

## ***Diseño de Experimentos – Diseño de Superficie de Respuesta***

### **Resumen**

Una vez que los factores importantes que afectan a un proceso han sido identificados, usualmente corridas adicionales son necesarias 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 por cada factor para poder estimar la curvatura. En STATGRAPHICS, diferentes tipos de diseños son populares para la optimización de la respuesta se han agrupado todos como *Diseños de Superficie de Respuesta*. Estos incluyen:

1. Diseño Central Compuesto
2. Diseño Factorial a Tres-Niveles
3. Diseño Box-Behnken
4. Diseño Draper-Lin

Este documento se concentra sobre la creación de los diseños de superficie de respuesta. Puesto que el análisis de los datos una vez recolectados se realiza en paralelo con los diseños para investigación, refiérase a la documentación de *DDE - Diseños para Investigación* para una discusión detallada de las salidas generadas de los procesos de análisis.

### **Ejemplo**

Myers y Montgomery (1995) describen un típico ejemplo de un experimento desarrollado para optimizar la resistencia de una envoltura de pan. El diseño implica los siguientes factores experimentales y el rango de valores sobre los cuales serán variados.

<b>Factor</b>	<b>Bajo</b>	<b>Alto</b>	<b>Unidades</b>
Temperatura de Lacre	225	285	°F
Temperatura de Conservación	46	64	°F
Polietileno	0.5	1.7	%

**Ejemplo StatFolio:** doe rsm.sgp

### **Creación del Diseño**

Los experimentos de superficie de respuesta son construidos seleccionando *Crear Diseño* del menú de diseño de experimentos y completando diversas cajas de dialogo.

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

La primera caja de dialogo se despliega durante la creación del tipo de diseño como se muestra abajo:

**Opciones de Creación de Diseños**

Clase de Diseño

- De Cribado
- Superficie de Respuesta
- Mezcla
- Factorial Multinivel
- Arreglos Interno/Externo
- Un Solo Factor Categórico
- Multi-Factor Categóricos
- Componentes de Varianza (jerárquicos)

No. de Variables de Respuesta:  
1

No. de Factores Experimentales:  
3

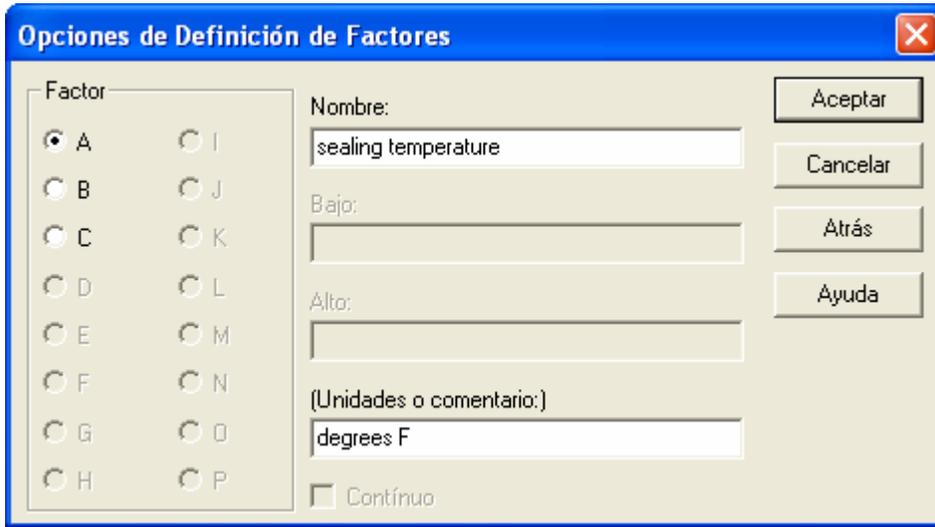
Comentario:  
Myers and Montgomery p. 300

Aceptar  
Cancelar  
Ayuda

- **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 8.
- **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:

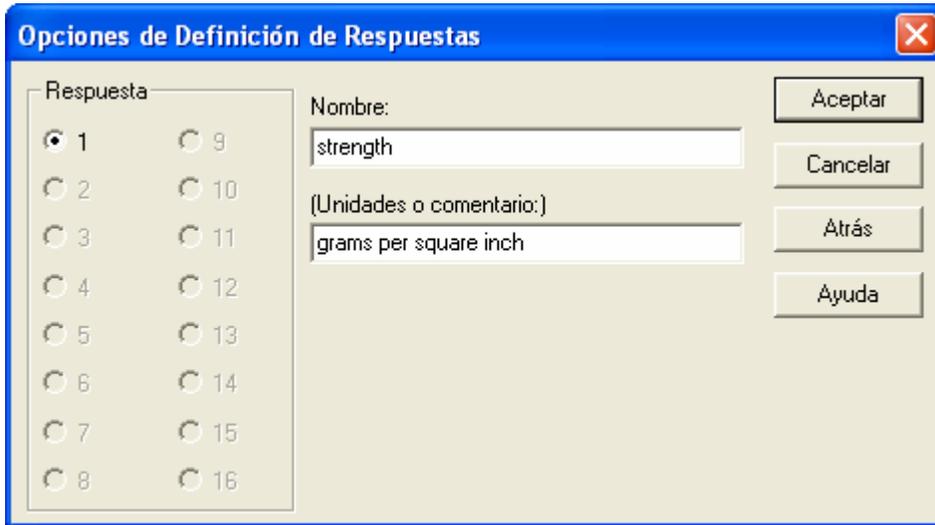
- **Nombre** – El nombre del factor es similar a un nombre de columna en una base de datos estándar de STATGRAPHICS, cada nombre debe tener hasta 32 caracteres y no contener cualquier carácter especial u operadores algebraicos.
- **Bajo** – El nivel bajo del factor.
- **Alto** – El nivel alto del factor.
- **Continuo** – Indica las variables que pueden variarse continuamente entre los niveles bajo y alto. Todas las variables en un diseño de superficie de respuesta deben ser continuas.
- **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 superficie de respuesta, las corridas experimentales se desarrollan en los niveles bajo y alto de cada factor, en el nivel central la mitad entre el nivel bajo y alto, y (para algunos diseños) en una distancia más pequeña arriba del nivel bajo y abajo del nivel alto. Para el ejemplo actual, 3 factores se definieron como:

Nombre	Bajo	Alto	Unidades
sealing temperature	225	285	°F
cooling bar temperature	46	64	°F
polyethylene	0.5	1.7	%

Caja de Dialogo #3 – Variable Respuesta

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

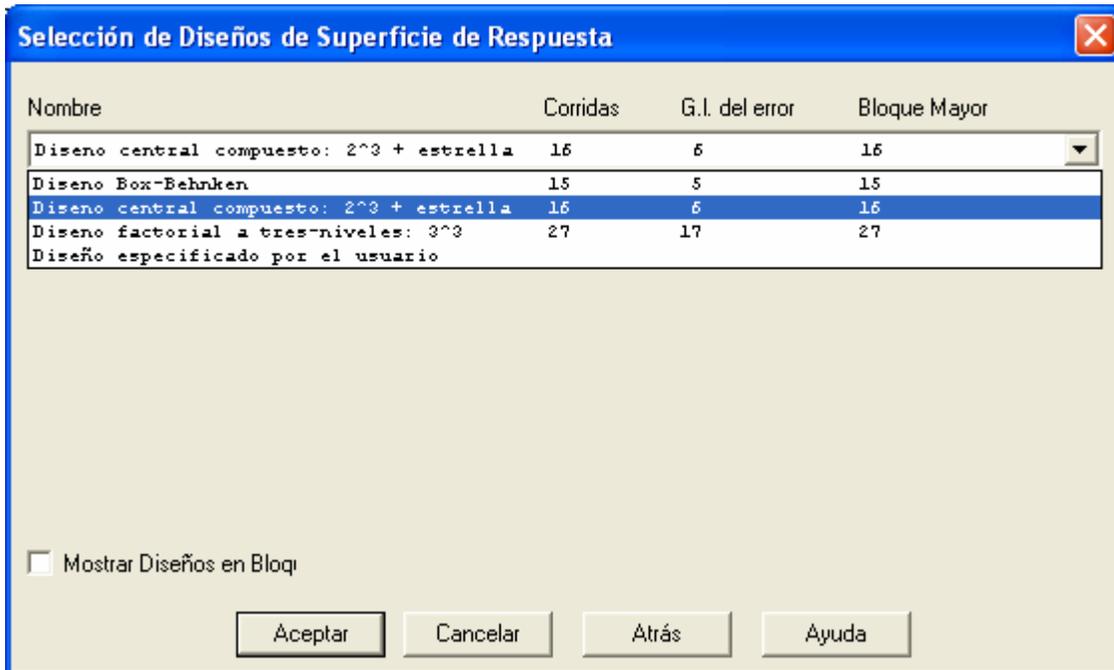


Clic sobre los números 1, 2, 3,..., solo una vez e ingrese la siguiente información para cada variable respuesta del experimento:

- **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 superficie de respuesta del catálogo en STATGRAPHICS para un número seleccionado de factores experimentales:



- **Nombre** – El nombre del diseño, incluyendo una abreviación como  $3^3$  si es relevante. Para diseños de superficie de respuesta, los siguientes tipos pueden aparecer en la lista, dependiendo del número de factores experimentales:
  1. *Diseño Central Compuesto* – Estos diseños consisten en un factorial completo con puntos al centro o un diseño factorial fraccionado con resolución V, más agregar puntos estrella usados para modelar la curvatura con respecto a cada factor.
  2. *Diseño Factorial a Tres-Niveles* – Estos diseños consisten de todas las combinaciones de 3 niveles de cada factor experimental.
  3. *Diseño Box-Behnken* – Diseños a 3 niveles incluyendo un subconjunto de corridas de un factorial completo a tres niveles.
  4. *Diseño Draper-Lin* – Un Diseño central compuesto más pequeño en el cual una porción central del diseño es menor que una resolución V.
- **Corridas** – El número de corridas en el diseño base, antes de sumar replicas o puntos al centro adicionales.
- **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.
- **Desplegar Diseños con Bloque** – Se Activa para incluir los diseños con más de un bloque.

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

La caja de dialogo final depende sobre el tipo de diseño seleccionado. Lo discutiremos por separado para cada tipo en las secciones siguientes:

### **Diseño Central Compuesto**

Un tipo popular de experimento para optimización empieza con un diseño factorial estándar a dos niveles y agrega dos corridas adicionales para cada factor. Las corridas adicionales, llamadas puntos estrella, son localizadas en distancias pequeñas abajo del nivel bajo de un factor y a la misma distancia arriba del nivel alto, esto permite la estimación de la curvatura con respecto a ese factor. Mientras cada factor se empieza variándose, los otros factores son fijados en sus valores centrales.

Por ejemplo, considere el diseño factorial  $2^3$  mostrado abajo:

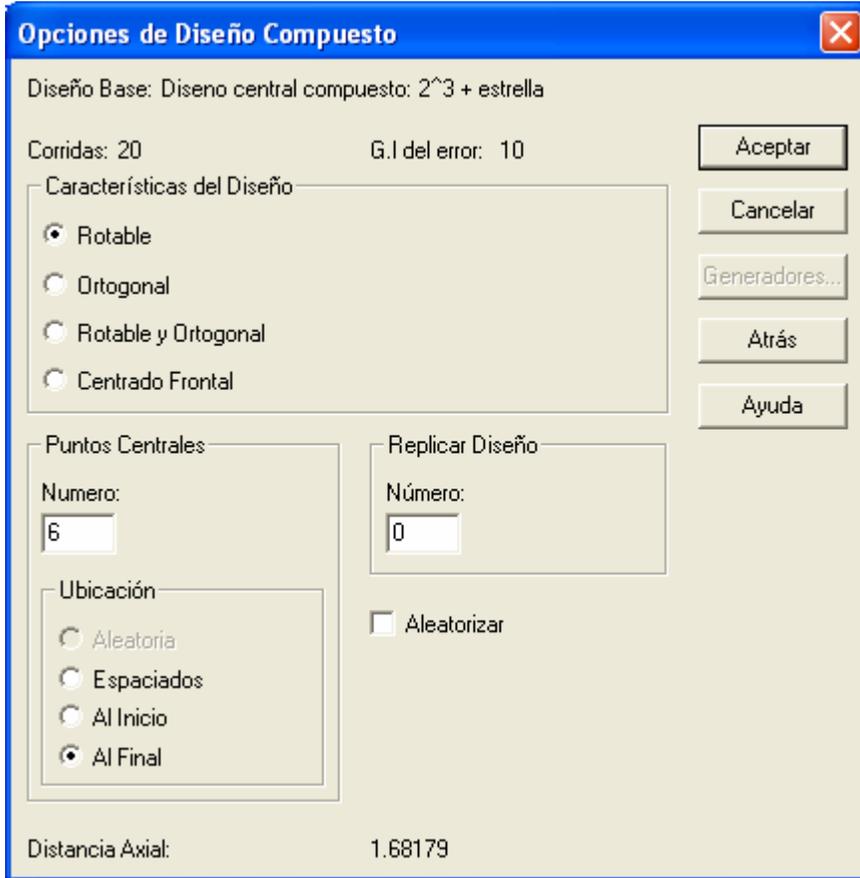
Corrida	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
1	-1.0	-1.0	-1.0
2	1.0	-1.0	-1.0
3	-1.0	1.0	-1.0
4	1.0	1.0	-1.0
5	-1.0	-1.0	1.0
6	1.0	-1.0	1.0
7	-1.0	1.0	1.0
8	1.0	1.0	1.0

El diseño es expresado en unidades estándar donde -1 representa el nivel bajo del factor y +1 representa el nivel alto. Los puntos estrella adicionales son localizados en:

Corrida	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
9	$-\alpha$	0	0
10	$\alpha$	0	0
11	0	$-\alpha$	0
12	0	$\alpha$	0
13	0	0	$-\alpha$
14	0	0	$\alpha$

donde  $\alpha$ , la *distancia axial*, es  $\geq 1$ . Determinando  $\alpha$  igual a valores específicos, las propiedades deseables pueden alcanzarse. Por lo menos un punto al centro también se debe agregar al diseño.

En la construcción de un diseño central compuesto, la quinta caja de dialogo tiene la siguiente forma:



- **Características del Diseño:** Especifica las propiedades del diseño de las siguientes opciones:

1. **Rotable** – Coloca las puntos estrella en una distancia haciendo que la varianza de la respuesta predicha sea la misma para todos los puntos lo cual es la misma distancia del centro del diseño (en unidades estándar). Esto es intuitivamente razonable, puesto que el experimentador típicamente no conoce la cual es la dirección probable de la localización del punto óptimo. La Rotabilidad es alcanzada fijando la distancia axial en:

$$\alpha = \sqrt[4]{F} \tag{1}$$

donde F es el número de corridas en la porción del factorial del diseño (no contando los puntos al centro). Cualquier número de puntos al centro pueden agregarse (3-5 puntos al centro son usualmente deseables), lo cual se pueden colocar en un bloque por separado si se desea. Para un factorial  $2^3$ , los puntos estrellas son colocados en  $\alpha = 1.68179$ .

2. **Ortogonal** – Coloca los puntos estrella en una distancia axial que asegura que todos los términos de segundo orden son ortogonales de cualquier otro. La propiedad de la distancia axial depende sobre el número de puntos al centro  $n_c$  agregados al diseño:

$$\alpha = \sqrt[4]{\left(\sqrt{2k + n_c + F} - \sqrt{F}\right)^2 F / 4} \tag{2}$$

Para un factorial  $2^3$  con dos puntos al centro por bloque axial, los puntos estrella con colocados en  $\alpha = 1.28719$ .

3. **Rotable y Ortogonal** – Alcanza ambas propiedades por la selección apropiada del número de puntos al centro. Esto se logra fijando la distancia axial igual a la requerida para alcanzar la rotabilidad, y entonces seleccionando un valor para  $n_c$  hace que la distancia axial para la ortogonalidad sea la misma que para la rotabilidad. Para un factorial  $2^3$ , los puntos estrella son colocados en  $\alpha = 1.68179$  y 9 puntos al centro son agregados el diseño.
  4. **Cara Centrada** - Coloca los puntos estrella en los niveles bajo y alto del diseño factorial, e.j., fija  $\alpha = 1.0$ . aunque es menos deseable desde una perspectiva estadística, puede ser más fácil de ejecutar puesto implica solamente los niveles bajo, medio y alto de cada factor. Para tales diseños, 1-2 puntos al centro son usualmente suficientes.
- **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.
  - **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 demás 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.

Para el ejemplo, un diseño central rotatable con 6 puntos al centro fue seleccionado. Este rindió un total de 20 corridas, como se despliega en la hoja de trabajo de abajo:

<b>Hoja de trabajo para C:\DocData\breadwrapper.sfx - Myers and Montgomery p. 300</b>				
<i>corrida</i>	<i>sealing temperature</i> <i>(degrees F)</i>	<i>cooling bar temperature</i> <i>(degrees F)</i>	<i>polyethylene</i> <i>(%)</i>	<i>strength</i> <i>(grams per square inch)</i>
1	225.0	46.0	0.5	
2	285.0	46.0	0.5	
3	225.0	64.0	0.5	
4	285.0	64.0	0.5	
5	225.0	46.0	1.7	
6	285.0	46.0	1.7	
7	225.0	64.0	1.7	
8	285.0	64.0	1.7	
9	204.546	55.0	1.1	
10	305.454	55.0	1.1	
11	255.0	39.8639	1.1	
12	255.0	70.1361	1.1	
13	255.0	55.0	0.0909243	
14	255.0	55.0	2.10908	
15	255.0	55.0	1.1	
16	255.0	55.0	1.1	
17	255.0	55.0	1.1	
18	255.0	55.0	1.1	
19	255.0	55.0	1.1	
20	255.0	55.0	1.1	

El diseño consiste de:

- Corridas 1-8: Un factorial 2<sup>3</sup>.
- Corridas 9-14: Los puntos estrella.
- Corridas 15-20: Los puntos al centro replicados.

Cuando STATGRAPHICS determina el diseño, los niveles de los puntos estrella son colocados en la distancia exacta para alcanzar la Rotabilidad. Estos niveles pueden redondearse a mano para tener valores más razonables cambiándolo directamente en la base de datos.

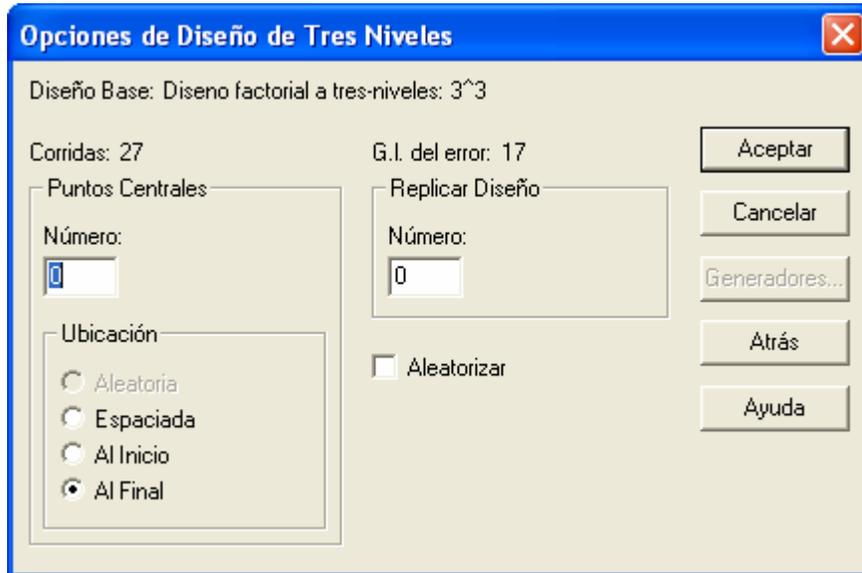
## Diseños Factoriales con Tres-Niveles

El diseño central compuesto implica cinco niveles de cada factor experimental. Diseños más simples pueden realizarse usando solamente tres niveles de por cada factor, un nivel bajo (-), nivel central (0), y un nivel alto (+). Los diseños factoriales a tres niveles generan corridas en cada combinación de estos tres niveles. Con 3 factores experimentales, obtenemos 27 corridas, como se presentan en la hoja de trabajo de abajo:

Corrida	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
1	-1	-1	-1
2	0	-1	-1
3	1	-1	-1
4	-1	0	-1
5	0	0	-1
6	1	0	-1
7	-1	1	-1
8	0	1	-1
9	1	1	-1
10	-1	-1	0
11	0	-1	0
12	1	-1	0
13	-1	0	0
14	0	0	0
15	1	0	0
16	-1	1	0
17	0	1	0
18	1	1	0
19	-1	-1	1
20	0	-1	1
21	1	-1	1
22	-1	0	1
23	0	0	1
24	1	0	1
25	-1	1	1
26	0	1	1
27	1	1	1

STATGRAPHICS incluye diseños factoriales a tres niveles para 2, 3, y 4 factores. En cada caso, sin embargo, el diseño central compuesto es igual o menor en tamaño que los factoriales a tres niveles. Usando puntos estrella con cara centrada, el diseño central compuesto puede transformarse en 3 niveles, bajo, medio y alto. A menos que el número de corridas no sea una preocupación, el diseño factorial a 3 niveles usualmente no es la mejor opción.

Si un diseño factorial a tres niveles es seleccionado, la quinta caja de dialogo tomara la siguiente forma:



- **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.
- **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 demás corridas experimentales, *espaciados* regularmente a través de las corridas, o localizarse *al inicio* o *al 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.

## Diseño Box-Behnken

Un tipo de diseño que presenta ser una alternativa razonable contra el central compuesto fue inventado por Box y Behnken para situaciones cuando el experimentador desea correr solamente 3 niveles de los factores pero se quiere que el diseño también sea rotatable. Mientras el diseño central compuesto con puntos al centro de cara centrada también tiene 3 niveles, este coloca la mayoría de los puntos del diseño en las esquinas de la región experimental. Si esa región define las condiciones viables para el proceso, el central compuesto con cara centrada es una opción natural. Sin embargo, si uno está comenzando en una combinación particular de factores y se está buscando simplemente las mejores condiciones, un arreglo más esférico de los puntos del diseño sería más eficiente.

El diseño Box-Behnken consiste de un conjunto de corridas donde cada par de factores es variado entre sus niveles bajo y alto, mientras los demás factores experimentales se fijan en el

nivel medio. Por ejemplo, un diseño Box-Behnken con 3 factores (con puntos al centro omitidos) se muestra abajo:

Corrida	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
1	-1	-1	0
2	1	-1	0
3	-1	1	0
4	1	1	0
5	-1	0	-1
6	1	0	-1
7	-1	0	1
8	1	0	1
9	0	-1	-1
10	0	1	-1
11	0	-1	1
12	0	1	1

Cada conjunto de 4 corridas consecutivas es un diseño experimental con 2 niveles implicando seleccionar un par de factores. Además, de 3-5 puntos al centro en (0,0,0) pueden agregarse al diseño. El diseño es comparable en tamaño al central compuesto y considerablemente más pequeño que un factorial a tres niveles. Para un diseño Box-Behnken, la quinta caja de dialogo es la misma que para un diseño factorial a tres niveles.

### Diseño Draper-Lin

Cuando los costos de experimentación son altos, el experimentador desearía reducir el número de corridas lo más pequeño posible. El diseño Draper-Lin del catalogo de diseño en STATGRAPHICS es un diseño compuesto reducido que consiste en un factorial fraccionado o Plackett-Burman con menor resolución V, mas unos puntos estrella. Por ejemplo, un diseño Draper-Lin para 4 factores se muestra abajo (omitiendo puntos al centro):

Corrida	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1	1	1	1	-1
2	1	1	-1	-1
3	1	-1	1	1
4	-1	1	-1	1
5	1	-1	-1	1
6	-1	-1	1	-1
7	-1	1	1	1
8	-1	-1	-1	-1
9	-α	0	0	0
10	α	0	0	0
11	0	-α	0	0
12	0	α	0	0
13	0	0	-α	0
14	0	0	α	0
15	0	0	0	-α
16	0	0	0	α

Las primeras 8 corridas abarcan una media fracción de un factorial  $2^4$ , tiene resolución IV. El segundo conjunto de 8 corridas son los puntos estrella, donde  $\alpha$  es la distancia axial como se definió para los diseños central compuesto. Diversos puntos al centro también son agregados. El diseño resultante es capaz de estimar el modelo de Segundo orden completo, aunque existirá cierta correlación entre los coeficientes del modelo estimado. Para los casos en el cual el costo de cada corrida es muy grande, sin embargo, el ahorro de 8 corridas sobre un diseño central compuesto complete puede ser significativo.

Para un diseño Draper-Lin, la quinta caja de dialogo es la misma que para el diseño central compuesto.

### 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:

Hoja de trabajo para C:\DocData\breadwrapper.sfx - Myers and Montgomery p. 300				
corrida	sealing temperature (degrees F)	cooling bar temperature (degrees F)	polyethylene (%)	strength (grams per square inch)
1	225.0	46.0	0.5	6.6
2	285.0	46.0	0.5	6.9
3	225.0	64.0	0.5	7.9
4	285.0	64.0	0.5	6.1
5	225.0	46.0	1.7	9.2
6	285.0	46.0	1.7	6.8
7	225.0	64.0	1.7	10.4
8	285.0	64.0	1.7	7.3
9	204.5	55.0	1.1	9.8
10	305.5	55.0	1.1	5.0
11	255.0	39.9	1.1	6.9
12	255.0	70.1	1.1	6.3
13	255.0	55.0	0.09	4.0
14	255.0	55.0	2.11	8.6
15	255.0	55.0	1.1	10.1
16	255.0	55.0	1.1	9.9
17	255.0	55.0	1.1	12.2
18	255.0	55.0	1.1	9.7
19	255.0	55.0	1.1	9.7
20	255.0	55.0	1.1	9.6

Note que la localización de los puntos estrellas pueden redondearse.

Para analizar el resultado, el menú DDE contiene un selección nombrada Analizar Diseño. Esta brinca a una caja de dialogo desplegando todas las columnas respuestas en el archive 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 del diseño de superficie de respuesta expresa la variable respuesta  $Y$  como una función lineal de los factores experimentales, interacciones entre los factores, términos cuadráticos y término del error. Existen dos tipos de modelos que generalmente se estiman, se ilustran abajo para 3 factores experimentales:

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

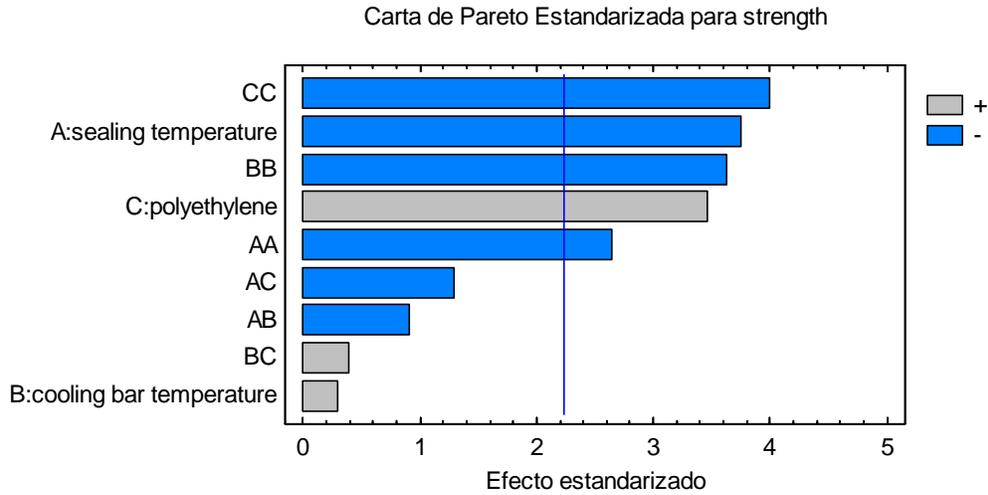
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad (3)$$

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

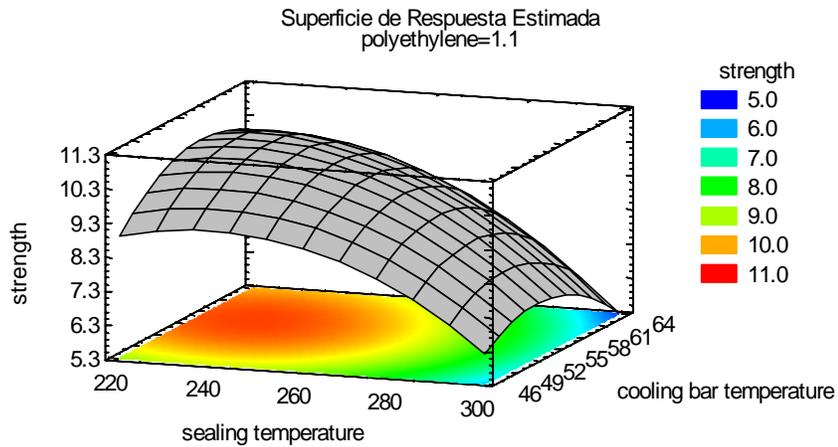
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \varepsilon \quad (4)$$

El error experimental  $\varepsilon$  se asume que fue aleatorio y que proviene de una distribución normal con media 0 y desviación estándar igual a  $\sigma$ .

La única diferencia en los análisis entre diseños de superficie de respuesta y diseños para investigación es la inclusión de los efectos cuadráticos. Estos aparecen en el Gráfico de Pareto y destacan otros efectos etiquetados como AA, BB, y CC:



El efecto cuadrático causa la estimación de la superficie de respuesta al exhibir la curvatura, la cual puede parecerse a una colina, un valle, una cresta, o alguna otra figura:



Para detalles sobre las diferentes tablas y gráficos disponibles en *Analizar Diseño*, ver la documentación para *DDE – Diseños para Investigación*.