

CAPITULO 2

METODOLOGIA DE SUPERFICIES DE RESPUESTA

En este capítulo hablaremos de qué es la Metodología de Superficies de Respuesta, su representación gráfica, el procedimiento a seguir hasta encontrar un óptimo y los diseños experimentales que pueden utilizar.

Para el desarrollo del capítulo fueron de gran utilidad Cornell [12] (1990) y Montgomery [6] (1991), de los cuales se tomó la teoría y fórmulas que se presentan a continuación.

2.1 Definición.

La Metodología de Superficies de Respuesta es un conjunto de técnicas matemáticas y estadísticas utilizadas para modelar y analizar problemas en los que una variable de interés es influenciada por otras. El objetivo es optimizar la variable de interés. Esto se logra al determinar las condiciones óptimas de operación del sistema.

2.2 Terminología.

A continuación se presenta la terminología que se utilizará a lo largo del capítulo.

2.2.1 Factores.

Son las condiciones del proceso que influyen la variable de respuesta. Estos pueden ser cuantitativos o cualitativos.

2.2.2 Respuesta.

Es una cantidad medible cuyo valor se ve afectado al cambiar los niveles de los factores. El interés principal es optimizar dicho valor.

2.2.3 Función de respuesta.

Al decir que un valor de respuesta Y depende de los niveles x_1, x_2, \dots, x_k de k factores, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, estamos diciendo que existe una función matemática de x_1, x_2, \dots, x_k cuyo valor para una combinación dada de los niveles de los factores corresponde a Y , esto es $Y=f(x_1, x_2, \dots, x_k)$.

2.2.4 Función de respuesta predicha.

La función de respuesta se puede representar con una ecuación polinomial. El éxito en una investigación de una superficie de respuesta depende de que la respuesta se pueda ajustar a un polinomio de primer o segundo grado.

Supongamos que la función de respuesta para los niveles de dos factores se puede expresar utilizando un polinomio de primer grado:

$$Y = b_0 + b_1x_1 + b_2x_2$$

donde b_0, b_1, b_2 son los coeficientes de regresión a estimar, x_1 y x_2 representan los niveles de \mathbf{x}_1 y \mathbf{x}_2 respectivamente. Suponiendo que se recolectan N^3 valores de respuesta (Y), con los estimadores b_0, b_1 y b_2 se obtienen \hat{b}_0, \hat{b}_1 y \hat{b}_2 respectivamente. Al remplazar los coeficientes de regresión por sus estimadores obtenemos:

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2$$

donde \hat{Y} denota el valor estimado de Y dado por x_1 y x_2 .

2.2.5 Superficie de respuesta

La relación $Y=f(x_1, x_2, \dots, x_k)$ entre Y y los niveles de los k factores x_1, x_2, \dots, x_k representa una superficie. Con k factores la superficie está en $k+1$ dimensiones. Por ejemplo cuando se tiene $Y=f(x_1)$ la superficie esta en dos dimensiones como se muestra en la figura 2.1 (Cornell [12] (1990)), mientras que si tenemos $Y=f(x_1, x_2)$ la superficie está en tres dimensiones, esto se observa en la figura 2.2 (Cornell [12] (1990)).

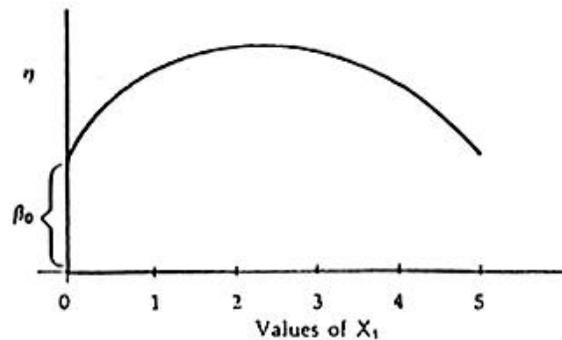


Figura 2.1 Superficie de respuesta en dos dimensiones

2.2.6 Gráfica de contornos.

La gráfica de contornos facilita la visualización de la forma de una superficie de respuesta en tres dimensiones. En ésta las curvas de los valores iguales de respuesta se grafican en un plano donde los ejes coordenados representan los niveles de los factores. Cada curva representa un valor específico de la altura de la superficie, es decir un valor

específico de \hat{Y} . Esto se muestra en la figura 2.3 (Cornell [12] 1990). Esta gráfica nos ayuda a enfocar nuestra atención en los niveles de los factores a los cuales ocurre un cambio en la altura de la superficie.

2.2.7 Región experimental.

La región experimental especifica la región de valores para los niveles de los factores. Esto se puede hacer empleando los niveles actuales de operación para cada factor; si se desea explorar el vecindario se incrementa y decrementa el valor del nivel en una cantidad determinada.

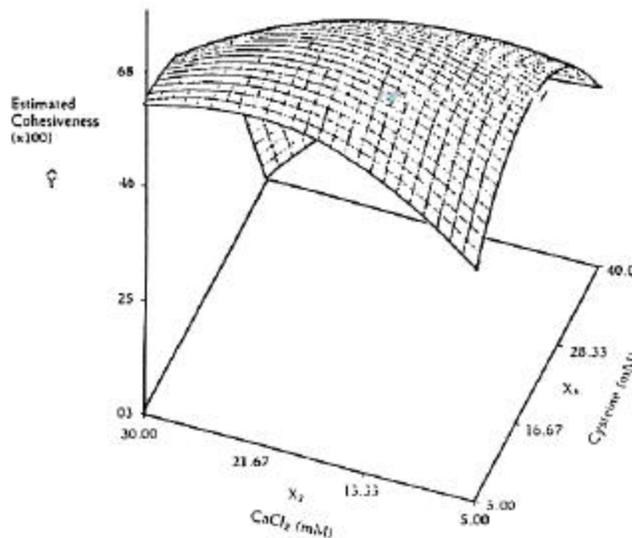


Figura 2.2 Superficie de respuesta tridimensional

2.3 Polinomio de primer orden.

Generalmente se desconoce la relación entre la respuesta y las variables independientes, por ello requerimos un modelo que aproxime la relación funcional entre Y y las variables independientes. Este modelo provee las bases para un nuevo experimento

que nos lleva hacia un nuevo modelo y el ciclo se repite. Si la respuesta se describe adecuadamente por una función lineal de las variables independientes se utiliza el modelo de primer orden (Cornell [12] (1990)):

$$Y = \mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{b}_2x_2 + \dots + \mathbf{b}_kx_k + \epsilon$$

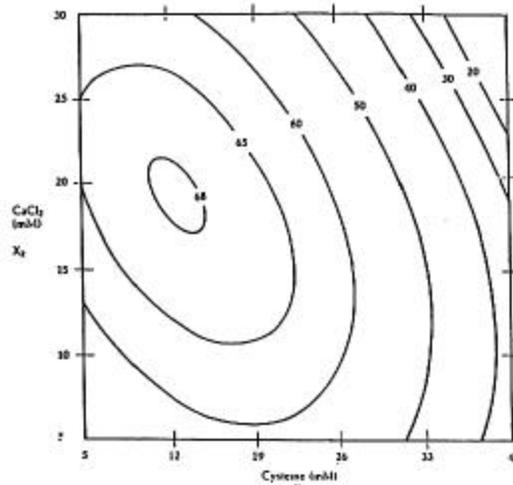


Figura 2.3 Gráfica de contornos

Los parámetros del modelo se estiman mediante el método de mínimos cuadrados. Una vez que se tienen los estimadores se sustituyen en la ecuación y obtenemos el modelo ajustado (Cornell [12] (1990)):

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

Este modelo se utiliza cuando queremos estudiar el comportamiento de la variable de respuesta únicamente en la región y cuando no conocemos la forma de la superficie.

2.3.1 Prueba de la significancia de los coeficientes estimados en el modelo ajustado.

De acuerdo a Cornell [12] (1990), para estimar los coeficientes se requieren $N^{\mathfrak{B}} k+1$ valores de respuesta (Y). El análisis de los datos de las corridas se presenta en una tabla de análisis de varianza. La tabla presenta las diferentes fuentes de variación que contribuyen a la variación total de los datos.

La variación total recibe el nombre de suma de cuadrados total SST, se calcula de la siguiente manera (Cornell [12] (1990)):

$$SST = \sum_{u=1}^N (Y_u - \bar{Y})^2$$

donde Y_u es el valor observado en la u -ésima corrida.

La suma de cuadrados se compone por la suma de cuadrados debido a la regresión y la suma de cuadrados no tomada en cuenta por el modelo ajustado. La fórmula de la suma de cuadrados debido a la regresión es (Cornell [12] (1990)):

$$SSR = \sum_{u=1}^N (\hat{Y}_u - \bar{Y})^2$$

La suma de cuadrados residual, que corresponde a la no tomada en cuenta, se calcula de la siguiente forma (Cornell [12] (1990)):

$$SSE = \sum_{u=1}^N (Y_u - \hat{Y}_u)^2$$

En la Tabla 2.1 se observa una tabla de análisis de varianza (Cornell [12] (1990)), en ella p representa el número de términos del modelo ajustado.

Tabla 2.1.- Análisis de Varianza.

Fuente	Grados de libertad	Suma de cuadrados	Media de Cuadrados
Regresión	p-1	SSR	SSR/p-1
Residuo	N-p	SSE	SSE/N-p
Total	N-1	SST	

La prueba de significancia de la ecuación de regresión ajustada tiene la siguiente hipótesis nula H_0 : Todas las \mathbf{b}_s (excluyendo \mathbf{b}_0) son cero contra la alternativa H_A : Al menos una de las \mathbf{b}_s (excluyendo \mathbf{b}_0) es diferente de cero. La prueba supone que el error se comporta normalmente, en ésta se utiliza el estadístico de prueba F , el cuál se calcula (Cornell [12] (1990)):

$$F = \frac{SSR/(p-1)}{SSE/(N-p)}$$

Este se compara con una $F_{(p-1, N-p)}$, si F calculada excede este valor la hipótesis nula se rechaza con un nivel de confianza de α . Esto significa que la variación explicada por el modelo es significativamente mayor que la variación inexplicable.

Además de esta prueba se puede hacer un análisis del ajuste del modelo con la R^2 , que es la proporción total de la variación de las $Y_{i,s}$ con respecto a la media que se puede explicar con la ecuación de regresión ajustada. Esta se calcula de la siguiente manera (Cornell [12] (1990)):

$$R^2 = \frac{SSR}{SST}$$

2.3.2 Prueba de falta de ajuste.

La falta de ajuste se presenta por la no planaridad o la curvatura de la superficie de respuesta, ésta no se detecta debido a la exclusión de los términos cuadráticos (o cúbicos) como son $b_{ii}x_i^2$ ($b_{iii}x_i^3$) o de los términos de producto cruzado ($b_{ijk}x_ix_jx_k$) que se refieren al efecto de la interacción entre los factores.

La prueba de falta de ajuste requiere que el diseño del experimento satisfaga (Cornell [12] (1990)):

1. El número de los distintos puntos del diseño, n , debe exceder el número de términos en el modelo ajustado, es decir $n > k+1$, y
2. Al menos 2 réplicas deben recolectarse en uno o más puntos del diseño para estimar la varianza del error.

Además, los valores del error aleatorio (\hat{I}_u) deben asumir una distribución normal e independiente con una varianza común S^2 .

Al cumplirse las condiciones 1 y 2 la suma de cuadrados residual se compone de dos fuentes de variación. La primera es la falta de ajuste del modelo ajustado (debido a la exclusión de términos de mayor orden) y la segunda es la variación del error puro. Para calcularlas necesitamos la suma de cuadrados calculada de las réplicas que recibe el nombre de error puro de la suma de cuadrados y sustraer de la suma de cuadrados residual éste para obtener la suma de cuadrados de la falta de ajuste. Es decir (Cornell [12] (1990)):

$$SSE_{ErrorPuro} = \sum_{l=1}^n \sum_{u=1}^{r_l} (Y_{lu} - \bar{Y}_l)^2$$

donde Y_{lu} es la u -ésima observación del l -ésimo punto del diseño.

$$u = 1, 2, \dots, r_l$$

$$l = 1, 2, \dots, n$$

\bar{Y}_l es el promedio de las r_l observaciones del l -ésimo punto del diseño.

$$SS_{Falta\ de\ Ajuste} = SSE - SS_{Error\ Puro}$$

$$SS_{Falta\ de\ Ajuste} = \sum_{l=1}^n r_l (\hat{Y}_l - \bar{Y}_l)^2$$

donde \hat{Y}_l es el valor predicho de la respuesta en el l -ésimo punto del diseño.

La prueba de adecuación del modelo ajustado es (Cornell [12] (1990)):

$$F = \frac{SS_{Falta\ de\ Ajuste} / (n - p)}{SSE_{ErrorPuro} / (N - n)}$$

La hipótesis de suficiencia de ajuste con un nivel α de significancia se rechaza cuando el valor calculado del estadístico es mayor a $F_{(n-p, N-n, \alpha)}$. Cuando la F calculada no es mayor el cuadrado medio residual es utilizado para estimar σ^2 y también se usa para probar la significancia del modelo ajustado.

Cuando la hipótesis de suficiencia de ajuste se rechaza, se debe de elevar el grado del modelo aumentando términos de producto cruzado y/o términos de mayor grado en x_1, x_2, \dots, x_k (Cornell [12] (1990)). Si se requieren puntos adicionales para estimar todos los coeficientes éstos se añaden. Se colectan los datos y se vuelve a hacer el análisis.

Si no se rechaza la hipótesis podemos inferir que la superficie es plana. Una vez que se tiene la ecuación y se ha probado el ajuste se buscan niveles que mejoren los valores de respuesta.

2.4 Método de máxima pendiente en ascenso.

Frecuentemente la estimación inicial de las condiciones de operación óptimas está alejada del óptimo real, en este caso se desea moverse rápidamente a la vecindad del óptimo. El método de máxima pendiente en ascenso es un procedimiento para recorrer secuencialmente la trayectoria de la máxima pendiente, que nos lleva en dirección del máximo aumento de la respuesta. Cuando se desea la minimización se habla de mínima pendiente en descenso.

De acuerdo a Montgomery [6] (1991), la dirección de ascenso máximo es en la que \hat{Y} aumenta más rápido, ésta es paralela a la normal de la superficie de respuesta ajustada. Los incrementos a lo largo de la trayectoria son proporcionales a los coeficientes de regresión b_1, b_2, b_3, \dots

Los experimentos se llevan a cabo hasta que deje de observarse un incremento en la respuesta, entonces se ajusta un nuevo modelo de primer orden con el que se determina una

nueva trayectoria y se continua con el procedimiento. Finalmente, se consigue llegar a la cercanía del óptimo, esto ocurre cuando existe falta de ajuste del modelo de primer orden.

2.4.1 Algoritmo para determinar las coordenadas de un punto en la trayectoria de máxima pendiente en ascenso

Un algoritmo propuesto por Montgomery [6] (1991) es el siguiente:

Supóngase que el punto $x_1 = x_2 = \dots = x_k = 0$

1. Se elige un tamaño de incremento o “escalón” en una de las variables del proceso, digamos Δx_j , usualmente se elige la variable de la que más se sabe, o la que tiene el mayor coeficiente de regresión absoluto $|b_j|$
2. El tamaño de incremento en las otras variables es

$$\Delta x_i = \frac{\hat{b}_i}{\hat{b}_j / \Delta x_j}$$

donde $i = 1, 2, \dots, k \quad i \neq j$

3. Se convierte Δx_i de variables codificadas a variables naturales.

2.5 Ejemplo del método de máxima pendiente en ascenso y pruebas de la falta de ajuste del modelo.

Para ejemplificar el proceso así como el uso de variables codificadas y la adecuación del modelo se emplea el siguiente problema tomado de Montgomery [6] (1991):

Un ingeniero químico desea determinar las condiciones de operación que maximicen el rendimiento de una reacción. Dos variables controlables influyen en éste: tiempo y temperatura de reacción. Actualmente el proceso opera con un tiempo de reacción de 35 minutos y una temperatura de 155° F, lo que produce un rendimiento del 40%. El

ingeniero decide que la región de exploración sea (30,40) minutos de reacción y (150,160)°F. Para simplificar los cálculos las variables independientes se codifican en un intervalo (-1,1). Las variables codificadas son:

$$x_1 = (\xi_1 - 35)/5 \quad y \quad x_2 = (\xi_2 - 155)/5 \quad \text{donde } \xi_1 = \text{variable natural tiempo}$$

$$\xi_2 = \text{variable natural temperatura}$$

En la Tabla 2.2 se muestran los datos, se utiliza un diseño factorial 2^2 aumentado en cinco puntos centrales. Las observaciones centrales sirven para estimar el error experimental y permiten probar la adecuación del modelo de primer orden.

Tabla 2.2 Datos del proceso para ajustar a un modelo de primer orden

<u>Variables naturales</u>		<u>Variables Codificadas</u>		<u>Respuesta</u>
ξ_1	ξ_2	x_1	x_2	Y
30	40	-1	-1	39.3
30	160	-1	1	40.0
40	150	1	-1	40.9
40	160	1	1	41.5
35	155	0	0	40.3
35	155	0	0	40.5
35	155	0	0	40.7
35	155	0	0	40.2
35	155	0	0	40.6

Con el método de mínimos cuadrados se obtiene:

$$\hat{Y} = 40.44 + 0.775x_1 + 0.325x_2$$

En la Tabla 2.3 se muestra el análisis de varianza, donde se observa que la F de regresión global es significativa al 1%.

Tabla 2.3 Análisis de Varianza para el modelo de primer orden

Fuente de variación	Suma de Cuadrados	Grados Libres	Media de Cuadrados	F ₀
Regresión (β_1, β_2)	2.8250	2	1.4125	47.83
Residuo	0.1772	6		
(Falta de Ajuste)	(0.0052)	2	0.0026	0.06
(Error puro)	(0.1720)	4	0.0430	
Total	3.0022	8		

El modelo indica que hay que desplazarse 0.775 unidades en dirección de x_1 por cada 0.325 unidades en dirección de x_2 . Sabemos que la trayectoria pasa por el punto ($x_1=0, x_2=0$) y tiene pendiente $0.0325/0.775$. En el ejemplo se decide usar 5 minutos como incremento en el tiempo de reacción lo que es equivalente a la variable codificada $\Delta x_1=1$. Los incrementos a lo largo de la trayectoria son: $\Delta x_1=1$ y $\Delta x_2=(0.325/0.775)=0.42$

El ingeniero calcula puntos a lo largo de esta trayectoria y observa el rendimiento en cada punto hasta notar un decremento en la respuesta. Los resultados aparecen en la Tabla 2.4 Los incrementos se muestran tanto para las variables codificadas como para las naturales, esto es porque las codificadas son más fáciles de manejar matemáticamente y las naturales son las que utilizamos para llevar a cabo el proceso.

Tabla 2.4 Experimento de máximo ascenso

Incrementos	Variables Codificadas		Variables Naturales		Respuesta Y
	X1	X2	ξ_1	ξ_2	
Origen	0	0	35	155	
Δ	1.00	0.42	5	2	
Origen + Δ	1.00	0.42	40	157	41.00
Origen + 2Δ	2.00	0.84	45	159	42.90
Origen + 3Δ	3.00	1.26	50	161	47.10
Origen + 4Δ	4.00	1.68	55	163	49.70

Origen + 5Δ	5.00	2.10	60	165	53.80
Origen + 6Δ	6.00	2.52	65	167	59.90
Origen + 7Δ	7.00	2.94	70	169	65.00
Origen + 8Δ	8.00	3.36	75	171	70.40
Origen + 9Δ	9.00	3.78	80	173	77.60
Origen + 10Δ	10.00	4.23	85	175	80.30
Origen + 11Δ	11.00	4.62	90	177	76.20
Origen + 12Δ	12.00	5.04	95	179	75.10

Se observa un aumento en la respuesta hasta el décimo incremento, a partir del undécimo se produce un decremento en el rendimiento. Por lo tanto se debe ajustar otro modelo de primer orden en la cercanía del punto ($\xi_1=85, \xi_2=175$).

Se ajusta un nuevo modelo de primer orden alrededor del punto ($\xi_1=85, \xi_2=175$). La región de exploración para ξ_1 es (80,90) y para ξ_2 es (170,180). Por lo tanto las variables codificadas son:

$$x_1 = (\xi_1 - 85) / 5 \quad \text{y} \quad x_2 = (\xi_2 - 175) / 5$$

Nuevamente se utiliza un diseño 2^2 con cinco puntos centrales. Los datos se muestran en la tabla 2.5

Tabla 2.5 Datos para el segundo modelo de primer orden

<u>Variables naturales</u>		<u>Variables Codificadas</u>		<u>Respuesta</u>
ξ_1	ξ_2	x_1	x_2	y
80	170	-1	-1	76.5
80	180	-1	1	77.0
90	170	1	-1	78.0
90	180	1	1	79.5
85	175	0	0	79.9
85	175	0	0	80.3
85	175	0	0	80.0
85	175	0	0	79.7
85	175	0	0	79.8

El modelo de primer orden ajustado es:

$$\hat{Y} = 78.97 + 1.00x_1 + 0.50x_2$$

En la Tabla 2.6 se muestra el análisis de varianza

Tabla 2.6 Análisis de Varianza para el segundo modelo de primer orden

Fuente de variación	Suma de Cuadrados	Grados Libres	Media de Cuadrados	F ₀
Regresión (β_1, β_2)	5.0000	2	2.5000	1.35
Residuo	11.1200	6		
(Falta de Ajuste)	(10.9080)	2	5.4540	102.91
(Error puro)	(0.2120)	4	0.0530	
Total	16.1200	8		

El resultado de la prueba de falta de ajuste implica que el modelo de primer orden no es una aproximación adecuada, por lo que se trata de una superficie con curvatura y logramos llegar a la cercanía del óptimo.

2.6 Polinomio de segundo orden.

El modelo de segundo orden es el siguiente (Cornell [12] (1990)):

$$Y = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i + \sum_{i=1}^k \mathbf{b}_{ii} x_i^2 + \sum_{i < j}^k \mathbf{b}_{ij} x_i x_j + \epsilon$$

En éste los \mathbf{b}_i son los coeficientes de regresión para los términos de primer orden, los \mathbf{b}_{ii} son los coeficientes para los términos cuadráticos puros, los \mathbf{b}_{ij} son los coeficientes para los términos de producto cruz y $\hat{\mathbf{I}}$ es el término del error aleatorio. Los términos

cuadráticos puros y los de producto cruz son de segundo orden. El número de términos en la ecuación esta dado por $p=(k+1)(k+2)/2$

Los parámetros del modelo se estiman mediante el método de mínimos cuadrados. Una vez que se tienen los estimadores se sustituyen en la ecuación y obtenemos el modelo ajustado en el vecindario del valor óptimo de la respuesta (Cornell [12] (1990)):

$$\hat{Y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i < j}^k b_{ij} x_i x_j$$

La significancia de los coeficientes estimados y el ajuste del modelo se prueban con el estadístico F , el cálculo de éste se presentó en la sección 2.3.1 y 2.3.2 respectivamente.

Una vez que se ha verificado que el modelo tiene suficiencia de ajuste y que los coeficientes son significativos, se procede a localizar las coordenadas del “punto estacionario” y se lleva a cabo un análisis más detallado del sistema de respuesta.

2.6.1 Localización del punto estacionario.

Suponiendo que se desea maximizar la respuesta, el máximo (si es que existe), de acuerdo a Montgomery [6] 1991, será el conjunto x_1, x_2, \dots, x_k tal que las derivadas parciales

$$\partial \hat{y} / \partial x_1 = \partial \hat{y} / \partial x_2 = \dots = \partial \hat{y} / \partial x_k = 0$$

Dicho punto $(x_{1,0}, x_{2,0}, \dots, x_{k,0})$ se denomina punto estacionario. El punto estacionario puede ser:

- a) Un punto de respuesta máxima
- b) Un punto de respuesta mínima
- c) Un punto silla.

Esto se muestra en la siguiente figura (Montgomery [6] (1991)).

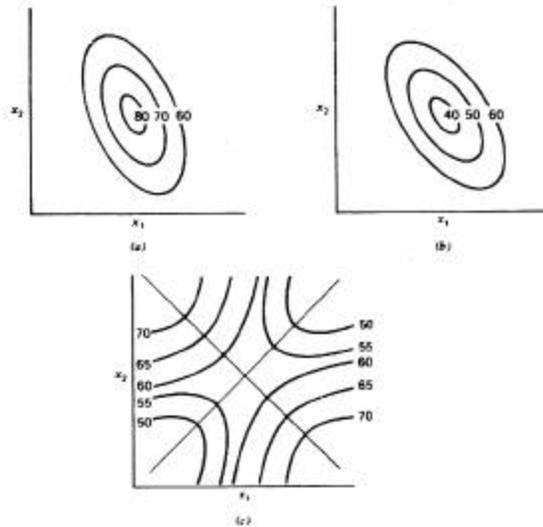


Figura 2.4 Puntos estacionarios en una superficie de respuesta de segundo orden ajustada.
 (a) Respuesta máxima. (b) Respuesta mínima. (c) Punto silla.

Podemos obtener el punto estacionario usando la notación matricial para el modelo de segundo orden (Montgomery [6] (1991)):

$$\hat{Y} = \hat{\mathbf{b}}_0 + x'b + x' Bx$$

donde

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad b = \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \vdots \\ \hat{\mathbf{b}}_k \end{bmatrix} \quad y \quad B = \begin{bmatrix} \hat{\mathbf{b}}_{11}, \hat{\mathbf{b}}_{12}, \dots, \hat{\mathbf{b}}_{1k} \\ \hat{\mathbf{b}}_{21}, \hat{\mathbf{b}}_{22}, \dots, \hat{\mathbf{b}}_{2k} \\ \vdots \\ \hat{\mathbf{b}}_{k1}, \hat{\mathbf{b}}_{k2}, \dots, \hat{\mathbf{b}}_{kk} \end{bmatrix}$$

En otras palabras, b es el vector $(k \times 1)$ de coeficientes de regresión de primer orden, y B es una matriz simétrica $(k \times k)$ cuya diagonal principal está formada por los coeficientes de los términos cuadráticos puros (\hat{b}_{ii}) , mientras que los elementos fuera de ésta corresponden a un medio del valor de los coeficientes cuadráticos mixtos $(\hat{b}_{ij}, i \neq j)$. La derivada de \hat{Y} con respecto al vector x igualada a cero es (Montgomery [6] (1991)):

$$\frac{\partial \hat{Y}}{\partial x} = b + 2Bx = 0$$

El punto estacionario es la solución de la ecuación es decir (Montgomery [6] (1991)):

$$x_0 = -\frac{1}{2}B^{-1}b$$

Sustituyendo ésta en la ecuación matricial para el modelo de segundo orden tenemos (Montgomery [6] (1991)):

$$\hat{Y}_0 = \hat{b}_0 + \frac{1}{2}x_0'b$$

2.6.2 Caracterización de la superficie de respuesta.

Habiendo encontrado el punto estacionario es necesario caracterizar la superficie de respuesta, es decir determinar si se trata de un punto de respuesta máximo, mínimo o silla. La forma directa de hacer esto es mediante la gráfica de contornos del modelo ajustado, sin embargo es útil un análisis más formal.

Como una alternativa se puede expresar la forma de la superficie de respuesta usando un nuevo conjunto de variables: W_1, W_2, \dots, W_k cuyos ejes representan los ejes principales de la superficie de respuesta, los cuales se interceptan en el punto estacionario como se observa en la figura 2.5 (Montgomery [6] 1991). Esto da por resultado el modelo ajustado (Montgomery [6] (1991)):

$$\hat{Y} = \hat{Y}_0 + \mathbf{I}_1 W_1^2 + \mathbf{I}_2 W_2^2 + \dots + \mathbf{I}_k W_k^2$$

donde las W_i son variables independientes transformadas y las \mathbf{I}_i son constantes.

Esta ecuación es llamada **forma canónica**.

Las \mathbf{I}_i son los valores propios (también conocidos como raíces características, autovalores o eigenvalores) y se toman de la matriz B.

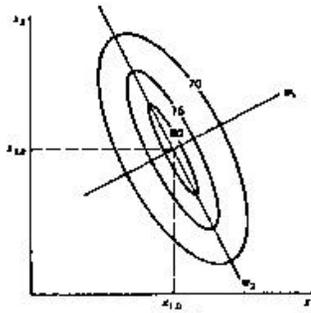


Figura 2.5 Superficie de respuesta en forma canónica.

La naturaleza de la superficie de respuesta puede determinarse a partir del punto estacionario y de el signo y magnitud de las \mathbf{I}_i . Si todas las \mathbf{I}_i son positivas, entonces es un punto de respuesta mínima, si todas las \mathbf{I}_i son negativas, entonces es un punto de respuesta máxima; y si las \mathbf{I}_i tienen signos distintos entonces es un punto de respuesta silla.

2.7 Diseños experimentales para ajustar superficies de respuesta.

El ajuste y análisis de una superficie de respuesta se facilita con la elección apropiada de un diseño experimental.

Un diseño es el conjunto específico de combinaciones de los niveles de las k variables que se utilizará al llevar a cabo el experimento.

2.7.1 Diseños para ajustar modelos de primer orden.

Una clase única de diseños que minimizan la varianza de los coeficientes de regresión ($\hat{\mathbf{b}}_i$) son los diseños ortogonales de primer orden. Por ortogonal se entiende que los elementos fuera de la diagonal de la matriz $(\hat{\mathbf{x}}\hat{\mathbf{x}})$ son iguales a cero, lo cual implica que los productos cruzados de las columnas de la matriz \mathbf{x} es igual a cero.

En esta clase de diseños ortogonales de primer orden se incluyen:

1. Diseños factoriales 2^k
2. Fracciones de la serie 2^k
3. Diseños simplex
4. Diseños Plackett-Burman

2.7.1.1 Diseños factoriales 2^k

En este diseño los k factores se codifican a los niveles estandarizados ± 1 . El diseño no permite la estimación del error experimental a menos que se repitan los experimentos, para lograr esto se aumenta el diseño con observaciones en el centro. La adición de los

puntos centrales no tiene influencia sobre las $\hat{\mathbf{b}}_i$ para $i \neq 1$, pero la estimación de \mathbf{b}_0 es el promedio general de todas las observaciones.

2.7.1.2 Fracciones de la serie 2^k

En programas experimentales se tienen dos razones para no llevar a cabo las 2^k combinaciones de un arreglo factorial completo:

1. A medida que el número k de factores incrementa crece rápidamente el número de combinaciones de niveles, haciéndose muy grande.
2. Sólo los primeros $k+1$ términos del modelo definen la ecuación de un hiperplano. Los restantes $2^k - (k+1)$ términos, consistentes en productos cruzados son una medida de la distorsión del hiperplano.

Como el nombre de este diseño lo indica es una fracción de un diseño 2^k . La fracción $1/2$ se denota como 2^{k-1} y contiene la mitad de las combinaciones de un 2^k mientras que la fracción $1/4$ se denota como 2^{k-2} y contiene la cuarta parte de las combinaciones de un 2^k . Las fracciones deben tener suficientes puntos para estimar los $k+1$ coeficientes.

Cabe señalar que al usar un diseño 2^{k-1} no podemos medir la posible falta de ajuste del modelo, a menos que se cuente con una estimación de la varianza del error haciendo réplicas del punto central; además, si el término $x_i x_j x_l$ realmente existe sesga la estimación del efecto principal asociado con x_k .

2.7.1.3 Diseño simplex

En este diseño los puntos se localizan en los vértices de una figura regular, ésta tiene $k+1$ vértices y está en k dimensiones. Para $k=2$ la figura geométrica es un triángulo equilátero y para $k=3$ es un tetraedro.

Como el número de puntos es igual al número de coeficientes del modelo se recomienda adicionar réplicas en el punto central para que sea posible obtener la varianza del error y/o llevar a cabo la prueba de falta de ajuste.

2.7.1.4 Diseños Plackett-Burman.

Estos diseños son fracciones de arreglos factoriales 2^k . Con este diseño los coeficientes se estiman con máxima precisión.

2.7.1.5 Comparación de los diseños ortogonales de primer orden.

Tabla 2.7 Comparación de los diseños ortogonales de primer orden

DISEÑO	NO. DE PUNTOS	VENTAJAS	DESVENTAJAS
Factoriales 2^k	$n=2^k$		Requiere observaciones en el centro. La estimación de \mathbf{b}_0 es el promedio general de todas las observaciones.
Fracciones de la Serie 2^k	$n=1/2$ de 2^k $n=1/4$ de 2^k	Contiene menos combinaciones que un 2^k	Requiere observaciones en el centro. Si el término $x_i x_j x_l$ existe sesga la estimación del efecto principal asociado con x_k .
Simplex	$n=K+1$		
Plackett-Burman	$n= K+1$, n es múltiplo de 4	Los coeficientes se estiman con máxima precisión.	

2.7.2 Diseños para ajustar modelos de segundo orden.

Un diseño experimental para ajustar un modelo de segundo orden debe tener al menos tres niveles de cada factor (-1, 0, +1). Así como en el diseño de primer orden se desea la ortogonalidad, en éste se desea que sea un diseño rotatable. Se dice que un diseño es rotatable cuando la varianza de la respuesta predicha en algún punto es función sólo de la distancia del punto al centro y no es una función de la dirección.

La rotabilidad es una propiedad importante, dado que la finalidad de la Metodología de Superficies de Respuesta es optimizar y desconocemos la localización del óptimo, tiene sentido utilizar un diseño que proporcione estimaciones precisas en todas direcciones.

Dentro de los diseños rotables de segundo orden se incluyen:

1. Diseño central compuesto
2. Diseño equirradial
3. Diseños Box-Behnken

2.7.2.1 Diseño central compuesto

Este diseño consiste en un factorial o factorial fraccionado 2^k , donde los factores son codificados de tal manera que el centro sea $(0,0,\dots,0)$, aumentado por 2^k puntos axiales $(\pm\alpha, 0, 0,\dots, 0)$, $(0, \pm\alpha, 0,\dots, 0)$, $(0, 0, \pm\alpha,\dots, 0)$, $(0, 0, 0,\dots,\pm\alpha)$, y n_c puntos centrales $(0,0,\dots,0)$.

De acuerdo a Montgomery [6] 1991 este diseño es probablemente el más usado.

Este diseño se convierte en rotatable mediante la elección de α , ésta se calcula de la siguiente manera (Montgomery [6] 1991):

$$a = (n_f)^{1/4}$$

Donde f es el número de puntos en la porción factorial del diseño.

Otra propiedad útil del diseño es que puede “crecer” a partir de un diseño 2^k de primer orden, agregando puntos axiales y quizá algunos puntos centrales (Montgomery [6] (1991)). Con la elección del número de puntos centrales (n_0), el diseño puede hacerse ortogonal o se puede transformar en uno de precisión uniforme.

En un diseño de precisión, la varianza de la respuesta predicha en el origen es igual a la predicha a una distancia unitaria del origen. Éste proporciona mayor protección que el ortogonal contra el sesgo de los coeficientes, debido a la presencia de términos de tercer y mayor orden.

2.7.2.2 Diseño equirradial.

Este diseño consiste en puntos igualmente espaciados sobre una circunferencia o una esfera. Para $k=2$, el diseño se obtiene combinando $n_2 = 5$ puntos igualmente espaciados sobre una circunferencia con $n_1 = 1$ puntos en su centro. El pentágono y el hexágono son útiles en este caso. Para $k=3$, los únicos arreglos que cuentan con puntos suficientes para estimar todos los parámetros son el icosaedro y el dodecaedro.

Según Montgomery [6] (1991), este diseño es ocasionalmente útil en problemas con dos o tres variables.

2.7.2.3 Diseño Box-Behnken.

Estos diseños se forman combinando factoriales 2^k con diseños de bloques incompletos. Los diseños resultantes suelen ser más eficientes en términos del número de corridas requerido. Además, son rotables (o casi rotables) y hace la estimación de los coeficientes de primer y segundo orden más eficiente.