**OPTIMIZACIÓN DE PROCESOS**

**1. REGRESION LINEAL**

En la búsqueda de mejoras o en la solución de problemas es necesario, frecuentemente, investigar la relación entre factores (o variables). Para lo cual existen varias herramientas estadísticas, entre los que se encuentran el diagrama de dispersión, el análisis de correlación y el análisis de regresión.

 En esta sección veremos el análisis de regresión el cual tiene como objetivo modelar matemáticamente el comportamiento de un factor o variable en función de otro factor o conjunto de factores. Por ejemplo, supongamos que el rendimiento de un proceso químico está relacionado con la temperatura de operación. Si mediante un modelo matemático se puede describir tal relación, entonces este modelo puede ser usado para propósitos de predicción, optimización o control.

 El análisis de regresión puede usarse para explicar la relación de un factor con otro(s). Para ello, son necesarios los datos, y estos pueden obtenerse de experimentos planeados, de observaciones de fenómenos no controlados o de registros históricos.

**1.1 Regresión lineal simple**

Sean dos variables X y Y. Supongamos que se quiere explicar el comportamiento de Y con el de X. Para esto, se mide el valor de Y sobre un conjunto de n valores de X, con lo que se obtienen n parejas de puntos (X1 ,Y1 ), (X2 ,Y2 ),...,(Xn ,Yn ). A Y se le llama la variable dependiente o la variable de respuesta y a X se le conoce como variable independiente o variable regresora. La variable X no necesariamente es aleatoria, ya que en muchas ocasiones el investigador fija sus valores, en cambio Y si es una variable aleatoria. Una manera de estudiar el comportamiento de Y respecto a X es mediante un modelo de regresión, que consiste en ajustar un modelo matemático de la forma

Y=f(X)

a las n parejas de puntos. Con lo cual se puede ver si dado un valor de la variable independiente X se puede predecir el valor promedio de Y.

 Supongamos que las variables X y Y están relacionadas linealmente y que para cada valor de X, Y es una variable aleatoria. Es decir, supongamos que cada observación de Y puede ser descrita por el modelo

(1.1) Y=ß0 +ß1X+e,

donde e es un error aleatorio con media cero y varianza σ2 . También supongamos que los errores aleatorios no están correlacionados. Los parámetros ß0 y ß1 son constantes desconocidas. La ecuación (1.1) es conocida como el modelo de Regresión lineal simple. Bajo el supuesto de que el modelo (1.1) es adecuado y como E(e)=0 se puede ver que la media de la variable Y, dado un valor de X, está dado por línea recta

(1.2) E(Y|X)=ß0 +ß1X.

 En esta ecuación se observa que el problema es ajustar una línea recta a las parejas de puntos (Xi ,Yi ), para lo cual es necesario estimar los parámetros ß0 y ß1 .

 Por ejemplo en una fábrica de pintura se desea investigar la relación entre la velocidad de agitación X y el porcentaje de impurezas en la pintura Y. Mediante un diseño experimental se obtienen los siguientes datos.

|  |  |
| --- | --- |
| Velocidad | Impurezas |
|  20 22 24 26 28 30 32 34 36 38 40 42 |  8.4 9.5 11.8 10.4 13.3 14.8 13.2 14.7 16.4 16.5 18.9 18.5 |

I

 Es claro que la variable de respuesta o variable dependiente es el porcentaje de impurezas, por eso se denota con Y.

 Para entender la relación que existe entre X y Y podemos representar los 12 pares de datos en un diagrama de dispersión; el cual se muestra en la figura 1.1.

 A partir de la figura 1.1 se ve que entre X y Y existe un correlación lineal. Por lo que es razonable suponer que la relación entre X y Y la explique un modelo de regresión lineal. Así, cada observación de Y, la podemos expresar como

 Yi =ß0 +ß1X +e i=1,2,...,12.

 Para estimar ß0 y ß1 ajustamos la recta que explique de mejor manera el comportamiento de los datos en el diagrama de dispersión de la figura 1.1. En otras palabras, debemos encontrar la recta que pasa más cerca de todos los puntos. Un procedimiento para ajustar la mejor recta y por lo tanto de estimar ß0 y ß1, es mediante el método de mínimos cuadrados.



Figura 1.1

 En el ejemplo, la línea recta que mejor explica la relación entre velocidad de agitación y porcentaje de impurezas, tendrá los siguientes parámetros

ß = 0.4566

ß =‑0.2893.

Con lo que, el modelo de regresión está dado por

 (1.3) Y=‑0.2893+0.4566X.

A continuación se muestra la tabla completa proporcionada por statgraphics

Regression Analysis ‑ Linear model: Y = a+bX

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

Dependent variable: PINTREG.impurezas Independent variable: PINTREG.veloc

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

 Standard T Prob.

Parameter Estimate Error Value Level

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

Intercept ‑0.289277 1.22079 ‑0.236959 .81747

Slope 0.456643 0.0384385 11.8798 .00000

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

 Analysis of Variance

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

Source Sum of Squares Df Mean Square F‑Ratio Prob. Level

Model 119.27524 1 119.27524 141.1304 .00000

Residual 8.4514219 10 .8451422

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

Total (Corr.) 127.72667 11

Correlation Coefficient = 0.96635 R‑squared = 93.38 percent

Stnd. Error of Est. = 0.919316

 En la figura 1.2 se muestra esta recta y como ajusta a los datos.



Figura 1.2

 Una vez que se ajusta un modelo de regresión a un conjunto de datos, se presenta la interrogante: ¿qué tanta de la variabilidad presente en Y fue explicada por el modelo? La respuesta a estas preguntas son importantes porque un objetivo básico del análisis de regresión es explicar la relación entre dos variables. Si el modelo explica bien la variabilidad o se ajusta satisfactoriamente a los datos, entonces se podrá decir que la variable X y Y están relacionadas por el modelo ajustado. En caso contrario, se tendrá que investigar más sobre la manera en que X y Y se relacionan. Un criterio elemental para investigar el ajuste del modelo es ver el ajuste gráficamente, como en la figura 1.2. En dicha figura se observa que el ajuste es bueno. Otro criterio muy usado es el coeficiente de determinación.

1.2 COEFICIENTE DE DETERMINACION

 El coeficiente de determinación para un modelo de regresión está dado por

(1.4) R2 =

Este se utiliza frecuentemente como un criterio para juzgar lo adecuado del modelo. Como se puede observar 0<R2<1. Este coeficiente es interpretado, con ciertas reservas, como el porcentaje de la variabilidad en los datos de Y explicada por el modelo de regresión. Para los datos de la fábrica de pintura es 0.9338, por lo que podemos decir que la recta (1.3) explica el 93.37% de la variabilidad en los datos de impurezas. Con lo que, el modelo es bueno, ya que "explica" un 93.37% de la relación entre velocidad de agitación y el porcentaje de impurezas.

**1.3 Regresión Lineal Múltiple**

En muchos problemas existen dos o más variables que están relacionadas y puede ser importante modelar y explorar esta relación. Por ejemplo, el rendimiento de una reacción química puede depender de la temperatura, presión y concentración del catalizador. En este caso se requiere al menos un modelo de regresión con tres variables.

 El problema general consiste en ajustar el modelo de primer orden



En forma matricial

Y=X +

Donde

Y= X=

 

Los parámetros desconocidos  se denominan coeficientes de regresión, los cuales se estiman mediante la siguiente ecuación:



**Ejemplo**

Un ingeniero químico se encuentra investigando el rendimiento de un proceso, del cual le interesan tres variables: temperatura, presión y concentración porcentual. Cada variable puede estudiarse a dos niveles, bajo y alto, y el ingeniero decide correr un diseño  con estas tres variables. El experimento y los rendimientos resultantes se muestran en la siguiente tabla,

Corrida x1(Temp.) X2(presion) X3(concentra) rendi

 ----------------------------------------------------------------------

 1 -1. -1. -1. 32

 2 - 1. -1. 1. 36

 3 -1. 1. -1. 57

 4 1. -1. -1. 46

 5 1. 1. -1. 65

 6 -1. 1. 1. 57

 7 1. -1. 1. 48

 8 1. 1. 1. 68

donde se utiliza la notación de variables codificadas +1, -1 normalmente empleada en diseños factoriales 2k para representar los niveles de los factores.

Supóngase que el ingeniero decide ajustar un modelo sólo de "efectos principales", digamos

Y=

Para este modelo se tiene

X´X= X´Y=

Como X´X diagonal, la inversa requerida simplemente es  y los estimadores de mínimos cuadrados son

=

Por lo que el modelo de regresión es el siguiente:

Y=

Estos coeficientes de regresión están estrechamente relacionados con las estimaciones de los efectos de cada factor, por ejemplo:

El efecto de la temperatura es:

Efec(temperatura)= =(46+65+48+68)/4

 -(32+36+57+57)/4 =56.75-45.5=11.25

Obsérvese que el coeficiente de regresión es la mitad del efecto= 11.25/2=5.625

**1-4 PRUEBAS DE HIPOTESIS EN LA REGRESION LINEAL MULTIPLE**

A menudo el experimentador desea probar la hipótesis que se refieren a los parámetros del modelo de regresión lineal múltiple.

Consideremos probar si la regresión es significativa. En la regresión lineal múltiple esto se logra probando las hipótesis



 al menos una i

El rechazo de Ho en esta ecuación implica que al menos una variable en el modelo contribuye significativamente en el ajuste. El procedimiento para probar las hipótesis es el siguiente:

La Suma total de cuadrados Syy se descompone en la suma de cuadrados de regresión SSr y en la suma de cuadrados del error SSe

# Syy=SSr + SSe

Y si la hipótesis nula es verdadera,  ~ , donde el numero de grados de libertad para  es igual al numero de variables de regresión en el modelo. También se puede mostrar que  ~ , y que SSr y SSe son independientes. Por lo que el procedimiento para probar las hipótesis consiste en calcular CMr= SSr/k, CMe=SSe/(n-k-1) y Fo= CMr/CMe.

Si Fo>entonces se rechaza Ho.

El análisis de varianza para el ejemplo es el siguiente:

Multiple Regression - y Analysis of Variance

-----------------------------------------------------------------------------

Source Sum of Squares Df Mean Square F-Ratio P-Value

-----------------------------------------------------------------------------

Model 1166.38 3 388.792 148.11 0.0001

Residual 10.5 4 2.625

-----------------------------------------------------------------------------

Total 1176.88 7

**Por consiguiente, se rechaza la hipótesis nula lo que significa que al menos una variable está contribuyendo en el modelo con un 95% de confianza.**

### 1-5 HIPOTESIS INDIVIDUALES

Con frecuencia es importante probar hipótesis con respecto a los coeficientes de regresión individuales. Tales pruebas son útiles para valorar cada variable de regresión en el modelo. Por ejemplo, el modelo puede ser más efectivo si se le introducen variables adicionales o quizás, si se desecha una o más variables que se encuentran en el mismo.

Introducir variables en el modelo de regresión siempre provoca que la suma de cuadrados de regresión aumente y que la del error disminuya.

 Las hipótesis para probar la significancia de cualquier coeficiente individual, por ejemplo  son

Ho: =0

Ha: 0

El estadístico de prueba es:



donde Cii es el (i+1)-esimo elemento de la diagonal de la matriz (X´X).

Ejemplo, para 

==9.81

Multiple Regression - **y** Multiple Regression Analysis

-----------------------------------------------------------------------------

Dependent variable: y

-----------------------------------------------------------------------------

 Standard T

Parameter Estimate Error Statistic P-Value

-----------------------------------------------------------------------------

CONSTANT 51.125 0.572822 89.2511 0.0000

x1 5.625 0.572822 9.81981 0.0006

x2 10.625 0.572822 18.5485 0.0000

x3 1.125 0.572822 1.96396 0.1210

----------------------------------------------------------------

Las variables X1, X2 son significativas en el modelo. La variable X3 no es significativa en el modelo.

El termino constante es significativo en el modelo.

### OTROS MODELOS DE REGRESION

**El modelo lineal Y=X + es un modelo general. Puede usarse par ajustarse cualquier relación que sea lineal en los parámetros desconocidos .**

**Ejemplos:**

1. ****

**b)**

**2. DISEÑOS 2 K CON PUNTOS AL CENTRO**

**\*Un motivo de preocupación potencial en el uso de diseños factoriales de dos niveles es la suposición de linealidad en los efectos de los factores.**

**\*El sistema 2k funciona bien incluso cuando la suposición de linealidad se cumpla sólo de manera aproximada.**

**\*Existe sin embargo un método para replicar ciertos puntos en un diseño factorial 2k, lo cual protegerá contra la curvatura (falta de linealidad) además de permitir obtener estimaciones de error de manera independiente.**

**\*Dicho método consiste en agregar puntos centrales al diseño 2k, para lo cual se hacen n réplicas en los puntos Xi = 0 ( i = 1,2 ...,k).**

**\*Un aspecto importante es que al agregar los puntos centrales, no se afectaran las estimaciones usuales de los efectos en el diseño.**

**Para ilustrar el método, considérese un diseño 22 con una observación en cada uno de los puntos factoriales (-,-), (+,-), (-,+) y (+,+) y nc observaciones en los puntos centrales (0,0). En la siguiente figura se ilustra la situación.**



**Sea** $\overbar{y\_{f}}$ **el promedio de las cuatro corridas en los cuatro puntos factoriales y sea** $\overbar{y\_{c}}$ **el promedio de las nc corridas en el punto central. Si la diferencia** $\overbar{y\_{f}}$ **-** $\overbar{y\_{c}}$ **es pequeña, entonces los puntos centrales se encuentran en el plano que pasa por los puntos factoriales (o cerca de él), y no hay curvatura.**

**Por otro lado, si** $\overbar{y\_{f}}$ **-** $\overbar{y\_{c}}$ **es grande, entonces existe curvatura. Una suma de cuadrados para la curvatura con un solo grado de libertad está dada por**

**Donde, en general, nF es el número de puntos en el diseño factorial. Esta cantidad puede compararse con el cuadrado medio de error para probar la curvatura.**

**Problema**

**Un ingeniero químico se encuentra estudiando el rendimiento de un proceso. Existen dos factores de interés, tiempo y temperatura de reacción. Debido a que tiene duda acerca de la suposición de linealidad en la región que explora, el ingeniero decide realizar un diseño 22 (con una sola replica de cada corrida factorial) aumentada con 5 puntos al centrales. Los datos de rendimiento se muestran a continuación:**

|  |  |  |
| --- | --- | --- |
| **A**  | **B**  | **RENDIMIENTO**  |
| **-1**  | **-1**  | **39.3**  |
| **1**  | **-1**  | **40.9**  |
| **-1**  | **1**  | **40**  |
| **1**  | **1**  | **41.5**  |
| **0**  | **0**  | **40.3**  |
| **0**  | **0**  | **40.5**  |
| **0**  | **0**  | **40.7**  |
| **0**  | **0**  | **40.2**  |
| **0**  | **0**  | **40.6**  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source  | Sum of Squares  | Df  | Mean Square  | F-Ratio  | P-Value  |
| A:A  | 2.4025 | 1 | 2.4025 | 68.75 | 0.0004 |
| B:B  | 0.4225 | 1 | 0.4225 | 12.09 | 0.0177 |
| AB  | 0.0025 | 1 | 0.0025 | 0.07 | 0.7998 |
| Total error  | 0.174722 | 5 | 0.034944 |    |    |
| Total (corr.)  | 3.00222 | 8 |    |    |    |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source**  | **Sum of Squares**  | **Df**  | **Mean Square**  | **F-Ratio**  | **P-Value**  |
| **A:A**  | **2.4025** | **1** | **2.4025** | **55.87** | **0.0017** |
| **B:B**  | **0.4225** | **1** | **0.4225** | **9.83** | **0.035** |
| **AB**  | **0.0025** | **1** | **0.0025** | **0.06** | **0.8213** |
| **Lack-of-fit**  | **0.0027222** | **1** | **0.0027222** | **0.06** | **0.8137** |
| **Pure error**  | **0.172** | **4** | **0.043** |  |  |
| **Total (corr.)**  | **3.00222** | **8** |  |  |  |

**SON SIGNIFICATIVOS LOS EFECTOS DE A Y B**

**NO HAY EFECTO DE CURVATURA**

**CON A(+) SE TIENE MAYOR RENDIMIENTO.**

**CON B(+) SE TIENE MAYOR RENDIMIENTO.**

**CON A(+) Y B(+) SE TIENE MAYOR RENDIMIENTO.**

**3. DISEÑOS PARA SUPERFICIE DE RESPUESTA**

\* Este tipo de diseños normalmente se emplea en las últimas fases de la experimentación

\* Su aplicación se hace indispensable, si después de haber identificado los factores significativos (a través de experimentos de diagnóstico), se considera necesario explorar la relación entre factor y la variable dependiente dentro de la región experimental, y no solamente en las fronteras (como se hace en los diseños factoriales).

\* Estos diseños y su optimización constituyen la fase final; por lo tanto, en algunos casos, no se requerirá de su utilización.

\* Tratar de emplear esta técnica antes de haber identificado (mediante el uso de experimentos de diagnóstico, experiencia o teoría) cuáles factores son relevantes, generalmente conducirá a resultados incompletos (imprecisos).

La metodología de superficie de respuesta es un junto de técnicas útiles para modelar y analizar problemas en los cuales una respuesta de interés es INFLUIDA por varias variables (factores, 2 a 6), y el objetivo es optimizar esta respuesta.

La MSR se emplean en las siguientes etapas:

1. MSR es secuencial. Si se está lejos del punto óptimo, la MSR guía eficaz y rápidamente a la cercanía del óptimo

2. Estando cerca del óptimo se usa un Diseño apropiado, para con los datos obtenidos determinar una ecuación general que se empleará para predicciones.

3. Aplicación de técnicas de regresión lineal múltiple para seleccionar la "mejor" ecuación que modele el comportamiento de los datos.

4. Análisis de la superficie ajustada mediante gráficas de contorno y otras técnicas matemáticas y numéricas.

**3-1 INTRODUCCIÓN A LA METODOLOGÍA DE SUPERFICIES DE REPUESTA**

La metodología de superficies de respuestas, (MSR o RSM, por sus siglas en inglés) es un conjunto de técnicas matemáticas y estadísticas útiles para modelar y analizar problemas en los cuales una respuesta de interés es influida por varias variables, y el objetivo es optimizar esta respuesta. Por ejemplo, supóngase que un ingeniero químico desea determinar los niveles de temperatura (x1) y presión (x2) que maximizan el rendimiento (**y**) de un proceso. En rendimiento del proceso es función de los niveles de temperatura y presión, digamos



donde  representa el ruido o error observado en la respuesta **y** . Si la respuesta esperada se denota por E(**y**) = f(x1,x2) =, entonces la superficie representada por



y se denomina superficie de respuesta.

Es posible representar gráficamente la superficie de respuesta como se muestra en la Fig. 1.1, donde  se gráfica contra los niveles de x1 y x2. Obsérvese que la respuesta se representa como una superficie sólida en un espacio tridimensional.

Figura 1.1 superficie de respuesta

****

 Para visualizar mejor la forma de una superficie de respuesta, a menudo se grafican los contornos de dicha superficie como se muestra en la Fig. 1-2. En esta gráfica de contornos, se trazan líneas de respuestas constante en el plano x1, x2.

Figura1.2 Gráfica de Contornos

****

O también se puede graficar la superficie y el contorno como se muestra en la siguiente figura 1-3.

****

En la mayoría de los problemas de RSM, la forma de la relación entre la variable de respuesta y las variables independientes se desconoce. Por ello, el primer paso en la RSM consiste en determinar una aproximación apropiada a la relación funcional real entre **y** el conjunto de variables independientes. Por lo general se emplea un polinomio de orden bajo sobre alguna región de las variables independientes. Si la respuesta es descrita adecuadamente por una función lineal de las variables independientes, la función de aproximación es el modelo de primer orden

****

Cuando existe curvatura, debe usarse un polinomio de mayor grado, por ejemplo el modelo de segundo orden,

****

Casi todos los problemas de RSM utilizan uno o ambos polinomios de aproximación. Por supuesto, es improbable que un modelo polinomial sea una aproximación razonable de la relación funcional real sobre todo el dominio de las variables independientes. Sin embargo, funcionan muy bien en regiones relativamente pequeñas de las variables independientes.

 El método de mínimos cuadrados, sirve para estimar los parámetros del polinomio de aproximación. El análisis de la superficie de respuesta se hace luego en términos de la superficie ajustada. Tal análisis será aproximadamente equivalente al análisis del sistema real, si la superficie ajustada es una aproximación adecuada a la función de respuesta real. La estimación de los parámetros del modelo se hace más eficazmente si se utilizan los diseños experimentales apropiados para recopilar los datos. A menudo, los diseños usados para ajustar superficies de respuesta se denominan diseños de superficie de respuesta.

 La MSR es una técnica secuencial. A menudo, cuando se considera un punto sobre la superficie de respuesta alejado del óptimo, como las condiciones de operación actuales de la Fig. 1-2, el polinomio de primer grado es apropiado porque existe poca curvatura en el sistema. En este caso, el objetivo consiste en guiar al experimentador rápida y eficientemente a la cercanía general del punto óptimo. Una vez que se ha determinado la región del punto óptimo, puede emplearse un modelo más elaborado, como por ejemplo una superficie de respuesta de segundo grado, y realizar un análisis para localizar el óptimo. A partir de la Fig. 1-2, se observa que el análisis de la superficie de respuesta puede interpretarse como el "ascenso de una loma", donde la cima representa el punto de la repuesta máxima. Si el óptimo real es un punto de respuesta mínima, se puede pensar en el "descenso hacia un valle".

 El objetivo eventual de la MSR consiste en determinar las condiciones de operación óptima para un sistema, o determinar la región del espacio de los factores en la que se satisfacen las condiciones de operación. La MSR no se usa principalmente para obtener un mayor entendimiento del mecanismo físico del sistema, a pesar de que la MSR puede ser útil para adquirir dicho conocimiento. Más aún, debe observarse que el "óptimo" de la MSR se utiliza en un sentido especial. Los procedimientos de la MSR para "escalar una loma" garantizan la convergencia sólo hacia un óptimo relativo.

**3-2 MÉTODO DE MÁXIMA PENDIENTE EN ASCENSO**

Con frecuencia, la estimación inicial de las condiciones de operación óptimas para un sistema estará alejada del óptimo real. En tales circunstancias, el objetivo del experimento es moverse rápidamente a la vecindad general del óptimo. Se desea usar un procedimiento experimental simple y económica­mente eficiente. En la lejanía de un óptimo, generalmente se supone que el modelo de primer orden es una aproximación adecuada a la superficie real en regiones pequeñas de las Xs.

****

8

0

 El método de máxima pendiente con ascenso es un procedimiento para recorrer secuencialmente a lo largo de la trayectoria de la máxima pendiente; en otras palabras, en la dirección del máximo incremento de la respuesta. Por supuesto, si se desea la minimización se hablará del método de máxima pendiente en descenso. El modelo de primer orden ajustado es

y la superficie de respuesta, o sea, las curvas de nivel de  consta de una serie de rectas paralelas, como se muestra en la Fig. 1-4. La dirección de ascenso máximo es aquella en la que  aumenta más rápidamente. Esta dirección es paralela a la normal de la superficie de respuesta ajustada. Por lo regular, la trayectoria de máxima pendiente en ascenso se toma como la recta que atraviesa el centro de la región de interés y es normal a la superficie ajustada. Por lo tanto, los incrementos a lo largo de la trayectoria son proporcionales a los coeficientes de regresión . El tamaño del incremento lo determina el experimentador con base a su experiencia con el proceso u otras consideraciones prácticas.

 Los experimentos se llevan acabo a lo largo de la trayectoria de máximo ascenso hasta que deje de observarse un incremento adicional en la respuesta. Entonces puede ajustarse un nuevo modelo de primer orden, determinar una nueva trayectoria de ascenso máximo y continuar con el procedimiento. Por último, el experimentador llegará a la cercanía del óptimo. Usualmente esto sucede cuando ocurre una falta de ajuste del modelo de primer orden. En tal momento los experimentos adicionales, que serán descritos en la en las siguientes secciones, se llevan a cabo para obtener una estimación más precisa del óptimo.

Figura1.4



**Ejemplo 3-1**

Un ingeniero quími­co está interesado en determinar las condi­ciones de operación que maximizan el rendimiento de una reacción. Dos varia­bles controlables influyen en este rendimiento: el tiempo y la temperatura de reacción. Actualmente ella opera sobre el proceso con un tiempo de reacción de 35 minutos y a una temperatura de 1550F. Esto produce un rendimiento de cerca de 40%. Ya que es poco probable que esta región contenga al óptimo, se ajustará un modelo de primer orden y se aplicará el método de ascenso máximo.

 El ingeniero decide que la región de exploración para ajustar el modelo de primer orden debe ser (30, 40) minutos de reacción y (150, 160)0F. Por lo que efectúa el siguiente diseño experimental:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables**  | **Naturales** | **Variables**  | **Codificadas** | **Respuesta** |
|  |  |  |  | **Y** |
| **30** | **150** | **-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** | **40.6** |
| **35** | **155** | **0** | **0** | **40.7** |
| **35** | **155** | **0** | **0** | **40.2** |
| **35** | **155** | **0** | **0** | **40.6** |

Para simplificar los cálculos, las variables independientes se codifican al intervalo (-1, 1). Por lo tanto, si representa la variable natural tiempo y  la variable natural temperatura, entonces las variables codificadas son  y  mediante las siguientes ecuaciones:

** y **

Nótese que el diseño experimental usado por la ingeniera es un diseño  con cinco puntos centrales. Las observaciones repetidas en el centro se utilizan para estimar el error experimental y permitir una prueba de adecuación del modelo de primer orden. Asimismo, el modelo está centrado alrededor de las condiciones actuales de operación del proceso.

Pero, antes de explorar la trayectoria de máximo ascenso, es necesario los siguientes pasos:

**1.- Realizar un análisis de varianza para ver la significancia de los factores (aquí podemos ver si hay o no curvatura).**

**2.- Encontrar el modelo de regresión de primer orden.**

**3.- Investigar la idoneidad del modelo de primer orden mediante el anova (aquí se puede utilizar el método de Forward o Backward).**

**Resultados del primer paso:**

**Analisis de Varianza para rendimiento**

**----------------------------------------------**

**Fuente SC G.L. C.Medios F P-Value**

----------------------------------------------

# **A:x1 2.4025 1 2.4025 55.87 0.0017**

**B:x2 0.4225 1 0.4225 9.83 0.0350**

**AB 0.0025 1 0.0025 0.06 0.8213**

**Curvatura0.002 1 0.0027 0.06 0.8137**

**error 0.172 4 0.043**

**----------------------------------------------**

**Total 3.00222 8**

**\*No Existe curvatura**

\*Tiempo y Temperatura son significativos

\*No hay efecto de interacción.

\*Por tanto se requiere un modelo de primer orden

**Resultados del segundo paso**

Coeficientes de Regresión para rendimiento

**------------------------------------**

**constant = 40.4444**

**A:x1 = 0.775**

**B:x2 = 0.325**

**------------------------------------**

**Resultados del tercer paso:**

 Es necesario investigar la idoneidad del modelo (utilizando el método de forward) encontramos que:

**Análisis de regresión múltiple**

**----------------------------------------------**

**Variable dependiente: rendimiento**

**----------------------------------------------**

 **Parámetro error estadístico**

 **Estimado estándar de prueba T P-Value**

**----------------------------------------------**

**CONSTANT 40.4444 0.0572878 705.987 0.0000**

**x1 0.775 0.0859317 9.01879 0.0001**

**x2 0.325 0.0859317 3.78207 0.0092**

**----------------------------------------------**

**Las variables x1 y x2 y la constante son significativas en el modelo con un 95% de confianza.**

**La variable más importante es x1.**

**Con respecto al modelo, revisando el anova**

##

##  Análisis de varianza para el modelo en general.

**----------------------------------------------**

**Fuente SC GL CM F-Ratio P-Value**

**----------------------------------------------**

**Model 2.825 2 1.412 47.82 0.0002**

**Residual 0.177 6 0.0295**

**----------------------------------------------**

**Total 3.00222 8**

 **Podemos concluir que el modelo es significativo al 95%.**

### 3.3 Trayectoria de máximo ascenso

**Para alejarse del centro del diseño a lo largo de la trayectoria de máximo ascenso es necesario desplazarse 0.775 unidades en la dirección de x1 por cada 0.325 unidades en la dirección de x2. Por consiguiente, la trayectoria de máximo ascenso pasa por el punto (x1=0, x2=0) y tiene una pendiente igual a 0.325/0.775.**

 **El ingeniero decide usar un incremento básico de tiempo de reacción de 5 minutos, lo que equivale pasar de 35 a 40 minutos. Usando la relación**

****

**tenemos que**

**=1**

**lo que significa que el incremento de tiempo de reacción de 5 minutos en la variable natural es equivalente a 1 en la variable codificada, es decir =1.**

 **De esta forma los incrementos a lo largo de la trayectoria de máximo ascenso son:**

**=1 y =(0.325/0.775) **

Nótese que el incremento de x2 depende del incremento en x1, como el incremento en x1 es igual a 1, el incremento de x2 es:

**=(0.325/0.775) (1)=0.4193**

a partir de esta relación, El ingeniero calcula los puntos a lo largo de esta trayectoria:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Variable****Codificada** | **Variables naturales** | Respuesta |
| **Incrementos** | **X1** | **X2** |  |  |  |
| **Origen** | **0** | **0** | **35** | **155** |  |
|  | **1** | **0.4193** | **5** | **2.09** |  |
| **Origen +**  | **1** | **0.4193** | **40** | **157.09** | **41.0** |
| **Origen +2** | **2** | **0.8386** | **45** | **159.18** | **42.9** |
| **Origen +3** | **3** | **1.2579** | **50** | **161.27** | **47.1** |
| **Origen +4** | **4** | **1.6772** | **55** | **163.36** | **49.7** |
| **Origen +5** | **5** | **2.0965** | **60** | **165.45** | **53.8** |
| **Origen +6** | **6** | **2.5158** | **65** | **167.54** | **59.9** |
| **Origen +7** | **7** | **2.9351** | **70** | **169.63** | **65.0** |
| **Origen +8** | **8** | **3.3544** | **75** | **171.72** | **70.4** |
| **Origen +9** | **9** | **3.7737** | **80** | **173.81** | **77.6** |
| **Origen +10** | **10** | **4.193** | **85** | **175.9** | **80.3** |
| **Origen +11** | **11** | **4.6123** | **90** | **177.99** | ***76.2*** |
| **Origen +12** | **12** | **5.0316** | **95** | **180.09** | ***75.1*** |

El ingeniero observa que con los niveles de 85 minutos de tiempo y 175 de temperatura obtiene un rendimiento del 80%, además que a partir de ahí se observa un descenso en la variable de respuesta. Por lo que decide efectuar otro diseño experimental donde los niveles de tiempo sean de 80 a 90 minutos y de la temperatura sean de 170 a 180 grados. Los resultados se muestran a continuación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables**  | **Naturales** | **Variables**  | **Codificadas** | **Respuesta** |
|  |  |  |  | **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** |

donde

** y **

Nótese que el diseño experimental usado es un diseño  con cinco puntos centrales. Las observaciones repetidas en el centro se utilizan para estimar el error experimental y permitir una prueba de adecuación del modelo de primer orden. Asimismo, el modelo está centrado alrededor de las nuevas condiciones de operación del proceso.

 El análisis de varianza para este diseño se muestra a continuación:

Analysis of Variance for rendimiento

-------------------------------------------------------------

Source Sum of Squares Df Mean Square F-Ratio P-Value

-------------------------------------------------------------

A:x1 4.0 1 4.0 75.47 0.0010

B:x2 1.0 1 1.0 18.87 0.0122

AB 0.25 1 0.25 4.72 0.0956

Lack-of-fit 10.658 1 10.658 201.09 0.0001

Pure error 0.212 4 0.053

-------------------------------------------------------------

Total (corr.) 16.12 8

**Conclusiones:**

**\*Existe curvatura**

\*Tiempo y Temperatura son significativos

\*No hay efecto de interacción.

\*Por tanto se requiere un modelo de segundo orden, lo que hace uso de un diseño de composición central.

**3-4 SELECCIÓN DEL ALGORITMO GENERAL PARA DETERMINAR LAS COORDENADAS**

Es facil formular un algoritmo general para determinar las coordenadas de un punto en la trayectoria de máxima pendiente en ascenso. Supongamos que el punto x1=x2=....xk=0 es la base o el punto origen. Entonces:

**1.- Se elige un tamaño de incremento o “escalón” en una de las variables del proceso, digamos un . Usualmente se eligiría la variable de la que mas se sabe, o la que tiene mayor coeficiente de regresión absoluto  .**

**2.- El tamaño de incremento en las otras variables es**

****

**3.- Se convierte el  de variables codificadas a variables naturales, mediante la siguiente relación:**

**.**

#### 3-5 DISEÑOS PARA AJUSTAR EL MODELO DE SEGUNDO ORDEN

Por lo general debido a la curvatura de la superficie real, el experimentador requiere un modelo cuyo grado sea mayor que o igual a 2.

**En la mayoría de los casos, el modelo de segundo orden:**

****

es adecuado.

Un diseño experimental para ajustar un modelo de segundo orden debe tener por lo menos tres niveles de cada factor. Existen muchos diseños que podrían emplearse para ajustar un modelo de segundo orden (ver STATG), así que elegir un diseño apropiado es útil con objeto de establecer un criterio de diseño. Para el modelo de primer orden, la ortogonalidad es la propiedad de diseño óptima ya que minimiza la varianza de los coeficientes de regresión. La ortogonalidad también es deseable en el caso de primer orden debido a que resulta ser una propiedad muy conveniente para la varianza de la respuesta predicha 

 Se dice que un diseño experimental es rotable o girable si la varianza de la respuesta predicha en algún punto x es función sólo de la distancia al punto desde el centro de diseño y no es una función de dirección. Además, un diseño con esta propiedad dejará sin cambio la varianza  cuando el diseño se haga rotar (o girar) alrededor del centro (0,0,...,0); de aquí el nombre de diseño rotado.

#### 3-6 DISEÑOS DE COMPOSICION CENTRAL

Este es uno de los diseños mas usados para propósitos de optimización, se conocen como diseño de composición central. Estos diseños se construyen con base en factoriales con dos niveles (lo cual permite la estimación de efectos principales e interacciones). Además, incluyen un conjunto de puntos en los ejes (llamados puntos estrella), los cuales junto con el punto central (por lo general, repetido) permiten estimar los términos cuadráticos puros.

 

 (-1,1) (1,1)

 (0,0)

  

 (-1,-1) (1,-1)

 

Figura. Diseño de composición central de dos factores.

 Un diseño compuesto central se convierte en rotable mediante la selección de . El valor de  para lograr la conversión a diseño rotable, depende del número de puntos de la porción factorial del diseño. De hecho,  =  proporciona un diseño compuesto central rotable, donde  es el número de puntos en la porción factorial del diseño (de esta forma para 2 factores alfa es 1.4142, para 3 es 1.6818, para 4 es 2.0, para 5 es 2.3784). En el ejemplo se ilustra el uso de un diseño central compuesto rotable. Ese ejemplo tiene dos factores y la porción factorial contiene  = 22 = 4 puntos. Por lo tanto, el valor de á para lograr la invariabilidad ante el giro es  = (4)1/4 = 1.414.

Otra propiedad útil del diseño central compuesto es que puede "crecer" a partir de un diseño de primer orden (el 2k) agregando los puntos axiales y quizás algunos puntos centrales.

Continuando con el ejemplo del ingeniero, al existir curvatura, se decidió realizar un diseño de composición central como se muestra a continuación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables**  | **Naturales** | **Variables**  | **Codificadas** | **Respuesta** |
|  |  |  |  | **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** |
| **92.07** | **175** | **1.414** | **0** | **78.4** |
| **77.93** | **175** | **-1.414** | **0** | **75.6** |
| **85** | **182.02** | **0** | **1.414** | **78.5** |
| **85** | **167.93** | **0** | **1.414** | **77.0** |

donde

** y **

 El análisis de varianza para este diseño es el siguiente:

**Analysis of Variance for rendimiento**

**----------------------------------------------**

**Fuente SC G.L. C.M. Fo P-Value**

**----------------------------------------------**

**A:x1 7.91 1 7.9198 111.93 0.0000**

**B:x2 2.12 1 2.1231 30.01 0.0009**

**AA 13.17 1 13.1761 186.21 0.0000**

**AB 0.25 1 0.25 3.53 0.1022**

**BB 6.97 1 6.97 98.56 0.0000**

**error 0.49 7 0.0707**

**----------------------------------------------**

**Total 28.74 12**

Nótese que son significativos los efectos simples del tiempo y de la temperatura, no hay efecto de interacción, y son significativos los cuadráticos del tiempo y temperatura.

**El modelo de regresión para el problema es:**

**Rendimiento= 79.94 + 0.994976 x1+0.515166x2**

 **-1.37625 x1x1+0.25x1x2**

 **-1.00125x2x2**

La superficie de respuesta se muestra a continuación:



La gráfica de contornos es



La gráfica de superficie con los contornos es:

****

#### ANALISIS DEL MODELO DE REGRESION

Ahora analizaremos si el modelo de regresión es significativo

**Rendimiento= 79.94 + 0.994976 x1+0.515166x2**

 **-1.37625 x1x1+0.25x1x2 -1.00125x2x2**

**Conforme al análisis de varianza el modelo de regresión es significativo con un 95% de confianza,**

#### 3-7 LOCALIZACIÓN DEL PUNTO ESTACIONARIO

 Supongamos que se desea determinar los niveles X1, X2, X3,.....,Xk que maximizan la variable de respuesta predicha.

**Este máximo, si existe, será el conjunto de X1, X2, X3,.....,Xk , tal que las derivadas parciales** . **Dicho punto, es decir  se denomina punto estacionario.**

**El punto estacionario puede ser:**

**1.- *un punto de respuesta máxima***

***2.- un punto de respuesta mínima***

***3.- un punto silla***

Una solución general para el punto estacionario, es la siguiente:

****

**en donde**

**X= b=B=**

**B es una matriz simétrica (k x k) cuya diagonal principal esta formada por los coeficientes de los términos cuadráticos puros  y los elementos fuera de la diagonal corresponden a un medio del valor de los coeficientes cuadráticos mixtos ; b es el vector (kx1) de coeficientes de regresión de primer orden.**

**La derivada de *y* con respecto *x* e igualada a cero es**

**b +2Bx=0**

**derivando esta ecuación y resolviendo para *x* tenemos que el punto estacionario es**

****

**Para encontrar el punto estacionario del ejemplo de la ingeniera, se procede así:**

** **

****

 **=-**

**Evaluando el punto estacionario en el modelo de regresión:**

**Rendimiento= 79.94 + 0.994976 x1+0.515166x2**

 **-1.37625 x1x1+0.25x1x2**

 **-1.00125x2x2**

**Rendimiento= 79.94 + 0.994976\*(0.3890)+0.515166\*(0.3056)**

**-1.37625\*(0.3890)(0.3890)-1.00125\*(0.3056)(0.3056)+**

**0.25\*(0.3890)(0.3056)**

**Rendimiento= 80.21**

**Se obtendría un rendimiento del 80%, Con lo que la ingeniera logra duplicar el rendimiento del proceso.**

**3-8 CARACTERIZACION DE LA SUPERFICIE DE RESPUESTA**

**Una vez que se ha hallado el punto estacionario, es necesario caracterizar la superficie de respuesta en la vecindad inmediata de ese punto. Por caracterizar se entiende determinar si el punto estacionario es un punto de respuesta máxima o mínima o punto silla.**

**La forma más directa de hacer esto consiste en examinar la gráfica de contornos del modelo ajustado. Si solo hay dos o tres variables del proceso, la interpretación de esta gráfica resulta fácil. Sin embargo, incluso cuando hay relativamente pocas variables, resulta útil hacer un análisis formal.**

**Es conveniente primero transformar el modelo en un nuevo sistema de coordenadas con el origen en el punto estacionario  y entonces rotar (girar) los ejes de este sistema hasta que sean paralelos a los ejes principales de la superficie de respuesta ajustada. Esta transformación se ilustra en la siguiente figura:**

****

**Es posible demostrar que esto da por resultado el siguiente modelo ajustado:**



**Donde las {wi} son las variables independientes transformadas y las  son constantes. A esta ecuación se le conoce como forma canónica del modelo. Además, las  son justamente los valores propios (también llamados raíces características, autovalores, o eigenvalores) de la matriz B.**

**La naturaleza de la superficie de respuesta puede determinarse a partir del punto estacionario y el signo y la magnitud de las . Primero, supóngase que el punto estacionario se encuentra dentro de la región de exploración para el ajuste del modelo de segundo orden. Si todas las  son negativas entonces  es un punto de respuesta máxima; y si las  son todas positivas entonces  es un punto de respuesta mínima; y si las  son positivas y negativas entonces  es un punto silla.**

**Continuando con el análisis del ejemplo anterior, emplearemos el análisis canónico descrito en esta sección para caracterizar la superficie de respuesta.**

**Primero es necesario expresar el modelo ajustado en forma canónica. Los valores propios son las raíces de la ecuación determinante:**

****

**Lo que significa que el Determinante de , esto es:**

****

**que se reduce a:**



**Las raíces de esta ecuación cuadrática son  y , las cuales son negativas, lo que significa que el punto estacionario**

** es un punto de respuesta máxima.**

**Usando**

** y **

**tenemos que  y  obteniendo un máximo de 80.21**

**Salida del Statgraphics**

**Optimize Response**

**-----------------**

**objetivo: maximizar rendimiento**

**Optimum value = 80.2124**

**Factor Low High Optimum**

**----------------------------------------------**

**x1 -1.41421 1.41421 0.389595**

**x2 -1.41421 1.41421 0.305137**

**4. PRACTICAS**

Practica 1.

La Resistencia a la tensión de un producto de papel se relaciona con la cantidad de madera dura en la pulpa. Se producen 10 muestras en la planta piloto y los datos se presentan en la siguiente tabla.

|  |  |
| --- | --- |
| RESISTENCIA | PORCENTAJE DE |
| Y | MADERA DURA, X |
| 160 | 10 |
| 171 | 15 |
| 175 | 15 |
| 182 | 20 |
| 184 | 20 |
| 181 | 20 |
| 188 | 25 |
| 193 | 25 |
| 195 | 28 |
| 200 | 30 |
|  |  |

1. Ajustar el modelo de regresión lineal que relacione la resistencia con el porcentaje de madera dura.
2. Probar si el modelo es significativo.

Practica 2.

En una planta se destila aire liquido para producir oxigeno, nitrógeno y argón. Se piensa que le porcentaje de impurezas en el oxigeno se relaciona linealmente con la cantidad de impurezas en el aire, medida por el “conteo de contaminación” en partes por millón (ppm). Una muestra de los datos de operación de la planta se presenta a continuación.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PUREZA%, Y |   | 93.3 | 92 | 92.4 | 91.7 | 94 | 94.6 | 93.6 | 93.1 |
| CONTEO DE | CONTAMINACION, X | 1.1 | 1.5 | 1.36 | 1.59 | 1.08 | 0.75 | 1.2 | 0.99 |
|  |  |  |  |  |  |  |  |  |  |
| PUREZA%, Y |   | 93.2 | 93 | 92.2 | 91.3 | 90.1 | 91.6 | 91.9 |  |
| CONTEO DE | CONTAMINACION, X | 0.83 | 1.2 | 1.47 | 1.81 | 2.03 | 1.75 | 1.68 |  |

1. Ajustar el modelo de regresión lineal.
2. Probar si el modelo es significativo.

Practica 3.

Se piensa que la potencia al freno desarrollada por el motor de un automóvil en un dinamómetro es una función de la rapidez del motor en revoluciones por minuto (rmp), el octanaje del combustible y la compresión del motor. Se llevo a cabo un experimento en el laboratorio y los datos colectados fueron:

|  |  |  |  |
| --- | --- | --- | --- |
| POTENCIA AL FRENO, Y | RPM, X1 | OCTANAJE, X2 | COMPRENSION, X3 |
| 225 | 2000 | 90 | 100 |
| 212 | 1800 | 94 | 95 |
| 229 | 2400 | 88 | 110 |
| 222 | 1900 | 91 | 96 |
| 219 | 1600 | 86 | 100 |
| 278 | 2500 | 96 | 110 |
| 246 | 3000 | 94 | 98 |
| 237 | 3200 | 90 | 100 |
| 233 | 2800 | 88 | 105 |
| 224 | 3400 | 86 | 97 |
| 223 | 1800 | 90 | 100 |
| 230 | 2500 | 89 | 104 |

1. Ajustar el modelo de regresión múltiple.
2. Probar si el modelo es significativo.

Practica 4. En el área de desarrollo de una empresa se pretende obtener un nuevo polímero de bajo peso molecular (Y1), de lograrse esto, se obtendrá un polímero que funcione como dispersante en la industria de la cerámica. De acuerdo a conocimientos técnicos que se tienen, se consideran que los factores críticos son: X1, Persulfato de Sodio (NaPS), X2, Acido hipofosforoso (H3PO2) y X3, Isopropanol (IPA). Para encontrar las condiciones optimas se realizo un experimento y se obtuvieron los siguientes datos (los valores de los factores están codificados). Además de la variable Y1, se midió la Viscosidad ( Y2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Y1 | Y2 | X1 | X2 | X3 |
| 8.392 | 1.08 | 0 | 0 | 0 |
| 9.895 | 2.33 | -1 | -1 | 0 |
| 9.204 | 1.58 | 1 | -1 | 0 |
| 7.882 | 0.69 | -1 | 1 | 0 |
| 7.105 | 0.42 | 1 | 1 | 0 |
| 8.939 | 1.19 | -1 | 0 | -1 |
| 8.548 | 0.93 | 1 | 0 | -1 |
| 8.598 | 0.92 | 0 | 0 | 0 |
| 9.152 | 1.28 | -1 | 0 | 1 |
| 8.992 | 0.86 | 1 | 0 | 1 |
| 10.504 | 5.6 | 0 | -1 | -1 |
| 7.462 | 5.4 | 0 | 1 | -1 |
| 9.368 | 1.23 | 0 | -1 | 1 |
| 7.772 | 0.62 | 0 | 1 | 1 |
| 8.44 | 1.02 | 0 | 0 | 0 |

1. Ajustar el modelo de regresión múltiple tanto para Y1 y Y2.
2. Probar si el modelo es significativo en ambas variables.

Practica 5. Se desea investigar la relación entre el peso de un individuo y su presión sanguínea sistólica. Para ello se seleccionan aleatoriamente 26 hombres cuyas edades fluctúan entre 25 y 30 años.

|  |  |  |  |
| --- | --- | --- | --- |
|  X |  Y |  X |  Y |
| 165167180155212175190210200149158169170 | 130133150128151146150140148125133135150 | 172159168174183215195180143240235192187 | 153128132149158150163156124170165160159 |

II

1. Mediante un diagrama de dispersión describa la relación entre ambas variables. ¿Qué tipo de relación observa?

2. Obtenga el coeficiente de correlación e interprételo.

3. Obtenga la mejor recta que modela la relación peso - presión sanguínea.

4. Si un hombre de entre 25 y 30 años de edad pesa 150 libras, según el modelo, ¿cuál sería su presión media? ¿La estimación es confiable? Argumente.

5. ¿El modelo obtenido sería útil para estimar la presión sanguínea de otro tipo de individuos, por ejemplo, mujeres, niños, ancianos, etc.?

Practica 6.

Se busca encontrar los niveles de tiempo (t) y temperatu­ra (T) que maximizan rendimien­to. Las condiciones actuales son t=75 min. y de T=1300C. Para explorar la superficie de respuesta entorno a estos valores, se corre el siguiente diseño experimental.

|  |  |  |
| --- | --- | --- |
|  Variables Natu­ra­les |  Variables Codifi­ca­das |  |
| T­iem­po | Tempe­rat. | x1 | x2 | Y |
| 70 | 127.5 | -1 | -1 | 54.3 |
| 80 | 127.5 | 1 | -1 | 60.3 |
| 70 | 132.5 | -1 | 1 | 64.6 |
| 80 | 132.5 | 1 | 1 | 68.0 |
| 75 | 130.0 | 0 | 0 | 60.3 |
| 75 | 130.0 | 0 | 0 | 64.3 |
| 75 | 130.0 | 0 | 0 | 62.3 |

V

a. Ajuste el modelo de primer or­den; ¿es adecuado (curvatura, R2)? explique.

b. Anote la ecuación del modelo con el que se enconara la trayec­to­ria de máximo ascenso.

c. Con base a la trayectoria de máximo crecimiento; proponga en qué niveles de t y T recomendaría experi­mentar.

|  |  |  |
| --- | --- | --- |
|  Variables Natu­ra­les |  Variables Codifica­das |   |
| Tiem­po | Tempe­rat. | x1 | x2 | Y |
| 80 | 140 | -1 | -1 | 78.8 |
| 100 | 140 | 1 | -1 | 84.5 |
| 80 | 150 | -1 | 1 | 91.2 |
| 100 | 150 | 1 | 1 | 77.4 |
| 90 | 145 | 0 | 0 | 89.7 |
| 90 | 145 | 0 | 0 | 86.8 |

VI

d. Siguiendo la trayectoria de máximo crecimiento, el último punto con el que se obtuvo una respuesta alta fue t=90 y T=1­45. Por lo que entorno a este se co­rrió un nuevo diseño. Ajuste un modelo de primer orden, y vea si este describe ade­cuada­mente la superficie de respuesta.

e. Se agregaron otras 6 para completar un diseño de composición central.e. Ajuste un modelo de 2do. orden, vea la calidad y encuentre las condiciones que optimizan.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 76 | 145 | -1.414 | 0 | 83.3 |
| 104 | 145 | 1.414 | 0 | 81.2 |
| 90 | 138 | 0 | -1.414 | 81.2 |
| 90 | 152 | 0 | 1.414 | 79.5 |
| 90 | 145 | 0 | 0 | 87.0 |
| 90 | 145 | 0 | 0 | 86.0 |

Practica 7. Los datos que se presentan en la si­guiente tabla fueron recopilados en un experimento para optimizar el creci­miento de cristales en función de tres variables x1, x2 y x3. Se desean valores altos de Y (rendimiento en gramos).

|  |  |  |  |
| --- | --- | --- | --- |
| x1  |  x2 | x3 | Y |
| -1 | -1 | -1 | 66 |
| -1 | -1 | 1 | 70 |
| -1 | 1 | -1 | 78 |
| -1 | 1 | 1 | 60 |
| 1 | -1 | -1 | 80 |
| 1 | -1 | 1 | 70 |
| 1 | 1 | -1 | 100 |
| 1 | 1 | 1 | 75 |
| -1.682 | 0 | 0 | 100 |
| 1.682 | 0 | 0 | 80 |
| 0 | -1.682 | 0 | 68 |
| 0 | 1.682 | 0 | 63 |
| 0 | 0 | -1.682 | 65 |
| 0 | 0 | 1.682 | 82 |
| 0 | 0 | 0 | 113 |
| 0 | 0 | 0 | 100 |
| 0 | 0 | 0 | 118 |
| 0 | 0 | 0 | 88 |
| 0 | 0 | 0 | 100 |
| 0 | 0 | 0 | 85 |

VII

a. Realice un análisis de varianza para un mode­lo de segundo orden. Determine los componentes signifi­cativos y más impor­tantes.

b. En qué condiciones se logra el óptimo y que valor se espera de Y.

c. Verifique la cali­dad del modelo (R2, resisuales).

Problema 2. Los datos que se presentan en la si­guiente tabla fueron recopilados en un experimento para optimizar el creci­miento de cristales en función de tres variables x1, x2 y x3. Se desean valores altos de Y (rendimiento en gramos).

|  |  |  |
| --- | --- | --- |
| Tempe­ra­tura (1) | Razón de pre­sión (2) | Pureza |
| -225 | 1.1 | 82.8 |
| -225 | 1.3 | 83.5 |
| -215 | 1.1 | 84.7 |
| -215 | 1.3 | 85.0 |
| -220 | 1.2 | 84.1 |
| -220 | 1.2 | 84.5 |
| -220 | 1.2 | 83.9 |
| -220 | 1.2 | 84.3 |

VIII

Practica 8. Una planta química produce oxíge­no median­te la licuefacción de aire y sepa­rándolo en sus gases componen­tes me­diante destilación fraccio­nada. La pure­za del oxígeno es función de la tempera­tura del condensador principal y de la relación de presión entre las columnas supe­rior e inferior. Las condicio­nes actuales de opera­ción son temperatura (11) = -220oC y relación de presión (22) =1.2. Usando los siguientes datos determine la trayectoria de ascenso máxi­mo.

Practica 9. Con el propósito de mejorar el rendimiento de un proceso, se decide correr un diseño de composición central, teniendo como facores temperatura (125 nivel actual) y presión (25 nivel actual). El diseño y los datos obtenidos se muestran a continuación.

Temp Presión Rendimiento

--‑‑‑‑- ‑‑‑‑‑‑‑ ‑‑‑‑‑‑‑‑-‑

 110.0 15.0 35.

 140.0 15.0 51.

 110.0 35.0 41.

 140.0 35.0 50.

 103.8 25.0 42.

 146.2 25.0 65.

 125.0 10.9 43.

 125.0 39.1 45.

 125.0 25.0 62.

 125.0 25.0 63.

 125.0 25.0 64.

 125.0 25.0 61.

1

a. Ajuste un modelo de segundo orden a los datos experimentales, con base al anova señale que términos son significativos.

b. ¿El modelo de segundo orden describe adecuada­mente la superficie? Explique.

c. Anote la ecuación del modelo ajustado.

d. Apoyándose en la gráfica superficie de respuesta y la gráfica de contornos de ésta, encuentre los niveles de temperatura y presión que maximizan la respuesta.

Practica 10. En el anodizado de artículos de aluminio mediante una solución de ácidos (sulfúrico, cítrico, bórico) y dicromato de aluminio; bajo ciertas condiciones de PH, temperatura, corriente y tiempo de permanencia. Se ha tenido el problema de un bajo grosor en el anodizado, lo cual genera piezas defectuosos, problemas de resistencia y durabilidad. Para abordar el problema se estudiar la influencia del PH y la temperatura, sobre el espesor del anodizado, mediante el siguiente diseño:

|  |  |  |
| --- | --- | --- |
|  PH | Temperatur | Espesor |
| 1.21.81.21.81.21.81.21.81.51.5 | -8-888-8-88800 | 914101981211201413 |

IX

a. Señale el nombre del diseño utilizado

b. ¿Un modelo de primer orden describe adecuadamente la superficie? (vea calidad de ajuste) Argumente.

c. Especifique el modelo de primer orden junto con sus coeficientes estimados.

d. Con el modelo anterior encuentre cinco puntos de la trayectoria de máximo crecimiento (utilice PH como factor y la longitud de paso igual a 0.2).

e. Explique que haría con los puntos encontrados en el inciso anterior.

Practica 11. Con el propósito de mejorar el rendimiento de un proceso, se decide correr un diseño de composición central, teniendo como factores temperatura (125 nivel actual) y presión (25 nivel actual).

Temp Presión Rendimiento

--‑‑‑‑- ‑‑‑‑‑‑‑ ‑‑‑‑‑‑‑‑-‑

 110.0 15.0 35.

 140.0 15.0 51.

 110.0 35.0 41.

 140.0 35.0 50.

 103.8 25.0 42.

 146.2 25.0 65.

 125.0 10.9 43.

 125.0 39.1 45.

 125.0 25.0 62.

 125.0 25.0 63.

 125.0 25.0 64.

 125.0 25.0 61.

2

a. Ajuste un modelo de segundo orden a los datos experi­mentales, ¿qué términos son significati­vos?

b. ¿El modelo de segundo orden describe adecua­damente la superficie? Explique.

c. Anote la ecuación del modelo ajustado.

d. Apoyándose en la gráfica superficie de res­puesta y la gráfica de contornos de ésta, encuen­tre los niveles de temperatura y presión que maximizan la respuesta.

Practica 12. En busca de optimizar el rendimiento de un proceso químico se decide correr un diseño de segundo orden, teniendo como factores a temperatura, presión y tiempo de residencia; y a la variable de respuesta a rendimiento. El diseño empleado y los resultados obtenidos se muestran a continuación.

|  |  |  |  |
| --- | --- | --- | --- |
| Temp. | Pre­sión | T­iem­po | Rend. |
|  90.0 120.0 90.0 120.0 90.0 120.0 90.0 120.0 79.8 130.2 105.0 105.0 105.0 105.0 105.0 105.0 105.0 105.0 |  10.0 10.0 30.0 30.0 10.0 10.0 30.0 30.0 20.0 20.0 3.2 36.8 20.0 20.0 20.0 20.0 20.0 20.0 |  25.0 25.0 25.0 25.0 35.0 35.0 35.0 35.0 30.0 30.0 30.0 30.0 21.6 38.4 30.0 30.0 30.0 30.0 |  36 52 45 47 44 58 41 50 45 66 42 46 30 60 59 60 62 58 |

X

(a) Señale el nombre del diseño utilizado

(b) ¿Este tipo de diseño es recomendable para las primeras fases de la optimización de un proceso (o en general de la investigación)?

(c) Realice el análisis de varianza para este experi­mento, de acuerdo a éste, ¿existe algún factor que no influye en el rendimiento? Argumente

(d) Ajuste un modelo de segundo orden, depurelo eliminando términos no significativos, y conteste:

 (d1) ¿El modelo describe adecuada­mente la superficie?

 (d2) ¿Cuál es la ecuación del modelo?

(e) Analice los residuales, ¿observa alguna anomalía?

(f) Con base a la gráfica de contornos, trate de encontrar los niveles de los factores que maximiza el rendimiento del proceso.

(g) En análisis de regresión encuentre el mejor modelo de segundo orden, utilizando el método Backward. Reporte ecuación del modelo y sus indicadores de calidad de ajuste (R2, SE, MAE), explicando el significado práctico de cada uno de ellos.

(h) Haga lo mismo que en el ejercicio anterior pero ahora con el método Forward.

(i) ¿Con cuál de los tres modelos (incisos d, g y h) se queda con propósitos de pronósti­co? Argu­mente.

(j) Utilizando el modelo seleccionado, vaya al menú de diseño de experimentos, depure el modelo hasta hacerlo coincidir y encuentre cinco puntos de la trayectoria de máximo crecimiento.

(k) ¿Qué haría con los puntos encontrados?

(l) Con el modelo seleccionado, calcule el rendimien­to del proceso en el punto central del diseño.