

Centro de Investigación en Matemáticas, A.C.

CIMAT

“MODELACIÓN CONJUNTA DE LA MEDIA
Y LA VARIANZA CON UN ENFOQUE NO
PARAMÉTRICO”

T E S I S

Que para obtener el grado de

Maestro en Ciencias

con Especialidad en

Probabilidad y Estadística

P r e s e n t a

Sara Silva Hernández

Director de Tesis:

Dr. Jorge Domínguez Domínguez

Guanajuato, Gto.. Diciembre de 2012

Índice general

1. Panorama general de diseño robusto de respuesta dual	1
1.1. Introducción	1
1.2. Metodología de superficie de respuesta	1
1.2.1. Diseños de primer orden	3
1.2.2. Diseños de segundo orden	4
1.2.3. Algunas propiedades comunes de los diseños	4
1.3. Diseño Robusto	5
1.3.1. Surgimiento y filosofía	5
1.3.2. Formulación del problema	6
1.4. Esquemas de optimización en la respuesta dual	8
1.4.1. Procedimiento Vining-Myers	9
1.4.2. Procedimiento deseabilidad	9
1.4.3. Procedimiento Lin-Tu	10
1.4.4. Copeland-Nelson	10
1.4.5. Vining-Bohn	11
1.4.6. Domínguez-Rocha	11
1.4.7. Procedimiento razón señal a ruido	11
1.4.8. El futuro de la MSR	12
1.4.9. Relación con la tesis	12
2. Elementos no paramétricos	13
2.1. Introducción	13
2.2. Definición	14
2.2.1. Ejemplos	14
2.3. Regresión kernel	14
2.3.1. Estimador Nadaraya Watson	15
2.3.2. Estimador Gasser Müller	15
2.3.3. Precisión	16
2.4. Estimación de densidades por kernel	16
2.4.1. Precisión	17

2.5. Caso multidimensional	17
2.6. Kernel óptimo	18
2.7. Elección del ancho de banda	18
2.7.1. Caso especial	19
2.8. Ejemplo	19
2.9. Discusión	20
2.10. Relación con la tesis	21
3. Resultados	22
3.1. Ejemplo de estudio	22
3.2. Análisis exploratorio	24
3.2.1. Dispersión de Desviación Estándar	25
3.2.2. Dispersión de Media	26
3.3. Enfoque paramétrico	27
3.3.1. Primera estrategia	27
3.3.2. Optimización	33
3.3.3. Otra estrategia de optimización	40
3.3.4. WLS	42
3.3.5. Método Gráfico	44
3.4. Enfoque con elementos no paramétricos	46
3.4.1. Enfoque semiparamétrico	47
3.4.2. Enfoque completamente no paramétrico	49
4. Conclusiones	50

Agradecimientos

Mis más sincero agradecimiento a CIMAT y CONACyT por el apoyo brindado durante la maestría y la elaboración de este trabajo.

Al Dr. Jorge Domínguez por ser el mejor asesor que hubiera podido tener y a mis sinodales: Dra. Angélica Hernández y Dr. Rogelio Ramos, por sus atinados comentarios y sugerencias.

A todo el personal de CIMAT y UAQ que colaboró de alguna manera en este trabajo.

No siempre se está listo para el futuro. Hubo un tiempo para sentirse anclado y perderse en círculos viciosos.

Por ello,

a todos los que estuvieron al inicio,

a los que permanecieron durante la rendición,

a los que hicieron posible una reconstrucción y un final.

Yo sólo puedo agradecerles.

Si trato de ser específica comenzaría a enumerar, padres, hermanos, Doctores, amigos de antes y después.

Villeda, Pepe, Job, gracias por estar ahí.

Abelardo, Aida, Ahtziri, Boryana, César, Ehyter, Diego, Karina, Lily, Lina, Manuel, Mario y los que seguramente estoy olvidando, gracias por compartir el drama.

Arceo, Félix, Plancarte, gracias por ayudarme a volver.

Los no-presenciales, Retro, Roberto, gracias por compartir sus lejanas vidas.

Está hecho.

Resumen

El peor enemigo de la llamada calidad total es la variabilidad, entendiendo como tal los inevitables, y en ocasiones casi imperceptibles cambios que modifican el proceso afectando la producción. Tales cambios pueden ser en tiempos, medidas o cualquier otra característica de interés.

En la optimización estadística de un problema industrial, ingenieros y estadísticos buscan acercarse al valor objetivo de una característica de calidad, en ocasiones un mínimo o máximo, al tiempo que se desea reducir la variación en los resultados o productos finales del proceso.

En el presente trabajo se plantea una serie de escenarios en el marco de la metodología de superficie de respuesta para modelar la media y la varianza, considerando el procedimiento clásico de Mínimos Cuadrados Ordinarios (OLS por sus siglas en inglés).

Los modelos de regresión para media y varianza se obtienen en el contexto del diseño robusto con el enfoque de respuesta dual, utilizando también las señales de razón a ruido correspondientes, tanto en la teoría como en los ejemplos a desarrollar.

Pero debido a que en muchos casos el método OLS no produce el ajuste deseable cuando se busca modelar la varianza, se propone como alternativa natural el uso de regresión no paramétrica para estimar tal variabilidad. Basándonos en la estimación kernel tendremos un enfoque semiparamétrico usando Mínimos Cuadrados Ponderados (o WLS por sus siglas en inglés) y uno completamente no paramétrico. Al optimizar cada uno de ellos se obtendrán el correspondiente conjunto de entradas óptimas.

Ante tal situación y como contribución de este trabajo, se desarrollan a detalle tales enfoques con su correspondiente determinación de pesos y estimadores, para ilustrar más adelante su aplicación al problema de respuesta dual.

Posteriormente se realizan las comparaciones pertinentes entre los diferentes enfoques a fin de evaluar la calidad de los valores obtenidos. Finalmente, se verán las ventajas que ofrece esta propuesta con herramientas no paramétricas.

Introducción

En un proceso industrial invariablemente se tendrán limitantes económicas, motivando el uso de un diseño experimental inteligente que permita aprovechar al máximo los recursos disponibles. Siendo algunos factores involucrados en el experimento más costosos o difíciles de controlar que otros. La metodología de Taguchi nos permite trabajar con ellos de manera óptima.

Una situación usual se presenta al acercar una característica de interés a su valor objetivo, tratando de que la variabilidad en torno a tal valor sea la menor posible. Este trabajo modela dicha situación para un ejemplo clásico de la literatura, presentado por Box y Draper y analizado posteriormente por diversos autores. Consiste en un proceso donde se busca la mayor calidad de la impresión posible, lo anterior llevando el resultado a una característica de calidad ya establecida y se tienen tres factores de control: velocidad, presión y distancia.

Para el análisis, se toma como punto de partida el trabajo de Vining y Myers (1990), con un análisis completamente paramétrico y usando modelos completos, para luego recurrir a una optimización por multiplicadores de Lagrange. Se incluirán también los modelos incompletos descartados en su trabajo que contienen sólo los términos significativos y usando otros métodos de optimización, como el error cuadrático medio o la función de deseabilidad, se obtendrán los valores que cumplen mejor con los objetivos del experimento. Posteriormente, se sigue el trabajo de Vining y Bohn (1998) con el ajuste de un kernel y posterior uso de Mínimos Cuadrados Ponderados (WLS por sus siglas en inglés), aunque usando una función kernel diferente a la presentada en su artículo.

Se presenta así el enfoque paramétrico con diferentes opciones de optimización y también un procedimiento incluyendo aspectos no paramétricos dado que la varianza es un sistema ruidoso por excelencia, logrando ventaja sobre el proceder tradicional, lo que en la práctica puede representar ahorro o mayor calidad.

Finalmente será de interés realizar una comparación de las metodologías.

Objetivos

General

Este trabajo tiene como objetivo general desarrollar un método para modelar la media y varianza en un contexto no paramétrico en contraste con la aproximación usual por regresión paramétrica y optimización por multiplicadores de Lagrange, que si bien provee de un enfoque muy fácil de calcular, ya explorado

y sencillo de interpretar deja con la necesidad de una metodología que rescate dicho aspecto positivo al tiempo que aminore sus desventajas.

Particulares

- Explorar las opciones de optimización en un caso paramétrico.
- Comparar las entradas óptimas para cada una de dichas opciones.
- Comparar la metodología no paramétrica con los métodos paramétricos.
- Evaluar la eficiencia del método no paramétrico.
- Realizar una revisión actualizada del avance del diseño robusto en el contexto de la respuesta dual.

Capítulo 1

Panorama general de diseño robusto de respuesta dual

En este apartado se hará una breve exposición de los fundamentos de la metodología de superficie de respuesta y el diseño robusto. Estas ideas nos permitirán plantear el caso paramétrico, así como la comprensión de los métodos posteriores.

1.1. Introducción

Al buscar las condiciones óptimas de un proceso industrial en pos de la calidad se tendrán dos respuestas a optimizar: una correspondiente a la característica de calidad que deseo llevar al valor objetivo y, adicionalmente, la inevitable variabilidad de dichos resultados, la cual se buscará minimizar. Es decir, si el proceso de interés consiste en la fabricación de una pieza de tamaño especificado, el tener como resultado una de menor o mayor tamaño puede hacerla inútil, y aunque es casi imposible lograr el tamaño exacto se tratará de que el error sea el menor posible.

Cada una de ellas influenciada por diferentes tipos de factores, desde los que podemos controlar fácilmente (factores de control) hasta aquellos cuyo cambio implique costos o tiempos excesivos (factores de ruido).

En los 80's Genichi Taguchi desarrolló una metodología para alcanzar mayor calidad reduciendo presupuestos, al centro de dicha metodología se encuentra aquel diseño no afectado severamente por los inevitables cambios de los factores de ruido, conocido como *diseño robusto*.

1.2. Metodología de superficie de respuesta

Se tiene una respuesta de interés y , y un conjunto de variables de control denotadas por x_1, \dots, x_k . Para desarrollar una adecuada relación funcional entre ellas son usadas las técnicas de la metodología de superficie de respuesta (RSM por sus siglas en inglés).

Aunque tal relación sea desconocida se considera que puede ser aproximada por un modelo polinomial de la forma

$$y = f^t(x) \beta + \varepsilon \quad (1.1)$$

donde $x = (x_1, \dots, x_k)^t$, $f(x)$ es un vector de funciones de p elementos que consiste en potencias y productos cruzados de potencias de x_1, \dots, x_k hasta el grado $d (\geq 1)$. β es un vector de p parámetros, ε es un error experimental, que supondremos con media cero. De obtener con este modelo una representación adecuada de la respuesta el valor esperado de la respuesta $\mu(x)$ estará dado por $f^t(x) \beta$.

Como indica Khuri (2010), los propósitos de considerar dicho modelo serían:

- Predicción: establecer una relación, aunque aproximada entre y y x_1, \dots, x_k que pueda ser usada para predecir valores de la respuesta ante valores dados de las variables de control.
- Inferencia: determinar mediante pruebas de hipótesis, la significancia de los factores cuyos valores están determinados por x_1, \dots, x_k .
- Optimización: determinar los valores óptimos de x_1, \dots, x_k que resultan en la máxima (o mínima) respuesta sobre una región de interés.

Dos casos especiales del modelo son el modelo de primer grado ($d = 1$),

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon \quad (1.2)$$

y el modelo de segundo grado ($d = 2$)

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + \varepsilon. \quad (1.3)$$

Ya en el problema específico, se procede a realizar una serie de n experimentos donde se mide la respuesta y para entradas especificadas de las variables de control. La totalidad de estas entradas constituye lo que referiremos como *diseño*, representado por la matriz de diseño $D_{n \times k}$.

$$D = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{nk} \end{pmatrix} \quad (1.4)$$

donde cada renglón de la matriz D representa un punto en un espacio euclidiano k -dimensional correspondiente a cada prueba experimental y x_{ui} denota la u -ésima entrada del diseño de x_i .

Sea y_u el valor de la respuesta obtenido al aplicar la u -ésima entrada de x . De 1.1 se tiene

$$y_u = f^t(x_u) \beta + \varepsilon_u, \quad u = 1, 2, \dots, n \quad (1.5)$$

considerando el vector $y = (y_1, \dots, y_n)^t$, $X_{n \times p}$ una matriz cuyo u -ésimo renglón es $f^t(x_u)$, y $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^t$ entonces 1.5 sería escrita en forma matricial

$$y = X\beta + \varepsilon. \quad (1.6)$$

Bajo el supuesto de que el error experimental ε tiene distribución normal con media cero y una matriz de varianzas-covarianzas dada por $\sigma^2 I_n$, el estimador de mínimos cuadrados para el vector de parámetros β es

$$\hat{\beta} = (X^t X)^{-1} X^t y. \quad (1.7)$$

Y la matriz de varianzas-covarianzas de $\hat{\beta}$ de la forma

$$\text{Var}(\hat{\beta}) = (X^t X)^{-1} X^t (\sigma^2 I_n) X (X^t X)^{-1} = \sigma^2 (X^t X)^{-1}. \quad (1.8)$$

Usando un estimado $\hat{\beta}$ para estimar la respuesta esperada en x_u se tendrá entonces

$$\hat{\mu}(x_u) = f^t(x_u) \hat{\beta}, \quad u = 1, 2, \dots, n.$$

En cualquier punto x , en una región experimental R la respuesta predicha $\hat{y}(x)$ será

$$\hat{y}(x) = f^t(x) \hat{\beta} \quad (1.9)$$

Sabemos que $\hat{\beta}$ es un estimador insesgado de β , por lo cual $\hat{y}(x)$ es un estimador insesgado de $f^t(x) \beta$ y por 1.8, la varianza de $\hat{y}(x)$ será

$$\text{Var}[\hat{y}(x)] = \sigma^2 f^t(x) (X^t X)^{-1} f(x). \quad (1.10)$$

Es importante notar que la varianza de la predicción determina la calidad de la misma, tratando entonces de hacer dicha variabilidad lo más pequeña posible mientras que el modelo siga representando un ajuste aceptable. Esta cantidad depende de la matriz D por lo que la elección de dicho diseño es crucial para el análisis.

1.2.1. Diseños de primer orden

Se describe en esta sección el diseño para ajustar modelos de primer orden ($d = 1$) más popular y de fácil interpretación presentado en Khuri (2010). En Montgomery (2001) se puede consultar a profundidad gran variedad de tales diseños.

Diseño factorial 2^k

En este diseño, cada variable de control es medida a dos niveles. El diseño consiste en todas las posibles combinaciones de dichos niveles de los k factores. Entonces, cada renglón de la matriz de diseño D consiste de una combinación de 1 y -1 y representa una combinación de tratamientos particular. En este caso, el número n de corridas experimentales es igual a 2^k si ningún punto en el diseño está replicado.

Si k es grande, el diseño 2^k requiere un gran número de puntos de diseño. Ya que el número de parámetros desconocidos es tan sólo $k + 1$, podemos considerar fracciones de 2^k para ajustar dicho modelo.

Por ejemplo, se puede considerar medio diseño que consiste en la mitad de puntos del diseño de 2^k o un cuarto de diseño correspondiente a la cuarta parte de 2^k .

En general, una fracción 2^{-m} -ésima de un diseño 2^k consiste en 2^{k-m} puntos de un diseño completo 2^k (con m tal que $2^{k-m} \geq k + 1$ de modo que puedan estimarse todos los parámetros en el modelo).

1.2.2. Diseños de segundo orden

El número de parámetros en el modelo de segundo orden en 1.3 es $p = 1 + 2k + \frac{1}{2}k(k-1)$, por lo que el número de puntos en los diseños de segundo orden debe ser al menos p . La codificación usual (que se rescatará en el capítulo 3) es $\frac{1}{n} \sum_{u=1}^n x_{ui} = 0$ y $\frac{1}{n} \sum_{u=1}^n x_{ui}^2 = 1$, donde n es el número de corridas experimentales.

Diseño factorial 3^k

Consiste en todas las combinaciones de niveles de las k variables de control, teniendo 3 niveles cada una. Si los niveles están igualmente espaciados, pueden ser codificados como $-1, 0, 1$ (como se hará en el ejemplo por trabajar).

El número de corridas experimentales es 3^k que crece rápidamente con k , y como en el diseño 2^k podemos considerar fracciones para reducir costos, el procedimiento para hacerlo está descrito a detalle en el capítulo 9 de Montgomery (2001).

1.2.3. Algunas propiedades comunes de los diseños

Como ya se mencionó, la elección del diseño es muy importante en cualquier investigación de superficie de respuesta, ya que la calidad de predicción medida por la varianza de la misma depende de D . Mediante esta elección podemos tener propiedades que nos otorgarán ventajas respecto a varianzas, correlaciones, etc. Entre las propiedades más deseables se consideran las siguientes:

Ortogonalidad

Si la matriz $X'X$ es diagonal. Lo que proporciona la ventaja de que los elementos de $\hat{\beta}$ estarán no correlacionados pues aquellos fuera de la diagonal de $Var(\hat{\beta})$ serán 0. Esto facilita las pruebas de significancia de los parámetros desconocidos en el modelo.

Todos los diseños de primer orden presentados comparten la propiedad de ortogonalidad. Para un diseño de primer orden, la ortogonalidad se alcanza si la matriz de diseño D cumple que $D'D$ es diagonal.

Puede demostrarse además que en un diseño ortogonal, las varianzas de los elementos de $\hat{\beta}$ tienen valores mínimos dando máxima precisión para estudiar los parámetros desconocidos y dichos elementos serán no correlacionados.

Rotabilidad

Como se menciona en Montgomery (2001), es importante para un modelo proveer de buenas predicciones en la región de interés, una manera de verlo es que el modelo tenga una varianza razonablemente estable para la respuesta a predecir en los puntos x .

Box y Hunter (1995) sugerían que el diseño de superficie de respuesta debía ser rotatable, es decir, que $Var[\hat{y}(x)]$ es la misma para todos los puntos x a la misma distancia del centro del diseño. Esto implica que la varianza de las predicciones es constante en una hiperesfera centrada en el origen (luego de una estandarización).

La ventaja de esta propiedad es que la varianza de la predicción permanece sin cambios bajo cualquier rotación de ejes.

Precisión uniforme

Noción introducida por Box y Hunter (1957), un diseño rotable tiene la propiedad adicional de precisión uniforme si $Var[\hat{y}(x)]$ en el origen es igual a su valor a una distancia de 1 del origen. Lo cual ayuda proporcionando estabilidad en la varianza de la predicción en la vecindad del centro del diseño.

Optimalidad

Diseños óptimos son aquellos construidos con base en algún criterio de cercanía de la predicción $\hat{y}(x)$ a la media de la respuesta $\mu(x)$ dentro de una región R .

A los criterios que buscan minimizar la varianza de los parámetros en el modelo 1.1 se les conoce como relativos a la varianza. El más importante es el criterio de D-optimalidad, que maximiza el determinante de la matriz X^tX , lo que conlleva la minimización de la región de confianza para el vector β en el modelo.

1.3. Diseño Robusto

Ante la creciente necesidad industrial de la mejora continua, en un proceso industrial tanto la media como la variabilidad del producto final son de interés y dependen de los diversos factores involucrados. Ajustar modelos de regresión para ambas características ha sido un tema de investigación estadística de la mitad de los años 80's a la fecha.

En esta sección se revisarán diferentes esquemas experimentales con la finalidad de seleccionar la combinación de los niveles en los factores que optimicen un proceso, con la menor varianza posible y por consiguiente, un aumento en la calidad.

1.3.1. Surgimiento y filosofía

En la manufactura, la calidad es esencial para la supervivencia de un proceso. Suponiendo que podemos obtener la respuesta apropiada para identificar la calidad, el objetivo del investigador es encontrar las condiciones operativas en que la respuesta sea maximizada, minimizada o alcanza un valor objetivo.

En los 80's Genichi Taguchi, popularizó la noción de calidad como una función de la variabilidad del proceso, pues frecuentemente la mayor parte de la variación asociada a la respuesta puede ser atribuida a la presencia de factores de ruido. Para estudiar la influencia de estas variables en el proceso, Taguchi impulsó y formalizó su uso en la etapa de diseño experimental como una metodología creada como un enfoque efectivo en costos para mejorar la calidad de productos y procesos haciéndolos robustos, es decir, poco sensibles a imperfecciones de manufactura, deterioro por aspectos ambientales, entre otros.

Partiendo del supuesto de que los factores en un sistema pueden ser divididos en factores de control (denotados por X) y factores de ruido (denotados por Z).

Los factores de control son aquellos cuyos niveles permanecen sin cambio en el proceso una vez que son seleccionados, mientras que los niveles de los factores de ruido son difíciles de controlar o a un costo alto y se supone que varían aleatoriamente dentro del proceso. Su falta de estabilidad puede transmitir variación no deseada a la respuesta. La frase diseño robusto de parámetros (RPD por sus siglas en inglés) se refiere al diseño de un proceso (no un diseño experimental) de modo tal que podamos operar en niveles de los factores de control que hagan el proceso robusto o insensible a los inevitables cambios en los factores de ruido.

Taguchi no sólo llamó la atención con la filosofía de RPD, sino que propuso el diseño experimental y las estrategias de análisis para identificar las entradas de X que alcancen el comportamiento robusto. Los diseños experimentales propuestos por Taguchi son conocidos como arreglos ortogonales. Los arreglos ortogonales cruzan un diseño ortogonal en X (conocido como arreglo interno) que describe los factores de control, con un diseño ortogonal en Z (conocido como arreglo externo) correspondiente a los factores de ruido.

Un arreglo experimental es una matriz donde en las columnas usualmente se colocan los factores de diseño. En los renglones se escribe la combinación de valores para cada factor y esto constituye un tratamiento. La región experimental se define por el área acotada por los valores extremos de los factores de control que se estudian.

Ya que se tienen dos respuestas de interés se tratará de un sistema de respuesta dual e involucrará optimizar para ambas. El modelo sobre la media permitirá obtener un mejor resultado sobre la característica de calidad de un producto, con valor objetivo T . El modelo sobre la varianza tratará de reducir la variabilidad en torno a tal valor. Para ello se utilizarán las técnicas de MSR.

El resultado óptimo será aquella combinación de los niveles en los factores que mayor calidad brinde al proceso.

1.3.2. Formulación del problema

Primera estrategia experimental

La estrategia experimental se plantea como un doble arreglo ortogonal, en el denominado diseño interno se colocan las combinaciones de los niveles de los k factores de control, denotados por X_1, \dots, X_k , y el denominado arreglo externo la combinación de los q factores de ruido, denotados por Z_1, \dots, Z_q .

Tomando los y_{ij} como las respuestas de la combinación de los niveles de los factores de control ($i = 1, \dots, n$) y los factores de ruido ($j = 1, \dots, q$) es posible visualizar el diseño como si se tratara de

un producto cartesiano.

Tabla 1.1

Primera estrategia experimental

			Z_1	z_{11}	...	z_{r1}			
			\vdots	\vdots		\vdots			
			Z_q	z_{1q}	...	z_{rq}			
X_1	...	X_k					Media	S	SR
x_{11}	...	x_{1k}		y_{11}	...	y_{r1}	\bar{y}_1	S_1	SR_1
\vdots		\vdots		\vdots		\vdots	\vdots	\vdots	\vdots
x_{n1}	...	x_{nk}		y_{n1}	...	y_{nr}	\bar{y}_n	S_n	SR_n

Es posible calcular para cada tratamiento la media \bar{y}_i y la desviación estándar S_i . Denotando estas respuestas de interés por Y_1 y Y_2 respectivamente, se modelarán a partir de los resultados experimentales, con los modelos cuadráticos completos.

$$Y_1 = \bar{y} = \beta_0 + X^t \beta + X^t B X + \varepsilon_1 \quad (1.11)$$

$$Y_2 = S = \gamma_0 + X^t \gamma + X^t D X + \varepsilon_2 \quad (1.12)$$

Donde $X^t = (X_1, \dots, X_k)$ son k factores, β_0 constante, $\beta^t = (\beta_1, \dots, \beta_k)$ un vector de parámetros lineales, $B = (\beta_{11}, \dots, \beta_{1k}, \beta_{k1}, \dots, \beta_{kk})$ la matriz de parámetros de segundo orden, y $\varepsilon_1 \sim N(0, \sigma_1^2)$. Similarmente para el modelo correspondiente a la variabilidad, γ_0 constante, $\gamma^t = (\gamma_1, \dots, \gamma_k)$ un vector de parámetros lineales, $D = (\gamma_{11}, \dots, \gamma_{1k}, \gamma_{k1}, \dots, \gamma_{kk})$ la matriz de parámetros de segundo orden, y $\varepsilon_2 \sim N(0, \sigma_2^2)$.

Al usar OLS se obtendrán los modelos ajustados, denotados por \hat{Y}_1 y \hat{Y}_2 .

El objetivo de la metodología de Taguchi es obtener la mejor combinación de los niveles en los factores de control en presencia de los factores de ruido, es decir, aquella que minimiza el ECM de la característica de interés, dando así el mejor valor de ambas respuestas. Considera tres situaciones básicas dentro de una región experimental R .

- *The target is best*, minimizar el ECM alrededor de un valor objetivo específico T .

$$\text{mín } \hat{Y}_2 \text{ s.a. } \hat{Y}_1 = T \quad (1.13)$$

- *The larger, the better*, donde se busca maximizar la característica de interés.

$$\text{máx } \hat{Y}_1 \text{ s.a. } \hat{Y}_2 = \sigma_0^2 \quad (1.14)$$

- *The smaller, the better*, donde se busca minimizar dicha característica.

$$\text{mín } \hat{Y}_1 \text{ s.a. } \hat{Y}_2 = \sigma_0^2 \quad (1.15)$$

Segunda estrategia experimental

Una alternativa al doble arreglo ortogonal consiste en combinar los factores de control y de ruido reduciendo así de manera importante el número de tratamientos (en el doble arreglo eran necesarios nq) representando una ventaja en costos. Tal como se presenta a continuación

Tabla 1.2

Segunda estrategia experimental

A	...	K	P	...	Q	
x_{11}	...	x_{1k}	z_{11}	...	z_{1q}	y_{11}
.
x_{n1}	...	x_{nk}	z_{n1}	...	z_{nq}	y_{n1}

Los resultados experimentales obtenidos con esta estrategia serán modelados en función tanto de los factores de control como de ruido.

$$Y_{xz} = \gamma_0 + x^t \beta + x^t B x + z^t \delta + x^t C z + \varepsilon \quad (1.16)$$

donde $x^t = (x_1, \dots, x_k)$ representa los k factores de control y $z = (z_1, \dots, z_q)$ los q factores de ruido, γ_0 constante, los vectores de los parámetros lineales $\beta^t = (\beta_1, \dots, \beta_k)$, $\delta^t = (\alpha_1, \dots, \alpha_q)$. De modo similar, $B = (\beta_{11}, \dots, \beta_{1k}, \beta_{k1}, \dots, \beta_{kk})$ y $C = (\alpha_{11}, \dots, \alpha_{1q}, \dots, \alpha_{k1}, \dots, \alpha_{kq})$ representan las matrices de parámetros de segundo orden y el error será tal que $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

Adicionalmente, se consideran los supuestos para los factores de ruido: $E(z) = 0$, matriz de varianzas y covarianzas V y $diag(V) = \sigma_z^2$.

Al calcular la media y varianza de la expresión 1.16 se tendrán los equivalentes en esta segunda estrategia a los modelos 1.11 y 1.12

$$Y_1 = E(Y_{xz}) = \gamma_0 + x^t \beta + x^t B x \quad (1.17)$$

$$Y_2 = Var(Y_{xz}) = (\delta + x^t C)^t V (\delta + x^t C) + \sigma_\varepsilon^2 \quad (1.18)$$

Los objetivos en esta estrategia son similares a los ya presentados. Primordialmente, encontrar las condiciones de operación óptimas dentro de una región R_x que acerquen Y_{xz} al valor objetivo, al tiempo que se estudian los factores de ruido dentro de una región R_z para reducir la variabilidad.

Siguiendo este esquema, ante ausencia de réplicas o factores de ruido no se tendrá una desviación estándar a modelar. En este caso la segunda respuesta de interés \hat{Y}_2 puede considerarse

$$\hat{Y}_2 = \log \left(abs \left(Y_{1i} - \hat{Y}_{1i} \right) \right) \quad (1.19)$$

donde Y_{1i} son los resultados experimentales de cada tratamiento.

1.4. Esquemas de optimización en la respuesta dual

Para obtener el conjunto de entradas óptimas que acerquen lo más posible al valor objetivo al tiempo que reducen variabilidad se considerarán diversos trabajos que han explorado los objetivos, naturaleza y

herramientas para atacar el problema. Algunos de ellos son brevemente descritos a continuación siguiendo lo expuesto por Barrios-Domínguez (2008).

1.4.1. Procedimiento Vining-Myers

Dentro del marco de la metodología de superficie de respuesta, Vining y Myers (1990) proponen el uso de la técnica de respuesta dual planteada por Myers y Carter (1973) para modelar dos respuestas Y_p y Y_s , que llaman primaria y secundaria respectivamente, con el uso de modelos completos de segundo orden (aunque contengan términos no significativos). Se busca optimizar alguna de las respuestas sujeta a una restricción de la otra. Así Y_p puede representar un modelo para la media, mientras que Y_s modela la varianza y el problema se puede formular mediante la estrategia siguiente: optimizar \hat{Y}_s sujeto a que la media tenga un valor T es decir $\hat{Y}_p = T$ como se planteó en 1.13, con varianza mínima. Y por el método de multiplicadores de Lagrange se tiene:

$$H(X) = \hat{Y}_s - \lambda (\hat{Y}_p - T) \quad (1.20)$$

Minimizando dicha expresión se llega el conjunto de entradas óptimas.

1.4.2. Procedimiento deseabilidad

Las funciones de deseabilidad son un método de optimización con criterios múltiples, como el problema que nos ocupa.

Partiendo de la idea de que el producto o proceso tiene múltiples características de calidad, y tener alguna de ellas fuera de los límites deseados es inaceptable.

Consiste en combinar las respuestas en una sola (usando la media geométrica) a la que se hará referencia como función de deseabilidad. La cual irá de 0, en el caso de que una o más de las características del producto son inaceptables, a 1, cuando todas las características de calidad estén en su valor objetivo.

Para lograr llegar a tal función se ligará a la variable de respuesta Y_i una función de deseabilidad individual d_i en el intervalo $[0, 1]$, cuyo valor se incrementa cuando la respuesta correspondiente se acerca a su valor objetivo. Se definen los valores LS y LI como los límites superior e inferior respectivamente para tal variable.

Retomando las situaciones básicas, para el caso *Target is Best*, la función individual está definida como

$$d_i(\hat{Y}_i) = \begin{cases} 0 & \text{si } \hat{Y}_i(x) < L_i \\ \left(\frac{\hat{Y}_i(x) - L_i}{T_i - L_i}\right)^s & \text{si } L_i \leq \hat{Y}_i(x) \leq T_i \\ \left(\frac{\hat{Y}_i(x) - U_i}{T_i - U_i}\right)^t & \text{si } T_i \leq \hat{Y}_i(x) \leq U_i \\ 0 & \text{si } \hat{Y}_i(x) > U_i \end{cases} \quad (1.21)$$

donde los exponentes s y t determinan la importancia de alcanzar el valor objetivo.

Cuando se desea maximizar la respuesta o caso *Larger the better*, es la definición correspondiente

$$d_i(\hat{Y}_i) = \begin{cases} 0 & \text{si } \hat{Y}_i(x) < L_i \\ \left(\frac{\hat{Y}_i(x) - L_i}{T_i - L_i}\right)^s & \text{si } L_i \leq \hat{Y}_i(x) \leq T_i \\ 0 & \text{si } \hat{Y}_i(x) > U_i \end{cases} \quad (1.22)$$

Y para el caso *Smaller the better* se tendrá

$$d_i(\hat{Y}_i) = \begin{cases} 0 & \text{si } \hat{Y}_i(x) < L_i \\ \left(\frac{\hat{Y}_i(x) - U_i}{T_i - U_i}\right)^s & \text{si } T_i \leq \hat{Y}_i(x) \leq U_i \\ 0 & \text{si } \hat{Y}_i(x) > U_i \end{cases} \quad (1.23)$$

De modo que este proceso consiste en realizar el experimento y ajustar modelos las k respuestas, posteriormente definir la función de deseabilidad individual para cada una.

Finalmente, se buscará maximizar la deseabilidad general D respecto a los factores de control.

Como nuestro problema involucra 2 variables de respuesta la función a optimizar será

$$D = (d_1 d_2)^{1/2} \quad (1.24)$$

1.4.3. Procedimiento Lin-Tu

Lin y Tu (1995) encontraron una posible debilidad del esquema de optimización de Vining-Myers en las restricciones que se impusieran a las respuestas. Plantearon así su propuesta dentro del contexto de diseño robusto, cuyo objetivo consiste en optimizar el error cuadrático medio o MSE, por sus siglas en inglés, compuesto de dos términos: sesgo y varianza, definido como:

$$ECM = (\hat{Y}_1 - T)^2 + \hat{Y}_2^2. \quad (1.25)$$

Se puede de este modo compensar la lejanía del valor objetivo con una varianza pequeña. De manera más general

$$ECM = \lambda (\hat{Y}_1 - T)^2 + (1 - \lambda) \hat{Y}_2^2$$

con $\lambda \in [0, 1]$.

1.4.4. Copeland-Nelson

Su propuesta añade una cota a la distancia entre la media y el valor objetivo, reformulando así la optimización del ECM.

$$\begin{aligned} & \text{mín } \hat{\sigma}(x) \\ & \text{s.a. } (\hat{\mu}(x) - T)^2 \leq \Delta^2 \end{aligned}$$

Para llegar a las condiciones óptimas basta optimizar la función

$$\hat{\sigma}(x) + \varepsilon \quad (1.26)$$

donde

$$\varepsilon = \begin{cases} (\hat{\mu}(x) - T)^2 & \text{si } (\hat{\mu}(x) - T)^2 > \Delta^2 \\ 0 & \text{si } (\hat{\mu}(x) - T)^2 \leq \Delta^2 \end{cases} \quad (1.27)$$

1.4.5. Vining-Bohn

Proponen la estimación conjunta con elementos no paramétricos y semiparamétricos.

Ante los malos ajustes resultantes para la varianza con la herramienta paramétrica, incluyen una regresión kernel para tal variabilidad y una regresión kernel separada para estimar la media. Luego buscan optimizar la función propuesta por Lin-Tu para finalmente minimizar el $ECM(y)$.

1.4.6. Domínguez-Rocha

Domínguez y Rocha (2004) proponen un enfoque basado en la optimización usando curvas de nivel. En un inicio se establecen en el plano el conjunto de restricciones (en el caso de la respuesta dual, la región queda determinada por un modelo ajustado de segundo orden y la región experimental). Posteriormente se sobrepone la función objetivo (el modelo de segundo orden a optimizar) en la región de restricciones.

1.4.7. Procedimiento razón señal a ruido

Taguchi definió la llamada razón de señal a ruido (SNR por sus siglas en inglés) para medir el comportamiento relativo al valor objetivo. Si y_1, \dots, y_m representa una muestra aleatoria de m observaciones, las razones correspondientes a las tres situaciones principales (es posible expresarlas en términos de media y varianza) serían:

- Target is Best

$$10 \log_{10} \left[\frac{\bar{y}^2}{s^2} \right] \quad (1.28)$$

- Larger the better

$$-10 \log_{10} \left[\frac{1}{m} \sum_{i=1}^m \frac{1}{y_i^2} \right] \quad (1.29)$$

- Smaller the better

$$-10 \log_{10} \left[\frac{1}{m} \sum_{i=1}^m y_i^2 \right] = -10 \log_{10} \left(\frac{(m-1)S^2}{m} + \bar{Y}^2 \right) \quad (1.30)$$

Se busca maximizar la razón de señal a ruido correspondiente a la situación por analizar.

Box (1985) calificó el enfoque de Taguchi como un avance significativo en la aplicación de diseños experimentales para incrementar la calidad de un producto. Sin embargo, señaló algunas posibles deficiencias como se menciona en Vining-Myers (1990):

- El método de Taguchi no explota una naturaleza secuencial de investigación.
- Los diseños propuestos son un tanto limitados y falla al enfrentarse a interacciones.
- Hay métodos de análisis más simples y eficientes.
- Resulta poco convincente su uso universal.

En este trabajo será de interés ver su comportamiento ante un problema donde el modelo para la varianza es sumamente deficiente.

1.4.8. El futuro de la MSR

En Khuri (2006) y Khuri (2010) se hace un completo recorrido por las bases de la MSR, sus herramientas, el surgimiento y planteamiento del diseño robusto, así como las estrategias de optimización mencionadas hasta el momento.

Además presentan los nuevos avances en dicha área; modelos con efectos aleatorios, la introducción de modelos lineales generalizados y novedosos procedimientos gráficos para comparar diseños o determinar su capacidad de predicción a lo largo de la región experimental.

Lo anterior siempre recalando su potencial en la búsqueda de la calidad total.

Referencias muy recomendables acerca de las que no se profundizará más en este trabajo pero que ameritan una lectura posterior.

1.4.9. Relación con la tesis

Utilizando los datos del ejemplo clásico de las tintas planteado por Box y Draper (1987) donde se busca optimizar la calidad de la impresión sobre etiquetas. Usando su diseño y resultados para tres réplicas, se tendrá la metodología de superficie de respuesta como el punto de partida para analizar el problema. Se procederá entonces con las estrategias y métodos descritos a lo largo de este capítulo a fin de encontrar el conjunto de entradas óptimas, para finalmente comparar o identificar posibles situaciones de aplicación para cada una.

Capítulo 2

Elementos no paramétricos

2.1. Introducción

Para entender los estimadores kernel de densidad primero necesitamos entender los histogramas, pues es en sus debilidades que surge la necesidad de una herramienta más poderosa con las mismas bondades.

Reconociendo al histograma como el estimador de densidad no paramétrico más sencillo y el que encontramos con más frecuencia cuando se comienza a usar estadística en las diversas áreas del conocimiento. Recordando que al construirlo es necesario considerar el ancho de los subintervalos en los cuales dividimos el rango de los datos y los puntos finales de dichos subintervalos. Surge así una primera debilidad: ser dependientes del ancho del subintervalo y, además, no ser suaves.

La estimación kernel ofrece solución a ambos problemas, siendo una manera de estimar funciones de densidad para una variable aleatoria. Es un problema fundamental de suavizamiento donde las inferencias sobre la población se hacen basados en una muestra finita. Los estimadores kernel pertenecen a una clase de estimadores llamados no paramétricos, caracterizados por no tener una estructura fija y depender del conjunto de datos para alcanzar un estimado, de modo contrario que los paramétricos donde el estimador tiene una estructura fija y hay un conjunto de parámetros para dicha forma funcional.

Se puede pensar en los estimadores kernel de densidad como la generalización y mejora de los histogramas, esto debido a que:

- Son suaves
- Sin puntos finales
- Dependen del ancho de banda.

En econometría el estimador kernel es también conocido como Método Parzen-Rosenblatt (en honor de Emanuel Parzen y Murray Rosenblatt a quienes se les atribuye el método en su forma actual).

En este capítulo se presentan las definiciones y propiedades básicas en el caso unidimensional, para luego extenderlo al caso multidimensional que requerirá especial atención, también se menciona la estrategia usual para elegir el ancho de banda, concluyendo con una breve discusión sobre sus ventajas y desventajas. Finalmente, se presentarán los fundamentos de la regresión no paramétrica por kernel.

2.2. Definición

En principio, un Kernel será una función suave K tal que

$$K(x) \geq 0 \quad (2.1)$$

$$\int K(x) dx = 1 \quad (2.2)$$

$$\int xK(x) dx = 0 \quad (2.3)$$

$$\sigma_K^2 = \int x^2 K(x) dx > 0. \quad (2.4)$$

Decimos que un kernel K es de orden k si

$$\int K(x) x^j dx = 0 \quad (2.5)$$

para todo $j < k$ con alguna posible excepción (usualmente el 0).

Tales funciones son usadas para tomar medias locales y tendrán un papel primordial en diferentes métodos de estimación.

2.2.1. Ejemplos

En Härdle (1994) y Wasserman (2006) se puede encontrar una detallada presentación de los kernel más populares y sus propiedades.

- Kernel Gaussiano o Normal

$$K(x) = (2\pi)^{-1/2} e^{-x^2/2} \quad (2.6)$$

El más conocido.

- Kernel Epanechnikov

$$K(x) = \begin{cases} \frac{3}{4} (1 - x^2/5) \sqrt{5}, & |x| < 5 \\ 0 & \text{elsewhere} \end{cases} \quad (2.7)$$

Usado comúnmente y que goza de algunas propiedades de optimalidad. Entre más suave sea el kernel usado más lo serán los estimados.

2.3. Regresión kernel

La regresión kernel tiene por objetivo encontrar una relación no lineal entre un par de variables. Bajo el supuesto de que se tienen n pares de observaciones $(x_1, Y_1), \dots, (x_n, Y_n)$ y que la respuesta Y está relacionada con x mediante la ecuación

$$Y_i = m(x_i) + \varepsilon_i \quad (2.8)$$

donde ε_i es un error con media 0 y m es la función de regresión. Buscamos estimar m bajo supuestos débiles y denotaremos a su estimador como $\hat{m}_n(x)$. Si tratamos a los x_i como variables aleatorias es posible interpretar m como

$$m(x) = E(Y|X = x) \quad (2.9)$$

es posible hacer el suavizamiento de manera lineal y obtener $\hat{m}_n(x)$ si para cada x existe un vector $l(x)$ tal que

$$\hat{m}_n(x) = \sum_{i=1}^n l_i(x) Y_i.$$

Se puede entonces definir el vector de valores ajustados

$$\mathbf{m} = (\hat{m}_n(x_1), \dots, \hat{m}_n(x_n))^T = LY$$

donde $L_{n \times n}$ es una matriz cuyo i -ésimo renglón es $l(x_i)^T$. Las entradas para tal renglón mostrarán los pesos dados a cada Y_i para formar el estimado.

Como menciona Wasserman (2006), es importante notar que los pesos deben ser tales que

$$\sum_{i=1}^n l_i(x) = 1$$

lo cual implicará preservar curvas constantes.

Definido lo anterior se procede con los estimadores de $m(x)$ que se obtienen al tomar promedios ponderados de Y_i (se da mayor peso a los puntos cercanos a x).

2.3.1. Estimador Nadaraya Watson

Propuesto por Nadaraya y Watson (1964), se considera $b > 0$ el parámetro de suavizamiento o ancho de banda tal que $b \rightarrow 0$ y $nb \rightarrow 0$ cuando $n \rightarrow \infty$. El estimador Nadaraya-Watson por kernel está definido como

$$\hat{m}_n(x) = \sum_{i=1}^n w_i(x) Y_i \quad (2.10)$$

donde la secuencia de pesos de los "smoothers" kernel (para dimensión 1) está definida de la manera presentada a continuación

$$w_i(x) = \frac{K_b(x - x_i)}{\frac{1}{n} \sum_{i=1}^n K_b(x - x_i)} \quad (2.11)$$

con

$$K_b(u) = \frac{1}{b} K\left(\frac{u}{b}\right).$$

2.3.2. Estimador Gasser Müller

Propuesto por Gasser y Müller (1979, 1984), dada una se toma un ancho de banda b y se define el estimador función kernel K

$$\hat{m}_n(x) = \sum_{i=1}^n w_i y_i \quad (2.12)$$

donde la secuencia de pesos está definida como

$$w_i = \frac{1}{b} \int_{I_{i-1}}^{I_i} K\left(\frac{x-u}{b}\right) du \quad (2.13)$$

y los límites de integración determinados por

$$I_i = \frac{x_{i-1} + x_i}{2}.$$

La elección del kernel no tiene demasiada importancia pues incluso usando funciones kernel diferentes los estimados son numéricamente muy similares.

2.3.3. Precisión

Usando el estimador Nadaraya-Watson, se tendrá que el riesgo, definido como

$$R(\hat{m}_n, m) = E \left(\frac{1}{n} \sum_{i=1}^n (\hat{m}_n(x_i) - m(x_i))^2 \right) \quad (2.14)$$

si se supone que los datos provienen de una densidad f se tendrá

$$R(\hat{m}_n, m) = \frac{b^4}{4} \left(\int x^2 K(x) dx \right)^2 \int \left(m''(x) + 2m'(x) \frac{f'(x)}{f(x)} \right)^2 dx \quad (2.15)$$

$$+ \frac{\sigma^2 \int K^2(x) dx}{nb} \int \frac{1}{f(x)} dx + o(nb^{-1}) + o(b^4).$$

Siendo el primer término el sesgo y el segundo la varianza. Como se muestra en Wasserman (2006), al optimizar el ancho de banda se obtiene

$$b_* = \left(\frac{1}{n} \right)^{\frac{1}{5}} \left(\frac{\sigma^2 \int K^2(x) dx \int dx/f(x)}{\left(\int x^2 K(x) dx \right)^2 \int \left(m''(x) + 2m'(x) \frac{f'(x)}{f(x)} \right)^2 dx} \right) \quad (2.16)$$

y así, $b_* = O(n^{-1/5})$. Con tal ancho de banda, el riesgo decrece hasta ser $O(n^{-4/5})$.

Esto es más lento que en la mayoría de los modelos paramétricos (cuyo riesgo para el estimador máximo verosímil decrece a 0 a una tasa $1/n$). Aunque en la práctica no podemos usar b_* pues depende de la función desconocida m .

2.4. Estimación de densidades por kernel

El objetivo de la estimación de densidades no paramétrica será estimar f con tan pocos supuestos como sea posible. El estimador para la densidad será denotado por \hat{f}_n . Que similarmente a lo visto en regresión kernel, dependerá de manera importante del parámetro de suavizamiento o ancho de banda b . Los histogramas son discontinuos mientras que los estimadores de densidad Kernel son más suaves y convergen más rápido a la verdadera densidad que los histogramas. Acorde a la definición, el kernel es una función real K continua, acotada, simétrica y que integra a 1.

Sean X_1, \dots, X_n datos observados en una muestra de f . Dado un kernel K y un ancho de banda b , el estimador de densidad kernel está definido por

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} K \left(\frac{x - x_i}{b} \right). \quad (2.17)$$

Al igual que en la regresión kernel la elección de K no es crucial pues la convergencia a la densidad igual se dará. En cambio, sí está bajo la influencia del valor elegido para b , pues valores grandes de b implican que más observaciones están siendo usadas para estimar $f(x)$, por lo que $\hat{f}_n(x)$ tendrá menor varianza pero mayor sesgo que usando un menor valor de b .

Es posible usar una gran variedad de funciones kernel, pero serán las consideraciones teóricas y prácticas las que limitan la elección.

Puede demostrarse que si f es continua en x , $b \rightarrow 0$ y $nb \rightarrow \infty$ cuando $n \rightarrow \infty$. Entonces

$$\hat{f}_n(x) \rightarrow f(x).$$

2.4.1. Precisión

Es necesario definir una medida global de desviación a fin de comparar los estimadores respecto a f y poder elegir aquel que más se le acerca (modificando el ancho de banda sería la opción más sencilla). Se evaluará la calidad de un estimador \hat{f}_n con el riesgo o error cuadrático medio integrado (IMSE por sus siglas en inglés), definiendo el riesgo en un punto x como

$$\begin{aligned} R_x &= E \left(f(x) - \hat{f}_n(x) \right)^2 \\ IMSE &= \int R_x dx \end{aligned} \quad (2.18)$$

y bajo el supuesto que de f'' es absolutamente continua y $\int (f'''(x))^2 dx < \infty$ se tendrá

$$R_x = \frac{1}{4} \sigma_K^4 b^4 (f''(x))^2 + \frac{f(x) \int K^2(x) dx}{nb} + O\left(\frac{1}{n}\right) + O(b^6). \quad (2.19)$$

Optimizando respecto a b se tendrá el ancho de banda óptimo

$$b_* = \left(\frac{c_2}{c_1^2 A(f) n} \right)^{1/5} \quad (2.20)$$

donde $c_1 = \int x^2 K(x) dx$, $c_2 = \int K(x)^2 dx$ y $A(f) = \int (f''(x))^2 dx$.

De modo que usando b_* se tendrá que $IMSE = O(n^{-4/5})$. Como se menciona en Wasserman (2006) y Härdle (1990), los histogramas convergen a una tasa $O(n^{-2/3})$ por lo que los estimadores kernel son superiores, además de que no hay estimador con convergencia más rápida que $O(n^{-4/5})$, una prueba detallada de este hecho puede encontrarse en Van der Vaart (1998).

Como ya se mencionó, la precisión de los estimados es sólo dependiente del par (K, b) principalmente del ancho de banda pues anchos de banda pequeños llevarán a estimados demasiado toscos mientras que anchos de banda grandes a estimados más suaves. Además, el error cuadrático medio se divide en dos partes: varianza y sesgo. El sesgo como función de b es creciente, mientras que la varianza es decreciente. Por su comportamiento cualitativo se tiene la idea previa de que el problema de suavizamiento consiste en un balance de la varianza contra el sesgo.

2.5. Caso multidimensional

En encontrarse con variables multidimensionales $X_i = (X_{i1}, \dots, X_{id})$ el estimador kernel puede ser fácilmente generalizado a d dimensiones. Es posible recurrir a una función producto kernel multidimensional de la siguiente forma

$$K(u_1, \dots, u_d) = \prod_{i=1}^d K(u_i). \quad (2.21)$$

Si se tienen d factores y el factor x_j tiene n_j niveles, se indexan los puntos del diseño con (i_1, \dots, i_d) y se particionan los intervalos del factor x_j con I_{i_j} . Sea $b = (b_1, \dots, b_d)$ un vector de anchos de banda y siguiendo lo presentado en Vining-Bohn (1998) se define el estimador para regresión kernel Gasser-Müller multivariado como

$$\hat{m}_n(x^*) = \sum_{i_d=1}^{n_d} \dots \sum_{i_1=1}^{n_1} \prod_{j=1}^d \frac{1}{b_j} \left(\int_{I_{i_j}} K\left(\frac{x_j^* - u}{b_j}\right) du \right) y_{i_1 \dots i_d}. \quad (2.22)$$

En el caso de la estimación de densidades, la curva estimada será

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - X_i), \quad (2.23)$$

donde

$$K_b(x - x_i) = \frac{1}{nb_1 \dots b_d} \left\{ \prod_{j=1}^d K\left(\frac{x_i - x_{ij}}{b_j}\right) \right\}. \quad (2.24)$$

Por lo general, las observaciones se toman a distancias regulares formando una red equidistante de puntos (ejemplos de ello son las observaciones de datos longitudinales o procesos análogos discretizados) pues aunque no sea un requisito del procedimiento, el estar uniformemente espaciados asegura tantos datos experimentales como sea posible en una vecindad de cada posible ubicación del valor de la predicción.

Un problema usual con los métodos de suavizamiento es la llamada "maldición de la dimensionalidad". Es decir, la estimación se vuelve más difícil conforme la dimensión de las observaciones aumenta.

2.6. Kernel óptimo

Entendiendo por optimización la minimización asintótica del IMSE, un kernel óptimo será aquel que minimice el IMSE expresado en 2.18. Gasser, Müller y Mammitzsch (1995) usaron métodos variacionales para minimizar la esperanza de R_x respecto a K . La respuesta son polinomios de grado p algunos de los cuales de presentan en la tabla 2.1.

Tabla 2.1

Kernel óptimo

k	p	Kernel $K(x)$
0	2	$\frac{3}{4}(1 - x^2)$
0	4	$\frac{15}{32}(3 - 10x^2 + 7x^4)$
1	3	$\frac{15}{4}(x^3 - x)$
1	5	$\frac{105}{32}(-5x + 14x^3 - 9x^5)$
2	4	$\frac{105}{16}(-1 + 6x^2 - 5x^4)$
3	6	$\frac{315}{64}(-5 + 63x^2 - 135x^4 + 77x^6)$

donde k es al grado del kernel ya definido en 2.5.

Vining y Bohn (1998) usan el segundo kernel de la tabla para su análisis por regresión kernel. En el presente trabajo, se usará principalmente el kernel óptimo

$$K(x) = \frac{35}{32}(1 - 3x^2 + 3x^4 - x^6). \quad (2.25)$$

2.7. Elección del ancho de banda

La elección de un buen parámetro de suavizamiento es de vital importancia pues provee información acerca de la razón de señal a ruido en los datos, tal vez sea el tema más controversial en cuanto al uso de funciones kernel pues tiene una fuerte influencia en la calidad del estimador.

Se han estudiado y descartado numerosos métodos. Desde los métodos subjetivos donde el investigador puede elegirlo a voluntad, hasta los automáticos o *data-driven* donde todo depende de los datos y no es necesaria experiencia alguna para un ajuste razonable, para información a detalle de los mismos se puede consultar Jones et. al (1996) y Ledl (2002).

El método más simple y posiblemente el más usado es elegir el ancho de banda basándose en la observación. Así, en un problema de suavizamiento de curva es recomendable analizar los estimadores obtenidos del uso de diferentes anchos de banda antes de pasar a un método más sofisticado, variando tal parámetro para ver la forma resultante, esto nos permite ver los datos a profundidad, proporcionando diferentes perspectivas de la posible estructura de la densidad desconocida. Lo anterior fundamentado en el hecho de que la curva verdadera está entre un estimador muy suave y uno muy oscilante.

En general, se deja que el ancho de banda dependa del tamaño de muestra, por lo que en ocasiones se le denota b_n .

Como ya se mencionó, revisando el error cuadrático medio, si el ancho de banda elegido es demasiado pequeño, la varianza del estimador es dominante, de otro modo el sesgo lo será.

En la práctica, si el interés estriba en evitar malas interpretaciones de fenómenos aleatorios en un análisis exploratorio de datos, es una buena política sobresuavizar un poco, lo cual hace menos posible la aparición de picos aleatorios o falsas oscilaciones.

2.7.1. Caso especial

Como se menciona en Härdle (1994), cuando se sabe que f es muy suave, se calcula b_* bajo el supuesto de una distribución de referencia, por lo general se supone que f es normal con parámetros μ y σ^2 , lo que sirve como guía para elegir el ancho de banda. Se define entonces la regla de referencia normal para densidades suaves y el kernel Normal definido en 2.6, se usará el ancho de banda

$$b \approx \frac{1,06\hat{\sigma}}{n^{1/5}} \quad (2.26)$$

donde

$$\hat{\sigma} = \min \left\{ s, \frac{Q}{1,34} \right\}.$$

Para hacer este estimador menos sensible a *outliers*, puede usarse el rango intercuartil Q en vez de $\hat{\sigma}$.

2.8. Ejemplo

Se revisará ahora un ejemplo de regresión kernel presentado en el trabajo de Vining y Bohn (1998), donde se suponen observaciones tomadas de $n = 5$ puntos de diseño igualmente espaciados, de modo que

$$x_1 = 0, \quad x_2 = 0,25, \quad x_3 = 0,5, \quad x_4 = 0,75, \quad x_5 = 1.$$

Las respuestas correspondientes serán y_1, y_2, y_3, y_4 y y_5 . Se considera el kernel óptimo

$$K(t) = \begin{cases} \frac{35}{32} (1 - 3t^2 + 3t^4 - t^6), & -1 \leq x_j \leq 1 \\ 0 & \end{cases} \quad (2.28)$$

Inicialmente se buscará predecir la respuesta en $x_0 = 0,48$ con un ancho de banda $b = 0,3$ la secuencia de interpolación

$$r_0 = -1/32, r_1 = 1/8, r_2 = 3/8, r_3 = 5/8, r_4 = 7/8, r_5 = 33/32.$$

Los pesos estarán determinados por la expresión

$$\begin{aligned} w_i &= \frac{1}{b} \int_{r_{i-1}}^{r_i} K\left(\frac{x_0 - u}{b}\right) du \\ &= \frac{1}{b} \left(\frac{35}{32} \int_{r_{i-1}}^{r_i} du - \frac{3(35)}{32b^2} \int_{r_{i-1}}^{r_i} (x_0 - u)^2 du \right) \\ &\quad + \frac{1}{b} \left(\frac{3(35)}{32b^4} \int_{r_{i-1}}^{r_i} (x_0 - u)^4 du - \frac{35}{32b^6} \int_{r_{i-1}}^{r_i} (x_0 - u)^6 du \right) \end{aligned} \quad (2.29)$$

para todo u en el intervalo (r_{i-1}, r_i) tal que $-1 \leq \frac{(x_0 - u)}{b} \leq 1$ y 0 en otro caso. De este modo se tendrá que

$$w_1 = 0 = w_5$$

pues los intervalos $[-1/32, 1/8]$ que contiene a $x_1 = 0$ y $[7/8, 33/32]$ que contiene a $x_5 = 1$, están fuera de $[x_0 - b, x_0 + b] = [0,18, 0,78]$.

Sustituyendo los valores restantes manteniendo los límites de integración entre $(0,18, 0,78)$ la secuencia de pesos como fue definida en 2.13 será

$$\begin{aligned} w_2 &= \frac{1}{0,3} \int_{0,18}^{/8} K\left(\frac{0,48 - u}{0,3}\right) du \\ &= 0,16074 \\ w_3 &= \frac{1}{0,3} \int_{0,375}^{0,625} K\left(\frac{0,48 - u}{0,3}\right) du \\ &= 0,76076 \\ w_4 &= \frac{1}{0,3} \int_{0,625}^{0,78} K\left(\frac{0,48 - u}{0,3}\right) du \\ &= 7.8504 \times 10^{-2}. \end{aligned}$$

Ahora, usando el estimador Gasser Müller definido en 2.12, considerando los valores de la respuesta y_1, \dots, y_5 en cada uno de los puntos de diseño, el estimador kernel en el punto $x_0 = 0,48$ será

$$\begin{aligned} \hat{m}_n(0,48) &= \sum_{i=1}^5 w_i y_i \\ &= 0y_1 + 0,1607y_2 + 0,7608y_3 + 0,0785y_4 + 0y_5 \\ &= 0,1607y_2 + 0,7608y_3 + 0,0785y_4. \end{aligned} \quad (2.30)$$

2.9. Discusión

Los métodos de regresión no paramétricos son a menudo de gran valor como herramientas exploratorias de datos en situaciones donde ya se agotaron los métodos paramétricos o es difícil su uso. Por mencionar algunas:

- Falta de información en el área objeto de aplicación para proponer un modelo paramétrico válido.

- La evolución temporal entre diferentes individuos o subgrupos varía significativamente de modo que no pueden ser descritos por uno o unos cuantos modelos de regresión paramétrica.
- Las evoluciones temporales por individuo son demasiado complejas para modelación paramétrica (las funciones paramétricas pueden no existir o requerir demasiados parámetros).
- No hay suficientes datos para estimar todos los parámetros que se requerirían para un ajuste adecuado (problemas de convergencia de algoritmos tipo Newton-Raphson).

Entre los métodos no paramétricos, tal vez el más flexible sea el método Kernel pues puede ser modificado con tan sólo elegir una función kernel diferente o cambiar el ancho de banda, ya sea global o localmente, a su vez dando la posibilidad de manipular sesgo y varianza.

Además, para grupos de datos pequeños ($n < 50$) los estimadores kernel son computacionalmente más rápidos que los *splines*.

2.10. Relación con la tesis

Ante resultados poco satisfactorios usando un modelo paramétrico para la varianza de un proceso industrial, o incluso para comparar enfoques, se recurre al empleo de técnicas no paramétricas. Se piensa entonces en la que tenga una convergencia más rápida, es decir, una regresión kernel, que se añadirá a la metodología de superficie de respuesta y al diseño robusto. Se modelará de acuerdo a dos enfoques:

- Semiparamétrico, usando regresión kernel para la varianza y posteriormente usando lo obtenido con WLS para la respuesta.
- No paramétrico, usando regresión kernel para varianza y respuesta.

Se buscará mantener el ECM al mínimo y obtener las condiciones óptimas en cada escenario que lleven a mayor calidad con menor variación del proceso.

Capítulo 3

Resultados

Se revisará ahora un ejemplo clásico para la aplicación de las técnicas de optimización antes mencionadas, visto desde tres diferentes enfoques. Primero en el contexto de diseño robusto y posteriormente con la introducción de elementos no paramétricos, ya sea parcial o completamente.

Finalmente, se realiza una serie de comparaciones entre ellos buscando el modelo que ofrezca mayores ventajas.

3.1. Ejemplo de estudio

Presentado como ejercicio en Box y Draper (1987), el ejemplo de las tintas ha sido analizado por Vining y Myers (1990), Del Castillo y Montgomery (1993), Lin y Tu (1995), Copeland y Nelson (1996), así como Vining y Bohn (1998).

El propósito del experimento fue estudiar el efecto de los factores

x_1 : Velocidad

x_2 : Presión

x_3 : Distancia

en la habilidad de una impresora para aplicar tintas de colores sobre etiquetas, con tres niveles disponibles para cada factor.

Todos estos análisis consideraron la situación donde se desea minimizar la variabilidad del proceso alrededor de un valor objetivo de 500 para la respuesta. El experimento original usó un factorial completo 3^3 con 3 réplicas (que se pueden ver como un factor de ruido) en cada punto de diseño para un total de 81 corridas.

La tabla 3.1 muestra los resultados experimentales.

Tabla 3.1

Resultados experimentales ejemplo de las tintas

i	x_1	x_2	x_3	y_{i1}	y_{i2}	y_{i3}	i	x_1	x_2	x_3	y_{i1}	y_{i2}	y_{i3}
1	-1	-1	-1	34	10	28	15	1	0	0	541	568	396
2	0	-1	-1	115	115	130	16	-1	1	0	288	192	312
3	1	-1	-1	192	186	263	17	0	1	0	432	336	513
4	-1	0	-1	82	88	88	18	1	1	0	713	725	754
5	0	0	-1	44	178	188	19	-1	-1	1	364	99	199
6	1	0	-1	322	350	350	20	0	-1	1	232	221	266
7	-1	1	-1	141	110	86	21	1	-1	1	408	415	443
8	0	1	-1	259	251	259	22	-1	0	1	182	233	182
9	1	1	-1	290	280	245	23	0	0	1	507	515	434
10	-1	-1	0	81	81	81	24	1	0	1	846	535	640
11	0	-1	0	90	122	93	25	-1	1	1	236	126	168
12	1	-1	0	319	376	376	26	0	1	1	660	440	403
13	-1	0	0	180	180	154	27	1	1	1	878	991	1161
14	0	0	0	372	372	372							

Tabla 3.2

Media y varianza de las observaciones

i	$\bar{y}_{i\bullet}$	s_i	i	$\bar{y}_{i\bullet}$	s_i
1	24.0	12.5	15	501.7	92.5
2	120.3	8.4	16	264.0	63.5
3	213.7	42.8	17	427.0	88.6
4	86.0	3.46	18	730.7	21.1
5	136.7	80.4	19	220.7	133.8
6	340.7	16.2	20	239.7	23.5
7	112.3	27.6	21	422.0	18.5
8	256.3	4.6	22	199.0	29.4
9	271.7	23.6	23	485.3	44.6
10	81.0	0.0	24	673.7	158.2
11	101.7	17.7	25	176.7	55.5
12	357.0	32.9	26	501.0	138.9
13	171.3	15.0	27	1010.0	142.5
14	372.0	0.0			

3.2. Análisis exploratorio

Primero es conveniente realizar un sencillo ejercicio exploratorio, graficando los valores de media y varianza muestral correspondientes a cada punto del diseño se tienen las primeras nociones para determinar la influencia de éstos en la dispersión de las respuestas de interés.

De igual manera, es útil para buscar indicios sobre lo que posteriormente se traducirá en un conjunto de entradas óptimas, también es recomendable analizar la influencia de cada nivel posible de las variables en ellas.

Tabla 3.3

Media por nivel de cada factor

	Media		
Nivel/Factor	x_1	x_2	x_3
-1	148.333	197.778	173.185
0	293.333	329.593	334.037
1	502.333	416.623	436.444
$\bar{Y}_m = 314,667$			

Varianza por nivel de cada factor

	Desv. Estándar		
Nivel/Factor	x_1	x_2	x_3
-1	37.868	32.232	24.396
0	45.191	48.871	36.808
1	60.922	62.878	82.777
$\bar{Y}_s = 47,994$			

Recordando que el valor objetivo para la característica de calidad está en 500 y la desviación estándar en el caso ideal sería 0, o en su defecto, la menor posible.

Al revisar la media y varianza global se tendrán valores sumamente lejanos a dichos ideales. Lo cual nos lleva a una serie de interrogantes iniciales:

- ¿Se deberá llevar cada uno de los factores a su nivel máximo para acercar la media a 500?
- ¿Se deberá llevar a cada uno de los factores a su nivel mínimo para que la variación se reduzca?
- ¿Cómo se logrará llegar a ambos objetivos si en esta primera visión parecen contradictorios?
- ¿Hay algún factor que favorezca estos objetivos?
- ¿Al momento de crear el modelo todos tendrán la misma importancia?

3.2.1. Dispersión de Desviación Estándar

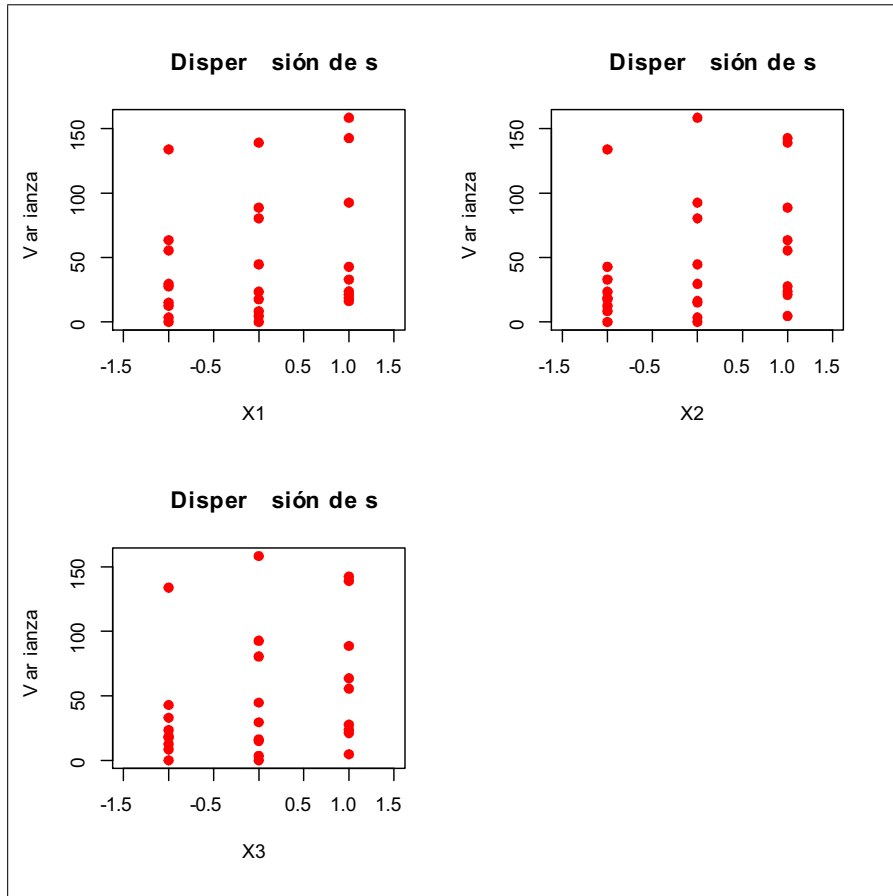


Figura 3.1

Es interesante notar que todas las entradas del diseño con desviaciones estándar muestrales menores a 10 tienen al menos uno de x_1 , x_2 y x_3 igual a cero. El centro exacto de este diseño tiene una desviación estándar de 0.

Ninguna de las entradas del diseño con una desviación estándar menor que 10 usa $x_3 = 1$. Además, todas las entradas del diseño con desviaciones estándar mayores que 100 involucran $x_3 = 1$.

Esto sugiere que para minimizar la desviación estándar, lo mejor será evitar en lo posible $x_3 = 1$ o al menos mantener este factor en valores pequeños.

Si bien s oscila entre 0 y 158, s^2 estará entre un rango de valores demasiado grande, lo que hará necesaria alguna transformación que suavice estas diferencias. La opción inmediata será pensar en la transformación $\ln s^2$. Sin embargo, al haber dos puntos con $s^2 = 0$ se limita la posibilidad de dicha transformación a menos que se incluya alguna constante que evite el problema con los valores no definidos de la función.

3.2.2. Dispersión de Media

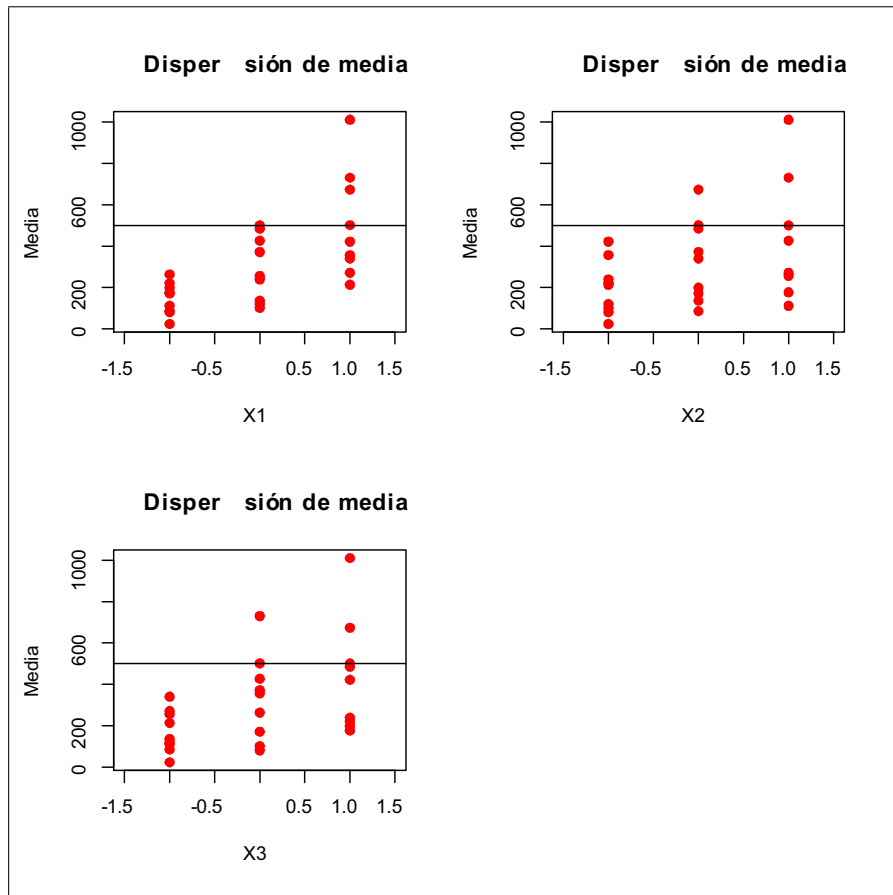


Figura 3.2

No es evidente un nivel de los factores que favorezca particularmente el acercamiento de la respuesta al valor objetivo de 500.

Pues al tener alguno de los 3 factores en su nivel más bajo nos lleva a valores considerablemente más pequeños que el 500 buscado, mientras que en el nivel más alto ocurre exactamente lo opuesto.

Si observamos los datos obtenidos, nos encontramos con que las entradas del diseño que poseen una media de respuestas que se acerca más al valor objetivo tienen algún factor igual a 0.

En el caso de la varianza, el hecho de que un solo factor sea la principal influencia podría resultar en falta de significancia de los demás factores y por ello, en un modelo con capacidades de predicción sumamente limitadas. Se espera que estas primeras impresiones se vean reflejadas en los conjuntos de entradas óptimas a obtener aplicando los diferentes enfoques.

3.3. Enfoque paramétrico

3.3.1. Primera estrategia

Un primer enfoque considera a x_1 , x_2 y x_3 como los factores de control y las réplicas z los factores de ruido, lo anterior acorde a la metodología de diseño robusto.

Se procede a un análisis similar al realizado por Vining y Myers (1990), ajustando primero modelos completos de segundo orden para media y varianza por separado, esto para tener una mejor comprensión del proceso completo.

La superficie de respuesta ajustada para la media de las observaciones fue

$$\hat{Y}_m = 327,6 + 177x_1 + 109,43x_2 + 131,46x_3 + 32x_1^2 - 22,39x_2^2 - 29,05x_3^2 + 66,02x_1x_2 + 75,47x_1x_3 + 43,58x_2x_3. \quad (3.1)$$

La tabla 3.4 resume el correspondiente análisis de varianza. Es importante notar la estadística F para la prueba de falta de ajuste.

Tabla 3.4

Análisis de varianza Y_m

Fuente	gl	SC	CM	F	Prob > F
Modelo	9	1248225.65	138691.739	23.94	< 0.0001
Residual	17	98495.29	5793.84		
$R^2 = 0,927$					
$R^2_{adj} = 0,888$					
$ECM = 5793.8$					

El valor F de 23.94 para el modelo implica que éste es significativo, ya que hay sólo una posibilidad del 0.01 % de que dicho valor sea tan grande debido al ruido. Revisando la significancia de cada factor

Tabla 3.5

Factores significativos

Variable	F	Prob > F
x_1	97.33	< 0.000
x_2	37.20	< 0.000
x_3	53.69	< 0.000
x_1^2	1.06	0.318
x_2^2	0.52	0.481
x_3^2	0.87	0.363
x_1x_2	9.03	0.008
x_1x_3	11.80	0.003
x_2x_3	3.93	0.064

Siguiendo lo realizado en dicho artículo, se probó añadir al modelo de segundo orden los términos *lineal × cuadrático* $(x_i x_j^2)$, sin embargo el ANOVA correspondiente, presentado en la tabla 3.6, hizo claro que no era justificado el uso de dicho modelo, ya que ninguno de los términos añadidos fue de interés y la interpretación se volvía más compleja.

Tabla 3.6

Análisis de varianza Y_m

Fuente	gl	SC	CM	F	Prob > F
Modelo	15	1248225.65	138691.739	23.94	< 0.0001
Residual	11	98495.29	5793.84		
$R^2 = 0,938$					
$R_{adj}^2 = 0,854$					
$ECM = 8954.1$					

Tabla 3.7

Factores significativos

Variable	Prob > F
x_1	0.003
x_2	0.006
x_3	0.024
x_1^2	0.387
x_2^2	0.541
x_3^2	0.431
$x_1 x_2$	0.023
$x_1 x_3$	0.012
$x_2 x_3$	0.111
$x_1^2 x_2$	0.711
$x_1^2 x_3$	0.670
$x_1 x_2^2$	0.989
$x_1 x_3^2$	0.951
$x_2^2 x_3$	0.973
$x_2 x_3^2$	0.222

Revisando la significancia de los factores se considera descartar también los términos cuadráticos. Al ajustar un modelo lineal e incluir las interacciones se tendrá

$$\begin{aligned} \hat{Y}_m = & 314,67 + 177,07x_1 + 109,43x_2 + 131,46x_3 \\ & + 66,03x_1x_2 + 75,47x_1x_3 + 43,58x_2x_3 \end{aligned} \quad (3.2)$$

Con el ANOVA correspondiente

Tabla 3.8

Análisis de varianza

Fuente	gl	SC	CM	F	Prob > F
Modelo	6	1234008.297	205668.049	36.494	< 0.000
Residual	20	112712.639	5635.632		
$R^2 = 0,916$					
$R^2_{adj} = 0,891$					
$ECM = 5635.6$					

El valor F de para el modelo indica que éste es significativo. Viendo cada factor de manera individual

Tabla 3.9

Factores significativos

Variable	F	Prob > F
x_1	100.05	< 0.000
x_2	38.24	< 0.000
x_3	55.20	< 0.000
x_1x_2	9.28	0.006
x_1x_3	12.13	0.002
x_2x_3	4.04	0.058

Se tienen entonces todos los factores significativos incluidos en el modelo con un valor R^2 mayor a 0,9.

Dicho modelo presenta ventajas en cuanto a interpretación y análisis aunque fue descartado por Vining y Myers por los requerimientos de su proceso de optimización. En procedimientos alternativos será considerado.

Nota:

Al ver la magnitud de la respuesta se consideró alguna transformación pertinente de Y_m , pero como puede observarse en la tablas 3.10 y 3.12 las transformaciones $\ln Y_m$ ni raíz cuadrada mejoraron el valor de R^2 .

Tabla 3.10

Análisis de varianza $\ln Y_m$

Fuente	gl	SC	CM	F	Prob > F
Modelo	6	14.967	2.495	24.441	< 0.000
Residual	20	2.041	0.102		
$R^2 = 0.880$					
$R^2_{adj} = 0.844$					
$ECM = 0,10207$					

Tabla 3.11

Factores significativos

Variable	F	Prob > F
x_1	73.661	< 0.000
x_2	28.692	< 0.000
x_3	42.271	< 0.000
x_1x_2	0.370	0.549
x_1x_3	0.569	0.459
x_2x_3	1.084	0.310

Tabla 3.12

Análisis de varianza $\sqrt[2]{Y_m}$

Fuente	gl	SC	CM	F	Prob > F
Modelo	6	899.474	149.912	32.399	< 0.000
Residual	20	92.542	4.627		
$R^2 = 0.907$					
$R^2_{adj} = 0.879$					
$ECM = 4.6271$					

Tabla 3.13

Factores significativos

Variable	F	Prob > F
x_1	98.982	< 0.000
x_2	36.514	< 0.000
x_3	54.587	< 0.000
x_1x_2	1.749	0.201
x_1x_3	2.275	0.147
x_2x_3	0.284	0.600

No se encuentra gran ventaja en las transformaciones de Y_m por lo que se descartan dichos modelos.

Para σ^2 se prefirió modelar inicialmente la desviación estándar pues es una característica importante por sí misma así como fácil de interpretar. El modelo ajustado para esta transformación fue

$$\hat{Y}_s = 34,88 + 11,52x_1 + 15,32x_2 + 29,19x_3 + 4,2x_1^2 - 1,315x_2^2 + 16,77x_3^2 + 7,7x_1x_2 + 5,1x_1x_3 + 14,08x_2x_3. \quad (3.3)$$

La tabla 3.14 muestra el análisis de varianza resultante.

Tabla 3.14

Análisis de varianza Y_s

Fuente	gl	SC	CM	F	Prob > F
Modelo	9	27168.51	3018.72	1.57	0.2017
Residual	17	32651.98	1920.70		
$R^2 = 0,454$					
$R^2_{adj} = 0,165$					
$ECM = 1920.7$					

El valor F para el modelo implica que éste es no significativo, hay una posibilidad del 20.17% de que dicho valor sea tan grande debido al ruido. Se procede a revisar la significancia de cada factor

Tabla 3.15

Factores significativos

Variable	F	Prob > F
x_1	1.25	0.280
x_2	2.20	0.156
x_3	7.99	0.012
x_1^2	0.055	0.817
x_2^2	5.409E-003	0.942
x_3^2	0.88	0.362
x_1x_2	0.37	0.549
x_1x_3	0.16	0.691
x_2x_3	1.24	0.281

Al tener términos no significativos se decidió que la afirmación de un modelo de segundo orden no es fuerte. Ajustándose entonces un modelo de primer orden

$$\hat{Y}_s = 47,99 + 11,53x_1 + 15,32x_2 + 29,19x_3 \quad (3.4)$$

La tabla 3.16 muestra el análisis de varianza correspondiente.

Tabla 3.16

Análisis de varianza

Fuente	gl	SC	CM	F	Prob > F
Modelo	3	21955.24	7318.41	4.45	0.0133
Residual	23	37865.26	1646.32		
$R^2 = 0,367$					
$R^2_{adj} = 0,285$					
$ECM = 1646.3$					

Este modelo es significativo, entonces se revisa la significancia de cada factor

Tabla 3.17

Factores significativos

Variable	F	Prob > F
x_1	1.45	0.240
x_2	2.57	0.123
x_3	9.32	0.006

Incluso ahora los factores x_1 y x_2 siguen siendo no significativos, por lo que se considera ajustar un modelo donde sólo se incluya x_3 , decisión que de algún modo respalda lo visto en el análisis exploratorio, pues para la desviación estándar el factor de mayor influencia sería x_3 . El modelo que surge al considerar sólo tal factor será

$$\hat{Y}_s = 47,99 + 29,19x_3 \quad (3.5)$$

Tabla 3.18

Análisis de varianza Y_s

Fuente	gl	SC	CM	F	Prob > F
Modelo	1	15337.32	15337.32	8.62	0.0070
Residual	25	44483.18	1779.33		
$R^2 = 0,256$					
$R^2_{adj} = 0,227$					
$ECM = 1779.3$					

Nuevamente, este modelo es significativo, así como el único factor presente

Tabla 3.19

Factores significativos

Variable	F	Prob > F
x_3	8.62	0.007

La estadística F resultante fue claramente más fuerte al ir retirando términos del modelo, pero R^2 fue más débil, pasando a 0,367 y luego cayendo hasta 0,256. Sin embargo, el R^2_{adj} refiere cierta inferioridad del modelo cuadrático.

Vining y Myers justificaron el uso del primer modelo argumentando que en la metodología de respuesta dual lo más importante es la habilidad del modelo para predecir, es decir, el modelo a usar será aquel con mayor R^2 (además de ser indispensable para su estrategia de optimización). Aunque mencionan la necesidad de una metodología que tome en cuenta directamente la habilidad de predicción del modelo permitiendo el uso de modelos de primer orden para la varianza. En este trabajo se considerarán alternativas al uso de modelos cuadráticos completos y será posible así no considerar términos no significativos.

Se consideró también la transformación común $\ln \sigma^2$, pero al haber algunas puntos con varianza muestral de 0 es recomendable usar $\ln(\sigma^2 + 1)$ para evitar dicho problema.

$$\hat{Y}_{\ln s^2} = 4,96 + 0,83x_1 + 0,87x_2 + 1,36x_3 + 0,596x_1^2 + 0,390x_2^2 + 1,27x_3^2 - 0,51x_1x_2 - 0,32x_1x_3 + 0,55x_2x_3 \quad (3.6)$$

Tabla 3.20

Análisis de varianza $\ln(Y_s^2 + 1)$

Fuente	gl	SC	CM	F	Prob > F
Modelo	9	80.034	8.893	1.317	0.298
Residual	17	114.754	6.750		
$R^2 = 0.411$					
$ECM = 6,750$					

Nuevamente, sólo x_3 resulta significativo. Al incluir solamente dicho término en el modelo

$$\hat{Y}_{\ln s^2} = 6,469 + 1,358x_3 \quad (3.7)$$

Tabla 3.21

Análisis de varianza $\ln(Y_s^2 + 1)$

Fuente	df	SC	CM	F	Prob > F
Modelo	1	33.189	33.199	5.134	0.032
Residual	25	161.599	6.464		
$R^2 = 0.170$					
$ECM = 6,464$					

3.3.2. Optimización

Teniendo un modelo para cada una de las respuestas de interés, resta encontrar la combinación óptima de los factores de control que nos lleven más cerca del valor objetivo con la menor variabilidad posible.

Recordando que el objetivo del enfoque de respuesta dual es optimizar una respuesta prioritaria sujeta a una restricción apropiada en el valor de una respuesta secundaria.

La decisión de hacer una respuesta primaria o secundaria depende del fin último del experimento, a considerar las tres situaciones básicas de Taguchi del modo siguiente:

- Mantener μ en el valor objetivo al tiempo que se minimiza σ^2 .
- Hacer μ tan grande como sea posible controlando σ^2 .
- Hacer μ tan pequeño como sea posible controlando σ^2 .

Las optimizaciones presentadas se hicieron con el software R, excepto por la función de deseabilidad, cuya optimización fue realizada con Design Expert 6.0.2.

Por multiplicadores de Lagrange

Se considera un escenario de minimización de varianza, manteniendo la media en el valor objetivo. Es decir

$$\text{Min } Y_s \quad \text{s.a.} \quad Y_m = 500. \quad (3.8)$$

Lo que se traduce a la función auxiliar

$$L(X) = \hat{Y}_s - \lambda (\hat{Y}_m - T) \quad (3.9)$$

donde el conjunto de entradas óptimas $x^* \in R$, una esfera (hiperesfera en el caso de tener más de tres factores) de radio 1 y el valor objetivo $T = 500$.

Para optimizarla, el uso de modelos completos es indispensable, lo que representa una limitante en el estudio de Vining y Myers (justifican dicha inclusión al elevar el valor de R^2 , situación engañosa pues de seguir añadiendo términos dicho valor seguiría en aumento pero no tendríamos un mejor modelo).

Los modelos a utilizar, obtenidos por OLS serán los presentados en 3.1 y 3.3. La tabla 3.22 indica que el conjunto de entradas óptimas para $\hat{\sigma} = 45,18$, será $x_1 = 1$, $x_2 = 0,124$, $x_3 = -0,261$ (conjunto similar a lo visto en el análisis exploratorio). Dichos valores permitirán alcanzar un valor de la característica de calidad de 500,692.

Tabla 3.22

Entradas óptimas de varianza mínima

λ	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$
-3	-0,525	-1	-1	69,027	21,736
-2	-0,511	-1	-1	69,059	21,657
-1	-0,473	-1	-1	69,214	21,449
-0,75	-0,449	-1	-1	69,354	21,329
-0,5	-0,408	-1	-1	69,690	21,125
-0,25	-0,311	-1	-1	70,896	20,711
0	-0,199	-1	-0,420	139,177	14,689
0,05	0,411	1	-0,309	207,322	16,811
0,1	1	-1	-0,098	322,392	24,437
0,12	1	0,124	-0,261	500,692	45,181

Comparando con el conjunto obtenido por V-M se encuentra una ligera mejora en términos de $\hat{\sigma}$.

Tabla 3.23

Entradas óptimas del estudio Vining-Myers

λ	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$
-0,25	0,614	0,228	0,100	500,0	51,778

Por función de Deseabilidad

Considerando los modelos establecidos al inicio del capítulo y la función de deseabilidad presentada en capítulos anteriores se tiene una alternativa a la optimización previa con la ventaja de poder omitir los términos no significativos.

Dando seguimiento al trabajo de Vining-Myers se utilizan primero los dos modelos cuadráticos completos. El conjunto de entradas óptimas (aquellas con deseabilidad más cercana a 1) correspondiente a la situación *Target is Best* obtenido con la ecuación 1.21, con límites superior e inferior establecidos como la mayor y menor observación respectivamente, se presenta en la tabla 3.24.

Tabla 3.24

Entradas óptimas con Y_s mínima

Escenario	D	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$
$Y_m = 500$	0.846	1.00	0.10	-0.25	499.017	44.991

La tabla 3.25 muestra lo obtenido fijando el valor deseado en σ y maximizando la media, como se establecía en 1.22, aunque los resultados son poco satisfactorios por lo que no se profundizará en tal perspectiva.

Tabla 3.25

Entradas óptimas para Y_m máximo

Escenario	D	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$
$Y_s = 60$	0.776	1.00	1.00	-0.22	633.882	62.619
$Y_s = 75$	0.830	1.00	1.00	0.05	703.243	75.024
$Y_s = 90$	0.869	1.00	1.00	0.33	768.788	90.027

Al revisar el conjunto de entradas óptimas se puede apreciar que x_3 se coloca lejos de su nivel máximo, lo esperado desde el análisis exploratorio, mientras que x_1 se mantiene en su máximo, al igual que en la optimización por multiplicadores de Lagrange.

Usando los modelos que contienen sólo términos significativos y repitiendo el análisis, se obtiene lo presentado en la Tabla 3.26.

Tabla 3.26

Entradas óptimas

Escenario	D	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$
$MinY_s$ sa $Y_m = 500$	0.906	1.00	1.00	-0.67	499.016	28.407
$MaxY_m$ sa $Y_s = 60$	0.871	1.00	1.00	0.60	817.428	65.508
$MaxY_m$ sa $Y_s = 75$	0.942	1.00	1.00	0.93	898.959	75.024
$MaxY_m$ sa $Y_s = 90$	0.882	1.00	1.00	1.00	917.617	77.182

Sólo el primer escenario lleva a resultados aceptables, permitiendo reducir la desviación estándar a casi la mitad, al tiempo que la media se mantiene en el valor objetivo. Sin embargo, dicho modelo presenta desventajas como residuales demasiado grandes para la varianza y un bajo R^2 .

Es posible cambiar límites superiores (LS) o inferiores (LI) a Y_m y Y_s observando así cómo se modifican los correspondientes conjuntos de entradas óptimas.

Tabla 3.27

Entradas óptimas con Y_s restringido

Modelos completos							
D	x_1	x_2	x_3	LI	LS	$\hat{\mu}$	$\hat{\sigma}$
0.637	1.00	-0.59	-0.12	0	60	402.595	33.6183
0.710	1.00	-0.05	-0.23	0	75	479.059	42.6057
0.771	1.00	0.08	-0.24	0	80	499.015	44.9946
Modelo sólo con términos significativos							
D	x_1	x_2	x_3	LI	LS	$\hat{\mu}$	$\hat{\sigma}$
0.786	1.00	1.00	-0.67	0	60	499.137	28.4204
0.836	1.00	1.00	-0.67	0	75	499.137	28.4204
0.867	1.00	1.00	-0.67	0	90	499.137	28.4204

Tabla 3.28

Entradas óptimas con mínimo Y_s y Y_m restringido de 450 a 550

Modelos completos								
Escenario	D	x_1	x_2	x_3	LI	LS	$\hat{\mu}$	$\hat{\sigma}$
$Y_s = 60$	1.000	0.89	0.18	0.41	450	550	588.32	60.019
$Y_s = 75$	1.000	0.96	0.89	0.92	450	550	863.01	74.986
$Y_s = 90$	0.926	0.14	0.92	1.00	450	550	630.283	77.184
Modelos con términos significativos								
Escenario	D	x_1	x_2	x_3	LI	LS	$\hat{\mu}$	$\hat{\sigma}$
$Y_s = 60$	1.000	1.00	-0.14	0.41	450	550	550.224	60.011
$Y_s = 75$	1.000	0.96	0.96	0.93	450	550	876.541	75.016
$Y_s = 90$	0.926	0.81	0.58	1.00	450	550	769.603	77.184

Nuevamente esta alternativa no da resultados aceptables pues se aleja excesivamente del valor objetivo.

Por función de pérdida o ECM

Otra opción consiste en minimizar la función de pérdida más común o ECM, expresada como

$$ECM = \left(\hat{Y}_s^2(x) \right)^2 + \left(\hat{Y}_m(x) - T \right)^2 \quad (3.10)$$

s.a. $X'X \leq 1$

donde T es el valor objetivo para la respuesta, igual a 500. Es recomendable el uso de modelos cuadráticos completos como en el caso de multiplicadores de Lagrange.

Realizando la optimización correspondiente, el conjunto de entradas óptimas con sus respectivos valores predichos se presentan en la tabla 3.29.

Tabla 3.29

Entradas óptimas con ECM mínimo

x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
1.000	0.074	-0.252	494.681	44.460	2005.027

Incluyendo sólo los términos significativos en ambos modelos se tiene

Tabla 3.30

Entradas óptimas con ECM mínimo

x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
0.713	0.599	-0.164	500	4.193e-08	6.814e-15

Coppelan-Nelson (1996) reformularon el problema de optimización por ECM estableciendo una cota Δ entre el valor objetivo de la característica de calidad y la respuesta del proceso, es decir:

$$\text{Min } \hat{\sigma}(x) \quad \text{sa} \quad (\hat{\mu}(x) - T)^2 \leq \Delta^2. \quad (3.11)$$

Usando los modelos cuadráticos presentados en 3.1 y 3.3 se tendrá la optimización presentada en la tabla 3.31.

Tabla 3.31

Entradas óptimas con ECM mínimo acotado

Δ	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
25	0.358	0.219	0.363	480.323	58.014	3752.8
50	0.287	0.196	0.300	451.313	53.463	5228.7

Esta optimización del ECM se rescatará en enfoques posteriores pues hace un balance entre lejanía del valor objetivo y variabilidad haciéndolo más flexible y capaz de cumplir con los objetivos.

Por razón de señal a ruido

Para combinar la información disponible referente a media y varianza en cada entrada de los factores de control y estudiar su comportamiento se puede emplear la llamada razón de señal a ruido (SNR por sus siglas en inglés).

Entre las opciones posibles, para el caso de interés *Target is best* con Y_m en el valor objetivo 500 es conveniente usar la razón:

$$SNR = 10 \log \left(\frac{\bar{y}_i^2}{s_i^2 + 1} \right) \quad (3.12)$$

donde s_i^2 es la varianza muestral de las corridas en la entrada i –ésima de los factores de control. Se buscará el conjunto de entradas de X que maximicen dicho SNR.

Al graficar el valor del SNR contra los niveles de los factores de control (Figura 3.3) se buscan los factores que tengan influencia sobre él, de modo similar a lo realizado en el análisis exploratorio.

Tabla 3.32

SNR en cada factor y nivel

Nivel	x_1	x_2	x_3
-1	5.672	3.579	4.576
0	5.379	4.437	4.123
1	5.812	3.113	4.039

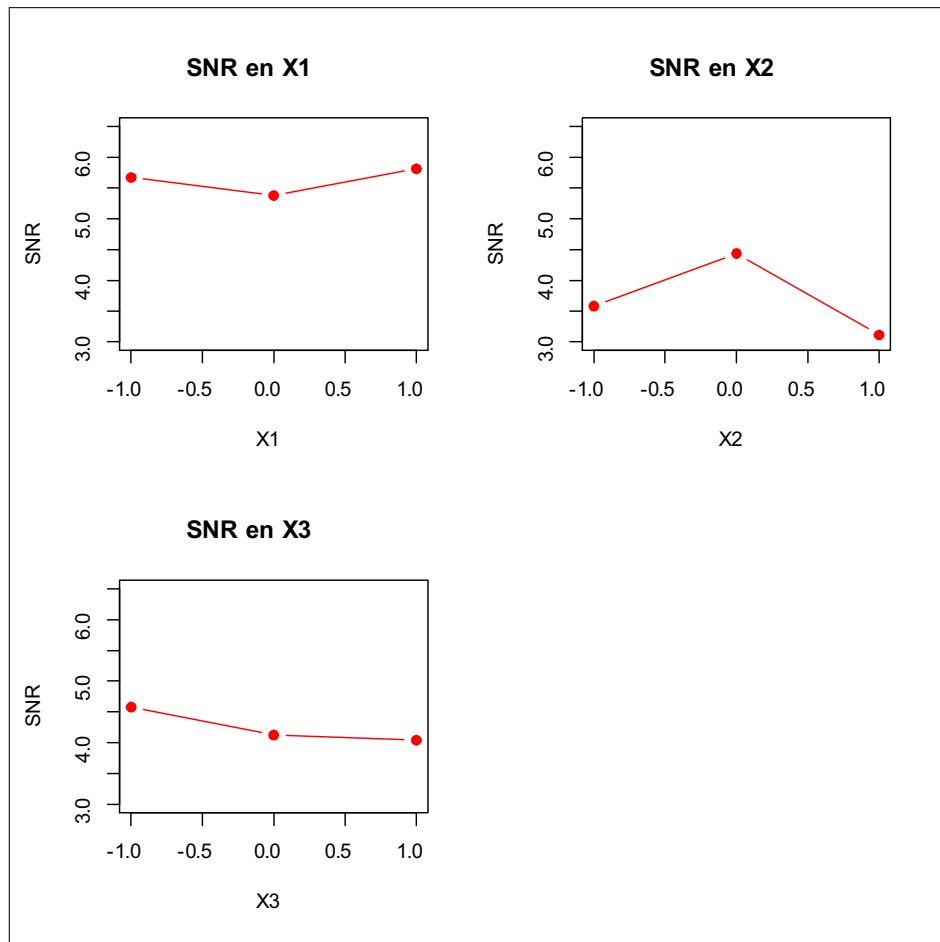


Figura 3.3

No hay un factor que tenga influencia claramente en el SNR, pero ya que buscaremos maximizar, se observa que los mayores valores se alcanzan cerca del máximo para x_1 , cercanos al valor intermedio de x_2 alrededor del mínimo para x_3 , aunque la diferencia no es contundente.

Tabla 3.33

Media en cada factor y nivel

	x_1	x_2	x_3
-1	148.333	197.778	173.519
0	293.333	329.593	334.037
1	502.333	416.629	436.444

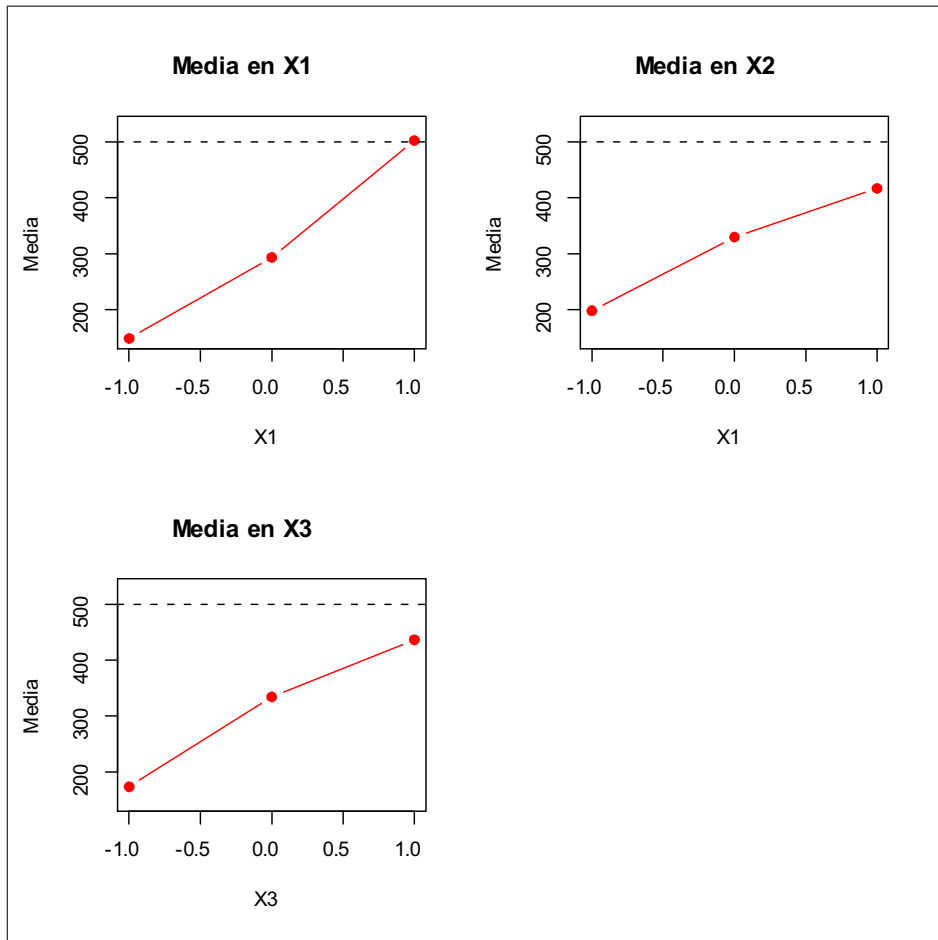


Figura 3.4

Pareciera que el factor x_1 es el más influyente, acercándose claramente al valor objetivo en su punto máximo. Es interesante notar que en las entradas óptimas obtenidas hasta el momento con las diferentes propuestas de optimización incluyen $x_1 = 1$.

Ya calculado el valor de SNR correspondiente a cada punto del diseño se procede a ajustar un modelo cuadrático a tal conjunto de valores

$$\begin{aligned}\hat{Y}_{snr} = & -28,56 - 2,03x_1 + 0,29x_2 + 1,64x_3 + 3,15x_1^2 \\ & + 3,33x_2^2 + 7,06x_3^2 - 1,73x_1x_2 - 0,79x_1x_3 \\ & + 3,24x_2x_3\end{aligned}\quad (3.13)$$

Aunque es importante notar que su $R^2 = 0,24$.

Y maximizando dicha función sujeta a las restricciones de cada factor se obtendrá el conjunto de entradas óptimas que se muestra en la Tabla 3.34.

Tabla 3.34

\hat{Y}_{snr}	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
28.98	0.34	0.10	-0.12	384.647	37.425	14707

Obteniendo así una media que puede considerarse demasiado alejada del valor objetivo y en consecuencia un ECM muy grande. Vale la pena mencionar que el SNR depende directamente de la varianza, y del mismo modo su modelo. Por lo cual, con las deficiencias vistas en el modelo que la describe, no puede esperarse que el SNR proporcione mejoras pues parte de una característica difícil de modelar y su modelo apenas alcanza a capturar parte de esa variabilidad, dando como resultado otro modelo deficiente y con poca capacidad de predicción.

3.3.3. Otra estrategia de optimización

Hasta el momento se ha trabajado el problema de respuesta dual ajustando un modelo a la media y por separado uno a la varianza del proceso, para posteriormente fijar uno y maximizar (o minimizar) el otro. Pero hay otras maneras de incluir el efecto de las réplicas (o la ausencia de ellas) en el modelo, como la que se presenta a continuación.

Ahora se considerará un enfoque ligeramente distinto, donde se ajustará un único modelo a la respuesta del proceso, de modo que contenga las variables tanto de control (x_1, x_2, x_3) como ruido (las réplicas). A partir de él se obtendrán expresiones para la media y varianza, tomando la esperanza y variabilidad del modelo y reemplazando los coeficientes con aquellos estimados para el modelo ajustado a la respuesta. Lo anterior siguiendo lo presentado por Welch et al. (1990) en su propuesta del arreglo combinado.

Considerando que tenemos ausencia de réplicas y en su lugar una variable de ruido z que tomará valores $-1, 0, 1$ acorde al número de réplica.

El modelo combinado será

$$Y_{xz} = \gamma_0 + x^t\beta + x^tBx + z^t\delta + x^tCz + \varepsilon \quad (3.14)$$

donde las matrices involucradas ya fueron descritas para la ecuación 1.16. Realizando el ajuste, se tendrá el ANOVA presentado en la tabla 3.35 y el modelo que se presenta a continuación

$$\begin{aligned}\hat{Y}_{xz} = & 327,63 + 177x_1 + 109,43x_2 + 131,46x_3 - 4,15z \\ & + 32x_1^2 - 22,39x_2^2 - 29,06x_3^2 + 66,03x_1x_2 + 75,47x_1x_3 \\ & + 11,36x_1z + 43,58x_2x_3 - 1,11x_2z - 15,97x_3z\end{aligned}\quad (3.15)$$

Tabla 3.35

Análisis de varianza

Fuente	gl	F	Prob > F
Modelo	13	36.93	0.0000
Residual	67		
$R^2 = 0.878$			

Que tiene como no significativos a todos los términos cuadráticos así como a aquellos que incluyan al factor z . En el análisis conservaremos todos los términos de modo similar que en enfoques previos. Ahora, los modelos para media y varianza se obtienen las expresiones con la notación de 1.16.

$$\hat{Y}_m = E(Y_{xz}) = \gamma_0 + x^t \beta + x^t Bx \quad (3.16)$$

$$\hat{Y}_s = Var(Y_{xz}) = (\delta + x^t C)^t V (\delta + x^t C) + \sigma^2 \quad (3.17)$$

Con los coeficientes ya estimados

$$\begin{aligned} \hat{Y}_m = & 327,63 + 177x_1 + 109,43x_2 + 131,46x_3 + 32x_1^2 - 22,39x_2^2 \\ & - 29,06x_3^2 + 66,03x_1x_2 + 75,47x_1x_3 + 43,58x_2x_3 \end{aligned} \quad (3.18)$$

$$\hat{Y}_s = (-4,15z + 11,36x_1z - 1,11x_2z - 15,97x_3z)^2 + 7831,54 \quad (3.19)$$

Entonces, se busca minimizar \hat{Y}_s sujeto a que $\hat{Y}_m = 500$. Usando el criterio de ECM se llega al conjunto de entradas óptimas presentado en la tabla 3.36.

Tabla 3.36

Entradas óptimas con segunda estrategia

Escenario	x_1	x_2	x_3	\hat{Y}_m	\hat{Y}_s	ECM
$Min \hat{Y}_s$ sa $\hat{Y}_m = 500$	0.548	0.333	0.107	500	7831.54	7831.54

En ausencia de réplicas, es posible usar una opción alternativa para la segunda respuesta. Ya ajustado Y_m se calcula el logaritmo de la diferencia absoluta entre lo predicho y lo observado.

$$Y_2 = \log |Y_m - \hat{Y}_m| \quad (3.20)$$

donde Y_m son los resultados experimentales para cada tratamiento.

Al hacer el ajuste correspondiente usando los datos predichos por 3.15 se convierte en

$$\begin{aligned} \hat{Y}_2 = & 3,21 + 0,335x_1 + 0,20x_2 + 0,12x_3 - 0,15x_1^2 + 0,63x_2^2 \\ & + 0,29x_3^2 + 0,08x_1x_2 - 0,08x_1x_3 + 0,25x_2x_3 \end{aligned} \quad (3.21)$$

Nuevamente, se busca minimizar \hat{Y}_s sujeto a mantener \hat{Y}_m en un valor objetivo de 500. Realizando dicha

optimización con el criterio se ECM se obtienen las entradas óptimas presentadas en la tabla 3.37.

Tabla 3.37

Entradas óptimas con 3.19

Escenario	x_1	x_2	x_3	\hat{Y}_m	\hat{Y}_2	ECM
$Min\hat{Y}_2$ <i>sa</i> $\hat{Y}_m = 500$	0.970	-0.215	0.046	500.001	3.363	11.311

Otra opción viable es modelar

$$Y_2 = \log \left(\left(Y_m - \hat{Y}_m \right)^2 \right) \quad (3.22)$$

que nos llevará al modelo

$$\begin{aligned} \hat{Y}_2 = & 6,424 + 0,669x_1 + 0,408x_2 + 0,246x_3 \\ & -0,301x_1^2 + 1,262x_2^2 + 0,583x_3^2 \\ & +0,152x_1x_2 - 0,151x_1x_3 + 0,495x_2x_3. \end{aligned} \quad (3.23)$$

Y si bien sólo los términos lineales son significativos y no nos sorprende su bajo valor de R^2 procedemos a optimizar.

Tabla 3.38

Entradas óptimas con 3.21

Escenario	x_1	x_2	x_3	\hat{Y}_m	\hat{Y}_2	ECM
$Min\hat{Y}_2$ <i>sa</i> $\hat{Y}_m = 500$	1	-0.111	-0.082	499.995	6.746	45.508

3.3.4. WLS

Al final del trabajo de Vining y Myers (1990) se hace la recomendación de incluir un análisis con WLS, para poder tomar en cuenta la importante heterogeneidad de las varianzas al momento de crear un modelo, para que la capacidad de predicción del mismo sea mayor.

Siguiendo tal enfoque, nuevamente se tendría para la respuesta del proceso

$$\hat{Y}_m = X^t \hat{\beta}$$

pero el vector de parámetros sería estimado como

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} Y_m \quad (3.24)$$

donde V es la matriz diagonal de varianzas y covarianzas de las respuestas obtenidas en el diseño.

Hay dos formas inmediatas de estimar esta matriz, la primera es usar los estimadores puntuales de la varianza (es posible ya que tenemos réplicas) como los elementos en la diagonal. Haciendo lo anterior tendremos

$$\hat{Y}_m = 366,12 + 166,21x_1 + 115,45x_2 + 146,34x_3. \quad (3.25)$$

Y el modelo anteriormente usado para Y_s presentado en 3.3.

Con estos dos modelos se procede a una optimización en el escenario *Target is Best* minimizando Y_s sujeto a mantener 3.25 en un valor objetivo de 500.

Tabla 3.39

Entradas óptimas WLS con 3.25

Escenario	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
$Min\hat{Y}_s$ <i>sa</i> $\hat{Y}_m = 500$	0.821	1.000	-0.842	494.823	39.428	1581.356

Repitiendo el mismo análisis con modelos lineales (es decir, omitiendo los términos cuadráticos no significativos en el modelo) se tendrán los resultados de la tabla 3.40.

Tabla 3.40

Entradas óptimas WLS modelos lineales

Escenario	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
$Min\hat{Y}_s$ <i>sa</i> $\hat{Y}_m = 500$	1.000	0.951	-1.000	495.783	31.779	1027.715

Una segunda manera de realizar el WLS será utilizando los valores predichos por el modelo 3.3 como pesos en la diagonal. de V . Se tendrá entonces

$$\hat{Y}_m = 283,36 + 127,22x_1 + 82,71x_2 + 107,90x_3. \quad (3.26)$$

Con las entradas óptimas presentadas en la siguiente tabla.

Tabla 3.41

Entradas óptimas WLS con 3.26

Escenario	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
$MinY_s$ <i>sa</i> $Y_m = 500$	1.000	1.000	-0.173	474.589	64.427	4796.609

Y repitiendo el proceso sin incluir los términos nos significativos.

Tabla 3.42

Entradas óptimas WLS sólo terminos significativos

Escenario	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
$MinY_s$ <i>sa</i> $Y_m = 500$	1.000	1.000	-0.086	483.984	59.204	3761.597

Finalmente, una tercera opción para la matriz V será definir sus elementos diagonales como

$$V_{ii} = \frac{\bar{y}_i}{\hat{\sigma}_i^2} \quad (3.27)$$

y repetir el análisis.

Tabla 3.43

Entradas óptimas WLS usando 3.27

Escenario	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
$MinY_s$ sa $Y_m = 500$	0.753	1.000	-0.667	494.164	41.780	1779.642

Ahora sólo con términos significativos

Tabla 3.44

Entradas óptimas WLS usando 3.27

Escenario	x_1	x_2	x_3	$\hat{\mu}$	σ	ECM
$MinY_s$ sa $Y_m = 500$	1.000	1.000	-0.888	494.103	35.792	1315.868

3.3.5. Método Gráfico

Es conveniente analizar las curvas de nivel correspondientes a los tres factores de control. Usando el modelo 3.2 para la media y 3.3 para la varianza se considerarán las restricciones fijando el valor de x_2 presentadas en la tabla 3.45. Para la creación de tales curvas se usó el paquete Design Expert 6.0.2.

Tabla 3.45

Curvas de nivel

\hat{Y}_m		\hat{Y}_s		
LI	LS	LI	LS	X_2
490	500	45	50	1
490	500	30	40	-1
490	500	43	45	-0.75
490	500	30	40	1
490	500	45	50	0
490	500	45	50	0.3

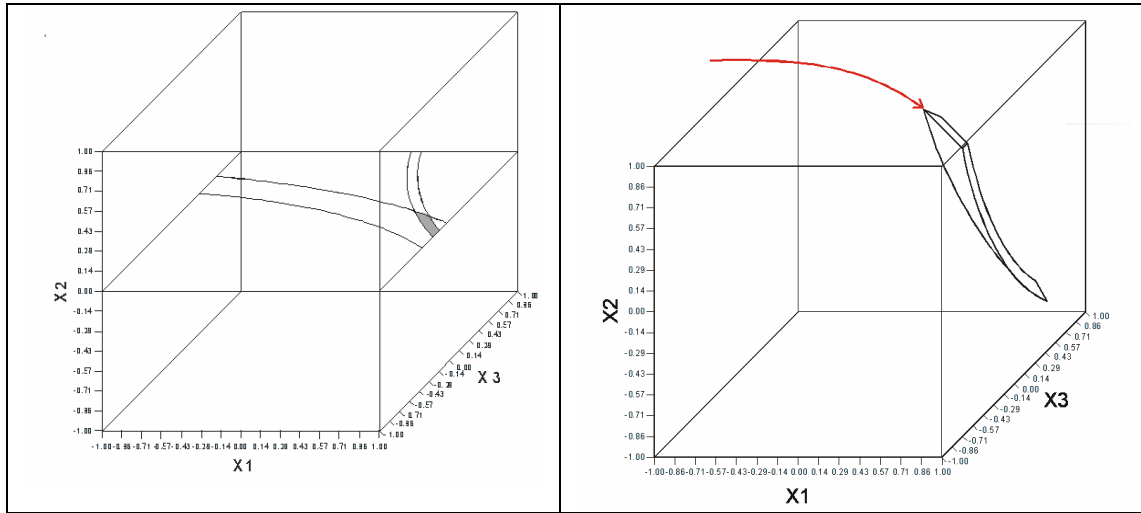


Figura 3.5

Ya delimitada la región en la cual se buscarán los valores óptimos de los factores de control, se analiza la zona donde la media se mantiene entre 490 y 500 y la desviación estándar dentro de lo exigido manteniendo fijo el valor de x_2 . Al sobreponer estas curvas se obtendrá la región de valores de los factores restantes (x_1 y x_3) que producen valores óptimos comunes para el promedio y la dispersión.

En la figura 3.6 se mantiene $x_2 = 1$ y s entre 45 y 50, posteriormente se fija $x_2 = -1$ manteniendo s entre 30 y 40.

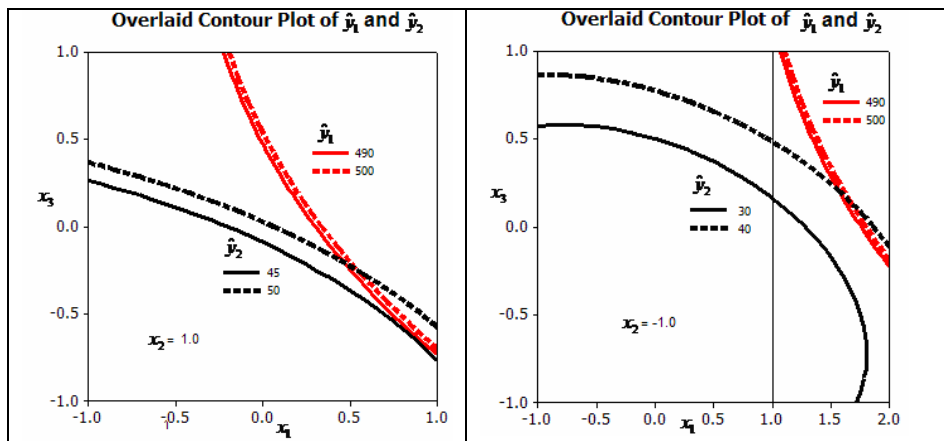


Figura 3.6

La figura 3.7 presenta las curvas de nivel cuando $x_2 = -0,75$ con s entre 43 y 45. Del lado derecho $x_2 = 1$ para s entre 30 y 40.

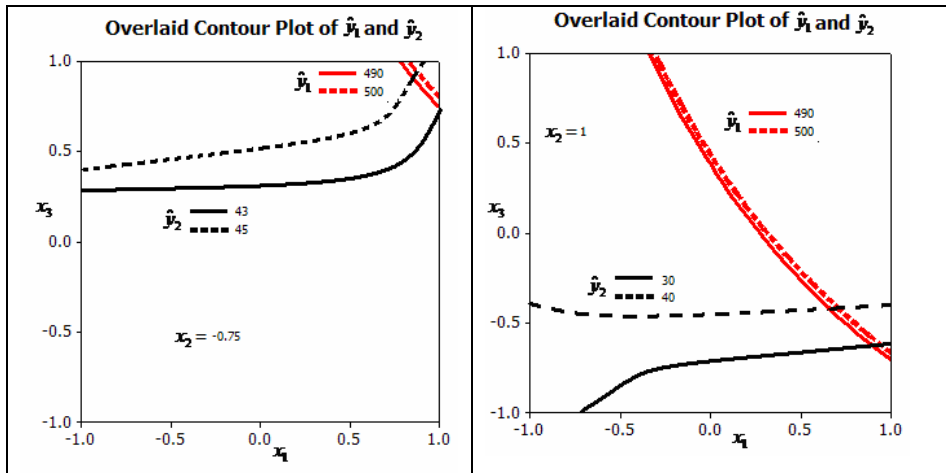


Figura 3.7

Finalmente, la figura 3.8 explora $x_2 = 0$ y $x_2 = 0,3$ manteniendo la desviación estándar entre 45 y 50.

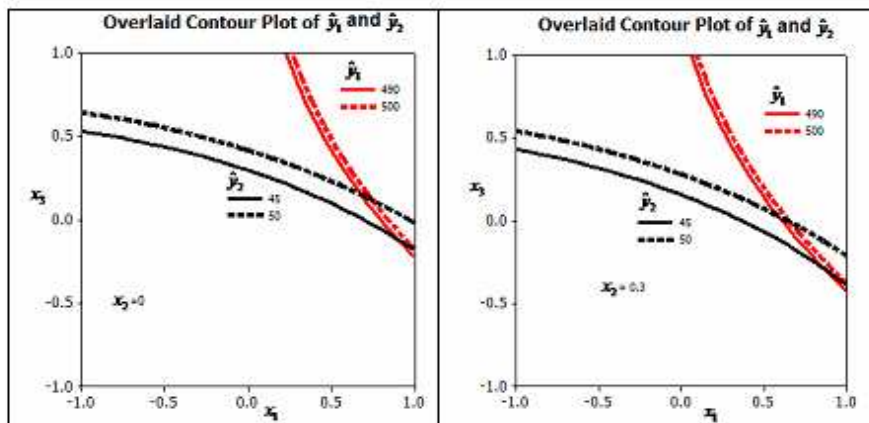


Figura 3.8

3.4. Enfoque con elementos no paramétricos

Hasta el momento se han explorado diferentes modelos dentro del terreno de lo paramétrico. Sin embargo, basta una mirada rápida a los errores al predecir la varianza del proceso para notar la insuficiencia de aquellos. Ante un valor realmente pobre de R^2 surge la necesidad de un modelo que represente de manera más adecuada la variabilidad.

Una de las mejores opciones para capturar esta gran variación en el proceso es el uso de técnicas no paramétricas.

Se procederá ahora con un enfoque semiparamétrico con un kernel para predecir la varianza que luego será incluida en un WLS para predecir la media y optimizando por ECM de modo similar a la sección anterior. Posteriormente, de modo completamente no paramétrico aplicando regresión no paramétrica tanto a media y varianza para luego optimizar su ECM.

Recordando lo expuesto en el Capítulo 2, en el punto x_0 con la función kernel escogida $K(\cdot)$ y ancho de banda b , teniendo 3 factores con 3 posibles niveles, se consideraron productos kernel para la región de diseño cuboidal y el mismo kernel y ancho de banda para cada factor. Usando el estimador Gasser-Müller para una función (suave) definido en 2.12 y 2.13.

$$\hat{m}(x^*) = \sum_{i_3=1}^3 \sum_{i_2=1}^3 \sum_{i_1=1}^3 w_{i_1, i_2, \dots, i_k} y_{i_1, \dots, i_k} \quad (3.28)$$

Con los pesos definidos como

$$w_{i_1, i_2, i_3} = \frac{1}{b^3} \left[\int_{I_{i_1}} K_1 \left(\frac{x_1^* - u_1}{b} \right) du_1 \right] \left[\int_{I_{i_2}} K_2 \left(\frac{x_2^* - u_2}{b} \right) du_2 \right] \left[\int_{I_{i_3}} K_3 \left(\frac{x_3^* - u_3}{b} \right) du_3 \right] \quad (3.29)$$

Con el factor x_j teniendo 3 niveles para $j = 1, 2, 3$, igualmente espaciados, redefinidos como 0, 1/2, 1. Si (i_1, i_2, i_3) , donde cada i_j indexa los puntos de diseño e $I_{i_j} = [r_{i_j-1}, r_{i_j}]$ serán los intervalos particionando los niveles del factor x_j . Donde $\{r_i : i = 0, 1, \dots, n\}$ es una secuencia de interpolación de los puntos del diseño trasformada de modo que $0 = x_1 < \dots < x_n = 1$ y

$$r_0 < x_1 < r_1 < x_2 < \dots < x_{n-1} < r_{n-1} < x_n < r_n.$$

cuidando que los rangos de integración sean apropiados para los niveles de los factores usados para la respuesta.

En este caso, la secuencia de interpolación tomada será

$$r_0 = -1/8, r_1 = 1/4, r_2 = 3/4, r_3 = 9/8$$

$$I_1 = (-0,125, 0,25)$$

$$I_2 = (0,25, 0,75)$$

$$I_3 = (0,75, 1,125).$$

3.4.1. Enfoque semiparamétrico

Para el enfoque semiparamétrico, se usará regresión no paramétrica para estimar la varianza como función de los 3 factores, incorporándolos en los elementos diagonales de V en 3.24 y entonces ajustar un modelo paramétrico lineal para la respuesta usando WLS,

En esta primera inclusión de herramientas no paramétricas motivada por los bajos valores de R^2 al modelar la respuesta con su contraparte paramétrica es importante considerar que el estimador Gasser-Müller requiere que la función a modelar sea suave.

Ya que se usará para estimar la varianza del proceso y ésta no lo es, como indican sus importantes diferencias en magnitud (oscila 0 y 10000 según vimos en el análisis exploratorio), la primera acción será una trasformación que la suavice.

La transformación más natural es el logaritmo de la varianza (más suave incluso que la desviación estándar usada por V-M que hemos modelado hasta el momento por su facilidad de interpretación). Así, para el i -ésimo punto de diseño se tomará como la variabilidad

$$Y_{s^2} = \log s_i^2 \quad (3.30)$$

Sin embargo, en algunos puntos muestrales s_i^2 es 0, por lo cual, para evitar problemas posteriores se le agrega una constante. Se tiene así la respuesta correspondiente a la varianza

$$Y_{s^2} = \log (s_i^2 + 2) \quad (3.31)$$

Siguiendo lo presentado por Vining y Bohn (1998), se elige un kernel óptimo, en particular en este trabajo se usará el que presentan como ejemplo y que ya fue explorado en el capítulo 2.

$$K(x_j) = \begin{cases} \frac{35}{32} (1 - 3x_j^2 + 3x_j^4 - x_j^6), & -1 \leq x_j \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Hecho lo anterior, se usa la regresión no paramétrica para estimar la varianza en cada punto de diseño.

Los pesos estarán definidos entonces por la expresión 3.29 redefiniendo previamente los niveles de cada factor como 0, 1/2, 1. Recordando siempre que $(x_0 - b, x_0 + b)$ debe contener al intervalo determinado por los puntos de la secuencia de interpolación.

Siguiendo el artículo de Vining y Bohn, se tomará un ancho de banda global. Pero como especifican, los valores menores a 0,25 o mayores a 0,75 incluirían datos sólo del nivel más cercano o de todos los niveles, respectivamente, por lo que lo más recomendable es tomar el valor de b entre 0,25 y 0,75. En el análisis se consideran tres casos, $b = 0,3, 0,5$ y $0,7$.

Tomando cada uno de los 27 puntos del diseño y usando como respuesta la varianza muestral ya transformada, se obtiene varianza estimada con 3.28. Colocando luego su inverso como elementos diagonales de V .

$$V_{ii} = \frac{1}{\hat{s}_i^2} \quad (3.32)$$

Y se estima el vector de parámetros como se indicó en 3.24. Con tales parámetros se obtiene un modelo paramétrico para la media del proceso

$$\hat{Y}_m = X^t \hat{\beta}.$$

Finalmente, se optimiza su ECM restringiéndose al cubo definido por el diseño.

$$ECM = \hat{Y}_{s^2}^2 + \left(\hat{Y}_m - 500 \right)^2 \quad (3.33)$$

Obteniendo así el conjunto de entradas óptimas expresadas en la métrica tradicional.

El proceso anterior se realizó usando el software R aplicando tal procedimiento al conjunto de datos descrito al inicio de este capítulo y se repitió para los anchos de banda preestablecidos. Los resultados se

presentan en la tabla 3.46.

Tabla 3.46

Entradas óptimas por método Semiparamétrico

b	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
0,3	0.447	0.467	0.474	500.023	1.553	2.412
0,5	0.482	0.475	0.518	499.867	4.605	21.223
0,7	0.481	0.397	0.574	498.806	11.265	128.329

3.4.2. Enfoque completamente no paramétrico

En el enfoque completamente no paramétrico, se usó regresión no paramétrica tanto para la varianza como para la respuesta. Con los mismos estimadores kernel y anchos de banda que en el caso anterior.

Es decir, tomando cada uno de los puntos del diseño y usando como respuestas los valores observados para la varianza muestral, ya transformada, se aplicó la ecuación 3.28 para obtener un modelo para \hat{Y}_{s^2} . Luego usando como respuesta los valores de la media muestral, se aplicó nuevamente 3.28 obteniendo un modelo independiente para \hat{Y}_m para este conjunto de datos.

Nuevamente se tomó un ancho de banda global con 3 posibles valores entre 0,25 y 0,75 y se procedió a optimizar el ECM correspondiente al escenario de minimización de dispersión con la media en un valor objetivo $T = 500$. Los resultados obtenidos se presentan en la tabla 3.47.

Tabla 3.47

Entradas óptimas por método No Paramétrico

b	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
0,3	0.750	0.725	0.521	496.303	19.996	413.509
0,5	0.713	0.657	0.617	495.165	21.557	488.075
0,7	0.777	0.725	0.719	477.129	24.167	1107.091

De lo cual puede observarse una variabilidad mucho menor al tiempo que la característica de calidad se mantiene a una corta distancia del valor objetivo, situación ideal del análisis. Y si bien, ya no se tienen dos modelos la calidad de la predicción para la respuesta secundaria aumenta de manera considerable.

Capítulo 4

Conclusiones

En todo proceso industrial o de manufactura se tendrán restricciones, así como factores difíciles de controlar, siendo necesaria la intervención en dicho proceso productivo para lograr los mayores beneficios a menores costos, siempre pensando en la máxima calidad del producto.

A lo largo del presente trabajo se abordó un mismo problema desde diferentes enfoques, con un mismo fin: lograr el máximo aprovechamiento de los recursos. Tal problema fue llevar los valores de los factores controlables a objetivos preestablecidos, pero evitando caer en el error común de ignorar la importancia de la variabilidad en los resultados. Se recurrió así a la metodología de superficie de respuesta, considerando tanto media como variabilidad características de interés y se añadieron elementos del diseño robusto para que nuestro proceso no se viera severamente afectado por los factores de ruido (en este caso, las réplicas).

Cada método revisado presenta ventajas así como limitaciones y podría responder a necesidades específicas, pues se dispone de un amplio abanico de opciones tanto en planteamiento como optimización.

Desde la idea inicial de Box y Draper, los modelos cuadráticos y multiplicadores de Lagrange de Vining-Myers, la función de pérdida de Lin-Tu, continuada en la idea de Coppeland-Nelson y por supuesto el SNR de Taguchi, creador de la metodología de diseño robusto, que ante una varianza difícil de describir resultó insuficiente y alejado de las exigencias del problema. También el modelo combinado de Welch fue revisado.

En todos ellos fue evidente que los modelos para la varianza resultaban poco acertados, pues sólo un factor era significativo, y al hacer los análisis para cada enfoque se corrió el riesgo de alejarse en exceso del valor objetivo para la característica de calidad.

Es decir, la variabilidad del proceso fue tan grande que los métodos anteriores apenas lograban describirla. Así fue como Vining y Bohn introducen la idea de una regresión no paramétrica, que tome en cuenta dicha dispersión, y aunque ya no se dispondrá de un modelo formal, la calidad de la predicción es mucho mayor.

El conjunto de entradas óptimas resultante de los enfoques con herramienta no paramétrica reportó varianzas relativamente pequeñas manteniendo su media en el valor objetivo, aunque no para todo ancho de banda elegido. Además, su capacidad de predicción fue mucho mayor, aunque es necesario probar con más de una función kernel para evitar varianzas negativas.

Así, se han explorado diversas opciones en este experimento de impresión, hablar en general acerca de qué método es el ideal, será determinado por las circunstancias del problema, pues a cada método se le puede hacer al menos una crítica; ya sea requerir modelos completos, no acotar la lejanía del valor objetivo, depender del modelo de la varianza, que de ser malo arrojará malos estimados, o la necesidad de recurrir a transformaciones previas al ajuste.

Ahora bien, pese a ser un ejemplo ya clásico en la literatura del tema, es ilustrativo de los diferentes procedimientos y las ventajas que se pueden alcanzar con ellos, permitiendo así un recorrido por diferentes autores y propuestas.

La tabla 3.48 presenta los resultados obtenidos con los procedimientos estudiados a lo largo de este trabajo. Todos ellos encaminados a llevar a la media a su valor objetivo al tiempo que minimizaban la varianza.

Tabla 3.48

Comparativo resultados obtenidos

Método	x_1	x_2	x_3	$\hat{\mu}$	$\hat{\sigma}$	ECM
VM	1	0.124	-0.261	500.692	45.181	2041.8
Deseabilidad	1.00	0.10	-0.25	499.017	44.991	2025.2
ECM	1.000	0.074	-0.252	494.681	44.460	2005.027
CN	0.358	0.219	0.363	480.323	58.014	3752.8
SNR	0.34	0.10	-0.12	384.647	37.425	14707
Welch	0.548	0.333	0.107	500	88.496	7831.54
WLS1	1.000	0.951	-1.000	495.783	31.779	1027.715
WLS2	1.000	1.000	-0.086	483.984	59.204	3761.597
Semiparamétrico	0.447	0.467	0.474	500.023	1.553	2.412
No paramétrico	0.750	0.725	0.521	496.303	19.996	413.509

Desde el análisis exploratorio se tenían indicios de entradas óptimas con niveles cercanos al máximo para el primer factor, cercanos al nivel medio para el segundo y lejanos del mayor para el tercero, tratando de controlar así la gran variabilidad del proceso.

Dicha dispersión afecta a los métodos paramétricos, alejándolos de una buena predicción y por consiguiente de un buen ECM.

Ante esta deficiencia se recurre a procedimientos con elementos no paramétricos, capaces de modelar mejor ese ruido (si se realiza previamente una transformación para suavizar la curva). Y en efecto, al comparar los resultados, dichas herramientas permiten alcanzar varianzas estimadas mucho menores que sus contrapartes paramétricas, manteniéndose en el valor deseado de la media aún en los casos más conservadores. Es decir, pese a sus desventajas y ser ligeramente más difíciles de interpretar, su implementación resultó exitosa.

Bibliografía

- [1] Barrios Zamudio Ernesto, Domínguez Domínguez Jorge. *Revisión de métodos de respuesta dual*. México, ITAM; 2008.
- [2] Box, G. D. W. Behnken. Some new three level designs for the study of quantitative variables. *Technometrics* 1960, 2: pp 455-475.
- [3] Box, G. E. P., Draper, N. R.. *Empirical Model Building and Response Surfaces*. NY, Wiley; 1987.
- [4] Box, G. S. Jones. Designing products that are robust to the environment. *Total Quality Management* 1992, 3: pp 265-282.
- [5] Carroll, R. J. and Ruppert, D. *Transformation and Weighting in Regression*. NY, Chapman and Hall; 1998.
- [6] Copeland, K. A. F. Nelson, P. R. Dual Response Optimization via Direct Function Minimization. *Journal of Quality Technology* 1996, 28: pp 331-336.
- [7] Del Castillo, E. Multiresponse process optimization via constrained confidence regions. *Journal of Quality Technology* 1996, 28: 61-70.
- [8] Del Castillo, E. and Montgomery, D.C. A Nonlinear Programming Solution to the Dual Response Problem. *Journal of Quality Technology* 1993, 25: pp 199-204.
- [9] Derringer, G. Suich, R. Simultaneous Optimization of several Response Variables. *Journal of Quality Technology* 1980, 12: pp 214-219.
- [10] Domínguez, D. J. Rocha, M. R. Optimización Multirrespuesta en Procesos Industriales: Una propuesta gráfica. *Memorias del 18 Foro Nacional de Estadística*, 2004.
- [11] Gasser, T. Müller, H. G. Kernel estimation and Regression Functions. En: *Smoothing Techniques for Curve Estimation*. Berlín, Springer-Verlag; 1979.
- [12] Gasser, T. Müller, H. G. Estimating Regression Functions and Their Derivatives by the Kernel Method. *Scandinavian Journal of Statistics* 1984, 11: pp 171-185.
- [13] Härdle, W. *Applied Nonparametric Regression*. Cambridge, UK, Cambridge University Press; 1994.

- [14] Härdle, W. Marron, J. S. Wand, M. P. Bandwith Choice for Density Derivatives. *Journal of the Royal Statistical Society* 1990, 52: 223-232.
- [15] Harrington, E. D. Jr. The Desirability Function. *Industrial Quality Control* 1965, 21: pp 494-498.
- [16] Kacker, R. N. Off-line quality control, parameter design, and the Taguchi method. *Journal of Quality Technology* 1985, 17: 176-188.
- [17] Khuri, A.I. A Test for Lack of Fit of a Linear Multiresponse Model. *Technometrics* 1985, 27: 213-218.
- [18] Khuri, A.I. *Advanced Calculus with applications in Statistics*. NY, Wiley: 1993.
- [19] Khuri, A.I. *Response Surface Methodology and Related Topics*. Singapur, World Scientific; 2006.
- [20] Khuri, A.I., and Cornell, J.A. *Response Surfaces*, New York, Dekker; 1996.
- [21] Khuri, A. I. and Mukhopadhyay, S. GLM Designs: The Dependence on Unknown Parameters Dilemma. En: Khuri AI, ed. *Response Surface Methodology and Related Topics*. Singapur: World Scientific; 2006, pp. 203-223.
- [22] Khuri, A.I. and Mukhopadhyay. *Response Surface Methodology*. Wiley Interdisciplinary Reviews - Computational Statistics, Volume 2, Number 2, E.J Wegman, Y.H. Said, and D.W. Scott, Eds., Hoboken New Jersey: Wiley, pp. 128-149.S. (2010).
- [23] Khuri, A.I. and Myers, R.H. Modified Ridge Analysis. *Technometrics* 1979, 21: 467-473.
- [24] Lin, D. K. Tu, W. Dual Response Surface Optimization. *Journal of Quality Technology* 1995, 27: pp. 34-39.
- [25] Montgomery, D. C. *Design and Analysis of Experiments*. NY, Wiley; 2001.
- [26] Müller, H. G. *Nonparametric Regression Analysis of Longitudinal Data*. Berlín, Germany, Springer-Verlag; 1998.
- [27] Myers, R. H. and Carter, W. H., Jr. Response Surface Techniques for Dual Response Systems. *Technometrics* 1973, 15; pp. 301-317.
- [28] Myers, R. H and Montgomery D.C. *Response Surface Methodology*. NY: Wiley; 2009.
- [29] Myers, R.H.; and Vining, G. Response Surface Alternatives to the Taguchi Robust Parameter Design Approach. *American Statistician* 1992, 46: pp- 131-139.
- [30] Phadke MS. *Quality Engineering Using Robust Design*. Englewood Cliffs: Prentice Hall; 1989.
- [31] Pickle, S.,Robinson T., Birch J. A Semi-parametric Approach to Robust Parameter Design. *Journal of Statistical Planning and Inference* 2008, 138 (1): 114-131.
- [32] Robinson, T. J., Wulff , S. S., Montgomery, D. C., and Khuri, A. I. Robust Parameter Design Using Generalized Linear Mixed Models. *Journal of Quality Technology* 2006, 38: 65-75.

- [33] Ruppert, D. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association* 1997, 92: 1049-1062.
- [34] Ruppert, D. Wand, M. Carroll, R. *Semiparametric Regression*. UK, Cambridge University Press; 2003.
- [35] Taguchi, G. *Taguchi Methods: Signal-to-Noise Ratio for Quality Evaluation*. Dearborn, MI., Quality Engineering Series, Vol. 3, ASI Press; 1991.
- [36] Taguchi, G. and Wu, Y. *Introduction to Off-Line Quality Control*. Nagoya, Japan: Central Japan Quality Control Association; 1985.
- [37] Van der Vaart, A. W. *Asymptotic Statistics*. UK, Cambridge University Press; 1998.
- [38] Vining, G.G., Bohn, L.L. Response surfaces for the mean and variance using a nonparametric approach. *Journal of Quality Technology* 1998, 30: 282-291.
- [39] Vining, G. G. and Myers, R. H. Combining Taguchi and Response Surface Philosophies: A Dual Response Approach. *Journal of Quality Technology* 1990, 22: pp 38-45.
- [40] Vining and R.H. Myers. A Graphical Approach for Evaluating Response of the Mean Squared Error of Prediction. *Technometrics* 1991, 33: 315-326.
- [41] Vining, G. G. and Schaub, D. Experimental Designs for Estimating Both Mean and Variance Functions. *Journal of Quality Technology* 1996, 28: pp. 135-147.
- [42] Wasserman, L. *All of Nonparametric Statistics*. Springer; 2006.
- [43] Welch WJ, Yu TK, Kang SM, Sacks J. Computer experiments for quality control by parameter design. *J Quality Technology* 1990, 22: 15-22.
- [44] Wu, C. F. J. Hamada, M. *Experiments: planning, analysis, and parameter design optimization*. NY, Wiley; 2000.