

01167

11
24



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE INGENIERIA
DIVISION DE ESTUDIOS DE POSGRADO

METODOLOGIA PARA MODELAR SUPERFICIES
DE RESPUESTA.

T E S I S

PRESENTADA POR:
GERARDO NUÑEZ MEDINA
PARA OBTENER EL GRADO DE
MAESTRO EN INGENIERIA
(PLANEACION)

DIRIGIDA POR:
RUBEN TELLEZ SANCHEZ

CIUDAD UNIVERSITARIA,

1996

**TESIS CON
FALLA DE ORIGEN**

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

A mis Padres:

Sergio Núñez y Rosa Medina

A mis Hermanos:

Sergio, Alejandro y Carmina

A Rosario E., y a mis amigos

METODOLOGÍA PARA MODELAR SUPERFICIES DE RESPUESTA

Gerardo Núñez Medina

Agosto, 96

Índice de figuras

1	Curva de respuesta en un espacio bidimensional	7
2	Superficie de respuesta	9
1	Diseño 2^3 factorial	49
1	Camino de máximo ascenso	62
1	Diseño Central Compuesto $2^3 + 7$	66
2	Punto silla sobre la superficie ajustada	68
3	Máximo sobre la superficie ajustada	69
4	Mínimo sobre la superficie ajustada	69
5	Sistema de cordillera estacionaria	69
6	Sistema de cordillera ascendente	69
7	Curvas de nivel con ejes w rotados	72
8	Sistema estacionario	72
9	Región de respuesta acotada	73
1	curvas de nivel de la función 7.1	78
2	curvas de nivel de la función 7.2	78

Contenido

1	Introducción	7
1.1	Modelos Empíricos y Mecanicistas	8
2	Antecedentes	11
2.1	Conceptos Matemáticos	11
2.2	Conceptos Estadísticos	14
2.3	Estimación Puntual	21
2.3.1	Estimación por Intervalo	25
2.4	Prueba de Hipótesis	25
3	Estimación de Parámetros, Análisis de Varianza y Diseños Experimentales	27
3.1	Supuestos del Modelo de Regresión Lineal Múltiple	27
3.2	Mínimos Cuadrados	28
3.3	Propiedades de los Coeficientes Estimados	29
3.3.1	Teorema de Gauss Markoff	30
3.4	Máxima Verosimilitud	31
3.5	Estimación por Intervalo	36
3.5.1	Intervalo de Confianza para σ^2	36
3.5.2	Intervalo de Confianza para β_i	36

3.6	Prueba de Hipótesis	36
3.7	La Respuesta Estimada	40
3.8	Los Residuos	41
3.9	Estimación de σ^2	42
3.10	La Tabla de Análisis de Varianza	42
3.11	El Coeficiente de Determinación	43
3.12	La Elección del Diseño	46
4	Modelos y Diseños de Primer Orden	47
4.1	Diseños 2^k Factoriales	48
4.1.1	Análisis del Diseño 2^k Factorial	49
4.2	Diseños 2^{k-p} Factoriales	50
4.2.1	Cuatro y Ocho Bloques	50
4.2.2	Análisis del diseño 2^{k-p} factorial	52
4.2.3	Resolución de un Diseño 2^{k-p} Factorial	53
4.3	Diseños Factoriales Fraccionados Aplicados a la Metodología de Superficies de Respuesta	54
4.3.1	Papel de Probabilidad Normal	55
4.3.2	Gráfica de Residuos Contra Valores Estimados \hat{Y}_i	55
5	Máxima Pendiente de Ascenso	57
5.1	¿Qué Hace el Paso de Máximo Ascenso?	58
5.2	Experimentación a lo Largo del Camino de Máximo Ascenso	59
5.2.1	La longitud de Paso t	62
5.2.2	Localización del Centro Para un Nuevo Diseño	63
5.3	Algoritmo de Máxima Pendiente de Ascenso	63

CONTENIDO	5
6 Diseños para Ajustar el Modelo de Segundo Orden	65
6.1 Diseño Central Compuesto	65
6.2 Optimización Sobre la Superficie de Respuesta	67
6.2.1 Determinación de un Punto Óptimo	67
6.3 Transformación Canónica	70
7 Estudio de Simulación	75
7.1 Generación de Muestras	76
7.2 Resultados de la Simulación	77
7.3 Conclusiones	79
7.4 Recomendaciones	80
8 Bibliografía	81

TESIS

COMPLETA

Capítulo 1 INTRODUCCIÓN

La metodología de superficies de respuesta (MSR) es un conjunto de técnicas matemáticas y estadísticas útiles para modelar y analizar fenómenos en los cuales una respuesta de interés es influenciada por un conjunto de factores controlables, de esta forma se busca modelar el comportamiento de alguna respuesta a través de variar los niveles de los diversos factores controlables. En términos matemáticos, podemos decir que el investigador está interesado en suponer una relación

$$\eta = f(\xi_1, \xi_2, \dots, \xi_k) \quad (1.1)$$

que exprese el nivel medio de respuesta η como función de las variables $(\xi_1, \xi_2, \dots, \xi_k)$, así la relación funcional entre la respuesta media y los niveles de las k variables de entrada quedará expresada como en (1.1).

Si hay sólo una variable de entrada ξ_1 , podemos relacionar la salida η a través de una curva de respuesta en un espacio bidimensional, como en la Figura 1. Si hay dos entradas ξ_1 y ξ_2 y se dibuja en un espacio tridimensional una gráfica de η contra ξ_1 y ξ_2 , se obtiene una superficie de respuesta como la que se observa en la figura 2. Cuando se tienen k variables de entrada, se hablará de una superficie de respuesta en un espacio $k + 1$ dimensional.

Obsérvese en la figura 2, que la superficie de respuesta se representa como una superficie sólida en un espacio $k + 1$ dimensional. Para visualizar mejor la forma de una superficie de respuesta, a menudo se grafican los contornos (curvas de nivel) de dicha superficie. Cada contorno corresponde a una altura específica de la superficie de respuesta. Esta gráfica es útil para estudiar los niveles de $\xi_1, \xi_2, \dots, \xi_k$ que dan por resultado cambios en la forma o altura de la superficie de respuesta.

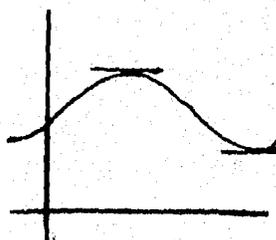


Figura 1 Curva de respuesta en un espacio bidimensional

La operación en un sistema con k variables de control ajustadas a un cierto nivel $\xi = (\xi_1, \xi_2, \dots, \xi_k)$, es llamada una corrida experimental. Si hacemos varias corridas experimentales a un nivel fijo de $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_k)$, la respuesta obtenida variará debido a errores de medición, de observación o de experimentación. Si consideramos η como la respuesta media en condiciones fijas $\hat{\xi}$, una respuesta obtenida en una corrida resultará ser una y , la cual caerá dentro de alguna distribución estadística alrededor de su valor medio η . Así el valor esperado de y es igual a $\eta = E(y)$. En una corrida particular, la diferencia $(y - \eta)$ observada entre el valor y y la media hipotética η será el error y se denotará por ε . Así, el objetivo será investigar el comportamiento de la relación funcional afectada por el error ε . Tal relación funcional se representara como:

$$y = f(\hat{\xi}) + \varepsilon.$$

En este punto se debe considerar que el investigador tratará con factores (variables) experimentales cuya naturaleza es necesariamente subjetiva. Tales factores deben ser definidos dentro de una región de exploración, esto es, una región en el cual los experimentos puedan ser conducidos. Las escalas de medición o métricas con o sin transformaciones de las variables, la localización de la región de interés, la especificación del modelo sobre la región, y el modelo experimental deben ser definidos al momento de explorar la región de interés.

1.1 Modelos Empíricos y Mecanicistas

Consideraremos algunas clases de problemas y la forma del modelo usado para resolver estos problemas. Se supone que en el estudio de algún fenómeno físico, se conoce lo suficiente de su mecanismo físico como para deducir la forma de la relación funcional que liga el valor medio η con los niveles de las entradas $(\xi_1, \xi_2, \dots, \xi_k)$ vía la expresión:

$$\eta = E(y) = f(\hat{\xi}; \bar{\theta}) \quad (1.2)$$

Donde $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_k)'$ representa un conjunto de parámetros físicos. Entonces se puede decir que (1.2) representa un modelo mecanicista (todos los parámetros involucrados en el modelo se encuentran perfectamente identificados).

A menudo, sin embargo, el conocimiento físico necesario del sistema está ausente o incompleto y no se puede disponer de un modelo mecanicista. En estas circunstancias, es más útil suponer que la relación entre η y $(\xi_1, \xi_2, \dots, \xi_k)$ es suave y por consiguiente, que $f(\hat{\xi}; \bar{\theta})$ puede ser aproximada localmente (sobre un rango limitado de $\hat{\xi}$) por una función de interpolación $g(\xi; \beta)$, tal como un polinomio, donde las β 's serán los coeficientes de la función de interpolación, las cuales serán consideradas de forma distinta a los parámetros $\bar{\theta}$ del modelo físico. La función de interpolación $g(\xi; \beta)$ puede proporcionar un modelo empírico local del sistema.

De este modo los modelos mecanicistas teóricos $\eta = f(\xi, \theta)$, y los modelos puramente empíricos $\eta \simeq g(\xi, \beta)$, representan dos casos extremos. El primero será apropiado en el caso

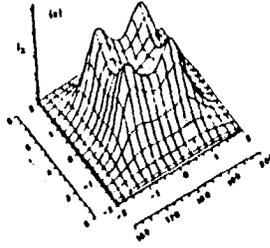


Figura 2 Superficie de respuesta

Paso	Desconocido	Objetivo	Metodología
Cuáles?	f ξ θ	Determinar el subconjunto de variables ξ potencialmente importantes para el problema	Búsqueda de variables
Cómo?	f θ	Determinar empíricamente el efecto de la variable de entrada conocida ξ	Construcción del modelo empírico
Porqué?	f θ	Estimar una función de interpolación local $g(\xi, \beta)$ para $f(\xi, \theta)$	Superficie de respuesta
	f θ	Determinar f	Construcción del modelo mecanicista
	θ	Determinar θ	Ajustar el modelo mecanicista

Tabla 1.1 Clasificación de problemas científicos

extremo en el que se tenga un conocimiento muy amplio y preciso acerca del sistema, y el último será apropiado en el otro caso extremo, donde nada puede ser supuesto, excepto que la superficie de respuesta es localmente suave. Debido a que los problemas reales ocurren casi siempre entre los puntos extremos antes mencionados, es necesaria una gran variedad de herramientas estadísticas para hacer frente a ellos.

Ambos, el estado de ignorancia en el cual se comienza a trabajar el experimento, y el estado de relativo conocimiento al cual se desea llegar, determinarán el grado de acercamiento. No obstante, es útil distinguir los tipos básicos de problemas, los cuales están categorizados en términos de lo que se desconoce sobre el verdadero modelo mecanicista en la Tabla 1.1. Para propósitos de referencia se asigna una columna de *paso* la cual indica el punto en el que se encuentra la investigación.

¿Cuáles? A menudo sucede que al inicio de una investigación hay demasiadas variables ξ_1, ξ_2, \dots ¿Cuáles serán de importancia para η ? Una forma de reducir la lista es sentarse con el investigador y preguntarle por las variables que él cree son las más importantes.

¿Cómo? Cuando las variables son cuantitativas y el error experimental no es demasiado grande comparado con el rango cubierto por la respuesta observada, puede ser provechoso intentar estimar la función respuesta dentro de alguna área de interés inmediata. En muchos problemas, la forma de la verdadera función respuesta $f(\xi, \theta)$ es desconocida y no puede ser obtenida de manera económica, pero quizá pueda ser aproximada localmente por un polinomio $g(\xi, \beta)$. La naturaleza iterativa de la metodología de superficies de respuesta asegurará que será posible aprender a cerca de: la cantidad necesaria de réplicas para alcanzar la precisión suficiente; la localización de la región experimental de mayor interés; apropiadas escalas y transformaciones para las variables de entrada y salida; y el grado de complejidad de una función de aproximación.

¿Porqué? Si fuera posible, usaríamos la verdadera forma funcional $f(\xi, \theta)$ para representar la respuesta, en vez de aproximarla por medio de un polinomio. En algunos problemas, se ajustan modelos mecánicos, los cuales, al menos, informan sobre las principales características de los mecanismos. Un modelo mecánico tiene las siguientes ventajas: contribuye al entendimiento científico del fenómeno bajo estudio; proporciona una mejor base para la extrapolación; tiende a usar menos parámetros y a encontrar mejores estimaciones de la respuesta.

La forma de la relación entre la variable de respuesta y las variables independientes en la mayoría de los problemas de MSR es desconocida. Por ello, el primer paso consiste en determinar una aproximación apropiada a la relación funcional real entre y y el conjunto de variables independientes $(\xi_1, \xi_2, \dots, \xi_k)$. Por lo general se emplea un polinomio de bajo orden sobre *alguna región* de las variables independientes, esto es, se ajustan polinomios de primero y segundo grado, que son llamados *diseños de primer orden y segundo orden*.

Capítulo 2 ANTECEDENTES

2.1 Conceptos Matemáticos

El propósito de este capítulo es revisar algunos conceptos básicos del álgebra matricial y de la estadística que serán necesarios para el desarrollo de diversos tópicos, de este modo se listarán algunas definiciones y propiedades:

Definición 2.1 Una matriz M de $p \times q$ es un arreglo rectangular de números

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1q} \\ m_{21} & m_{22} & \cdots & m_{2q} \\ \vdots & \vdots & & \vdots \\ m_{p1} & m_{p2} & \cdots & m_{pq} \end{bmatrix}$$

que contiene p renglones y q columnas y se denota como $M_{p \times q}$.

Se dice que dos matrices son iguales si y sólo si tienen la misma dimensión y son idénticas entrada a entrada. Así, la igualdad de matrices implica tantas igualdades como entradas haya lugar por lugar.

Definición 2.2 Una matriz cuadrada es una matriz en la que el número de renglones es igual al número de columnas.

Definición 2.3 La transpuesta de una matriz M de $p \times q$ es una matriz M' de $q \times p$, obtenida de intercambiar los renglones por las columnas

$$\begin{bmatrix} m_{11} & m_{21} & \cdots & m_{q1} \\ m_{12} & m_{22} & \cdots & m_{q2} \\ \vdots & \vdots & & \vdots \\ m_{1p} & m_{2p} & \cdots & m_{qp} \end{bmatrix}$$

Definición 2.4 Una matriz es simétrica si es igual a su transpuesta, esto es, si $M = M'$, entonces, M es simétrica.

Teorema 2.1 La matriz transpuesta de M' es igual a M ; es decir, $(M')' = M$.

Definición 2.5 Una matriz diagonal, es una matriz cuadrada cuyos elementos m_{ij} , $i \neq j$ fuera de la diagonal son iguales a cero

$$M = \begin{bmatrix} m_{11} & 0 & \cdots & 0 \\ 0 & m_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & m_{kk} \end{bmatrix} = \text{diag}(m_{11}, m_{22}, \dots, m_{kk}).$$

Definición 2.6 Un escalar es un número (una matriz de 1×1).

Definición 2.7 La matriz identidad I es una matriz diagonal con $m_{ii} = 1$ para toda $i = 1, 2, \dots, k$.

Definición 2.8 Si A es una matriz de $k \times k$, entonces la traza de A es la suma de los elementos de la diagonal y se denota por $\text{tr}(A) = \sum_{i=1}^k a_{ii}$.

Definición 2.9 El determinante de toda matriz M cuadrada es un escalar denotado por $|M|$ y definido como:

$$|m_{11}| = m_{11}$$

$$\begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix} = m_{11}m_{22} - m_{12}m_{21}$$

$$\begin{vmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{vmatrix} = m_{11}m_{22}m_{33} + m_{12}m_{23}m_{31} + m_{13}m_{21}m_{32} - m_{13}m_{22}m_{31} - m_{11}m_{23}m_{32} - m_{12}m_{21}m_{33}$$

Teorema 2.2 La matriz inversa de A^{-1} es igual a A ; es decir, $(A^{-1})^{-1} = A$.

Teorema 2.3 Los símbolos de la matriz transpuesta e inversa pueden ser permutados, es decir, $(A^{-1})' = (A')^{-1}$.

Teorema 2.4 $(AB)' = B'A'$.

Teorema 2.5 $(AB)^{-1} = B^{-1}A^{-1}$.

Teorema 2.6 Si A es cualquier matriz, entonces $AI = IA = A$.

Teorema 2.7 Si A es una matriz de $n \times n$ y $\text{rango}(A) < n$, entonces las columnas de A son linealmente dependientes.

Teorema 2.8 Si A es una matriz de $n \times n$ y $\text{rango}(A) = m < n$, entonces el número de columnas linealmente independientes de A es m .

Teorema 2.9 Sean A, B dos matrices, entonces $\text{Tr}(AB) = \text{Tr}(BA)$.

Definición 2.10 Una matriz es idempotente si y sólo si $A = A'$ y $A = AA = A^2$.

Teorema 2.10 Cualquier matriz simétrica $B_{k \times k}$ puede escribirse como

$$B = U\Lambda U' = \sum_{i=1}^k \delta_i U_{(i)} U'_{(i)},$$

donde $\Lambda = \text{Tr}(\delta_1, \delta_2, \dots, \delta_k)$; $\delta_1, \delta_2, \dots, \delta_k$ son los valores propios de la matriz B y U es una matriz ortogonal cuyas columnas son los vectores propios de B .

Prueba 2.1 Por definición, se tiene que $U_{(i)}$ es el vector propio asociado al valor propio δ_i de la matriz B , si $BU_{(i)} = \delta_i U_{(i)}$. La manera de calcular los vectores y valores propios, es resolver el polinomio característico en δ , $P(\delta)$, que resulta de calcular el determinante de la matriz $(B - \delta I)$ igualado a cero. Dado que B es una matriz simétrica, por álgebra lineal se sabe que todos sus valores propios son reales. Ahora para cada valor propio δ_i se busca el vector $U_{(i)}$ que resuelva la ecuación $(B - \delta_i I)U_{(i)} = 0$, y se elige cada $U_{(i)}$ de norma unitaria y así para todo $i = 1, 2, \dots, k$. Además se pide que $U'_{(i)}U_{(i)} = 0$, para todo $i \neq j$, $i, j = 1, 2, \dots, k$, lo cual se logra mediante el proceso de ortogonalización de Gram-Schmit. De este modo, definimos

$$\Lambda = \text{Tr}(\delta_1, \delta_2, \dots, \delta_k) \text{ y } U = [U_{(1)}, U_{(2)}, \dots, U_{(k)}].$$

Así

$$\begin{aligned} BU &= [BU_{(1)}, BU_{(2)}, \dots, BU_{(k)}] \\ &= [\delta_1 U_{(1)}, \delta_2 U_{(2)}, \dots, \delta_k U_{(k)}], \end{aligned}$$

por definición. De este modo

$$\begin{aligned} U'BU &= \begin{bmatrix} U'_{(1)} \\ U'_{(2)} \\ \vdots \\ U'_{(k)} \end{bmatrix} [\delta_1 U_{(1)}, \delta_2 U_{(2)}, \dots, \delta_k U_{(k)}] \\ &= \text{Tr}(\delta_1, \delta_2, \dots, \delta_k), \end{aligned}$$

dado que $U'_{(i)}U'_{(j)} = 0$ para toda $i \neq j$ y $U_{(i)}$ es el vector de norma unitaria, lo que implica que $U'_{(i)}U_{(i)} = 1$, tenemos que

$$B = U \Lambda U' \text{ con } U'U = I \quad \square$$

2.2 Conceptos Estadísticos

Definición 2.11 El espacio muestral de un experimento es el conjunto de todos los posibles resultados del experimento y se denotará por Ω .

Definición 2.12 Un evento es cualquier subconjunto de Ω , esto es, si $A \subset \Omega$, entonces, A es un evento.

Definición 2.13 Una colección de subconjuntos de Ω es llamada un sigma álgebra, denotada por σ , si satisface:

1. - $\emptyset \in \sigma$
2. - Si $A \in \sigma$, entonces $A^c \in \sigma$
3. - Si $A_1, A_2, \dots \in \sigma$, entonces $\bigcup_{i=1}^{\infty} A_i \in \sigma$.

Definición 2.14 Dos eventos A y B son disjuntos a pares si $A \cap B = \emptyset$. Los eventos A_1, A_2, \dots son disjuntos a pares si $A_i \cap A_j = \emptyset$ para toda $i \neq j$.

Definición 2.15 Dado un espacio muestral Ω y un sigma álgebra σ asociado a Ω , una función de probabilidad P con dominio Ω satisface

1. - $P(A) \geq 0 \forall A \in \sigma$
2. - $P(\Omega) = 1$
3. - Si $A_1, A_2, \dots \in \sigma$ son disjuntos a pares, entonces $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Teorema 2.11 Si P es una función de probabilidad y $A \in \sigma$, entonces

1. - $P(\emptyset) = 0$
2. - $P(A) \leq 1$
3. - $P(A^c) = 1 - P(A)$.

Definición 2.16 Si $A, B \in \Omega$, y $P(B) > 0$, entonces la probabilidad condicional de A dado B , denotada por $P(A|B)$, es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Definición 2.17 Una colección de eventos A_1, A_2, \dots, A_n son mutuamente independientes si para toda subcolección $A_{i_1}, A_{i_2}, \dots, A_{i_k}$, tenemos que:

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

Definición 2.18 Dado un espacio muestral Ω , un sigma álgebra σ asociado a Ω y una función de probabilidad P con dominio en Ω . Una variable aleatoria denotada por X o $X(\cdot)$ es una función con dominio en Ω y contradominio en la recta real.

Definición 2.19 La función de distribución* acumulada de una variable aleatoria X , denotada por $F_X(x)$, es definida por:

$$F_X(x) = P_X(X \leq x), \text{ para toda } x \in \mathfrak{R}.$$

Teorema 2.12 La función $F_X(x)$ es una función de distribución acumulada si y sólo si cumple las siguientes tres condiciones:

1. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ y $\lim_{x \rightarrow \infty} F_X(x) = 1$
2. $F_X(x)$ es una función no decreciente de x
3. $F_X(x)$ es continua por la derecha.

Definición 2.20 Una variable aleatoria X es continua si $F_X(x)$ es una función continua de x . Una variable aleatoria X es discreta si $F_X(x)$ es una función escalonada de x .

Definición 2.21 Las variables aleatorias X y Y son idénticamente distribuidas si para todo conjunto A , $P(X \in A) = P(Y \in A)$.

Definición 2.22 La función de masa de probabilidad $f_X(x)$ de una variable aleatoria discreta X esta dada por

$$f_X(x) = P(X = x) \text{ para toda } x.$$

Definición 2.23 La función de densidad de probabilidad $f_X(x)$ de una variable aleatoria X continua, es una función que satisface

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \text{ para toda } x.$$

*la expresión X tiene una distribución dada por $F_X(x)$ es abreviada por " $X \sim F_X(x)$ ", donde el símbolo " \sim " se lee "se distribuye como".

Teorema 2.13 Una función $f_X(x)$ es una función de densidad de probabilidad de una variable aleatoria X si y sólo si

1. - $f_X(x) \geq 0$ para toda x
2. - $\sum_x f_X(x) = 1$ o $\int_{-\infty}^{\infty} f_X(x) = 1$.

Definición 2.24 Si X es una variable aleatoria con función de densidad de probabilidad

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-(x-\mu)^2/2\sigma^2} \quad \text{donde } -\infty < \mu < \infty \text{ y } \sigma > 0$$

entonces X tiene una función de distribución normal.

Teorema 2.14 Si Y esta distribuida como una normal con media 0 y matriz de covarianza $\sigma^2 I$, entonces $E[Y'AY] = \sigma^2 \text{tr}(A)$.

Teorema 2.15 Si Y es un vector de $p \times 1$ y es distribuido normalmente con media μ y matriz de covarianza V y si B es una matriz de $q \times p$ donde $q \leq p$ y rango(B) = q entonces el vector $Z = BY$ es distribuido como una normal.

Teorema 2.16 Si Y es distribuido como $N(\mu, \sigma^2 I)$ entonces $\frac{Y'AY}{\sigma^2}$ es distribuido como $\chi^2(k, \lambda)$ donde $\lambda = \frac{\mu' A \mu}{2\sigma^2}$ si solo si A es una matriz idempotente de rango(A).

Teorema 2.17 Si B es una matriz de $q \times n$, A es una matriz de $n \times n$ y Y es un vector distribuido como $N(\mu, \sigma^2 I)$ entonces la forma lineal BY es independiente de la forma cuadrática $Y'AY$ si $BA = 0$.

Teorema 2.18 Sea Y una variable aleatoria distribuida como $N(\mu, \sigma^2 I)$ y

$$\sum_{i=1}^k Y' A_i Y = Y' Y$$

donde el rango(A_i) = n_i si se cumplen las siguientes condiciones:

1.- A_i es una matriz idempotente para $i = 1, \dots, k$

2.- $A_i A_j = 0$ para $i \neq j$

3.- Si $\sum_{i=1}^k n_i = n$ entonces

a.- $\frac{Y'AY}{\sigma^2} \sim \chi^2(n, \lambda)$ donde $\lambda = \frac{\mu' A \mu}{2\sigma^2}$

b.- $Y' A_i Y$ y $Y' A_j Y$ son independientes.

Teorema 2.19 Si X_1, \dots, X_k son variables aleatorias distribuidas normalmente e independientes con media μ_i y varianza σ_i^2 para $i = 1, \dots, k$ entonces

$$u = \sum_{i=1}^k \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

tiene una distribución Ji-Cuadrada con k grados de libertad.

Prueba Si $Z_i = \frac{x_i - \mu_i}{\sigma_i}$ entonces Z_i tiene una distribución normal standard. Con lo cual,

$$\begin{aligned} m_u(t) &= E[e^{tu}] \\ &= E\left[\exp\left(t \sum_{i=1}^k Z_i^2\right)\right] \\ &= E\left[\prod_{i=1}^k e^{tZ_i^2}\right] \text{ y por ser variables independientes} \\ &= \prod_{i=1}^k E[e^{tZ_i^2}] \end{aligned}$$

por otro lado se tiene que:

$$\begin{aligned} E[e^{tZ^2}] &= \int e^{tZ^2} \left(\frac{1}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}Z^2} dz \\ &= \int \left(\frac{1}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}(1-2t)Z^2} dz \\ &= \frac{1}{\sqrt{1-2t}} \int \frac{\sqrt{1-2t}}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2t)Z^2} dz \end{aligned}$$

como $\mu = 0$ y $\sigma^2 = \frac{1}{1-2t}$ entonces $\frac{1}{\sigma} = \sqrt{1-2t}$ y de esto se obtiene que:

$$\begin{aligned} E[e^{tZ^2}] &= \frac{1}{\sqrt{1-2t}} \int \left(\frac{1}{\sqrt{2\pi\sigma}}\right) e^{-Z^2/2\sigma^2} \\ &= \frac{1}{\sqrt{1-2t}} \text{ para } t < \frac{1}{2} \end{aligned}$$

entonces

$$\begin{aligned} \prod_{i=1}^k E[e^{tZ_i^2}] &= \prod_{i=1}^k \left(\frac{1}{\sqrt{1-2t}}\right)^k \\ &= \left(\frac{1}{1-2t}\right)^{k/2} \end{aligned}$$

y por el teorema anterior, se observa que la variable aleatoria U tiene la misma función generadora de momentos que la distribución Ji-Cuadrada con k grados de libertad. Por lo tanto U se distribuye como una Ji-Cuadrada. \square

Teorema 2.20 Si X es una variable aleatoria distribuida como una Ji-Cuadrada con k grados de libertad entonces su función generadora de momentos esta dada por:

$$m_x(t) = \left(\frac{1}{1-2t}\right)^{k/2}$$

Prueba Como

$$m_x(t) = E[\exp(tx)] = \int e^{tx} \left(\frac{1}{\Gamma(k/2)} \right) \left(\frac{1}{2} \right)^{k/2} x^{k/2-1} e^{-x/2} dx$$

$$m_x(t) = \left(\frac{1/2}{1/2-t} \right)^{k/2} \int \frac{(1/2-t)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-(1/2-t)x} dx$$

al efectuar un cambio de variable de manera tal que

$$u = (1/2 - t)x$$

entonces

$$x = \frac{u}{1/2-t} \quad y \quad dx = \frac{du}{1/2-t}$$

por lo tanto,

$$m_x(t) = \left(\frac{1/2}{1/2-t} \right)^{k/2} \left(\frac{1}{\Gamma(k/2)} \right) \int u^{k/2-1} e^{-u} du$$

y como $\Gamma(\alpha) = \int u^{\alpha-1} e^{-u} du$ entonces

$$\begin{aligned} m_x(t) &= \left(\frac{1/2}{1/2-t} \right)^{k/2} \\ &= \left(\frac{1}{1-2t} \right)^{k/2} \quad \square \end{aligned}$$

Definición 2.25 Si X es una variable aleatoria con función de densidad de probabilidad

$$f_x(x) = \frac{1}{\Gamma(k/2)} \left(\frac{1}{2} \right)^{k/2} x^{k/2-1} \exp(-x/2) I_{(0,\infty)}(x)$$

entonces X tiene una distribución Ji-Cuadrada con k grados de libertad.

Definición 2.26 Si X es una variable aleatoria con función de densidad de probabilidad

$$f_x(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n} \right)^{m/2} \frac{x^{(m-2)/2}}{[1+(m/n)x]^{(m+n)/2}} I_{(0,\infty)}(x)$$

entonces X tiene una distribución F con m y n grados de libertad.

Teorema 2.21 Sean U y V dos variables aleatorias con función de distribución Ji-Cuadrada con m y n grados de libertad respectivamente y además U y V son independientes entonces la variable aleatoria

$$X = \frac{U/m}{V/n}$$

está distribuida como una función de distribución F con m y n grados de libertad.

Prueba Como U y V son independientes entonces su función de densidad conjunta es

$$\begin{aligned} f_{u,v}(u, v) &= f_u(u) f_v(v) \\ &= \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} u^{(m-2)/2} v^{(n-2)/2} e^{-\frac{1}{2}(u+v)} I_{(0,\infty)}(u) I_{(0,\infty)}(v) \end{aligned}$$

y para encontrar la función de distribución de X , haremos la transformación

$$X = \frac{U/m}{V/n} \quad y \quad Y = V$$

con lo cual se obtiene que

$$u = \left(\frac{m}{n}y\right)x$$

por lo tanto, el Jacobiano de la transformación es

$$\frac{du}{dx} = \left(\frac{m}{n}y\right)$$

y como

$$f_u(u) = \frac{1}{\Gamma(m/2)} \left(\frac{1}{2}\right)^{m/2} u^{m/2-1} e^{-u/2} I_{(0,\infty)}(u)$$

entonces

$$f_{x|y}(x|y) = \left(\frac{m}{n}y\right) \left[\frac{1}{\Gamma(m/2)} \left(\frac{1}{2}\right)^{m/2} \left(\frac{m}{n}xy\right)^{(m-2)/2} e^{-\frac{1}{2}\left(\frac{m}{n}xy\right)} \right]$$

Con lo cual la función de densidad conjunta de X y Y es

$$\begin{aligned} f_{x,y}(x, y) &= f_{x|y}(x|y) f_y(y) \\ &= \left(\frac{m}{n}y\right) \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} \left(\frac{m}{n}xy\right)^{(m-2)/2} y^{(n-2)/2} e^{-\frac{1}{2}\left[\left(\frac{m}{n}xy\right)+y\right]} \end{aligned}$$

y la función de densidad marginal de X es

$$\begin{aligned} f_x(x) &= \int f_{x,y}(x, y) dy \\ &= \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} \left(\frac{m}{n}\right)^{m/2} x^{(m-2)/2} \int y^{(m+n-2)/2} e^{-\frac{1}{2}\left[\left(\frac{m}{n}\right)x+1\right]y} dy \end{aligned}$$

Al efectuar el cambio de variable $u = \frac{1}{2} \left[\left(\frac{m}{n}\right)x + 1 \right] y$ entonces $y = \frac{2u}{\left[\left(\frac{m}{n}\right)x + 1\right]}$ y

$$dy = \frac{2du}{\left[\left(\frac{m}{n}\right)x + 1\right]}$$

se obtiene lo siguiente

$$\begin{aligned} f_x(x) &= \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} \left(\frac{m}{n}\right)^{m/2} x^{(m-2)/2} \int \left(\frac{2u}{\left[\left(\frac{m}{n}\right)x + 1\right]}\right)^{(m+n-2)/2} e^{-u} \left(\frac{2}{\left[\left(\frac{m}{n}\right)x + 1\right]}\right) du \\ &= \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} \left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{\left[\left(\frac{m}{n}\right)x + 1\right]} \int u^{(m+n)/2-1} e^{-u} du \end{aligned}$$

Como

$$\Gamma(\alpha) = \int u^{\alpha-1} e^{-u} du$$

entonces

$$f_x(x) = \frac{\Gamma[(m+n)-2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{[(m/n)x+1]} I_{(0,\infty)}(x) \square$$

Definición 2.27 Si X es una variable aleatoria con función de densidad de probabilidad

$$f_x(x) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+x^2/k)^{(k+1)/2}}$$

entonces X tiene una distribución t de Student.

Teorema 2.22 Sean Z y U dos variables aleatorias que se distribuyen como una normal standard y como una Ji-Cuadrada con k grados de libertad, donde Z y U son independientes entonces la variable aleatoria

$$X = \frac{Z}{\sqrt{U/k}}$$

tiene una distribución t de Student con k grados de libertad.

Prueba 2.2 Puesto que Z y U son independientes entonces su función de densidad conjunta es:

$$\begin{aligned} f_{z,u}(z, u) &= f_z(z) f_u(u) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} u^{k/2-1} e^{-u/2} e^{-z^2/2} I_{(0,\infty)}(u) \end{aligned}$$

Para encontrar la función de distribución de X se hará la siguiente transformación

$$X = \frac{Z}{\sqrt{U/k}} \text{ y } Y = U$$

con lo cual se obtiene que:

$$Z = X\sqrt{Y/k}$$

por lo tanto, el Jacobiano de la transformación es

$$\frac{dz}{dx} = \sqrt{Y/k}$$

y como

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

entonces

$$f_{x|y}(x | y) = \sqrt{y/k} \left(\frac{1}{\sqrt{2\pi}} e^{-x^2 y/k} \right)$$

Con lo cual, la función de densidad conjunta de X y Y es

$$\begin{aligned} f_{x,y}(x,y) &= f_{x|y}(x|y) f_y(y) \\ &= \sqrt{y/k} \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} y^{k/2-1} e^{-y/2} e^{-x^2 y/k} \\ &= \frac{1}{\sqrt{k\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{(k+1)/2} y^{(k+1)/2-1} e^{-\frac{1}{2}(1+x^2/k)y} \end{aligned}$$

Así, se puede obtener la función de densidad marginal de X como

$$\begin{aligned} f_x(x) &= \int f_{x,y}(x,y) dy \\ &= \frac{1}{\sqrt{k\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{(k+1)/2} \int y^{(k+1)/2-1} e^{-\frac{1}{2}(1+x^2/k)y} dy \end{aligned}$$

Al efectuar el cambio de variable

$$w = \frac{1}{2} (1 + x^2/k) y$$

entonces

$$y = \frac{2w}{1 + x^2/k} \quad y \quad dy = \frac{2}{1 + x^2/k} dw$$

Con lo cual se obtiene lo siguiente

$$\begin{aligned} f_x(x) &= \frac{1}{\sqrt{k\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{(k+1)/2} \int \left(\frac{2w}{1 + x^2/k}\right)^{(k+1)/2} e^{-w} \left(\frac{2}{1 + x^2/k}\right) dw \\ &= \frac{1}{\sqrt{k\pi}} \frac{1}{\Gamma(k/2)} \frac{1}{(1+x^2/k)} \int w^{(k+1)/2-1} e^{-w} dw \\ &= \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \square \end{aligned}$$

2.3 Estimación Puntual

El problema de estimación consiste en obtener a partir de una muestra X_1, \dots, X_n de variables aleatorias con función de densidad de probabilidad $f_x(\cdot; \theta) = f(\cdot; \theta)$ conocida el valor del parámetro θ desconocido o el valor de alguna función de θ , es decir $\tau(\theta)$. La estimación puntual presenta dos problemas fundamentales que son:

- i) Obtener un estimador de θ .
- ii) Seleccionar un criterio para encontrar el mejor estimador de θ .

El problema de estimación puntual consiste en encontrar alguna estadística $t(X_1, \dots, X_n)$, la cual estime el valor desconocido de θ .

La estimación por intervalo consiste en encontrar dos estadísticas $t_1(X_1, \dots, X_n)$ y $t_2(X_1, \dots, X_n)$ donde $t_1(X_1, \dots, X_n) < t_2(X_1, \dots, X_n)$, con lo cual $(t_1(X_1, \dots, X_n), t_2(X_1, \dots, X_n))$ forman un intervalo, donde se determina la probabilidad de contener el valor desconocido de $\tau(\theta)$.

Definición 2.28 Una estadística es una función de los valores de las variables aleatorias observadas x_1, \dots, x_n , la cual no contiene ningún parámetro desconocido.

Definición 2.29 Un estimador es cualquier estadística cuyos valores son usados para conocer θ .

Definición 2.30 Un estimador (T_1, \dots, T_r) , donde $T_j = t_j(X_1, \dots, X_n)$ para $j = 1, \dots, r$, es estimador insesgado de $(\tau_1(\theta), \dots, \tau_r(\theta))$ si y sólo si $E[T_j] = \tau_j(\theta)$ para $j = 1, \dots, r$ para todo $\theta \in \Theta$.

Definición 2.31 La función de verosimilitud de n variables aleatorias X_1, \dots, X_n es la función de densidad conjunta de las n variables, es decir $f_{x_1, \dots, x_n}(x_1, \dots, x_n; \theta)$, la cual es considerada como función de θ .

Ejemplo 1 Si X_1, \dots, X_n es una muestra aleatoria con función de densidad $f(x; \theta)$ entonces la función de verosimilitud es $L(\theta) = f(x_1; \theta) \dots f(x_n; \theta)$.

Definición 2.32 Sea $L(\theta) = L(\theta; x_1, \dots, x_n)$ la función de verosimilitud para las variables aleatorias X_1, \dots, X_n . Si $\hat{\theta}$ es el valor de $\theta \in \Theta$, el cual maximiza $L(\theta)$ entonces $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ es el estimador máximo verosímil de θ para la muestra x_1, \dots, x_n .

Definición 2.33 Sea $T_1^*, \dots, T_n^*, \dots$ una sucesión de estimadores de $\tau(\theta)$, donde $T_n^* = t_n(X_1, \dots, X_n)$.

La sucesión de estimadores $\{T_n^*\}$ es eficiente si cumple con las siguientes condiciones:

i) La distribución de $\sqrt{n}[T_n^* - \tau(\theta)]$ es asintóticamente normal con media cero y varianza $\sigma^2(\theta)$.

ii) Si $\{T_n\}$ es cualquier otra sucesión de estimadores para los cuales $\sqrt{n}[T_n - \tau(\theta)]$ es asintóticamente normal con media cero y varianza σ^2 , entonces

iii) $\sigma^{*2}(\theta) < \sigma^2(\theta)$ para todo θ en cualquier intervalo abierto.

Definición 2.34 Sea T_1, \dots, T_n, \dots una sucesión de estimadores de $\tau(\theta)$, donde $T_n = t_n(X_1, \dots, X_n)$. La sucesión de estimadores $\{T_n\}$ es consistente. Si para cada $\epsilon > 0$ entonces

$$\lim_{n \rightarrow \infty} P[\tau(\theta) - \epsilon < T_n < \tau(\theta) + \epsilon] = 1 \quad \text{para } \theta \in \Theta.$$

Definición 2.35 Una sucesión de estimadores $T_1^*, \dots, T_n^*, \dots$ de $\tau(\theta)$ son los mejores asintóticamente si y sólo si cumplen las siguientes condiciones:

i) La distribución de $\sqrt{n}[T_n^* - \tau(\theta)]$ es aproximadamente una distribución normal con media cero y varianza $\sigma^{*2}(\theta)$ cuando n tiende a infinito.

ii) Para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P \left[|T_n^* - \tau(\theta)| > \epsilon \right] = 0 \text{ para todo } \theta \in \Theta.$$

iii) Sea $\{T_n\}$ cualquier otra sucesión de estimadores consistentes para los cuales la distribución de $\sqrt{n}[T_n - \tau(\theta)]$ es aproximadamente una distribución normal con media cero y varianza $\sigma^2(\theta)$.

iv) $\sigma^2(\theta) > \sigma^{*2}(\theta)$ para todo θ en cualquier intervalo abierto.

Teorema 2.23 Si la función de densidad $f(x; \theta)$ satisface las condiciones de regularidad y si $\hat{\Theta}_n = \hat{\vartheta}(X_1, \dots, X_n)$ es el estimador máximo verosímil de θ para una muestra de tamaño n entonces se cumplen las siguientes condiciones:

i) $\hat{\Theta}$ es distribuida asintóticamente normal con medio θ y varianza

$$\frac{1}{n E \left[\left[\frac{\partial}{\partial \theta} \log f(x; \theta) \right]^2 \right]}$$

ii) La sucesión de estimadores máximo verosímiles $\hat{\Theta}_1, \dots, \hat{\Theta}_n$ son los mejores asintóticamente normales.

Definición 2.36 Sean X_1, \dots, X_n una muestra aleatoria de una función de densidad de probabilidad $f(\cdot; \theta)$, entonces cualquier estadística $S = s(X_1, \dots, X_n)$ es una estadística suficiente si y sólo si la distribución condicional de cualquier estadística T dado S no depende de θ .

Definición 2.37 Sean X_1, \dots, X_n una muestra aleatoria de una función de densidad de probabilidad $f(\cdot; \theta)$. Las estadísticas S_1, \dots, S_r son estadísticas suficientemente conjuntas si y sólo si la distribución condicional de X_1, \dots, X_n dado $S_1 = s_1, \dots, S_r = s_r$ no depende de θ .

Teorema 2.24 Sea X_1, \dots, X_n una muestra aleatoria con función de densidad de probabilidad $f_x(\cdot; \theta)$, donde el parámetro puede ser un vector. Un conjunto de estadísticas $S_1 = s_1(X_1, \dots, X_n), \dots, S_r = s_r(X_1, \dots, X_n)$ son conjuntamente suficientes si y sólo si la función de densidad de probabilidad conjunta de X_1, \dots, X_n puede ser factorizada como

$$\begin{aligned} f_{x_1, \dots, x_n}(x_1, \dots, x_n; \theta) &= g(s_1(X_1, \dots, X_n), \dots, s_r(X_1, \dots, X_n); \theta) h(x_1, \dots, x_n) \\ &= g(s_1, \dots, s_r; \theta) h(x_1, \dots, x_n) \end{aligned}$$

donde la función $h(x_1, \dots, x_n) > 0$ y $g(s_1, \dots, s_r; \theta) > 0$ depende de x_1, \dots, x_n solo a través de las funciones $s_1(x_1, \dots, x_n), \dots, s_r(x_1, \dots, x_n)$.

Definición 2.38 Un conjunto de estadísticas suficientes es suficiente minimal si y sólo si es función de cualquier otra estadística suficiente.

Definición 2.39 Si una familia de funciones de densidad de probabilidad $f(\cdot; \theta)$ pueden expresarse como

$$f(x; \theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k) b(x) \exp \left(\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k) d_j(x) \right)$$

para una apropiada selección de funciones $a(\cdot, \dots, \cdot)$, $b(\cdot)$, $c(\cdot, \dots, \cdot)$ y $d_j(\cdot)$ para $j = 1, \dots, k$ entonces pertenece a la familia exponencial.

Definición 2.40 Sea X_1, \dots, X_n una muestra aleatoria con función de densidad de probabilidad $f(\cdot; \theta)$. Un estimador $T^* = t^*(X_1, \dots, X_n)$ de $\tau(\theta)$ es un estimador insesgado de varianza mínima uniforme de $\tau(\theta)$ si y sólo si:

- i) $E[T^*] = \tau(\theta)$, es decir, T^* es un estimador insesgado.
- ii) $Var[T^*] \leq Var[T]$ para cualquier estimador $T = t(X_1, \dots, X_n)$ de $\tau(\theta)$ el cual satisface que $E[T] = \tau(\theta)$.

Definición 2.41 Sea (T_1, \dots, T_r) estimadores insesgados de $(\tau_1(\theta), \dots, \tau_r(\theta))$. Sea $\sigma^{ij}(\theta)$ el ij -ésimo elemento de la matriz inversa de la covarianza de (T_1, \dots, T_r) , donde ij -ésimo elemento de la matriz de covarianza es $\sigma_{ij}(\theta) = Cov[T_i, T_j]$. La concentración elipsoidal de (T_1, \dots, T_r) es el interior y la elipsoide acotada por

$$\sum_{i=1}^r \sum_{j=1}^r \sigma^{ij}(\theta) [T_i - T_i(\theta)] [T_j - T_j(\theta)] = r + 2.$$

Definición 2.42 Sea X_1, \dots, X_n una muestra aleatoria con función de densidad de probabilidad $f(x; \theta_1, \dots, \theta_k)$ y sea (T_1, \dots, T_m) un conjunto de estadísticas. T_1, \dots, T_m son estadísticas completas si y sólo si $E\{z(T_1, \dots, T_m)\} \equiv 0$ para todo $\theta \in \Theta$ implica que $p\{z(T_1, \dots, T_m) = 0\} \equiv 1$ para todo $\theta \in \Theta$, donde $z(T_1, \dots, T_m)$ es una estadística.

Teorema 2.25 Sea X_1, \dots, X_n una muestra aleatoria con función de densidad de probabilidad

$$f(x; \theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k) b(x) \exp \left(\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k) d_j(x) \right)$$

es decir, $f(x; \theta_1, \dots, \theta_k)$ es miembro de la familia exponencial, entonces

$$\left(\sum_{i=1}^n d_1(x_1), \dots, \sum_{i=1}^n d_k(x_k) \right)$$

es un conjunto de estadísticas suficientes minimales y completas conjuntamente.

Teorema 2.26 Sea X_1, \dots, X_n una muestra aleatoria con función de densidad de probabilidad $f(x; \theta_1, \dots, \theta_k)$ y sea $S_1 = s_1(X_1, \dots, X_n), \dots, S_m = s_m(X_1, \dots, X_n)$ un conjunto de estadísticas suficientes conjuntamente. Sea (T_1, \dots, T_r) un estimador insesgado de $(\tau_1(\theta), \dots, \tau_r(\theta))$

donde $\theta = (\theta_1, \dots, \theta_k)$. Definamos $T_j^* = E[T_j | S_1, \dots, S_m]$ para $j = 1, \dots, k$ entonces:

- i) (T_1^*, \dots, T_r^*) es un estimador insesgado de $(\tau_1(\theta), \dots, \tau_r(\theta))$ y $T_j^* = t_j^*(S_1, \dots, S_m)$ es decir T_j^* es una función de las estadísticas suficientes S_1, \dots, S_m , para $j = 1, \dots, r$.
- ii) $Var [T_j^*] \leq Var [T_j]$ para cada $\theta \in \Theta$, para $j = 1, \dots, r$.
- iii) La concentración elipsoidal de (T_1^*, \dots, T_r^*) esta contenida en la concentración elipsoidal de (T_1, \dots, T_r) para cada $\theta \in \Theta$.

2.3.1 Estimación por Intervalo

Sea X_1, \dots, X_n una muestra de una función de densidad de probabilidad $f(\cdot; \theta)$ y sea $T_1 = t_1(X_1, \dots, X_n)$ y $T_2 = t_2(X_1, \dots, X_n)$ dos estadísticas tales que $T_1 \leq T_2$ para las cuales $P[T_1 < \tau(\theta) < T_2] = \gamma$, donde $\gamma \in (0, 1)$ entonces el intervalo (T_1, T_2) es llamado intervalo de confianza del 100γ ; γ es llamado coeficiente de confianza; T_1 y T_2 son llamados limite inferior y superior de confianza respectivamente para $\tau(\theta)$.

Al valor (t_1, t_2) del intervalo aleatorio (T_1, T_2) es llamado también intervalo de confianza del 100γ por ciento para $\tau(\theta)$.

2.4 Prueba de Hipótesis

Definición 2.43 Una hipótesis estadística es una afirmación o conjetura acerca de una distribución de una o más variables.

Definición 2.44 Una prueba de hipótesis estadística H es una regla o procedimiento para decidir si se rechaza H .

Definición 2.45 Sea Υ una prueba de una hipótesis estadística H . Una región crítica es el subconjunto para el cual la hipótesis H sería rechazada.

Definición 2.46 Rechazar H_0 dado que H_0 es verdadera, se define como error del tipo I.

Definición 2.47 Aceptar H_0 dado que H_0 es falsa, se define como error del tipo II.

Definición 2.48 La probabilidad de rechazar H_0 dado que H_0 es cierta, se define como la probabilidad (o tamaño) del error tipo I y se denota por α , $0 \leq \alpha \leq 1$.

Definición 2.49 La probabilidad de aceptar H_0 dado que H_0 es falsa, se define como la probabilidad (o tamaño) del error tipo II y se denota por β , $0 \leq \beta \leq 1$.

Definición 2.50 Sea Υ una prueba de la hipótesis nula H_0 . La función potencia de la prueba Υ , denotada por $\pi_{\Upsilon}(\theta)$ es la probabilidad de rechazar H_0 .

Capítulo 3

ESTIMACIÓN DE PARÁMETROS, ANÁLISIS DE VARIANZA Y DISEÑOS EXPERIMENTALES

3.1 Supuestos del Modelo de Regresión Lineal Múltiple

El modelo general de regresión lineal plantea los siguientes supuestos:

1. $E[\varepsilon] = 0$ este supuesto explica que el valor promedio de ε_i dado cualquier valor x_{1i}, \dots, x_{ik} es igual a cero.
2. $Cov[\varepsilon] = E[\varepsilon\varepsilon'] = \sigma^2 I$, no existe autocorrelación entre las ε_i y tienen la misma varianza, así las perturbaciones asociadas a alguna observación no están siendo influenciadas por el término de perturbación asociado a cualquier otra observación; es decir, las perturbaciones ε_i y las ε_j si $i \neq j$ no están correlacionadas. También postula que la varianza de ε_i para cada x_{1i}, \dots, x_{ik} es un número positivo constante igual a σ^2 . Esto se conoce como homocedasticidad o igual dispersión de varianza.
3. La matriz X es no estocástica lo que implica que está formada por números fijos. Esto es obvio ya que en el análisis de regresión los valores de Y dependen de los valores fijos de X .
4. El rango de la matriz X es k , y k es el número de columnas de la matriz. Lo que significa que las columnas de la matriz son linealmente independientes.
5. ε está distribuida como $N(0, \sigma^2 I)$.

Como ε_i representa la influencia combinada sobre la variable dependiente de un gran número de variables independientes que no se introducen explícitamente en el modelo de regresión lineal, se espera que la influencia de estas variables, omitidas o no tenidas en cuenta, fuera pequeña y en el mejor de los casos aleatoria. Por el teorema del límite central se demuestra que sí existe un número grande de variables aleatorias independientes e idénticamente distribuidas, entonces la distribución de su suma tenderá a seguir una distribución normal a medida que el número de estas variables aumente. Es precisamente este teorema del límite central el que proporciona una justificación teórica para el supuesto de normalidad ε_i .

Las pruebas de significancia (*estadísticas - t y - F*), y los procedimientos de estimación de intervalos de confianza así como el método de estimación máximo verosímil, requieren de un supuesto adicional:

1. Los errores ε_i tiene distribución normal.

El método de mínimos cuadrados selecciona como estimadores de los parámetros del modelo de regresión (3.1), aquellos valores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ que minimizan:

$$\sum_{i=1}^n \varepsilon_i^2 = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon' \varepsilon$$

Como el modelo es de la forma $Y = X\beta + \varepsilon$ tenemos que $\varepsilon = Y - X\beta$ con lo cual:

$$\sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (Y - X\beta)' (Y - X\beta)$$

Derivando con respecto a β obtenemos:

$$\frac{\partial (\varepsilon' \varepsilon)}{\partial \beta} = 2X'Y - 2X'X\beta = 0$$

$$X'X\beta = X'Y$$

Como X es de rango completo entonces $X'X$ es de rango k . Con lo cual existe la matriz inversa de $X'X$. Por lo tanto:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

3.3 Propiedades de los Coeficientes Estimados

Las propiedades de los estimadores $\hat{\beta}$ son heredadas de los supuestos hechos sobre ε , tales supuestos pueden ser escritos como:

1. $E(\varepsilon) = 0$.
2. $Var(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I$

Así, la esperanza del vector $\hat{\beta}$ está dada por:

$$E(\hat{\beta}) = E[(X'X)^{-1} X'Y] = E[(X'X)^{-1} X'(X\beta + \varepsilon)] = \beta + E[(X'X)^{-1} X'\varepsilon] = \beta$$

lo que implica que si los supuestos son ciertos, el estimador $\hat{\beta}$ de β es insesgado.

La matriz de varianza-covarianza del vector de estimaciones $\hat{\beta}$ es

$$\text{Var}(\hat{\beta}) = \text{Var}[(X'X)^{-1} X'Y] = (X'X)^{-1} X' \text{Var}(Y) X (X'X)^{-1}$$

ya que $\text{Var}(Y) = \text{Var}(\varepsilon) = \sigma^2 I$, entonces

$$\text{Var}(\hat{\beta}) = (X'X)^{-1} \sigma^2. \quad (3.2)$$

Sobre la diagonal de la matriz $C = (X'X)^{-1} \sigma^2$, los elementos $c_{ii} \sigma^2$ representan la varianza de $\hat{\beta}_i$. El elemento ij -ésimo de la matriz C , $c_{ij} \sigma^2$ es la covarianza entre los estimadores $\hat{\beta}_i$ y $\hat{\beta}_j$. Si, los errores ε se distribuyen conjuntamente normales, entonces $\hat{\beta}$ se distribuye como:

$$\hat{\beta} \sim N(\beta, (X'X)^{-1} \sigma^2)$$

3.3.1 Teorema de Gauss Markoff

Teorema 3.1 Gauss-Markoff. Si el modelo general de regresión lineal de rango completo $Y = X\beta + \varepsilon$ que cumple las siguientes condiciones:

1. $E[\varepsilon] = 0$.
2. $\text{Cov}[\varepsilon] = E[\varepsilon\varepsilon'] = \sigma^2 I$,

entonces el estimador por mínimos cuadrados $\hat{\beta} = (X'X)^{-1} X'Y$ es el mejor estimador lineal insesgado (MELI) de β .

Prueba 3.1 Sea A cualquier matriz de $k \times n$ de valores conocidos y sea β^* cualquier otro estimador lineal insesgado de β tal que $\beta^* = AY$. como debemos especificar los elementos de A , tal que β^* sea el mejor estimador insesgado de β .

Sea $A = (X'X)^{-1} X' + B$ y como $(X'X)^{-1} X'$ es conocido debemos encontrar B para poder especificar A .

$$\begin{aligned} E[\beta^*] &= E(AY) = E[((X'X)^{-1} X' + B)Y] \\ &= ((X'X)^{-1} X' + B) E[Y] = ((X'X)^{-1} X' + B) E[X\beta + \varepsilon] \\ &= ((X'X)^{-1} X' + B) [E[X\beta] + E[\varepsilon]] \\ &= ((X'X)^{-1} X' + B) (X\beta) \\ &= \beta + BX\beta \end{aligned}$$

Como β^* es insesgado entonces $BX\beta = 0$ para toda β , por lo tanto $BX = 0$.

Para demostrar que es el mejor estimador lineal insesgado, se debe encontrar la matriz B

tal que $Var(\beta^*)$ sea mínimo para $i = 1, \dots, k$ sujeto a la restricción $BX = 0$.

$$\begin{aligned}
 Cov(\beta^*) &= E\left\{(\beta^* - \beta)(\beta^* - \beta)'\right\} \\
 &= E\left\{\left[\left((X'X)^{-1}X' + B\right)Y - \beta\right]\left[\left((X'X)^{-1}X' + B\right)Y - \beta\right]'\right\} \\
 &= E\left\{\left[\left((X'X)^{-1}X' + B\right)(X\beta + \varepsilon) - \beta\right]\left[\left((X'X)^{-1}X' + B\right)(X\beta + \varepsilon) - \beta\right]'\right\} \\
 &= E\left\{\left[\beta + BX\beta + (X'X)^{-1}X'\varepsilon + B\varepsilon - \beta\right]\left[\beta + BX\beta + (X'X)^{-1}X'\varepsilon + B\varepsilon - \beta\right]'\right\} \\
 &= E\left\{\left[BX\beta + (X'X)^{-1}X'\varepsilon + B\varepsilon\right]\left[BX\beta + (X'X)^{-1}X'\varepsilon + B\varepsilon\right]'\right\} \\
 &= E\left\{\left[(X'X)^{-1}X'\varepsilon + B\varepsilon\right]\left[(X'X)^{-1}X'\varepsilon + B\varepsilon\right]'\right\} \\
 &= E\left\{\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} + B\varepsilon\varepsilon'B' + (X'X)^{-1}X'\varepsilon\varepsilon'B' + B\varepsilon\varepsilon'X(X'X)^{-1}\right]\right. \\
 &= \left.(X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1} + BE[\varepsilon\varepsilon']B' + (X'X)^{-1}X'E[\varepsilon\varepsilon']B' \right. \\
 &= \left. + BE[\varepsilon\varepsilon']X(X'X)^{-1}\right\} \\
 &= \sigma^2(X'X)^{-1} + \sigma^2BB' \\
 &= \sigma^2(X'X)^{-1} + \sigma^2BB'
 \end{aligned}$$

Sea $BB' = G = g_{ij}$. Como $Cov(\beta^*) = \sigma^2(X'X)^{-1} + \sigma^2BB'$ la diagonal de los elementos de la $Cov(\beta^*)$ deben ser iguales a la varianza de β_i^* .

Para minimizar cada $Var(\beta_i^*)$ es necesario por lo tanto minimizar los elementos de la diagonal $Cov(\beta^*)$.

Como $\sigma^2(X'X)^{-1}$ son constantes, se debe encontrar una matriz g tal que los elementos de la diagonal sean mínimos; pero como $G = BB'$ es semidefinida positiva se tiene que $g_{ij} \geq 0$ con lo cual los elementos de la diagonal serán mínimos cuando $g_{ij} = 0$ para $i = 1, \dots, k$ pero como $B = b_{ij}$, entonces, $g_{ij} = 0 = \sum b_{ij}^2$ y por lo tanto, $b_{ij} = 0$, para toda i y j , por lo tanto $B = 0$.

Con lo cual $A = (X'X)^{-1}X'$ y así $\beta^* = \beta$. \square

3.4 Máxima Verosimilitud

Analizaremos el caso que supone que ε se distribuye como $N(0, \sigma^2 I)$ con varianza desconocida. En este caso utilizaremos el método de estimación máximo verosímil, para estimar β_1, \dots, β_k y σ^2 . Ya que ε esta distribuida como $N(0, \sigma^2 I)$ la función de verosimilitud es:

$$L(\varepsilon; \beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\left[\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right]}$$

usando logaritmos se obtiene:

$$\ln L(\varepsilon; \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)$$

donde el espacio parametral de Ω es:

$$\Omega = \left\{(\beta, \sigma^2) : \sigma^2 > 0, -\infty < \beta_i < \infty \forall i = 1, \dots, k\right\}$$

con lo cual, para encontrar los valores de β y σ^2 en Ω , tal que la función de verosimilitud tome su valor máximo, se debe derivar parcialmente con respecto a β y σ^2 e igualar las ecuaciones a cero, es decir:

$$\frac{\partial}{\partial \beta} \ln L(\varepsilon; \beta, \sigma^2) = \frac{2}{2\sigma^2} (X'Y - X'X\beta) = 0$$

$$\frac{\partial}{\partial \sigma^2} \ln L(\varepsilon; \beta, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^4} = 0$$

Si $\hat{\beta}$ y $\hat{\sigma}^2$ son soluciones de las anteriores ecuaciones, se tiene que:

$$X'X\hat{\beta} = X'Y$$

y

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n}$$

Las ecuaciones:

$$X'X\hat{\beta} = X'Y$$

son llamadas *ecuaciones normales*.

Como X es de rango completo entonces $X'X$ es de rango k . Con lo cual existe la matriz inversa de $X'X$. Por lo tanto:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

y

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n}$$

son las estimaciones máximo verosímil de β y σ^2 respectivamente.

Teorema 3.2 Sea $Y = X\beta + \varepsilon$ el modelo de regresión lineal múltiple de rango completo. Si ε está distribuido como $N(0, \sigma^2 I)$ entonces los estimadores:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

y

$$\hat{\sigma}^2 = \frac{Y'(I - X(X'X)^{-1}X')Y}{n - p}$$

cumplen las siguientes propiedades:

- i) Consistentes.
- ii) Eficientes.
- iii) Insesgados.

iv) Suficientes.

v) Completas.

vi) $\hat{\beta}$ está distribuido como una $N(\beta, \sigma^2(X'X)^{-1})$.

vii) Insesgados de varianza mínima uniforme.

viii) $\frac{\hat{\sigma}^2(n-p)}{\sigma^2}$ está distribuido como una distribución χ^2_{n-p} .

ix) $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes.

Prueba i) y ii) Como $\hat{\beta}$ y $\hat{\sigma}^2$ son estimadores máximo verosímiles y por el teorema (2.23), entonces $\hat{\beta}$ y $\hat{\sigma}^2$ son estimadores consistentes y eficientes.

iii) Insesgados

$$\begin{aligned} E[\hat{\beta}] &= E[(X'X)^{-1}X'Y] \\ &= (X'X)^{-1}X'E[Y] \\ &= (X'X)^{-1}X'E[X\beta + \varepsilon] \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'E[\varepsilon] \text{ por hipótesis } E[\varepsilon] = 0 \\ &= \beta \end{aligned}$$

y

$$\begin{aligned} E[\hat{\sigma}^2] &= E\left[\frac{Y'(I - X(X'X)^{-1}X')Y}{n-p}\right] \\ &= \frac{1}{n-p} E\left[(X\beta + \varepsilon)'(I - X(X'X)^{-1}X')(X\beta + \varepsilon)\right] \text{ por (2.3)} \\ &= \frac{1}{n-p} E\left[(\beta'X' + \varepsilon')(I - X(X'X)^{-1}X')(X\beta + \varepsilon)\right] \\ &= \frac{1}{n-p} E\left[\varepsilon'(I - X(X'X)^{-1}X')\varepsilon\right] \end{aligned}$$

como

$$X'(I - X(X'X)^{-1}X') = 0 \text{ y } (I - X(X'X)^{-1}X')X = 0$$

Por lo tanto por el teorema (2.14):

$$E[\hat{\sigma}^2] = \frac{\sigma^2}{n-p} \text{tr}(I - X(X'X)^{-1}X')$$

Y como

$$\text{tr}(I - X(X'X)^{-1}X') = \text{tr}(I) - \text{tr}(X(X'X)^{-1}X')$$

donde I es la matriz identidad de $n \times n$, con lo cual $\text{tr}(I) = n$ y por el teorema (2.9)

$$\text{tr}(X(X'X)^{-1}X') = \text{tr}(X'X(X'X)^{-1}) = \text{tr}(I) = p$$

Puesto que $X'X(X'X)^{-1} = I$ es la matriz identidad de $p \times p$, por lo tanto:

$$E[\hat{\sigma}^2] = \frac{\sigma^2}{n-p} (n-p) = \sigma^2$$

iv) *Suficiencia*. La función de densidad conjunta de Y es

$$f(\beta, \sigma^2) = \left(\frac{1}{(2\pi\sigma^2)^{n/2}} \right) \exp \left[-\frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta) \right]$$

y como

$$\begin{aligned} (Y - X\beta)' (Y - X\beta) &= \left[(Y - X\hat{\beta}) - X(\beta - \hat{\beta}) \right]' \left[(Y - X\hat{\beta}) - X(\beta - \hat{\beta}) \right] \\ &= (Y - X\hat{\beta})' (Y - X\hat{\beta}) - (Y - X\hat{\beta})' X(\beta - \hat{\beta}) \\ &\quad - (\beta - \hat{\beta})' X' (Y - X\hat{\beta}) + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \\ &= (Y - X\hat{\beta})' (Y - X\hat{\beta}) + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \\ &= (n - p) \hat{\sigma}^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \end{aligned}$$

Puesto que

$$(Y - X\hat{\beta})' X = Y' X - \hat{\beta}' X' X = Y' X - Y' X (X' X)^{-1} X' X = 0.$$

y

$$X' (Y - X\hat{\beta}) = X' Y - X' X \hat{\beta} = X' Y - X' X (X' X)^{-1} X' Y = 0.$$

Con lo cual

$$\begin{aligned} f(\beta, \sigma^2) &= \left(\frac{1}{(2\pi\sigma^2)^{n/2}} \right) \exp \left[-\frac{1}{2\sigma^2} \left[(n - p) \hat{\sigma}^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right] \right] \\ &= g(\hat{\beta}, \hat{\sigma}^2, \beta, \sigma^2) h(Y) \end{aligned}$$

donde $h(Y) = 1$.

Por lo tanto, por el teorema (2.24)

$$\begin{aligned} \hat{\beta} &= (X' X)^{-1} X' Y \\ \hat{\sigma}^2 &= \frac{Y' (I - X(X' X)^{-1} X') Y}{n - p} \end{aligned}$$

son estadísticas suficientes.

v) *Completas*. Como Y se distribuye como una $N(X\beta, \sigma^2 I)$ y está pertenece a la familia exponencial entonces por el teorema (2.25)

$$\begin{aligned} \hat{\beta} &= (X' X)^{-1} X' Y \\ \hat{\sigma}^2 &= \frac{Y' (I - X(X' X)^{-1} X') Y}{n - p} \end{aligned}$$

son estadísticas suficientes y completas.

vi) $\hat{\beta}$ está distribuido como una $N(\beta, \sigma^2(X'X)^{-1})$. Como $\hat{\beta} = (X'X)^{-1}X'Y$, entonces por el teorema (2.15) tenemos que $\hat{\beta}$ está distribuido como una normal y su media es $E[\hat{\beta}] = \beta$ esto fue calculado en el inciso iii) y la matriz de covarianza de $\hat{\beta}$ es:

$$\begin{aligned} \text{Cov}[\hat{\beta}] &= E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right] \\ &= E\left[((X'X)^{-1}X'Y-\beta)((X'X)^{-1}X'Y-\beta)'\right] \end{aligned}$$

Si se sustituye $Y = X\beta + \varepsilon$ se obtiene que:

$$\begin{aligned} \text{Cov}[\hat{\beta}] &= E\left\{[(X'X)^{-1}X'(X\beta + \varepsilon) - \beta][(X'X)^{-1}X'(X\beta + \varepsilon) - \beta]'\right\} \\ &= E\left\{[(X'X)^{-1}X'\varepsilon][(X'X)^{-1}X'\varepsilon]'\right\} \\ &= E\left\{(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right\} \\ &= (X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Con lo cual β está distribuido como una $N(\beta, \sigma^2(X'X)^{-1})$.

viii) $\frac{\hat{\sigma}^2(n-p)}{\sigma^2}$ se distribuye como una $\chi^2_{(n-p)}$. Puesto que

$$(n-p)\hat{\sigma}^2 = Y'(I - X(X'X)^{-1}X')Y$$

y

$$I - X(X'X)^{-1}X'$$

son matrices idempotentes de rango $n-p$ por el teorema (2.16)

$$\frac{\hat{\sigma}^2(n-p)}{\sigma^2} \sim \chi^2_{(n-p)}$$

donde

$$\lambda = \frac{1}{2\sigma^2}\beta'X'(I - X(X'X)^{-1}X')X\beta = 0.$$

Con lo cual

$$\frac{\hat{\sigma}^2(n-p)}{\sigma^2} \sim \chi^2_{(n-p)}$$

ix) β y σ^2 son independientes. Como

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\sigma}^2 = \frac{1}{n-p} [Y'(I - X(X'X)^{-1}X')Y]$$

y por el teorema (2.17) se sabe que para que $\hat{\beta}$ y $\hat{\sigma}^2$ sean independientes el producto de la matriz lineal y la matriz de la forma cuadrática debe ser igual a cero.

Es decir, $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes si y sólo si

$$(X'X)^{-1}X'(I - X(X'X)^{-1}X') = 0$$

y esto es claro, ya que

$$X'(I - X(X'X)^{-1}X') = 0. \square$$

3.5 Estimación por Intervalo

3.5.1 Intervalo de Confianza para σ^2

Puesto que $\frac{\hat{\sigma}^2(n-p)}{\sigma^2}$ se distribuye como una $\chi^2_{(n-p)}$, un intervalo de confianza para σ^2 es el siguiente:

Si α_0 y α_1 son dos constantes tales que

$$P \left[\alpha_0 \leq \frac{\hat{\sigma}^2(n-p)}{\sigma^2} \leq \alpha_1 \right] = 1 - \alpha$$

Por lo tanto:

$$P \left[\frac{\hat{\sigma}^2(n-p)}{\alpha_1} \leq \sigma^2 \leq \frac{\hat{\sigma}^2(n-p)}{\alpha_0} \right] = 1 - \alpha$$

El ancho del intervalo de confianza es:

$$\frac{\hat{\sigma}^2(n-p)}{\alpha_0} - \frac{\hat{\sigma}^2(n-p)}{\alpha_1} = \hat{\sigma}^2(n-p) \left(\frac{1}{\alpha_0} - \frac{1}{\alpha_1} \right).$$

3.5.2 Intervalo de Confianza para β_i

Puesto que $\hat{\beta}_i$ se distribuye como $N(\beta_i, c_{ii}\sigma^2)$, donde c_{ij} es el ij -ésimo elemento de $S^{-1} = C'$.

Por lo tanto $\frac{(\hat{\beta}_i - \beta_i)}{\sigma\sqrt{c_{ii}}}$ se distribuye como $N(0, 1)$ y es independiente de $\frac{\hat{\sigma}^2(n-p)}{\sigma^2}$, la cual se distribuye como $\chi^2_{(n-p)}$.

Por el teorema (2.22), se tiene que $u = \frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{c_{ii}}} \sqrt{\frac{\sigma^2}{\hat{\sigma}^2}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 c_{ii}}}$ se distribuye como $t_{(n-p)}$.

Con lo cual:

$$P \left[-t_{\alpha/2} \leq \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 c_{ii}}} \leq t_{\alpha/2} \right] = 1 - \alpha$$

Por lo tanto:

$$P \left[\hat{\beta}_i - t_{\alpha/2} \sqrt{\hat{\sigma}^2 c_{ii}} \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2} \sqrt{\hat{\sigma}^2 c_{ii}} \right] = 1 - \alpha$$

Con lo cual, el ancho del intervalo de confianza es:

$$\hat{\beta}_i + t_{\alpha/2} \sqrt{\hat{\sigma}^2 c_{ii}} - \left(\hat{\beta}_i - t_{\alpha/2} \sqrt{\hat{\sigma}^2 c_{ii}} \right) = 2t_{\alpha/2} \sqrt{\hat{\sigma}^2 c_{ii}}$$

3.6 Prueba de Hipótesis

La prueba de hipótesis $H_0 : \beta = \beta^*$ donde β^* es un vector conocido, es equivalente a la prueba de hipótesis de cada coeficiente $\beta_i = \beta_i^*$

En realidad para evaluar la función potencia de la distribución, también será importante conocer cuando la hipótesis alternativa $H_1 : \beta \neq \beta^*$ es verdadera.

Para esto se usará la razón de verosimilitud L .

La función de verosimilitud es:

$$f(\beta, \sigma^2) = \left(\frac{1}{(2\pi\sigma^2)^{n/2}} \right) \exp \left[-\frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta) \right]$$

El criterio de prueba es $L = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$, que a continuación se explicará lo que significa $L(\hat{\omega})$ y $L(\hat{\Omega})$.

La función de verosimilitud es una función de $p + 1$ parámetros donde Ω es de dimensión $p + 1$, sujeto a las siguientes restricciones $0 < \sigma^2 < \infty$ y $-\infty < \beta_i < \infty$ para $i = 1, \dots, p$.

El espacio $\hat{\Omega}$ serán los valores de los parámetros β_1, \dots, β_p y σ^2 en Ω tal que la función de verosimilitud sea máximo y $L(\hat{\Omega})$ será el valor máximo.

El espacio ω , estará sujeto a las siguientes restricciones $0 < \sigma^2 < \infty$ y $\beta_1 = \beta_1^*, \dots, \beta_p = \beta_p^*$, es decir ω es de una dimensión.

Para encontrar $L(\hat{\omega})$ y $L(\hat{\Omega})$ se usarán logaritmos en la función de verosimilitud.

Para encontrar $L(\hat{\omega})$ se procederá como sigue:

$$\log f(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta)$$

Como β_i^* son conocidos, la función solamente tiene un parámetro, que es σ^2 , y por lo tanto, el valor que maximiza la función de verosimilitud es:

$$\frac{d}{d\sigma^2} (\log f(\beta, \sigma^2)) = \frac{1}{2\sigma^4} (Y - X\beta^*)' (Y - X\beta^*) - \frac{n}{2\sigma^2} = 0$$

La solución es:

$$\hat{\sigma}^2 = \frac{(Y - X\beta^*)' (Y - X\beta^*)}{n}$$

entonces

$$L(\hat{\omega}) = \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} [(Y - X\beta^*)' (Y - X\beta^*)]^{n/2}}$$

De manera similar se obtendrá $L(\hat{\Omega})$, para obtener el valor máximo $L(\hat{\Omega})$ de la función de verosimilitud se resolverán las siguientes ecuaciones:

$$\frac{d}{d\beta} (\log f(\beta, \sigma^2)) = \frac{X'Y}{\sigma^2} - \frac{X'X\beta}{\sigma^2} = 0$$

$$\frac{d}{d\sigma^2} (\log f(\beta, \sigma^2)) = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{2\hat{\sigma}^4} - \frac{n}{2\hat{\sigma}^2} = 0$$

con lo cual

$$L(\hat{\Omega}) = \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} \left[(Y-X\hat{\beta})'(Y-X\hat{\beta}) \right]^{n/2}}$$

por lo tanto, la razón de verosimilitud es

$$L = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{(Y-X\beta^*)'(Y-X\beta^*)} \right]^{n/2}$$

Si $g(L, \beta^*)$ es la distribución de L bajo $H_0 : \beta = \beta^*$, la región crítica es $0 \leq L \leq A$, donde A es tal que:

$$\int g(L, \beta^*) dL = \alpha$$

donde α es la probabilidad del error tipo I.

Para determinar A , se debe encontrar $g(L, \beta^*)$ y la distribución de L bajo $H_0 : \beta = \beta^*$ entonces L sería determinada de la elección de los datos y H_0 sería rechazada si $L \leq A$.

Para determinar L , se deben estudiar las cantidades que están envueltas en L , es decir,

$$(Y-X\beta^*)'(Y-X\beta^*) \text{ y } (Y-X\hat{\beta})'(Y-X\hat{\beta})$$

con lo cual

$$\begin{aligned} (Y-X\beta^*)'(Y-X\beta^*) &= \left[(Y-X\hat{\beta}) - X(\beta^* - \hat{\beta}) \right]' \left[(Y-X\hat{\beta}) - X(\beta^* - \hat{\beta}) \right] \\ &= (Y-X\hat{\beta})'(Y-X\hat{\beta}) + (\beta^* - \hat{\beta})' X' X (\beta^* - \hat{\beta}) \\ &= (Y-X\hat{\beta})'(Y-X\hat{\beta}) + (\hat{\beta} - \beta^*)' X' X (\hat{\beta} - \beta^*) \end{aligned}$$

puesto que

$$X'(Y-X\hat{\beta}) = X'Y - X'X\hat{\beta} = X'Y - X'X((X'X)^{-1}X'Y) = X'Y - X'Y = 0$$

de igual manera

$$(Y-X\hat{\beta})X = 0$$

Ahora sí se substituye $\hat{\beta} = (X'X)^{-1}X'Y$ en

$$(Y-X\beta^*)'(Y-X\beta^*) = (Y-X\hat{\beta})'(Y-X\hat{\beta}) + (\hat{\beta} - \beta^*)' X' X (\hat{\beta} - \beta^*)$$

se obtiene lo siguiente:

$$\begin{aligned} (Y-X\beta^*)'(Y-X\beta^*) &= (Y-X(X'X)^{-1}X'Y)'(Y-X(X'X)^{-1}X'Y) \\ &\quad + ((X'X)^{-1}X'Y - \beta^*)' X' X ((X'X)^{-1}X'Y - \beta^*) \\ &= Y'(I - X(X'X)^{-1}X')Y + Y'X(X'X)^{-1}X'Y - Y'X\beta^* \\ &\quad - (\beta^*)' X'Y + (\beta^*)' X'X\beta^* \\ &= Y'(I - X(X'X)^{-1}X')Y + Y'X(X'X)^{-1}X'Y - Y'X\beta^* \\ &\quad - (X\beta^*)'Y + (X\beta^*)'X\beta^* \end{aligned}$$

como

$$\begin{aligned}(X\beta^*)'(I-X(X'X)^{-1}X')(-X\beta^*) &= 0 \\ Y'(I-X(X'X)^{-1}X')(-X\beta^*) &= 0 \\ (-X\beta^*)'(I-X(X'X)^{-1}X')Y &= 0\end{aligned}$$

por lo tanto

$$Y'(I-X(X'X)^{-1}X')Y = (Y-X\beta^*)'(I-X(X'X)^{-1}X')(Y-X\beta^*).$$

También como

$$\begin{aligned}Y'(X(X'X)^{-1}X')(X\beta^*) &= Y'X\beta^* \\ (X\beta^*)'(X(X'X)^{-1}X')Y &= (X\beta^*)'Y \\ (X\beta^*)'(X(X'X)^{-1}X')(X\beta^*) &= (X\beta^*)'X\beta^*\end{aligned}$$

tenemos que

$$\begin{aligned}Y'X(X'X)^{-1}X'Y - Y'X\beta^* - (X\beta^*)'Y + (X\beta^*)'X\beta^* &= \\ = Y'(X(X'X)^{-1}X')Y - Y'(X(X'X)^{-1}X')(X\beta^*) - \\ - (X\beta^*)'(X(X'X)^{-1}X')Y + (X\beta^*)'(X(X'X)^{-1}X')(X\beta^*) &= \\ = (Y-X\beta^*)'X(X'X)^{-1}X'(Y-X\beta^*).\end{aligned}$$

Ahora si $Z = (Y-X\beta^*)$ se obtiene que

$$Z'Z = (Y-X\beta^*)'(Y-X\beta^*) = Z'A_1Z + Z'A_2Z$$

donde

$$A_1 = I - X(X'X)^{-1}X' \quad A_2 = X(X'X)^{-1}X'$$

son matrices idempotentes.

Como Y está distribuido como $N(X\beta, \sigma^2 I)$ entonces Z está distribuido como $N(X\beta - X\beta^*, \sigma^2 I)$.

Como A_1 y A_2 son matrices idempotentes aplicando el teorema (2.18) tenemos que

$$\begin{aligned}i) \frac{Z'A_1Z}{\sigma^2} &\sim \chi^2_{(n-p)}, \\ ii) \frac{Z'A_2Z}{\sigma^2} &\sim \chi^2_{(p,\lambda)} \text{ donde } \lambda = \frac{(\beta-\beta^*)'X'A_2X(\beta-\beta^*)}{2\sigma^2}, \\ iii) \frac{Z'A_1Z}{\sigma^2} \text{ y } \frac{Z'A_2Z}{\sigma^2} &\text{son independientes.}\end{aligned}$$

Se ve que

$$\frac{Z'A_1Z}{\sigma^2} \sim \chi^2_{(k_1,\lambda_1)} \text{ donde el rango}(A_1) = k_1$$

y

$$\lambda_1 = \frac{(\beta-\beta^*)'X'A_1X(\beta-\beta^*)}{2\sigma^2}.$$

Puesto que A_1 es una matriz idempotente entonces

$$\begin{aligned}\text{rango}(A_1) &= \text{traza}(A_1) \\ &= \text{tr}(I) - \text{tr}(X(X'X)^{-1}X') = n - p.\end{aligned}$$

y $\lambda_1 = 0$, ya que

$$A_1X = (I - X(X'X)^{-1}X')X = 0.$$

Así que

$$\chi_{(n-p,0)}^2 = \chi_{(n-p)}^2 \cdot \frac{Z'A_2Z}{\sigma^2} \sim \chi_{(k_2, \lambda_2)}^2 \text{ donde el rango}(A_2) = k_2$$

y

$$\lambda_2 = \frac{(\beta - \beta^*)' X' (X(X'X)^{-1}X') X (\beta - \beta^*)}{2\sigma^2}$$

De la misma manera como A_2 es una matriz idempotente de $\text{rango}(A_2) = \text{tr}(A_2) = p$ y

$$\lambda_2 = \frac{(\beta - \beta^*)' X' X (\beta - \beta^*)}{2\sigma^2}$$

Como $X'X$ es semidefinida positiva entonces $\frac{Z'A_2Z}{\sigma^2}$ se distribuye como Ji-Cuadrada Central si sólo si $(\beta - \beta^*) = 0$, es decir si sólo si $H_0: \beta = \beta^*$ es verdadero.

Como

$$\frac{Z'A_1Z}{\sigma^2} \sim \chi_{(n-p)}^2$$

y

$$\frac{Z'A_2Z}{\sigma^2} \sim \chi_{(n-p)}^2$$

se sigue que

$$u = \frac{Z'A_2Z}{Z'A_1Z} \left(\frac{n-p}{p} \right) \sim F'$$

si sólo si H_0 es verdadera.

3.7 La Respuesta Estimada

Uno de los propósitos al ajustar un modelo es usarlo para pronosticar una respuesta dentro de la región de experimentación. Sea x'_p un vector cuyos elementos corresponden a los elementos de un renglón de la matriz X , donde $p = k + 1$. Si $\hat{\beta}$ representa el vector de elementos estimados de β , entonces una expresión para el valor pronosticado de la respuesta en cualquier punto x_1, x_2, \dots, x_k de la región experimental es:

$$\hat{Y} = x'_p \hat{\beta}. \quad (3.3)$$

El valor estimado de la respuesta correspondiente a la u -ésima observación se expresa como:

$$\hat{Y}_u = x'_{up} \hat{\beta},$$

donde x'_{up} es el u -ésimo renglón de X .

Una medida de la precisión del modelo esta dada por la varianza de \hat{Y} , la cual se expresa como:

$$\text{Var}(\hat{Y}) = \text{Var}(x'_p \hat{\beta}) = x'_p \text{Var}(\hat{\beta}) x_p = x'_p (X'X)^{-1} x_p \sigma^2. \quad (3.4)$$

Así, la matriz inversa $(X'X)^{-1}$ usada para estimar $\hat{\beta}$ es también usada para obtener las varianzas y covarianzas de los elementos de $\hat{\beta}$, además de la varianza de \hat{Y} .

3.8 Los Residuos

Siempre que se realiza un análisis estadístico es importante comprobar que los datos observados cumplan las hipótesis en que se basa el análisis. En el modelo de regresión lineal una manera eficaz de descubrir posibles deficiencias en el modelo o violaciones de las suposiciones radica en llevar a cabo un análisis de residuos.

Los residuos son definidos como la diferencia entre el valor observado de la respuesta y el valor estimado de la misma al emplear para ello el modelo ajustado. En otras palabras, la diferencia entre el valor observado de la respuesta en el u -ésimo experimento y el valor estimado por el modelo ajustado en el u -ésimo experimento, es llamado el u -ésimo residuo, y se expresará como

$$r_u = Y_u - \hat{Y}_u, \quad u = 1, 2, \dots, n.$$

Si los valores estimados son calculados para cada uno de los n ensayos, entonces el vector de residuos será

$$r = (r_1, r_2, \dots, r_n)' = \hat{Y} - X\hat{\beta}.$$

Con el análisis de residuos se puede detectar lo siguiente:

1. Si los residuos se distribuyen normalmente con una varianza constante.
2. Observar que una variable explicativa que ejerce influencia importante puede no estar incluida en el modelo.
3. Si se ha definido la ecuación de una manera correcta y no existe ninguna deficiencia.
4. Si hay observaciones aberrantes.
5. Si hay variables omitidas.
6. Si los residuos están correlacionados en vez de ser independientes como se supuso.

Es importante distinguir entre el error ε_u en Y_u respecto de $Y_u = \eta + \varepsilon_u$ y el residuo r_u respecto de \hat{Y}_u . Los residuos individuales son utilizados para probar qué tan bien se ajusta el modelo propuesto a los datos, además de que éstos sirven para probar la normalidad, independencia y varianza constante de los errores ε .

Vale la pena mencionar algunas propiedades de los residuos. Cuando el modelo ajustado contiene el término β_0 , la suma de los n residuos es igual a cero, esto es, $\sum_{u=1}^n r_u = 0$. Además, la suma de productos $\sum_{u=1}^n r_u x_{ui}$ es igual a cero, para toda $i = 1, 2, \dots, k$. En notación matricial, estas propiedades pueden escribirse:

1. $-1'r = 0$, donde $1'$ es un vector de unos.
2. $-\hat{Y}'r = 0$.
3. $-X'r = 0$.

3.9 Estimación de σ^2

La varianza de los errores σ^2 , es supuestamente conocida en la ecuación (3.2) para la matriz de varianza-covarianza de $\hat{\beta}$, además de en la fórmula de la ecuación (3.4) para la varianza de \hat{Y} . Esta suposición es raras veces cierta, sin embargo, generalmente es estimada. La estimación es obtenida a partir del análisis de los datos. Para el caso general, cuando el modelo ajustado contiene p parámetros y el número total de observaciones es n ($n > p$), la estimación de σ^2 es calculada como:

$$s^2 = \frac{1}{n-p} \sum_{u=1}^n r_u^2 = \frac{1}{n-p} (Y - X\hat{\beta})' (Y - X\hat{\beta}) = \frac{SSE}{n-p} = MSE$$

donde SSE es la suma de los residuos al cuadrado, y MSE es la suma media de residuos al cuadrado. El divisor $n-p$ representa los grados de libertad del estimador s^2 . Cuando el verdadero modelo está dado por $Y = X\beta + \varepsilon$, entonces s^2 es un estimador insesgado de σ^2 .

3.10 La Tabla de Análisis de Varianza

Los datos son analizados y los resultados del análisis son desplegados en forma de tabla, la cual es llamada tabla de análisis de varianza. Las entradas de la tabla representan medidas de información concernientes a las fuentes separadas de variación de los datos.

La variación total en un conjunto de datos es llamada la suma de cuadrados total (SST). La cantidad SST es calculada al sumar los cuadrados de las desviaciones de las respuestas observadas Y_u alrededor de su valor promedio, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$,

$$SST = \sum_{u=1}^n (Y_u - \bar{Y})^2. \quad (3.5)$$

La cantidad SST está asociada con $n-1$ grados de libertad, ya que la suma de las desviaciones $Y_u - \bar{Y}$, es igual a cero.

La suma de cuadrados total puede ser particionada en dos partes: la suma de cuadrados debida a la regresión (o suma de cuadrados explicada por el modelo ajustado) y la suma de cuadrados de los residuos (la suma de cuadrados no explicada por el modelo). La fórmula para el cálculo de la suma de cuadrados debida a la regresión (SSR) es:

$$SSR = \sum_{u=1}^n (\hat{Y}_u - \bar{Y})^2. \quad (3.6)$$

Si el modelo ajustado contiene p parámetros, entonces el número de grados de libertad asociados con SSR es $p-1$. La suma de cuadrados de los residuos está dada por:

$$SSE = \sum_{u=1}^n (Y_u - \hat{Y}_u)^2. \quad (3.7)$$

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados Medios
Regresión	$p-1$	SSR	$SSR/(p-1)$
Residuos	$n-p$	SSE	MSE
Total	$n-1$	SST	

Tabla 3.1 Tabla de análisis de varianza

El número de grados de libertad para SSE fue definido como $n - p$, que es la diferencia entre $(n - 1) - (p - 1) = n - p$.

Es posible simplificar las fórmulas para SST , SSR y SSE utilizando notación matricial. Sea $1'$ un vector de $1 \times n$ de unos, entonces:

$$\begin{aligned} SST &= Y'Y - n\bar{Y}^2. \\ SSR &= \hat{\beta}'X'Y - n\bar{Y}^2. \\ SSE &= Y'Y - \hat{\beta}'X'Y. \end{aligned}$$

La partición de la suma de cuadrados total en SSR y SSE es desplegada en la tabla de análisis de varianza Tabla 3.1

La prueba más común de significancia de ajuste de la regresión es una prueba de hipótesis nula dada por:

$$H_0 : \beta_i = 0 \forall i = 1, 2, \dots, k \quad \text{vs} \quad H_a : \beta_i \neq 0 \text{ para al menos una } i$$

suponiendo normalidad de los errores. Si H_0 es verdadera entonces el cociente

$$F = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{MSR}{MSE} \quad (3.8)$$

seguirá una distribución F con $p - 1$ y $n - p$ grados de libertad en el numerador y en el denominador respectivamente. El segundo paso de la prueba es comparar el valor de F dado por la ecuación (3.8) con el valor en tablas de la distribución $F_{\alpha, p-1, n-p}$, el cual es el porcentaje superior 100α de la distribución F con $p - 1$ y $n - p$ grados de libertad. Si el valor de F en la ecuación (3.8) excede el valor de tablas $F_{\alpha, p-1, n-p}$, entonces la hipótesis nula H_0 es rechazada al nivel de significancia α .

Una estadística que acompaña a la estadística F de la ecuación (3.8) es el coeficiente de determinación, el cual es una medida de la proporción de la variación total de los valores de Y_u alrededor de la media \bar{Y} explicada por el modelo ajustado.

3.11 El Coeficiente de Determinación

Al considerar la bondad de ajuste de la línea de regresión ajustada a un conjunto de datos, se intentará encontrar en qué medida se ajusta la línea de regresión muestral a los datos. Si todas las observaciones coincidieran con la línea de regresión, se obtendría un ajuste

“perfecto”, lo que raras veces ocurre. Generalmente tienden a presentarse r_i positivos y negativos con la esperanza de que los residuos localizados alrededor de la línea de regresión sean lo más pequeño posible.

En este sentido el coeficiente de determinación R^2 es una medida que nos dice que también la línea de regresión se ajusta a los datos. Utilizando el análisis de varianza se obtendrá R^2 . El análisis de varianza divide la variación total de las observaciones en sus partes componentes de acuerdo al modelo propuesto. Para esto se analizará la desviación de la observación Y_i de la media de las observaciones \bar{Y} . Considérense los siguientes casos:

a) Si se supone que todas las observaciones Y_i son iguales entre sí, así que las $\beta_j = 0$, para $j = 1, \dots, p$, $r_i = 0$ y $Y_i = \bar{Y}$ para toda i .

b) Por otro lado, si la magnitud de la desviación $Y_i - \bar{Y}$ es diferente de cero, ésta deberá atribuirse a las componentes del modelo.

Para la magnitud de la desviación

$$Y_i - \bar{Y}$$

al sumar y restar el estimador \hat{Y}_i entonces

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

De aquí, la desviación total de la observación Y_i con respecto a la media \bar{Y} es la suma de la desviación de \hat{Y}_i estimada de la media \bar{Y} y la desviación de Y_i con respecto a \hat{Y}_i .

Las desviaciones $Y_i - \hat{Y}_i$ representan la contribución al componente del error y a la variación total. Si la magnitud de la desviación de $\hat{Y}_i - \bar{Y}$ es grande, entonces se tiene un efecto lineal.

Siguiendo el análisis de varianza se elevan al cuadrado ambos miembros de la identidad

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

y se sumarán para todas las observaciones. Entonces se tiene

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + 2 \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i).$$

Puesto que

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i) = \sum_{i=1}^n \hat{Y}_i (Y_i - \hat{Y}_i) - \sum_{i=1}^n \bar{Y} (Y_i - \hat{Y}_i)$$

y como

$$\sum_{i=1}^n \hat{Y}_i (Y_i - \hat{Y}_i) = \sum_{i=1}^n (\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \varepsilon_i = 0$$

dado que

$$\beta_j \sum_{i=1}^n x_{ij} r_i = 0$$

puesto que los residuos no están correlacionados con x_{ij} . Por lo tanto, la ecuación del análisis de regresión es:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

El primer término representa la variación total de las observaciones con respecto a la media

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

El segundo término representa la variación total de los valores estimados de Y con respecto a la media

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2.$$

El último término representa la variación residual o la variación de las observaciones con respecto a los valores estimados

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Con lo cual $SST = SSR + SSE$, donde se muestra que la variación de los valores observados de Y_i alrededor de su media pueden dividirse en dos componentes, el primero en la línea de regresión y el segundo en los errores aleatorios. Al dividir ambos miembros por SST , se obtiene:

$$1 = \frac{SSR}{SST} + \frac{SSE}{SST}$$

se tiene

$$\begin{aligned} SSE &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n r^2 = (Y - X\hat{\beta})' (Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - (X\hat{\beta})'Y + (X\hat{\beta})'X\hat{\beta} \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - \hat{\beta}'X'Y \end{aligned}$$

puesto que

$$\begin{aligned} \hat{\beta}'X'X\hat{\beta} &= ((X'X)^{-1}X'Y)'X'X\hat{\beta} \\ &= Y'X\hat{\beta} \end{aligned}$$

$$SST = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = Y'Y - n\bar{Y}^2$$

y como

$$\begin{aligned} SSR &= SST - SSE \\ &= Y'Y - n\bar{Y}^2 - (Y'Y - \hat{\beta}'X'Y) \\ &= \hat{\beta}'X'Y - n\bar{Y}^2 \end{aligned}$$

Así, el coeficiente de determinación queda definido como:

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

3.12 La Elección del Diseño

La elección del diseño (i. e. la combinación de factores donde se toma la muestra) debe responder a los diferentes intereses del investigador, como por ejemplo: la economía (esto es tanto por el número de observaciones, como por la cantidad de cálculos implicados) y la precisión de la estimación (i. e. que la varianza $V(\hat{\eta})$ sea pequeña).

De esta manera, es conveniente contar con criterios para elegir aquellos diseños que sean *económicos* y *precisos*. Con la finalidad de establecer dichos criterios se discuten las características del modelo lineal para ajustar un polinomio de grado r a los datos, ($r=1$ y 2); con k factores.

Capítulo 4 MODELOS Y DISEÑOS DE PRIMER ORDEN

Si la respuesta es descrita adecuadamente por una función lineal, el modelo de primer orden (grado $r=1$) está dado por:

$$Y = \beta_0 + \beta_1 x_{1u} + \beta_2 x_{2u} + \cdots + \beta_k x_{ku} + \varepsilon_u \quad \forall u = 1, \dots, n \quad (4.1)$$

el cual puede ser tratado como un caso particular de regresión lineal múltiple, donde los estimadores de los parámetros $\hat{\beta}_i$ están dados por el método de mínimos cuadrados como:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad y \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

donde el modelo puede ser escrito como:

$$Y = X\beta + \varepsilon$$

y la estimación de los parámetros dada en forma matricial es:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

donde Y es la variable de respuesta observada, $\beta_0, \beta_1, \dots, \beta_k$ son parámetros desconocidos, y ε es un término de error aleatorio. Si ε tiene media cero, entonces la porción no aleatoria del modelo en la ecuación (4.1) representa la verdadera respuesta media η , esto es:

$$\eta = \beta_0 + \sum_{i=1}^k \beta_i X_i \quad (4.2)$$

Si, sin embargo, el modelo en la ecuación (4.2) no representa debidamente la respuesta media (lo cual ocurre cuando la relación entre la respuesta y las variables de entrada está sobresimplificada), entonces ε contiene, en adición al error experimental, un error no aleatorio (o error sistemático). El último error es atribuido a la ausencia de términos x_1, x_2, \dots, x_k de grado mayor que uno, omitidos en el modelo (4.1), además de la ausencia de otras variables las cuales tienen influencia en la respuesta. Este error adicional es llamado error de falta de ajuste.

El estimador $\hat{\beta}$ es el mejor estimador lineal insesgado de β (cuando se cumplen los supuestos de normalidad y varianza constante).

4.1 Diseños 2^k Factoriales

Existe una clase única de diseños que minimizan la varianza de los coeficientes de regresión $\{\hat{\beta}_i\}$. Estos son los diseños ortogonales de primer orden. Un diseño de primer orden es ortogonal si los elementos fuera de la diagonal de la matriz $(X'X)$ son iguales a cero. Esto implica que la suma de los productos cruzados de las columnas de la matriz X son iguales a cero.

Esta clase de diseños ortogonales de primer orden incluyen diseños factoriales 2^k , en los que los efectos principales no son alias entre sí. Al usar estos diseños se considera que los k factores se codifican en los niveles estandarizados ± 1 .

El diseño 2^k no permite la estimación del error experimental a menos que se repitan algunos ensayos. Un método común para incluir la repetición de ensayos en el diseño 2^k consiste en aumentar el diseño con algunas observaciones en el centro (el punto $x_i = 0$, $i = 1, 2, \dots, k$). La adición de puntos centrales al diseño 2^k no tiene influencia sobre las $\{\hat{\beta}_i\}$ para $i \geq 1$, pero como resultado, la estimación de β_0 es el promedio general de todas las observaciones. Además, agregar puntos centrales no altera las propiedades de ortogonalidad del diseño.

En un diseño factorial 2^k cada variable de entrada o factor es medida en dos niveles alto y bajo, los cuales pueden codificarse de tal manera que cuando la variable se encuentre en su nivel bajo, tome el valor de -1 y cuando se encuentre en su nivel alto tome el valor de $+1$. Al considerar todas las posibles combinaciones de los niveles de los k factores, obtendremos una matriz de diseño D cuyas entradas serán las variables codificadas. En el u -ésimo renglón, los elementos serán iguales a $+1$ o -1 y representarán las coordenadas del punto de diseño en la u -ésima corrida experimental. Cada uno de los 2^k renglones de D representará una posible combinación de niveles de los factores o tratamientos.

El uso de variables codificadas en lugar de las variables originales facilita la construcción del diseño experimental, ya que al codificar las variables de entrada se remueven las unidades de medida, se estandarizan así las distancias entre las distintas variables y sus niveles, ya que tendrán la misma métrica, de este modo la gráfica definirá un hipercubo.

Para codificar una variable se debe aplicar la siguiente transformación:

$$x_{ui} = \frac{2(\xi_{ui} - \bar{\xi}_i)}{R_i} \quad i = 1, 2, \dots, k \quad (4.3)$$

donde X_{ui} es el valor original de la variable de entrada en el i -ésimo factor para la u -ésima corrida experimental \bar{X}_i es el promedio entre el nivel alto y bajo del i -ésimo factor, y R_i es el rango o la diferencia entre el nivel alto y bajo.

En general, un diseño factorial 2^k consta de todos los puntos 2^k (corridas) con los niveles:

$$(x_1, x_2, \dots, x_k) = (\pm 1, \pm 1, \dots, \pm 1)$$

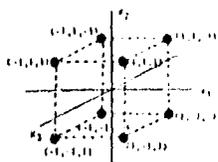


Figura 1 Diseño 2^3 factorial

I	1	2	3	12	13	23	123	Y
+	-	-	-	+	+	+	-	2.83
+	+	-	-	-	-	+	+	3.56
+	-	+	-	-	+	-	+	2.23
+	+	+	-	+	-	-	-	3.06
+	-	-	+	+	-	-	+	2.47
+	+	-	+	-	+	-	-	3.30
+	-	+	+	-	-	+	-	1.95
+	+	+	+	+	+	+	+	2.56
8	4	4	4	4	4	4	4	

Tabla 4.1 Tabla de diseño experimental

donde toda posible combinación de signos \pm es seleccionada. Geométricamente el diseño consta de los vértices de un hipercubo en k dimensiones. La Figura 1 es un ejemplo de un diseño 2^3 factorial. Para propósitos de análisis es conveniente listar las corridas en orden standard, no en el orden en el cual fueron tomadas (aleatorio); este orden standard se obtiene al escribir alternadamente $-$ y $+$ en la columna encabezada por 1, los signos $--$ y $++$ alternadamente en la columna encabezada por 2, escribir alternadamente cuatro $---$ y $+++$, en la columna 3, y así sucesivamente ver Tabla 4.1.

La construcción de una tabla para un diseño 2^k factorial se realiza de la siguiente manera: se inicia con una columna de $+1$'s de longitud 2^k , a continuación se escriben las k columnas de las variables de entrada en el orden standard, después se escriben las $(2^k - k - 1)$ columnas de interacciones de 12, 13, ..., 123, ..., 123... k obtenidas de multiplicar los elementos renglón por renglón en la forma indicada por el encabezado de la columna, al pie de cada columna se escribe su divisor, esto es el número de signos $+$ en cada columna, y finalmente se anota la columna de respuesta, cuyo encabezado será y .

4.1.1 Análisis del Diseño 2^k Factorial

El efecto principal de una variable es definido como la diferencia promedio en el nivel de respuesta al pasar del nivel bajo al nivel alto de esa variable. Una propiedad importante de los diseños factoriales, es que hacen posible el calculo de todos efectos principales de las variables y de los efectos de todas las interacciones entre las variables.

Dos variables se dicen que interactúan (en su efecto sobre la respuesta), si el efecto de una de las variables es diferente en los dos niveles diferentes de la otra variable. La interacción entre la variable x_1 y la variable x_2 se denotará por 12.

El efecto principal de x_1 puede calcularse multiplicando la columna de datos y , por la columna de $- y +$ encabezada por x_1 y dividiendo la por el total de signos $+$ en la columna.

Para calcular la interacción entre dos o más variables, se toma la columna de la tabla que contiene dicha combinación de variables y se multiplica por la columna de datos y .

4.2 Diseños 2^{k-p} Factoriales

Observe que cuando recopilamos todos los datos de las 2^k corridas experimentales del diseño 2^k factorial, es posible estimar todos los coeficientes del modelo

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \sum_{j=1}^k \beta_{ij} x_i x_j + \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \beta_{ijl} x_i x_j x_l + \beta_{12\dots k} x_1 x_2 \cdots x_k + \varepsilon \quad (4.4)$$

donde los términos $\beta_{ij} x_i x_j$, $\beta_{ijl} x_i x_j x_l$, ..., $\beta_{12\dots k} x_1 x_2 \cdots x_k$ en la ecuación (4.4) son asociados con las interacciones entre los k factores.

Si se quisieran dividir las 2^k corridas experimentales en dos bloques (I y II) cada uno con 2^{k-1} corridas (la mitad de las corridas en cada bloque), y si se repartieran las corridas para las cuales el producto de las columnas 123456 toman el signo menos para el bloque I, y toman el signo más para el bloque II, entonces, según Fisher, diremos que el efecto del bloque (denotado por B) está completamente confundido con la interacción 123456, esto es:

$$B = 123456.$$

Así se pierde la capacidad para estimar (al menos con la misma exactitud) la interacción 123456. Sin embargo, en muchos contextos prácticos, esta interacción de alto orden será probablemente de tamaño insignificante y de poco interés.

El diseño 2^k factorial tiene una importante propiedad ortogonal por lo cual una sucesión de signos correspondiente a un efecto particular es ortogonal a cualquier otra sucesión.

4.2.1 Cuatro y Ocho Bloques

Ahora, se trata de ordenar un diseño en cuatro bloques, cada uno de los cuales contendrá la cuarta parte del número total de corridas del experimento. Esto puede lograrse confundiendo dos interacciones de alto orden con bloques de contraste. Si se asocian los contrastes

123456 y 23456 con bloques (esto puede parecer razonable, ya que se quieren confundir interacciones del más alto orden posible). Así los bloques generadores para el diseño serán:

$$B_1 = 123456, \quad B_2 = 23456, \quad (4.5)$$

Las corridas serán entonces repartidas en los cuatro bloques de acuerdo al siguiente esquema:

$$\begin{array}{cc} & - & + \\ - & I & II \\ + & III & IV \end{array}$$

Así, las corridas serán asignadas a los bloques I, II, III, IV de acuerdo a los signos asociados con las columnas 123456 y 23456 por parejas $(-, -)$, $(-, +)$, $(+, -)$, y $(+, +)$. Desafortunadamente, una seria dificultad ocurre con esta forma de arreglo. Hay, por supuesto 3 grados de libertad asociados entre los cuatro bloques. Si se asocian B_1 y B_2 con dos de estos contrastes, el tercero debe ser la interacción $B_1 \times B_2 = B_1 B_2$. Sin embargo, entonces la interacción $B_1 B_2$ se confundirá con la interacción entre 123456 y 23456. Para ver el significado de esta interacción, considere la siguiente tabla de signos en orden standard

123456	+	-	-	+	-	+	+	-	...	-	+
23456	-	-	+	+	+	+	-	-	...	+	+
Producto	-	+	-	+	-	+	-	+	...	-	+

Vemos que la interacción entre 123456 y 23456 es el efecto principal 1. Este arreglo sería así muy pobre, debido a que el efecto principal de la primera variable se confundiría con bloques diferentes. Para tomar una mejor elección, es necesario entender como calcular las interacciones entre efectos complejos. Para hacer esto, hay que tomar cualquier conjunto de signos con la finalidad de obtener algún (digamos 1) y multiplicar aquellos signos por los signos del mismo efecto, y así se obtendrá un renglón de +/s, el cual será denotado por la identidad I.

1	-	+	-	+	-	+	-	+	...	-	+	-	+
1	-	+	-	+	-	+	-	+	...	-	+	-	+
$1 \times 1 = I$	+	+	+	+	+	+	+	+	...	+	+	+	+

Así, la multiplicación de cualquier contraste con la identidad I no afectará el contraste

$$1 \times I = 1, \quad 23 \times I = 23, \quad 12345 \times I = 12345.$$

Aplicando esta regla al diseño anterior, tenemos que:

$$B_1 B_2 = 123456 \times 23456 = 12^2 3^2 4^2 5^2 6^2 = 1 \times I^5 = 1$$

lo que indica que $B_1 B_2$ y 1 están confundidos.

Un arreglo de cuatro bloques para seis factores, de modo que las interacciones confundidas con bloques sean del más alto orden posible puede ser obtenido usando como generadores dos interacciones de cuatro factores en las cuales sólo dos símbolos se traslapen. Si elegimos, por ejemplo:

$$B_1 = 1234$$

$$B_2 = 3456$$

entonces

$$B_1 B_2 = 1256.$$

Para dividir el experimento en ocho bloques, elegiríamos los generadores

$$B_1 = 1234, \quad B_2 = 3456, \quad B_3 = 136,$$

obteniendo así, el siguiente patrón de confusiones para los siete contrastes entre los ocho bloques:

$$\begin{aligned} B_1 &= 1234 \\ B_2 &= 3456 \\ B_1 B_2 &= 1256 \\ B_3 &= 136 \\ B_1 B_3 &= 246 \\ B_2 B_3 &= 145 \\ B_1 B_2 B_3 &= 235. \end{aligned}$$

Bajo la suposición de que los bloques tienen sólo efectos aditivos, esto es, no contienen efectos principales o interacciones de dos factores confundidas con bloques.

4.2.2 Análisis del diseño 2^{k-p} factorial

Para ilustrar la utilidad de los diseños, conviene considerar el problema de dividir en dos bloques el diseño 2^k de seis factores, usando para ello la interacción 123456 ($=B$) como bloque de contraste. Para cada una de las 32 corridas en el primer bloque, los signos de los seis factores se multiplicarán para dar -1 , de este modo tendremos:

$$\begin{aligned} 123456 &= -I, \quad \text{para el bloque I,} \\ 123456 &= I, \quad \text{para el bloque II} \end{aligned}$$

Si supone ahora que, en lugar de ejecutar las 64 corridas, se hacen sólo 32, esto es el bloque II, estas son las corridas para las cuales los elementos de 123456 son $+$. Así se muestra la tabla 4.2 de signos con el cálculo de los efectos principales y las interacciones de dos, tres, cuatro, cinco y seis factores, justo como un diseño factorial completo, pero usando únicamente las corridas del bloque II.

Se puede utilizar la tabla de signos para calcular los 63 efectos principales y las interacciones. No obstante, algo parece andar mal, ya que difícilmente es posible esperar calcular 63 efectos independientes con sólo 32 observaciones. La explicación se encuentra en la duplicidad que ocurre en la tabla 4.2. Como era de esperarse, los signos de la columna 123456 son idénticos a los de la columna I. Se observa también, sin embargo, que los signos de la columna de interacciones 23456 son idénticos a aquellos de la columna I, o simbólicamente $I=23456$. Entonces, así, se obtiene no sólo una estimación para I , o la interacción 23456, sino también de su suma:

$$I_1 \rightarrow I + 23456.$$

corrida	I	1	2	3	4	5	6	12	13	...	23456	123456
1	+	-	-	-	-	-	-	+	+	...	-	+
4	+	+	+	-	-	-	-	+	-	...	+	+
6	+	+	-	+	-	-	-	-	+	...	+	+
7	+	-	+	+	-	-	-	-	-	...	-	+
10	+	+	-	-	+	-	-	-	-	...	+	+
...			...									
...			...									
64	+	+	+	+	+	+	+	+	+	...	+	+
divisor	8	16	16	16	16	16	16	16	16	...	16	16

Tabla 4.2 Diseño factorial fraccionado 2^{6-1}

La treta para identificar todos estos efectos confundidos (alias), es proporcionada por el generador de este diseño fraccionado, el cual puede escribirse como:

$$I = 123456 \tag{4.6}$$

La ecuación simplemente nos dice cómo escoger las corridas del diseño completo cuyo producto 123456 sea positivo, igual que la columna I. Multiplicando ambos lados de la ecuación (4.6) por 1, se tiene que $1=23456$. Todas las demás relaciones de confusión (alias) pueden ser obtenidas de forma similar.

Las interacciones 12, 345, ..., para variables cuantitativas, reflejan los efectos de las derivadas conjuntas de la función respuesta η , tal como:

$$\frac{\partial^2 \eta}{\partial x_1 \partial x_2}, \frac{\partial^3 \eta}{\partial x_3 \partial x_4 \partial x_5}, \dots$$

Parecería razonable suponer que las derivadas de orden mayor a dos se ignorarán. Esto es, en la expansión de la serie de Taylor de la función respuesta, los términos de tercer orden y mayores se considerarán como insignificantes dentro de la región local de experimentación. En estas circunstancias se podrían ignorar las interacciones de tres o más factores en el análisis de los datos. Así, para que un menor número de corridas puedan proporcionar una exactitud adecuada comparable a la del diseño completo es necesario que σ^2 no sea demasiado grande, de este modo, los diseños factoriales fraccionados son muy útiles en situaciones experimentales con mas de tres o cuatro variables.

4.2.3 Resolución de un Diseño 2^{k-p} Factorial

La resolución R de un diseño factorial fraccional 2^{k-p} es la longitud de la palabra más corta en la relación de definición. Por ejemplo, un diseño 2^{6-1} con la relación de definición $I=123456$ es de resolución VI (la resolución es expresada en números romanos). El diseño 2^{6-2} con la relación de definición $I=1234=3456=1256$ tiene resolución IV.

Para ajustar una superficie de respuesta deberemos usar aproximaciones polinomiales a la función respuesta de primero y segundo grado. Los diseños de primer grado

deberán tener al menos resolución III, para poder obtener estimaciones de los coeficientes en la ecuación. Esto asegurará que ningún efecto principal estará confundido con otro efecto principal, aunque quizá algún efecto principal pueda estar confundido (si $R=III$) con una interacción de dos factores. Un diseño de resolución IV proporcionará seguridad adicional debido a que entonces ningún efecto principal será confundido con otro efecto principal o interacciones de dos factores.

4.3 Diseños Factoriales Fraccionados Aplicados a la Metodología de Superficies de Respuesta

Puede suponerse que la respuesta esperada $E(y) = \eta$ es una función de k variables de predicción x_1, x_2, \dots, x_k , codificadas de tal manera que la región de interés está en el origen $(0, 0, \dots, 0)$. Al considerar la expansión de la serie de Taylor se tiene:

$$E(y) = \eta = \eta_0 + \sum_{i=1}^k \left[\frac{\partial \eta}{\partial x_i} \right]_0 x_i + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \left[\frac{\partial^2 \eta}{\partial x_i \partial x_j} \right]_0 x_i x_j + \dots \quad (4.7)$$

donde el subíndice cero indica la evaluación en el origen. Si ignoramos los términos de orden mayor uno, la expansión producirá la aproximación de primer grado (primer orden o plano):

$$\eta = \beta_0 + \sum_{i=1}^k \beta_i x_i \quad (4.8)$$

Si además conservamos los términos de segundo grado obtendremos la aproximación de segundo grado (de segundo orden o cuadrática):

$$\eta = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \sum_{j \geq i}^k \beta_{ij} x_i x_j \quad (4.9)$$

Si los diseños factoriales fraccionados son usados como diseños de primer orden, éstos deben tener una resolución al menos III, de modo que si el modelo (4.8) es válido, el contraste de los efectos principales del diseño, proporcionarán estimaciones de todos los β_i que no serán confundidos con otros.

Cuando un diseño de nivel dos es utilizado como un block de construcción en diseños compuestos de segundo orden, se requiere normalmente de resolución al menos V. Pero si el modelo (4.9) es válido, los efectos principales y las interacciones de dos factores del diseño de nivel dos proporcionarán estimaciones de todos los coeficientes principales β_i y de todos los coeficientes de interacción β_{ij} ($i \neq j$), que tampoco se confundirán con otros.

Debido a que nunca se podrá asegurar nada acerca de la adecuación de la aproximación (interpolación) elegida, esto es, del grado del polinomio seleccionado, será necesario verificar la falta de ajuste, además de los parámetros estimados del modelo. La adición de puntos centrales en el diseño hace posible verificar la curvatura general. Un diseño de

resolución IV proporcionará seguridad en los diseños de primer orden, ya que los coeficientes no serán confundidos con interacciones de dos factores.

Las estimaciones del diseño factorial fraccionado pueden ser graficadas en un papel de probabilidad normal, para distinguir los efectos reales del ruido.

4.3.1 Papel de Probabilidad Normal

Un método simple para checar la suposición de normalidad es graficar los residuos en papel de probabilidad normal. Una gráfica de este tipo es la representación de la distribución acumulada de los residuos sobre papel de probabilidad normal. Si la distribución de los errores es normal, esta gráfica parecerá una línea recta. Al visualizar dicha línea hay que poner más énfasis en los valores centrales que en los extremos de la gráfica. Aunque pequeñas desviaciones de normalidad no afectan el modelo. Si llegamos al problema de que los residuos no siguen una distribución normal entonces los estimadores por mínimos cuadrados siguen siendo estimadores insesgados óptimos, pero todas las pruebas de significancia que se apliquen carecen de valor, ya que las estadísticas t o F y los intervalos de confianza y predicción dependen de la suposición de normalidad.

4.3.2 Gráfica de Residuos Contra Valores Estimados \hat{Y}_i

Esta gráfica sirve para detectar si la varianza del error es o no constante. Si el modelo es correcto y las suposiciones se satisfacen entonces la gráfica de los residuos contra los valores estimados \hat{Y}_i tenderán a encontrarse dentro de una banda horizontal centrada alrededor del valor cero, sin ninguna tendencia sistemática. Cualquier desviación con respecto a este comportamiento indicará la existencia de un problema. Un defecto que en ocasiones revela la gráfica es el de una varianza no constante.

Capítulo 5

MÁXIMA PENDIENTE DE ASCENSO

Si se intenta ajustar una función empírica sobre la *región operativa* completa (esto es, sobre la región entera dentro de la cual el sistema estudiado puede ser operado) sería necesaria una función muy compleja. El ajuste de tal función involucraría un número excesivo de corridas experimentales. La exploración de la *región operativa* entera es, sin embargo, un objetivo poco sensato o pocas veces factible. Primero, la extensión de esta región es pocas veces conocida y, segundo el investigador puede a menudo rechazar áreas completas de la región por considerarlas poco eficientes. Lo que generalmente se busca hacer, es explorar una más pequeña subregión de interés, la cual puede cambiar al desarrollarse la investigación. Se puede esperar que, dentro de esta más pequeña subregión, una función de interpolación represente adecuadamente los datos.

El objetivo del método de máximo ascenso es encontrar rápidamente la vecindad del óptimo, usando un procedimiento experimental simple, económico y eficiente. En la lejanía del óptimo, generalmente se supone que el modelo de primer orden es una aproximación adecuada a la superficie de respuesta en regiones pequeñas de las x 's. Este método es un procedimiento que recorre secuencialmente la superficie de respuesta a lo largo de la trayectoria de máxima pendiente de ascenso; en otras palabras, en la dirección de máximo incremento de la respuesta.

La dirección de ascenso máximo es aquella en la que \hat{Y} aumenta más rápidamente. Esta dirección es paralela a la normal de la superficie de respuesta ajustada, por lo regular, la trayectoria de máxima pendiente de 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 $\{\hat{\beta}_i\}$.

El procedimiento matemático de ajuste seleccionado, de todas las posibles superficies del grado ajustado, es aquel que aproxima la respuesta a los puntos experimentales más cercanos (en el sentido de mínimos cuadrados). Las características de la superficie ajustada en puntos lejanos de la región del diseño experimental tienen algún parecido, aunque no necesariamente con las características de la superficie actual. Se puede esperar que la aproximación sea útil sólo en la vecindad inmediata de la región actual de experimentación. Si, por ejemplo, se estuviera cerca de un máximo sería posible representar de manera aproximada sus principales características ajustando para un conjunto localmente aceptable de puntos experimentales, tal vez con variables transformadas, una ecuación de

segundo grado. Sin embargo, si el sistema es investigado por primera vez, es común que las condiciones iniciales se encuentren lejos del máximo. Así, el investigador debe realizar primero un procedimiento preliminar para obtener un punto aceptable, donde la ecuación de segundo grado puede emplearse de manera exitosa.

5.1 ¿Qué Hace el Paso de Máximo Ascenso?

En el primer paso de la investigación, cuando pueden lograrse una gran cantidad de progresos, entonces es posible esperar que las ecuaciones de primer grado (contornos planos), proporcionen una buena aproximación alrededor del punto inicial de experimentación, el cual en principio está lejos del óptimo. Una dirección correcta sobre este plano es la dirección de máximo ascenso. En la práctica, la ecuación de primer grado debe ser estimada vía experimentos. Esto se logra corriendo un diseño alrededor de un punto inicial.

Generalmente se encontrará que, después de una o quizá dos aplicaciones del método de máximo ascenso, los efectos de primer orden desaparecerán y la aproximación de primer orden será inadecuada. En este paso, deberán aplicarse otros métodos más sofisticados. La técnica de máximo ascenso es particularmente efectiva para la investigación inicial de nuevos procesos.

Si se busca por ejemplo, maximizar el rendimiento, sería demasiado costoso investigar en detalle las regiones de baja producción. Generalmente, las principales características de estas regiones muestran efectos de primer orden, los cuales apuntan a la dirección de ascenso de la superficie. Una vez que los efectos de primer orden se vuelven pequeños comparados con los efectos de segundo orden o con el error, o ambos, será necesario aproximar un modelo de segundo orden. Para verificar la adecuación del modelo de primer orden o, equivalentemente, la necesidad de un modelo de segundo orden, necesitamos:

1. Examinar las interacciones, que no son utilizadas para estimar los efectos de primer orden.
2. Estimar la curvatura total.

Una prueba útil de curvatura es proporcionada al añadir algunos puntos centrales al diseño factorial. La forma en que el diseño es completado, añadiendo puntos centrales, es aleatoriamente. Así, un conjunto de corridas experimentales repetidas aleatoriamente, proporcionan una estimación del error puro de la variación básica de los datos. La prueba de la curvatura total es proporcionada por la diferencia

$$c = \{ \text{respuesta promedio de las corridas factoriales en dos niveles} \} \\ - \{ \text{respuesta promedio de las corridas centrales} \}.$$

Puede mostrarse que si la verdadera función respuesta es de segundo grado, c proporciona

un estimador insesgado de la suma de los efectos cuadráticos puros, esto es:

$$c \rightarrow \sum_{i=1}^k \beta_{ii}.$$

Para una superficie que contiene un minimax, los β_{ii} serían de signos diferentes y teóricamente $E(c)$ sería cero aún cuando hubiera una gran curvatura. Sin embargo, las superficies minimax son raras en la práctica y generalmente, los β_{ii} serán del mismo signo.

5.2 Experimentación a lo Largo del Camino de Máximo Ascenso

El procedimiento de máximo ascenso consiste en ejecutar una sucesión de experimentos a lo largo del camino de máximo incremento de la respuesta (recuerde que la dirección depende de la escala de las variables codificadas). El procedimiento comienza al ajustar la superficie de respuesta para usar una ecuación de primer orden. La información que proporciona el modelo estimado se utiliza para determinar la dirección hacia la cual se incrementa la respuesta. Mientras se sube por la superficie de respuesta y se avanza en dirección a la región donde se presenta la curvatura, el incremento de la respuesta eventualmente se estabiliza en el punto más alto de la superficie. Si se continúa en esa dirección y la altura de la superficie decrece, un nuevo conjunto de experimentos deberá ser ejecutado, y otra vez el modelo de primer orden deberá ser ajustado. Una nueva dirección de incremento de la respuesta deberá a ser determinada y otra sucesión de experimentos a lo largo de esa dirección deberá realizarse. Esta sucesión de pruebas continuará hasta que sea evidente que poco o ningún incremento adicional en la respuesta puede ser logrado por este método.

Para describir el método de máximo ascenso matemáticamente, es preciso suponer que la verdadera superficie de respuesta puede ser aproximada localmente por el modelo de primer orden:

$$\eta = \beta_0 + \sum_{i=1}^k \beta_i x_i.$$

Los datos que se obtienen de los puntos de un diseño de primer orden y se utilizan para calcular las estimaciones de los coeficientes del modelo de primer orden, son:

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i. \quad (5.1)$$

El siguiente paso es alejarse una distancia de r unidades del centro del diseño en la dirección de máximo incremento de la respuesta. Al elegir el centro de diseño de las variables codificadas x_1, x_2, \dots, x_k éste será denotado por $(0, 0, \dots, 0)$ entonces el movimiento del centro r unidades afuera, está dado por la solución de:

$$\max_{x_i} \left(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i \right)$$

$$s.a. \sum_{i=1}^k x_i^2 = r^2.$$

La restricción se establece por las siguientes razones:

1. Maximizar η sin ninguna restricción produce soluciones insatisfactorias, cuando el modelo es un plano horizontal todos los vectores son óptimos y cuando el plano tiene cierta inclinación η es no acotada.
2. Maximizar η sobre toda la región de diseño puede resultar analíticamente muy complicado.
3. El valor de r es variable de tal manera que se pueden considerar todos los valores de x que determinan la región experimental.

La maximización de la función respuesta (objetivo) es realizada a través del método de multiplicadores de Lagrange. Sea:

$$Q(x_1, x_2, \dots, x_k) = \left(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i \right) - \lambda \left(\sum_{i=1}^k x_i^2 - r^2 \right)$$

donde λ es el multiplicador de Lagrange. Para maximizar \hat{Y} en la ecuación (5.1) sujeta a la restricción mencionada, se calculan las siguientes derivadas parciales:

$$\frac{\partial Q(x_1, x_2, \dots, x_k)}{\partial x_i} = b_i - 2\lambda x_i \quad \forall i = 1, 2, \dots, k$$

y

$$\frac{\partial Q(x_1, x_2, \dots, x_k)}{\partial \lambda} = - \left(\sum_{i=1}^k x_i^2 - r^2 \right)$$

y se igualan a cero:

$$\hat{\beta}_i - 2\lambda x_i = 0 \quad \forall i = 1, 2, \dots, k \quad (5.2)$$

y

$$- \sum_{i=1}^k x_i^2 + r^2 = 0. \quad (5.3)$$

despejando x_i de (5.2) tenemos

$$x_i = \frac{\hat{\beta}_i}{2\lambda} \quad \forall i = 1, 2, \dots, k \quad (5.4)$$

elevando al cuadrado ambos lados de la igualdad y sumando sobre $i = 1, 2, \dots, k$ se tiene

$$\sum_{i=1}^k x_i^2 = \frac{\sum_{i=1}^k \beta_i^2}{4\lambda^2}$$

usando la restricción se obtiene

$$r^2 = \frac{\sum_{i=1}^k \beta_i^2}{4\lambda^2}$$

despejando λ resulta

$$\lambda = \frac{\pm \sqrt{\sum_{i=1}^k \beta_i^2}}{2r}$$

Los posibles valores de λ corresponden a un máximo y a un mínimo, ya que η es una función continua y se optimiza sobre un conjunto cerrado y acotado.

Si

$$\lambda = \frac{\sqrt{\sum_{i=1}^k \beta_i^2}}{2r}$$

entonces,

$$x_i = \frac{\beta_i r}{\sqrt{\sum_{i=1}^k \beta_i^2}}$$

sustituyendo en la función objetivo se tiene

$$\hat{\eta}_{op} = \beta_0 + \sum_{i=1}^k \beta_i \left[\frac{\beta_i r}{\sqrt{\sum_{i=1}^k \beta_i^2}} \right] = \beta_0 + r \sqrt{\sum_{i=1}^k \beta_i^2}$$

Si

$$\lambda = -\frac{\sqrt{\sum_{i=1}^k \beta_i^2}}{2r}$$

entonces,

$$x_i = -\frac{\beta_i r}{\sqrt{\sum_{i=1}^k \beta_i^2}}$$

y

$$\hat{\eta}_{op} = \beta_0 - r \sqrt{\sum_{i=1}^k \beta_i^2}$$

de aquí que se alcanza un máximo si $\lambda > 0$ y un mínimo si $\lambda < 0$.

Así, para un radio r dado, el óptimo de la función denotado por Δ esta dado por el vector:

$$\left(\frac{r\beta_1}{\sqrt{\sum_{i=1}^k \beta_i^2}}, \frac{r\beta_2}{\sqrt{\sum_{i=1}^k \beta_i^2}}, \dots, \frac{r\beta_k}{\sqrt{\sum_{i=1}^k \beta_i^2}} \right).$$

Los experimentos se llevan a cabo a lo largo de trayectoria Δ 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 existe una falta de ajuste del modelo de primer orden. Entonces debe aplicarse un procedimiento para obtener el óptimo de manera más precisa.

Las corridas exploratorias ejecutadas a lo largo del camino de máximo ascenso (indicadas en la Figura 1) indicarán por lo general un incremento en la respuesta a lo largo del camino. El mejor punto encontrado, o el punto máximo estimado por interpolación sobre el camino, será la base para un nuevo diseño de primer orden, dentro del cual otros avances podrán realizarse.

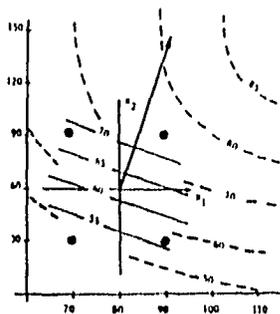


Figura 1 Camino de máximo ascenso

5.2.1 La longitud de Paso t

Suponga que la tendencia local de η , esta correctamente estimada en una región circundante al diseño mediante un polinomio de primer grado; suponga también que el vector Δ indica la dirección de máximo ascenso, entonces cerca del diseño se cumplirá que

$$\eta(it\Delta) < \eta((i+1)t\Delta).$$

Dado que las observaciones están sujetas a un error aleatorio, que la respuesta Y_i resulte ser mayor que Y_{i+1} , no necesariamente implicará que la tendencia de η ha cambiado de ascendente a descendente; este hecho podría deberse únicamente a la aleatoriedad del proceso. Es por esto, que tiene sentido pedir que la longitud del paso sea tal que permita detectar la tendencia de η con poca influencia del error. De este modo se desea encontrar la magnitud del paso tal que si la tendencia de η en dirección Δ es creciente, la probabilidad de que Y_i resulte ser mayor que Y_{i+1} no sea mayor que un cierto valor fijo, así tenemos que:

$$\begin{aligned} P(Y_i > Y_{i+1}) &= P[\eta(it\Delta) + \varepsilon_i > \eta((i+1)t\Delta) + \varepsilon_{i+1}] \\ &= P[\eta((i+1)t\Delta) - \eta(it\Delta) < \varepsilon_i - \varepsilon_{i+1}] \end{aligned} \quad (5.5)$$

con $\varepsilon_i \sim N(0, \sigma^2)$ independientes, se sigue que $\varepsilon_i - \varepsilon_{i+1} \sim N(0, 2\sigma^2)$; y que

$$P(Y_i > Y_{i+1}) = \alpha$$

cuando

$$\eta((i+1)t\Delta) - \eta(it\Delta) = z_\alpha \sqrt{2\sigma^2}. \quad (5.6)$$

Por ejemplo: si la probabilidad de detectar equivocadamente un cambio de ascendente a descendente es igual a 0.025, la constante z_α en (5.6) es 1.96, y si dicha probabilidad es igual a 0.0013 entonces $z_\alpha = 3$.

Dado que tanto σ como η son desconocidos, se utilizarán estadísticas para estimarlos, así tenemos que

$$\frac{SSE}{\sigma^2} = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2$$

y por lo tanto

$$\frac{(\varepsilon_i - \varepsilon_{i+1})}{\sqrt{2\sigma^2}} \sim T_{n-2}$$

de esto se sigue que por (5.6)

$$\eta((i+1)t\Delta) - \eta(it\Delta) = T_{n-2,\alpha} \sqrt{2\sigma^2}.$$

Con esta ecuación puede encontrarse un estimador para la longitud de paso t . al sustituir η por $\hat{\eta}$ y despejar t de la ecuación resultante, se llega a

$$\hat{t} = \frac{T_{n-2,\alpha} \sqrt{2\sigma^2}}{\beta_1 x_1 + \dots + \beta_k x_k}.$$

Otra manera de determinar el escalor t es considerar el criterio de O'Reilly[28]. El escalor t es el número real positivo que cumple con:

$$(1, t\Delta') (X'X)^{-1} \begin{pmatrix} 1 \\ t\Delta \end{pmatrix} = 1. \tag{5.7}$$

5.2.2 Localización del Centro Para un Nuevo Diseño

Para estimar el punto donde η alcanza el óptimo en dirección de Δ , se propone tomar k nuevas observaciones, Y_1, Y_2, \dots, Y_k en $t\Delta, 2t\Delta, 3t\Delta, \dots, kt\Delta$ y con ellas estimar la función

$$f(u) = \beta_0 + \beta_1 u + \beta_2 u^2$$

que define la tendencia cuadrática de η en dirección de Δ , donde la variable u determina el factor real que multiplica al vector $t\Delta$.

Si $\hat{\beta}_2 < 0$ entonces en el valor de

$$u_0 = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

es donde $f(u)$ alcanza el máximo; así el máximo estimado de η en dirección de Δ es el punto

$$c = u_0 t \Delta = -\frac{\hat{\beta}_1}{2\hat{\beta}_2} t \Delta.$$

Por consiguiente c resulta ser el centro del nuevo diseño.

Si $\hat{\beta}_2 > 0$ se toma un nuevo valor para el escalor t (igual a kt) y se estima nuevamente la tendencia cuadrática de η en dirección de Δ .

5.3 Algoritmo de Máxima Pendiente de Ascenso

El método de máxima pendiente de ascenso consta de los siguientes pasos:

1. Se elige un diseño para el ajuste de un polinomio de primer grado (por ejemplo un diseño factorial), y se obtienen las observaciones en los puntos del diseño.
2. Estimar con base en los datos la tendencia local de η mediante la función lineal $\hat{\eta}$.
3. Se hace una prueba de falta de ajuste para determinar si el plano se ajusta bien a los datos, o si existe ya una evidente curvatura. Si hay evidencia de curvatura se pasa al punto 7.
4. Se efectúa la prueba de hipótesis $H_0 : \beta_i = 0$, para toda $i = 0, 1, \dots, k$ contra $H_a : \beta_i \neq 0$, para alguna $i = 0, 1, \dots, k$; con el fin de determinar si existe evidencia de que los factores influyen en el resultado de la respuesta. Si no se rechaza H_0 implica que se tienen evidencias de que η se ajusta bien a un hiperplano horizontal y, por lo tanto, se está cerca de un punto crítico, en este caso se pasa al punto 7.
5. Encontrar el vector de dirección de máxima pendiente en $\hat{\eta}$; para ello se maximiza la función η sujeta a la restricción $\|x\| = 1$. Al vector encontrado se le denota por Δ .
6. Tomar en $t\Delta$, $2t\Delta$, $3t\Delta$, ..., $it\Delta$, ... nuevas observaciones de tal manera que si $Y_i < Y_{i+1}$, se toma la observación Y_{i+2} , en caso contrario se detiene el proceso porque se considera que en el punto $it\Delta$ la función de respuesta η en dirección de Δ cambia de ser una función creciente a ser decreciente, entonces con centro en el punto $it\Delta$, que se denota por $c = itv$, se elige un nuevo diseño. Con los datos de esta nueva muestra se reinicia el proceso desde el paso 1.
7. Completar el diseño y estimar una función cuadrática η . El punto donde esta función alcanza el óptimo se denota como \hat{x}_{op} y es el estimador del óptimo.

Si $\hat{\eta}$ no tiene un máximo, sino un punto silla, es conveniente hacer un análisis de cordillera posterior al análisis de ascenso por máxima pendiente.

Cuando $\hat{\eta}$ tiene más de un máximo relativo y el diseño inicial no está localizado en una región adecuada, el método de máxima pendiente de ascenso no garantiza que el punto óptimo obtenido sea un máximo global, este puede ser un máximo local, pero por experiencia se sabe que en la mayoría de los casos $\hat{\eta}$ presenta un único máximo.

Capítulo 6

DISEÑOS PARA AJUSTAR EL MODELO DE SEGUNDO ORDEN

Para estimar los parámetros de un polinomio de segundo grado es necesario contar con más observaciones, esto es debido a que dichos polinomios contienen un mayor número de parámetros. Por lo que es común usar, en este caso, los diseños utilizados para el ajuste de un polinomio de primer grado, agregándole convenientemente algunos puntos más para tener el número necesario de observaciones.

El modelo lineal general de segundo grado ($r=2$) es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \beta_{11} x_{1i}^2 + \beta_{12} x_{1i} x_{2i} + \cdots + \beta_{kk} x_{ki}^2 + \varepsilon_i, \quad (6.1)$$

el cual tiene: un parámetro independiente β_0 ; k parámetros en los términos de primer grado β_i ; k parámetros en los términos cuadráticos β_{ii} ; y por último C_2^k parámetros en los términos cruzados β_{ij} ($1 \leq i < j \leq k$). En total hay

$$1 + k + k + C_2^k = C_0^k + C_1^k + C_1^k + C_2^k = C_1^{k+1} + C_2^{k+1} = C_2^{k+2}$$

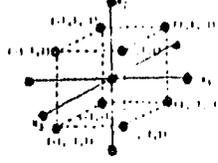
parámetros*. Entonces para estimar un polinomio de segundo grado se necesitan al menos C_2^{k+2} observaciones (por lo menos uno más si se desea estimar también σ^2).

6.1 Diseño Central Compuesto

Es el diseño más utilizado; está formado por los puntos de un diseño factorial 2^k , o de un factorial fraccionado 2^{k-p} , más los puntos de la forma $\pm \alpha e_i$, donde (e_1, e_2, \dots, e_k) es la base canónica de \mathfrak{R}^k y α es una constante de escalamiento. Este diseño tiene $2^{k-p} + 2k$ puntos.

Los puntos que se agregan al diseño factorial, permiten estimar los términos de segundo orden de la superficie. El aumento hecho a la matriz de un diseño factorial 2^k es

*La notación C_n^m se refiere a las combinaciones de m en n , con $m \geq n$, esto es $C_n^m = \binom{m}{n} = \frac{m!}{(m-n)!n!}$

Figura 1 Diseño Central Compuesto $2^3 + 7$

como sigue:

$$\begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ 0 & 0 & 0 & \cdots & 0 \\ -\alpha & 0 & 0 & \cdots & 0 \\ +\alpha & 0 & 0 & \cdots & 0 \\ 0 & -\alpha & 0 & \cdots & 0 \\ 0 & +\alpha & 0 & \cdots & 0 \\ 0 & 0 & -\alpha & \cdots & 0 \\ 0 & 0 & +\alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha \\ 0 & 0 & 0 & \cdots & +\alpha \end{pmatrix}$$

donde el primer renglón representa el punto central y los demás puntos auxiliares. Usualmente en el origen se hace más de una observación. La figura 1 es una representación de los 15 puntos de un diseño central compuesto con $k = 3$. Los primeros 8 puntos corresponden a un diseño factorial 2^3 , el punto 9 es el centro del diseño y los 6 restantes son puntos auxiliares.

$$\begin{pmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \\ 0 & 0 & 0 \\ -\alpha & 0 & 0 \\ +\alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ 0 & +\alpha & 0 \\ 0 & 0 & -\alpha \\ 0 & 0 & +\alpha \end{pmatrix}$$

El primer criterio para seleccionar α es que resulte un diseño ortogonal, esto es, que $X'X$ sea diagonal.

6.2 Optimización Sobre la Superficie de Respuesta

El problema fundamental es determinar bajo que niveles las variables independientes proporcionan una respuesta óptima, ésto puede lograrse con la ayuda de un modelo el cual se construye a partir de un diseño experimental, de tal modelo se estima un óptimo.

6.2.1 Determinación de un Punto Óptimo

Una vez que se ha localizado una región en la cual es factible encontrar un punto que sea óptimo, se deberá ajustar un modelo de grado $r=2$, ya que se considera que el comportamiento de los datos en una vecindad cerca de un óptimo es cuadrático. Así se tiene que para k variables independientes el modelo de segundo grado esta dado por:

$$Y = \eta + \varepsilon$$

donde

$$\eta = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j}^k \sum_{j=1}^k \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_j x_j^2 \quad (6.2)$$

A partir de un diseño experimental puede obtenerse por el método de mínimos cuadrados el estimador $\hat{\beta}$ de β . El objeto es maximizar

$$\hat{\eta} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i + \sum_{i < j}^k \sum_{j=1}^k \hat{\beta}_{ij} x_i x_j + \sum_{i=1}^k \hat{\beta}_j x_j^2$$

en términos de x_1, \dots, x_k . Donde el modelo $\hat{\eta}$ no presenta falta de ajuste. Esto quiere decir que η esta bien aproximada por $\hat{\eta}$. Ya que la función (6.2) es lineal en los parámetros, se aplica la teoría del modelo de regresión lineal para verificar la calidad del ajuste.

En términos matriciales el modelo (6.2) puede escribirse como:

$$\hat{\eta} = \hat{\beta}_0 + b'x + x'Bx \quad (6.3)$$

donde $\hat{\beta}_0$, es un real; b es un vector de dimensión k y B es una matriz de orden $k \times k$ negativa definida

$$b = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad B = \begin{pmatrix} \beta_{11} & \beta_{12}/2 & \beta_{13}/2 & \cdots & \beta_{1k}/2 \\ \beta_{21}/2 & \beta_{22} & \beta_{23}/2 & \cdots & \beta_{2k}/2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{k1}/2 & \beta_{k2}/2 & \beta_{k3}/2 & \cdots & \beta_{kk} \end{pmatrix}$$

Dado que η es derivable sobre todo \mathfrak{R}^k , los puntos óptimos se encontraran entre aquellos que resuelvan $\nabla \hat{\eta} = 0$. Los puntos que satisfacen tal condición se conocen como puntos críticos o estacionarios. El término estacionario se debe al hecho de que los valores que toma la función η no varían en vecindades pequeñas de x' . Los puntos críticos se dividen en máximos, mínimos y sillars.

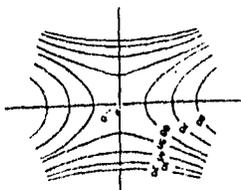


Figura 2 Punto silla sobre la superficie ajustada

Ya que la derivada es un operador lineal, se tiene que

$$\frac{\partial \hat{\eta}}{\partial x} = \frac{\partial \hat{\beta}_0}{\partial x} + \frac{\partial(b'x)}{\partial x} + \frac{\partial(x' Bx)}{\partial x}$$

entonces se tiene que

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

igualando a cero se obtienen los puntos estacionarios

$$b + 2Bx = 0$$

lo que implica que

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

es punto crítico de $\hat{\eta}$. Para saber si se trata de un máximo, un mínimo o un punto silla, se debe estudiar la matriz de segundas derivadas o matriz Hessiana. Si la matriz Hessiana es definida positiva se tiene un mínimo, si es negativa definida se tiene un máximo y si es cero se tiene un punto silla. Resolver el problema presenta las siguientes dificultades:

1. A medida que el número de variables crece el cálculo del Hessiano se hace cada vez más difícil.
2. No se puede analizar el comportamiento de la función η en las cercanías de x_{op} .

Así se discutirá una estrategia que permita caracterizar x_{op} de modo más sencillo. Esta es la llamada transformación canónica.

Si el punto estacionario x_{op} es un máximo, en la medida en que se aleje de éste, la respuesta decrecerá. Si x_{op} es un mínimo en la medida en que se aleje de éste la respuesta se incrementará. Sin embargo, en el caso de un punto silla el experimentador podría conseguir tanto un incremento como un decremento en la respuesta dependiendo en la dirección en la cual se aleje. Por lo cual, en este caso es de interés conocer en que dirección se deben mover las coordenadas para optimizar la respuesta, esto se ilustra mediante curvas de nivel en las figuras 2, 3, 4.

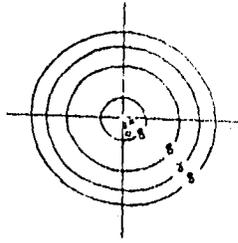


Figura 3 Máximo sobre la superficie ajustada

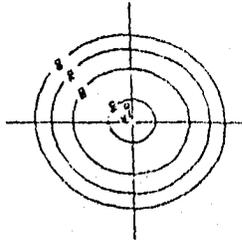


Figura 4 Mínimo sobre la superficie ajustada

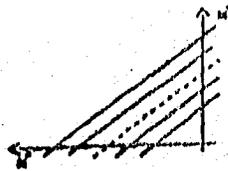


Figura 5 Sistema de cordillera estacionaria

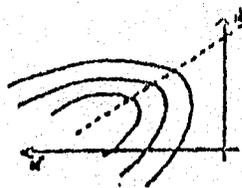


Figura 6 Sistema de cordillera ascendente

Otras situaciones que pueden ocurrir en la práctica son las presentadas en las figuras 5 y 6. La primera muestra lo que se conoce como sistema de cordillera estacionaria, mientras que la segunda muestra una cordillera ascendente. En figura 6 puede verse que hay una región más que un punto en el que la respuesta se maximiza (en este caso B^{-1} no existe). Esta región proporciona condiciones alternativas para la optimalidad. La cordillera ascendente indica la dirección en la cual el investigador debe mover la zona experimental para obtener un óptimo dentro de dicha región.

La detección de sistemas de cordillera es una parte importante del análisis de superficies de respuesta.

6.3 Transformación Canónica

Esta transformación consiste de una translación seguida de una rotación, lo que tiene por objeto simplificar la forma funcional de la superficie ajustada. Para ello es necesario conocer previamente un punto estacionario. La translación se realiza moviendo el origen del sistema al punto estacionario ubicado en $x_{op} = -\frac{B^{-1}b}{2}$, obteniéndose así la siguiente expresión

$$\hat{\eta} = \hat{\eta}_0 + z' B z$$

mediante el siguiente análisis:

De la ecuación (6.3) el valor estimado de la respuesta en el punto estacionario está dado por:

$$\hat{\eta} = \beta_0 + x'_{op} b + x'_{op} B x'_{op}$$

dado que

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

se tiene que

$$\begin{aligned} \eta_{op} &= \beta_0 + \left[-\frac{B^{-1}b}{2}\right]' b + \left[-\frac{B^{-1}b}{2}\right]' B \left[-\frac{B^{-1}b}{2}\right] \\ &= \beta_0 - \frac{b' B^{-1} b}{2} + \frac{b' B^{-1} b}{4} \\ &= \beta_0 - \frac{1}{4} b' B^{-1} b \end{aligned}$$

y nuevamente como usando x_{op} tenemos que

$$\eta_{op} = \beta_0 + \frac{x'_{op} b}{2}$$

Definiendo $z = x - x_{op}$, esto es $x = z + x_{op}$ (donde z son las variables correspondientes a los ejes translados), en (6.3) se tiene

$$\begin{aligned} \eta &= \beta_0 + (z' + x'_{op}) b + (z' + x'_{op}) B (z + x_{op}) \\ &= \beta_0 + x'_{op} b + x'_{op} B x_{op} + z' b + z' B x_{op} + x'_{op} B z + z' B z \\ &= \eta_{op} + z' b + z' B x_{op} + x'_{op} B z + z' B z \end{aligned}$$

dado que $z'Bx_{op}$ y $x'_{op}Bz$ son equivalentes, ya que uno es el transpuesto del otro. B es simétrica y el producto es un escalar, se tiene que

$$\begin{aligned}\eta &= \eta_{op} + z'b + 2z'Bx_{op} + z'Bz \\ &= \eta_{op} + z'(b + 2Bx_{op}) + z'Bz \\ &= \eta_{op} + z'\left[b - 2B\frac{B^{-1}b}{2}\right] + z'Bz \\ &= \eta_{op} + z'(b - b) + z'Bz \\ &= \eta_{op} + z'Bz.\end{aligned}$$

Al hacer la translación de ejes al punto x_{op} , se transforma el modelo original a uno equivalente que tiene una forma analítica más simple en la que no aparecen términos lineales.

El siguiente paso es realizar una rotación de ejes que elimine los productos cruzados de la forma cuadrática $z'Bz$. Esto se logra aplicando el teorema (2.10) a la matriz B

$$z'Bz = z'U\Lambda U'z = w'\Lambda w \quad \text{con } w = U'z, \quad (6.4)$$

donde U es una matriz ortogonal. Como Λ es una matriz diagonal de valores propios, entonces

$$w'\Lambda w = \sum_{i=1}^k \delta_i \omega_i^2,$$

y se tiene que

$$\eta = \eta_{op} + z'Bz = \eta_{op} + \sum_{i=1}^k \delta_i \omega_i^2. \quad (6.5)$$

De donde es más fácil saber si en realidad el punto estacionario x_{op} corresponde a un máximo, a un mínimo o a un punto silla. Además, las variables ω_i determinarán los ejes principales de las curvas de nivel de la función respuesta.

Ahora se analizará la ecuación (6.5) por casos para poder especificar la naturaleza del punto estacionario. Recuerde que el efecto del punto crítico se encuentra encerrado únicamente en la función $\hat{\eta}$, por lo cual sólo será necesario analizar el término $\sum_{i=1}^k \delta_i \omega_i^2$.

Caso I: Si en la ecuación (6.5) $\delta_i < 0$ para toda $i = 1, 2, \dots, k$ entonces x_{op} corresponde a un valor máximo de $\hat{\eta}$, pues los términos $\delta_i \omega_i^2 < 0$ para toda $i = 1, 2, \dots, k$ y $\sum_{i=1}^k \delta_i \omega_i^2 < 0$ lo que implica que

$$\eta_{op} + \sum_{i=1}^k \delta_i \omega_i^2 < \eta_{op}.$$

Es decir, el alejarse de x_{op} , sin importar la magnitud del desplazamiento provoca un decremento en el valor de la respuesta. Si $k = 2$ las curvas de nivel en términos de las variables ω , se verán como en la figura 7. Si $|\delta_2|$ es considerablemente mayor que $|\delta_1|$, en la dirección del eje ω_2 habrá un decremento más acelerado de la respuesta comparado con el descenso ocurrido al avanzar sobre ω_1 . Esto provoca que las curvas de nivel se alarguen en dirección de ω_2 .

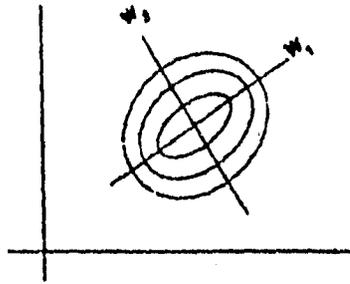
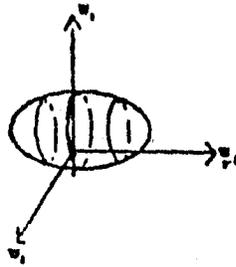
Figura 7 Curvas de nivel con ejes w rotados

Figura 8 Sistema estacionario

En general $|\delta_i|$ mide la magnitud de decremento de la respuesta η en la dirección de ω_i .

La expresión (6.5) presenta una manera muy sencilla para lograr un análisis de sensibilidad. Si existe al menos un subíndice $j \in \{1, 2, \dots, k\}$ tal que δ_j es aproximadamente cero y además $\delta_i < 0$ para toda $i = 1, 2, \dots, k$, entonces x_{op} corresponde a un máximo, pero para propósitos prácticos no es único, sino que habrá una infinidad de puntos que dan un valor máximo, todos ellos en el hiperplano definido por las ω_i tales que su correspondiente δ_j es casi cero. Cuando se presentan estos casos, se dice que hay una cordillera estacionaria.

Para $k = 2$, en el límite, esto es, cuando $\delta_1 = 0$ y $\delta_2 < 0$, la región de posible máximos tiende a una recta, ver figura 5. Cuando se está cerca de dicho límite al desplazarse sobre el eje ω_1 no hay cambios significativos en la respuesta η . Así, es posible encontrar valores de x_i cercanos al óptimo.

Para $k = 3$, en el caso $\delta_2 = \delta_3 = 0$, $\delta_1 < 0$, se tiene un plano, como se ve en la figura 8. Cuando $\delta_1, \delta_2 < 0$, y $\delta_3 = 0$ la región de posibles máximos tiende a estar sobre el eje ω_3 encerrado por las curvas de nivel de forma cilíndrica, como puede verse en la figura 9. La importancia de hacer $\delta_j = 0$ radica en la facilidad para entender los casos en que δ_j es muy cercana a cero. Para realizar la transformación canónica se requiere que B^{-1} exista, lo que ocurre si y sólo si todos los valores propios de B son diferentes de cero.

Caso II: Si en la ecuación (6.5), $\delta_i > 0$ para toda $i = 1, 2, \dots, k$ entonces x_{op} corresponde a un mínimo de η , ya que $\delta_i \omega_i^2 > 0$ para toda $i = 1, 2, \dots, k$ lo cual indica que

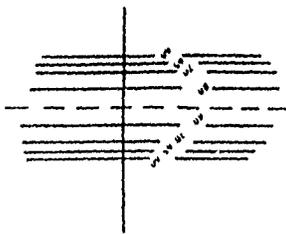


Figura 9. Región de respuesta acotada

$\sum_{i=1}^k \delta_i \omega_i^2 > 0$ y por lo tanto

$$\eta_{op} < \eta_{op} + \sum_{i=1}^k \delta_i \omega_i^2$$

con un análisis similar al caso I.

También se observa que si la ecuación (6.4) al menos un subíndice $j \in \{1, 2, \dots, k\}$ es aproximadamente cero y $\delta_i > 0$ para toda $i = 1, 2, \dots, k$, entonces se tiene una cordillera estacionaria de donde es posible escoger los puntos que más convengan.

Caso III: Si la ecuación (6.5) hay algunos valores propios positivos y otros negativos, entonces x_{op} es un punto silla. Esto quiere decir que a partir de x_{op} existen direcciones en las que al desplazarse la función crece, pero también existen otras direcciones en las cuales decrece.

Para obtener el punto que corresponde al máximo se debe avanzar sobre las direcciones en las que η crece. Estas direcciones están dadas por las variables ω_i , gracias a la ecuación (6.4).

Capítulo 7

ESTUDIO DE SIMULACIÓN

En este capítulo se presentan las características y los resultados de una serie de simulaciones, realizadas en el *sistema para el diseño de superficies de respuesta*, cuya finalidad es sondear la bondad de ajuste de los métodos propuestos anteriormente.

Es razonable pensar que la bondad de ajuste de los estimadores obtenidos con los diferentes métodos de búsqueda, no es igual en todos los casos. Para cada función de respuesta η , y para cada valor de σ^2 , debe existir un diseño, con el que se obtengan mejores estimadores.

Estudiar analíticamente el sesgo y la varianza de los estimadores resulta inaccesible por la naturaleza de sus distribuciones. Bajo estas circunstancias, se propone sondear la bondad de ajuste de los métodos propuestos utilizando simulación digital; y así, posteriormente, elegir el mejor estimador con base en los criterios relacionados con la bondad del procedimiento.

El estudio no incluye al diseño aleatorio, ya que su confiabilidad es pobre. Los métodos incluidos en el análisis son: el método de máxima pendiente de ascenso y el de diseño central compuesto.

Siguiendo la idea de Myers y Khuri [24], la cual consiste medir el acercamiento empírico de \hat{x}_{op} y $\hat{\eta}_{op}$ al punto óptimo real mediante simulación, se mide la bondad de ajuste de los estimadores en base a los siguientes criterios:

1. La distancia empírica promedio entre \hat{x}_{op} y x_{op} , la cual está dada por la expresión:

$$s_{\hat{x}_{op}} = \frac{1}{100} \sum \|\hat{x}_{op} - x_{op}\|$$

2. La distancia empírica promedio entre $\hat{\eta}_{op}$ y η_{op} , dada por la expresión:

$$s_{\hat{\eta}_{op}} = \frac{1}{100} \sum |\hat{\eta}_{op} - \eta_{op}|$$

3. El tamaño de muestra promedio requerido para converger; la muestra incluye todas las observaciones, desde las del primer diseño, hasta el último, incluyendo las obser-

vaciones que se tomaron en dirección de Δ , dado por:

$$\hat{n} = \frac{\sum n_i}{100}$$

4. El tiempo de proceso computacional.

7.1 Generación de Muestras

La simulación se realizó en base a la generación de muestras aleatorias tomadas de una función de segundo grado del tipo:

$$Y = -(x_{op} - x)'A'(x_{op} - x) + \varepsilon_i \quad \text{donde } \varepsilon_i \sim N(0, \sigma^2).$$

El cálculo de x_{op} y de η_{op} se realizó usando el método de búsqueda de máxima pendiente de ascenso.

Se considero el siguiente esquema de diseño: primero se construyo el factorial 2^2 para la estimación del plano, y después se completo el diseño central compuesto con $\alpha = 1$ para estimar la función cuadrática.

Para generar la muestra se consideró que la función η pertenece a la familia de funciones cuadráticas cuya gráfica tiene un máximo único y bien definido, esto es, que es de la forma:

$$\begin{aligned} \eta(x_1, \dots, x_k) &= -\sum (a_{i0} + \sum a_{ij}x_j)^2 \\ &= -(x_{op} - x)'A'(x_{op} - x) \end{aligned}$$

con $A \in M_{k \times k}$ por lo que η tiene un máximo único y bien definido, si y sólo si, A es invertible.

De este modo, se consideró que las observaciones son de la forma:

$$y = 50 + x_1 + 5x_2 - \frac{1}{2}x_1^2 - x_2^2 + x_1x_2 + \varepsilon$$

donde

$$\varepsilon \sim N(0, \sigma^2).$$

El óptimo de ésta función y el punto en el que se alcanza son:

$$\eta_{op} = 86.5 \quad \text{y} \quad x_{op} = (11, 8).$$

Las curvas de nivel son elipses concéntricas al óptimo, con una marcada diferencia entre la magnitud de sus ejes. Así entonces, la evidencia de curvatura se manifiesta de manera diferente desde las distintas direcciones por la que se asciende al óptimo, además los factores se relacionan a través de la respuesta.

Para la simulación se utilizó el escalar que cumple con la ecuación (5.7), y el número de observaciones en dirección del vector Δ , se tomó como $k = 5$.

El diseño utilizado fue un factorial 2^2 centrado en el origen, con cuatro repeticiones en el centro. El proceso de búsqueda en la dirección Δ , al igual que lo hicieron Myers y Khuri, se realizó 4 veces.

7.2 Resultados de la Simulación

En la tabla de resultados se reportan en cada renglón los valores de la observación y_i y las coordenadas del vector Δ , tanto codificadas como sin codificar.

y	x_1	x_2	ξ_1	ξ_2
62.62	0.657	1.253	1.315	2.506
64.35	0.767	1.462	1.534	2.924
64.67	0.876	1.671	1.753	3.342
67.29	0.986	1.879	1.973	3.759
66.16	1.096	2.088	2.192	4.177
76.27	1.404	-0.318	5.001	3.539
75.68	1.685	-0.382	5.562	3.411
76.41	1.966	-0.446	6.124	3.284
74.73	2.247	-0.510	6.686	3.156
72.86	2.528	-0.573	7.248	3.029
81.74	0.040	0.920	6.205	5.125
83.99	0.051	1.150	6.226	5.585
86.24	0.061	1.381	6.246	6.046
80.59	0.071	1.611	6.267	6.506
79.30	0.081	1.841	6.286	6.966
82.42	0.643	-0.282	8.285	6.381
83.51	0.858	-0.376	8.715	6.493
84.60	1.073	-0.470	9.144	6.504
85.69	1.287	-0.564	9.573	6.716
83.59	1.502	-0.658	10.00	6.928

El tamaño de muestra reportado incluye las ocho observaciones de cada diseño utilizado para encontrar el vector Δ , más las observaciones en dirección de Δ . El resultado final con 52 observaciones es:

$$\hat{\eta}(x_1, x_2) = 85.69 \quad x_1 = 9.573 \quad x_2 = 6.116$$

Con la finalidad de obtener resultados confiables, se propuso realizar la simulación

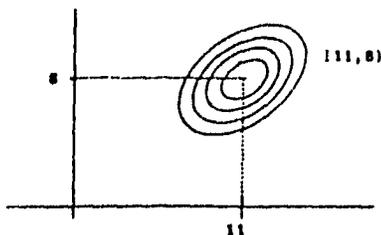


Figura 1 curvas de nivel de la función 7.1

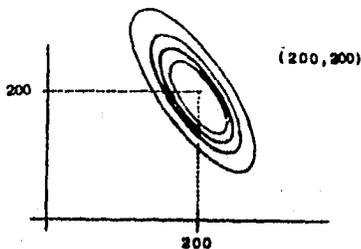


Figura 2 curvas de nivel de la función 7.2

anterior 100 veces, y con base en los resultados obtener la variación empírica promedio de x_{op} , el punto final que se alcanza con el acercamiento, y el valor de η_{op} .

Se consideraron dos funciones de respuesta:

$$y = 50 + 3x_1 + 5x_2 - \frac{1}{2}x_1^2 - x_2^2 + x_1x_2 + \varepsilon \quad (7.1)$$

en este caso el óptimo se encuentra en $x_{op} = (11, 8)$ y $\eta_{op} = 86.5$. Las curvas de nivel de la función (7.1) pueden verse en la figura 1.

La segunda función esta dada por:

$$y = -16000 + 800x_1 + 800x_2 - 1.008x_1^2 - 1.008x_2^2 - 1.983x_1x_2 + \varepsilon \quad (7.2)$$

esta función alcanza el óptimo en $x_{op} = (200, 200)$ y $\eta_{op} = 0.0$. Las curvas de nivel función (7.2) pueden verse en la figura 2.

En cada caso se consideraron dos valores para σ^2 :

$$\sigma^2 = 1 \quad \text{y} \quad \sigma^2 = 0.1.$$

El diseño utilizado fue un factorial 2^2 con cuatro repeticiones en el centro, el diseño fue previamente estandarizado.

Para las estimaciones se utilizó el escalador t que cumple con la ecuación (5.7) y el valor de $k = 5$.

El método propuesto consiste en estimar el vector Δ (con 8 observaciones en el diseño) y determinar el punto en el que se produce el cambio de dirección de Δ (5 observaciones en $t\Delta, 2t\Delta, 3t\Delta, 3t\Delta, 5t\Delta$), requiere de 13 elementos muestrales, lo que implica que con cuatro cálculos del vector Δ , la muestra siempre resultará ser de 52 elementos. Con la finalidad de variar el tamaño de muestra, se estimó el vector de dirección Δ , tantas veces como fue necesario, hasta que dos estimaciones sucesivas del punto en el cual se produce el cambio de dirección difirieran en menos de uno. se eligió uno por que las coordenadas de los puntos del diseño codificadas son 1 o -1.

Los resultados obtenidos son de la función 7.1 para $\sigma^2 = 1.0$ y 0.1 respectivamente son:

$$\begin{array}{ll} s_{\hat{\eta}_{op}} & 0.7533 \\ s_{\hat{x}_{op}} & 1.6989 \\ \hat{n} & 72 \end{array}$$

y

$$\begin{array}{ll} s_{\hat{\eta}_{op}} & 0.3242 \\ s_{\hat{x}_{op}} & 1.2754 \\ \hat{n} & 78 \end{array}$$

Los resultados obtenidos de la función 7.2 para $\sigma^2 = 1.0$ y 0.1 respectivamente son:

$$\begin{array}{ll} s_{\hat{\eta}_{op}} & 0.0618 \\ s_{\hat{x}_{op}} & 0.5202 \\ \hat{n} & 37 \end{array}$$

y

$$\begin{array}{ll} s_{\hat{\eta}_{op}} & 0.0003 \\ s_{\hat{x}_{op}} & 0.0121 \\ \hat{n} & 26 \end{array}$$

7.3 Conclusiones

Los resultados obtenidos de las simulaciones son muy claros al señalar que el método de máxima pendiente de ascenso proporciona *buenos* estimadores tanto de x_{op} como de η_{op} .

Los criterios utilizados para calificar la bondad de ajuste de los estimadores fueron:

1. La desviación media estimada, tanto de x_{op} como de η_{op} .
2. El tamaño de muestra requerido para obtener las estimaciones.
3. El tiempo de cómputo consumido.

Los resultados registrados son:

Con respecto a la exactitud, la desviación media empírica de \hat{x}_{op} es de orden menor a 2 unidades y aun menor en los casos en los que la varianza disminuye. En el caso de $\hat{\eta}_{op}$ la distancia es aun más pequeña, menor de una unidad, y menor que media unidad si la varianza decrece. Esto se observó aun en los casos en que el óptimo real se encuentra lejos del punto inicial de búsqueda; o bien, cuando las curvas de nivel de la función respuesta son elipses.

Los resultados de la simulación parecen indicar que con un diseño de simetría regular con respecto al origen, como el factorial 2^2 o el fraccionado 2^{2-p} , se obtiene mejores estimaciones cuando las curvas de nivel de $\hat{\eta}$ también son concéntricas (sobre todo circunferencias) al óptimo.

Debe considerarse que para llevar a cabo las simulaciones se consideró únicamente el caso en el que la matriz A es invertible; si ocurre que la matriz \hat{B} proporciona evidencia de ser singular se recomienda hacer un análisis anterior, por ejemplo: de componentes principales, para reducir el número de factores en la ecuación y tener una matriz no singular.

El tiempo de proceso computacional resultó ser directamente proporcional al tamaño de la muestra, así entre mayor sea n , mayor es el tiempo que se requirió para obtener el estimador.

7.4 Recomendaciones

Para finalizar se recomienda para trabajos futuros:

- Realizar mayor número de simulaciones, incrementando el número de factores y cambiando la forma funcional de η .
- Investigar otro tipo de diseños, los cuales mejoren las posibilidades de obtener mejores estimaciones del óptimo, estos pueden ser por ejemplo: diseños rotables.
- Aplicar el sistema desarrollado aquí, en casos reales, para constatar su eficiencia, y para determinar las limitaciones que puedan presentarse en problemas particulares.
- Investigar nuevas maneras de determinar la dirección Δ de máxima pendiente de ascenso.
- Generar una manera no subjetiva de determinar el rango del diseño.

BIBLIOGRAFÍA

1. Aho A. V, R. Sethi and J. D. Ullman
Compiladores (Principios, técnicas y herramientas)
Addison-Wesley Iberoamericana, 1990.
2. Anderson V. L, R. A. Mclean
Design of experiments: A realistic approach
Marcel Dekker, New York, 1974.
3. Box G. E. P. and N. R. Draper
A basis for the selection of response surface design
Journal of American Statistical Association, 54, 1959.
4. Box G. E. P. and N. R. Draper
Empirical Model-Building and response surface
John Wiley and Sons, 1987.
5. Box G. E. P, W. G. Hunter, and J. S. Hunter
Statistics for experiments
New York, John Wiley and Sons, 1978.
6. Box G. E. P. and K. B. Wilson
On the experimental attainment of optimum condition
Journal of the royal statistical society, ser B. 13, 1951.
7. Carter Jr. W. H, V. M. Chinchilli, R. H. Myers and E. D. Campbell
Confidence interval and an improved ridge analysis of response surfaces
Technometrics, 28, No 4, 339-346, 1986.
8. Cochran W. G. and G. M. Cox
Diseños Experimentales
Trillas México, 1980.
9. Dennis J. E. and R. B. Schnabel
Numerical methods for Unconstrained Optimization and nonlinear equations
Prentice-Hall, 1983.
10. Draper N. R.
Ridge analysis of response surfaces
Technometrics, 5, No 4, 469-479, 1963.
11. Draper N. R. and H. Smith
Applied regression analysis
John Wiley, Second Edition, 1981.
12. Fletcher R.
Practical methods of optimization
John Wiley and Sons, Second Edition, 1988.

13. Friedberg S. H, A. J. Insel and L. E. Spence
Algebra lineal
Publicaciones cultural S. A. 1982.
14. Gill P. E, W. Murray and M. Wright
Practical optimization
Academic Press, London 1981.
15. Graybill A. F.
Theory and application of the linear model
Duxbury Press, 1976.
16. Hoerl, A. E. and R. W. Kennard
Ridge regression: Biased estimation for nonorthogonal problems
Technometrics, 1970.
17. Khuri A. I. and R. H. Myers
Modified ridge analysis
Technometrics, 4, 469-579, 1979.
18. Khuri A. I. and J. A. Cornell
Response Surfaces: Designs and analyses
Marcel Dekker, Inc, 1987.
19. Mardia K. V, J. T. Kent and J. M. Bibby
Multivariate analysis
Academic Press, 1979.
20. Marsden J. E. y A. J. Tromba
Cálculo vectorial
Fondo educativo interamericano, 1981.
21. Méndez I.
Metodología de superficies de respuesta
Comunicaciones Técnicas, I.I.M.A.S, U.N.A.M, 1977.
22. Montgomery D. C. and E. A. Peck
Introduction to linear regression analysis
Jhon Wiley and Sons, Second Edition, 1992.
23. Mood, A. M.; F. A. Graybill, and D. Boes
Introduction to the theory of statistics
Mc graw-Hill, 1974.
24. Myers R. H, A. I. Khuri, and W. H. Carter
Response surface methodology:1966-1988
Technometrics, 31, No 2, 137-157, 1989.
25. Noble B.
Applied linear algebra
Prentice Hall, Inc, 1969.
26. O'reilly F. J.
On a criterion for extrapolation in normal regression
Ann. Statistics, 3, No 1. 219-222, 1975.
27. Pérez S. B.
Regiones de predictibilidad en superficies de respuesta
Tesis doctoral de matemáticas, Facultad de ciencias U.N.A.M, 1991.
28. Pérez S. B. y F. J. O'reilly
Sobre el ascenso por pendiente máxima en superficies de respuesta

- Reportes de investigación, I.I.M.A.S, U.N.A.M, 3, No 26, 1993.
29. Press W. H, S. A. Teukolsky, W. T. Vetterling and B. P Flannery
Numerical recipes in C: The art of scientific computing
Cambridge university press, 1992.
 30. Scarle S. R.
Linear models
John Wiley and Sons, 1971.
 31. Seber G. A.
Multivariate observations
John Wiley and Sons, 1984.