – El límite de confianza superior para el valor promedio de  $Y$  en cada uno de  $n$  valores de  $X$ .
8. *Residuos* – Los  $n$  residuales.
9. *Residuos Estudentizados* – Los  $n$  residuales estandarizados.
10. *Leverages* – Los valores ponderados correspondientes a  $n$  valores de  $X$ .
11. *Estadístico DFITS* – El valor del estadístico DFITS correspondientes a  $n$  valores de  $X$ .
12. *Distancias de Mahalanobis* – La distancia de Mahalanobis correspondientes a  $n$  valores de  $X$ .

### Cálculos

Detalles de los cálculos desarrollados pueden encontrarse en la documentación para el procedimiento de *Regresión Múltiple*.