Yordan Garcia. 2024. Optimización de una superficie de respuesta empleando R estudio. Revista Sigma, 20 (2). Páginas 16–24.

REVISTA SIGMA

Departamento de Matemáticas y Estadística

Universidad de Nariño

Volumen XX № 2 (2024), páginas 16–24

Optimización de una superficie de respuesta empleando R estudio

Yordan Nicolay Garcia Murcia[1]

Abstract: The Response Surface Methodology (RSM), also known as Response Surface Methodology, is a set of methods designed with the objective of optimizing processes during the study of the relationship between one or more response variables and a set of independent variables. This approach is based on sequential experimentation, which allows it to be applied iteratively to approximate the region of interest.

In the development of a system, the Response Surface methodology is adapted through sequential experiments, using increasingly complex designs that depend on the information obtained at each stage. This facilitates a systematic and efficient exploration of the design space, optimizing the process conditions.

In addition to its application in process optimization, RSM has significant implications in the design, development and formulation of new products. It allows variables to be carefully adjusted to improve the quality and efficiency of existing products, thus contributing to innovation and continuous improvement in various fields.

The rsm library provides specific functions for the construction and analysis of response surfaces in R, with which an application problem will be analyzed and how to perform this analysis will be explained.

Keywords. Response surface methodology, optimization, processes.

Resumen: La metodología de Superficie de Respuesta (RSM), también conocida como Response Surface Methodology, constituye un conjunto de métodos diseñados con el objetivo de optimizar procesos durante el estudio de la relación entre una o más variables de respuesta y un conjunto de variables independientes. Este enfoque se basa en experimentación secuencial, lo que permite su aplicación de manera iterativa para aproximarse a la región de interés.

En el desarrollo de un sistema, la metodología de Superficie de Respuesta se adapta a través de experimentos secuenciales, utilizando diseños cada vez más complejos que dependen de la información obtenida en cada etapa. Esto facilita una exploración sistemática y eficiente del espacio de diseño, optimizando las condiciones del proceso.

[1] Maestría en Estadística Aplicada, Facultad Ciencias Exactas y Naturales, Universidad de Nariño. Correo: yordangarcia@hotmail.es

Además de su aplicación en la optimización de procesos, la RSM tiene notables implicaciones en el diseño, desarrollo y formulación de nuevos productos. Permite ajustar variables de manera cuidadosa para mejorar la calidad y eficiencia de los productos existentes, contribuyendo así a la innovación y mejora continua en diversos campos.

La biblioteca rsm proporciona funciones específicas para la construcción y análisis de superficies de respuesta en R, con el cual se analizará un problema de aplicación y se explicará cómo efectuar este análisis.

Palabras Clave. Metodología de superficie de respuesta, optimización, procesos.

1. Introducción

George E. P. Box es un estadístico británico conocido por sus contribuciones significativas al diseño de experimentos y a la estadística aplicada, es especialmente conocido por el desarrollo de la metodología de la superficie de respuesta junto con su esposa, J. Stuart Hunter, y su colega William G. Hunter. El diseño de experimentos y la metodología de superficie de respuesta desarrollados por Box y sus colaboradores son fundamentales en la optimización de procesos y la mejora de productos, su trabajo influyó en el campo de la calidad industrial y en la aplicación de técnicas estadísticas en la toma de decisiones.

La exploración de cómo la variación de factores cuantitativos impacta en una variable dependiente sugiere que los experimentos buscan comprender las relaciones causales entre estos factores y la variable de interés. Este enfoque requiere un diseño experimental cuidadoso para aislar y medir el impacto específico de los factores cuantitativos seleccionados en la variable de respuesta, garantizando resultados confiables y significativos.

La inclusión de factores cuantitativos destaca la naturaleza numérica y medible de la investigación, facilitando una evaluación precisa de la relación entre variables y proporcionando una base sólida para el análisis estadístico.

Estos experimentos son esenciales en campos como la investigación científica, la ingeniería y la tecnología, permitiendo comprender cómo las variaciones afectan una variable y contribuyendo a la optimización de procesos y la toma de decisiones informada. Además de su valor académico, la exploración de estas relaciones ofrece información práctica aplicable en diversas áreas, permitiendo a empresas, investigadores y profesionales mejorar la eficiencia, calidad y rentabilidad de sus procesos y productos.

2. Optimización de superficies de respuesta

La expresión matemática de los modelos de superficie de respuesta (MSR) puede adoptar diversas formas, entre ellas:

Modelos	Expresión matemática
Modelo lineal sin interacciones.	$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + e$
Modelo lineal con interacciones.	$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{j=2}^{k} \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + e$
Modelo cuadrático sin interacciones.	$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x^2 + e$
Modelo cuadrático con interacciones.	$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{j=2}^{k} \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x^2 + e$

Tabla 1: Modelos de Primer y Segundo Orden en Superficies de Respuesta.

En la **Metodología de Superficies de Respuesta (RSM)**, la variable de respuesta y depende de los niveles de los factores cuantitativos, representados por las variables $x_1, x_2, \ldots, x_k$. Los **modelos polinomiales** relacionan la variable de respuesta y dichos factores, estos modelos son pertinentes a la hora de trabajar cuando se limita los factores cuantitativos.

El modelo $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$ permite estudiar cómo los factores x_1 y x_2 afectan la respuesta y de manera lineal, en este caso la gráfica de **superficie de respuesta** es un plano en el espacio tridimensional y sus **curvas de nivel** que son líneas con valores iguales de respuesta, donde cada línea de contorno representa una altura determinada de la superficie de respuesta, lo que facilita la interpretación de cómo varía la respuesta en función de los factores o lo que es lo mismo, distintas combinaciones de niveles de factores que provocan una misma respuesta. En este caso el gráfico de contornos resultante está compuesto por líneas paralelas.

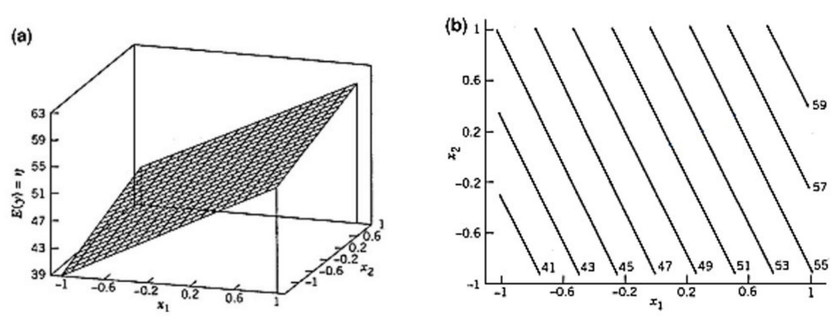

Figura 1: Superficie de respuesta (a) y gráfico de contorno (b) de un modelo de primer orden. Tomado de Metodología de Superficie de Respuesta (RSM), por Cristina Gil Martínez, 2019.

Una de las aplicaciones de las derivadas es el gradiente de una función cuyo objetivo es buscar la dirección a la cual la función crece más rápidamente, así nace El **Método de Máxima**

Pendiente en Ascenso, que es una técnica utilizada en la Metodología de Superficies de Respuesta (RSM), que busca mejorar el valor de una función a través de experimentos encontrando la dirección en la cual la función de respuesta mejora más rápidamente, es decir, la dirección de la máxima pendiente o gradiente en el espacio de los factores. Para ello primero, se ajusta un modelo lineal de primer orden luego se determina la dirección del gradiente en el punto actual, que corresponde a la dirección en la cual la respuesta aumenta más rápidamente por último se realizan experimentos moviéndose en esa dirección, haciendo cambios graduales en los niveles de los factores hasta que no se observe una mejora significativa en la respuesta.

Este método no requiere modelos complejos, pues permite hacer mejoras rápidas cuando se está lejos del óptimo, reduciendo el número de experimentos necesarios para mejorar la respuesta. Claro está que funciona bien solo en regiones donde la respuesta es aproximadamente lineal. Si la respuesta es más compleja (por ejemplo, presenta una curvatura significativa), este método puede no llevar directamente al óptimo y se requerirá un modelo cuadrático o de segundo orden.

Un algoritmo utilizado para determinar las coordenadas de la trayectoria de máxima pendiente en ascenso es el siguiente propuesto por (Gil Martínez, 2019):

1. Se escoge un tamaño de incremento para uno de las variables independientes o factores Δx_i. Normalmente elegimos la variable con mayor coeficiente de regresión absoluto $|\beta_i|$ o que más conozcamos.
2. Aplicamos el incremento en el resto de variables

$$\Delta x_i = \frac{\widehat{\beta}_i}{\dfrac{\widehat{\beta}_j}{\Delta x_j}} = \frac{\dfrac{\widehat{\beta}_i}{1}}{\dfrac{\widehat{\beta}_j}{\Delta x_j}} = \frac{\Delta x_j \widehat{\beta}_i}{\widehat{\beta}_j} \qquad i = 1, 2, 3, \ldots, k \qquad i \neq j$$

3. Convertimos Δx_i a unidades naturales.

Por otro lado, el modelo $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_{11} x_1^2 + \widehat{\beta}_{22} x_2^2 + \widehat{\beta}_{12} x_1 x_2$ es un modelo de segundo orden y se considera que es representado gráficamente de la siguiente manera.

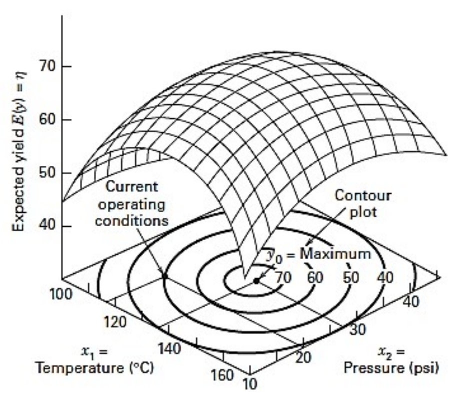

Figura 2: Superficie de respuesta tridimensional y gráfico de contornos correspondiente. Tomado de Metodología de Superficie de Respuesta (RSM), por Cristina Gil Martínez, 2019.

Se evidencia una gráfica de superficie tridimensional. En este modelo la forma de la superficie puede variar considerablemente dependiendo de los coeficientes del modelo.

Coeficientes	**Superficie**	**Curvas de Nivel**	**Conclusión**
Todos los coeficientes cuadráticos son positivos	Convexa	Elípticas	Indica un mínimo global. Las curvas de nivel representan combinaciones de factores que optimizan la respuesta.
Algunos coeficientes cuadráticos tienen signos opuestos	Silla de montar	Hiperbólicas	La relación entre los factores es más compleja. Las curvas de nivel indican un punto de silla de montar en la superficie de respuesta.
Coeficientes de términos cuadráticos idénticos	Simétrica	Circulares	Sugiere que el efecto de los factores en la respuesta es uniforme. Las curvas de nivel presentan simetría, lo que puede facilitar la interpretación del modelo.

Tabla 2: Descripción de las gráficas del Modelos de Segundo Orden.

La función de deseabilidad es una herramienta estadística que permite optimizar varias respuestas al mismo tiempo, esta herramienta es útil en el contexto Metodología de Superficies de Respuesta (RSM) pues hay contextos en el que las variables pueden estar muy ligadas y esto presenta conflicto entre ellas y es necesario establecer la optimización entre ellas basándose en las ideas propuestas por el investigador.

Derringer y Suich (1980), hace uso de funciones de deseabilidad. Se basa en convertir cada respuesta y_i en una función de deseabilidad individual d_i que varíe entre 0 y 1 y que represente la cercanía de la respuesta a un valor ideal o deseado: (Gil Martínez, 2019):

$$0 \leq d_i \leq 1$$

2.1. Función de Deseabilidad Global

Esta función permite optimizar conjuntamente todas las respuestas, su cálculo es la media geométrica de las deseabilidades individuales:

$$D = (d_1 \cdot d_2 \cdot \cdots \cdot d_n)^{1/n}$$

Donde n es el número de respuestas optimizadas y d_i es la deseabilidad individual de la respuesta i. son considerados el conjunto óptimo de condiciones para el proceso pues busca maximizar el valor de D, y los valores de los factores que produzcan la mayor deseabilidad global.

<table>
<tr><th>Objetivo</th><th>Función deseabilidad individual</th><th>Descripción</th></tr>
<tr><td>Maximizar una respuesta y</td><td>$$d = \begin{cases} 0 & si \quad y < L \\ \left(\dfrac{y-L}{T-L}\right)^{r} & si \quad L \leq y \leq T \\ 1 & si \quad y > T \end{cases}$$</td><td rowspan="3">L y T son el límite mínimo y el valor objetivo que establezcamos oportunos o deseables, y r un valor de peso asociado.
Si $r = 1$, la función de deseabilidad es lineal. Escoger.
Si $r > 1$ pone énfasis en acercarse lo máximo posible al valor objetivo, es decir, hace más difícil alcanzar una deseabilidad alta.
Si los Valores de $r < 1$ hace menos estricto satisfacer la condición.</td></tr>
<tr><td>Si el objetivo es mantener en un rango la respuesta y</td><td>$$y = \begin{cases} 0 & si \quad y < L \\ \left(\dfrac{y-L}{T-L}\right)^{r_1} & si \quad L \leq y \leq T \\ \left(\dfrac{U-y}{U-T}\right)^{r_2} & si \quad T \leq y \leq U \\ 0 & si \quad y > U \end{cases}$$</td></tr>
<tr><td>Si el objetivo es minimizar una respuesta y</td><td>$$d = \begin{cases} 1 & si \quad y < T \\ \left(\dfrac{y-L}{T-L}\right)^{r} & si \quad T \leq y \leq U \\ 0 & si \quad y > U \end{cases}$$</td></tr>
</table>

Tabla 3: Descripción de la función de deseabilidad tomado de (Gil Martínez, 2019)

La ortogonalidad entre dos vectores se define matemáticamente como la relación en la que el producto punto de ambos vectores es igual a cero

$$u \cdot v = 0$$

(Gil Martínez, 2019) Menciona que es conveniente que el diseño escogido sea ortogonal. La ortogonalidad garantiza que se pueda estimar de forma independiente los efectos de los factores, ya que las estimaciones de los parámetros del modelo a ajustar, no presentarán correlación. En este caso, la varianza de la respuesta esperada es expresable como la suma ponderada de las varianzas de los parámetros estimados.

Que el diseño sea invariante por rotación resulta también interesante ya que asegura que la varianza de la respuesta esperada dependa solamente de la distancia de un punto al centro del diseño, y no de la dirección. Esto es importante ya que la localización del óptimo es desconocida.

2.2. Diseños para modelos de primer y segundo orden

Diseño	Modelo	Número de niveles	Propósito	Ventajas
Factorial Completo 2^k	Primer orden	2	Explorar efectos lineales	Sencillo de analizar, permite interacciones
Factorial Fraccionado 2^{k-p}	Primer orden	2	Reducir número de experimentos	Ahorra tiempo y recursos
Máxima Pendiente en Ascenso	Primer orden	2	Identificar dirección de mejora	Rápido ajuste hacia la región óptima
Diseño Central Compuesto (CCD)	Segundo orden	2 (con puntos axiales)	Modelar curvatura, optimización	Ajuste de modelos cuadráticos, rotabilidad
Diseño Box-Behnken	Segundo orden	3	Modelar interacciones y cuadratura	Menos ensayos que CCD
Diseño de Doehlert	Segundo orden	3	Ajuste preciso de superficies de segundo orden	Flexibilidad y eficiencia en el número de ensayos

Tabla 4: Cuadro comparativo de Diseños para modelos de primer y segundo orden.

3. Aplicación en R

Ejemplo ejercicio propuesto por Robert Kuehl: Se desea estudiar la relación entre el crecimiento de los pollos y el metabolismo de metionina (un aminoácido azufrado) y caroteno (vitamina A). Se ha encontrado, por experimentos previos, que los niveles óptimos de metionina y caroteno son 0.9 % de metionina en la dieta y 50 microgramos de caroteno al día. Se usó un diseño central compuesto rotatorio para el experimento. Se asignaron al azar ocho pollos a cada tratamiento dietético y se registraron sus aumentos de peso después de 38 días. El aumento promedio de peso para los tratamientos es el siguiente:

Factores originales		*Factores codificados*		
Metionina	*Caroteno*	x_1	x_2	*Aumento de peso*
1.183	85.36	+1	+1	445
1.183	14.64	+1	−1	331
0.617	85.36	−1	+1	443
0.617	14.64	−1	−1	336
1.183	50.00	$\sqrt{2}$	0	414
0.500	50.00	$-\sqrt{2}$	0	389
0.900	100.00	0	$\sqrt{2}$	435
0.900	0.00	0	$-\sqrt{2}$	225
0.900	50.00	0	0	442
0.900	50.00	0	0	412
0.900	50.00	0	0	418
0.900	50.00	0	0	440
0.900	50.00	0	0	441

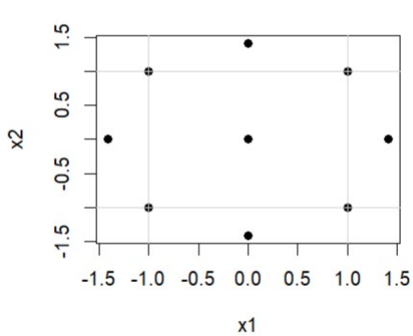

Figura 3: Ejercicio propuesto por Robert Kuehl, tomado de (Figueroa Preciado 2003).

```
Call:
rsm(formula = Aumento_de_Peso ~ SO(x1, x2), data = datos)
            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 430.6000     8.9360 48.1873 4.337e-10 ***
x1            4.0444     7.0645  0.5725 0.5848945
x2           64.7481     7.0645  9.1653 3.789e-05 ***
x1:x2         1.7500     9.9907  0.1752 0.8659109
x1^2         -8.8000     7.5758 -1.1616 0.2834748
x2^2        -44.5500     7.5758 -5.8805 0.0006114 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La ecuación de la superficie de respuesta de segundo orden, estimada para los datos de aumento de peso es:

$$\widehat{y} = 430.6 + 4.04x_1 + 64.75x_2 - 8.80x_1^2 - 44.55x_2^2 + 1.75x_1x_2$$

```
Multiple R-squared:  0.9445,  Adjusted R-squared:  0.9048
F-statistic: 23.82 on 5 and 7 DF,  p-value: 0.0002928

Analysis of Variance Table

Response: Aumento_de_Peso
            Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2  33669 16834.7 42.1650 0.0001247
TWI(x1, x2)  1     12    12.2  0.0307 0.8659109
PQ(x1, x2)   2  13870  6934.9 17.3696 0.0019317
Residuals    7   2795   399.3
Lack of fit  3   1964   654.5  3.1498 0.1483597
Pure error   4    831   207.8
```

El siguiente resultado nos muestra que R encontró un máximo en la superficie de respuesta y los valores codificados de Metionina y Caroteno donde se alcanza este máximo son 0.3026436 y 0.7326345, respectivamente.

```
Stationary point of response surface:
       x1        x2
0.3026436 0.7326345
```

Se puede llegar a la solución anterior explorando con el cursor la gráfica de contornos que R ofrece. En esta gráfica se puede observar que los valores de la respuesta van aumentando conforme los círculos se van cerrando en el centro, lo que nos indica que la solución es un máximo.

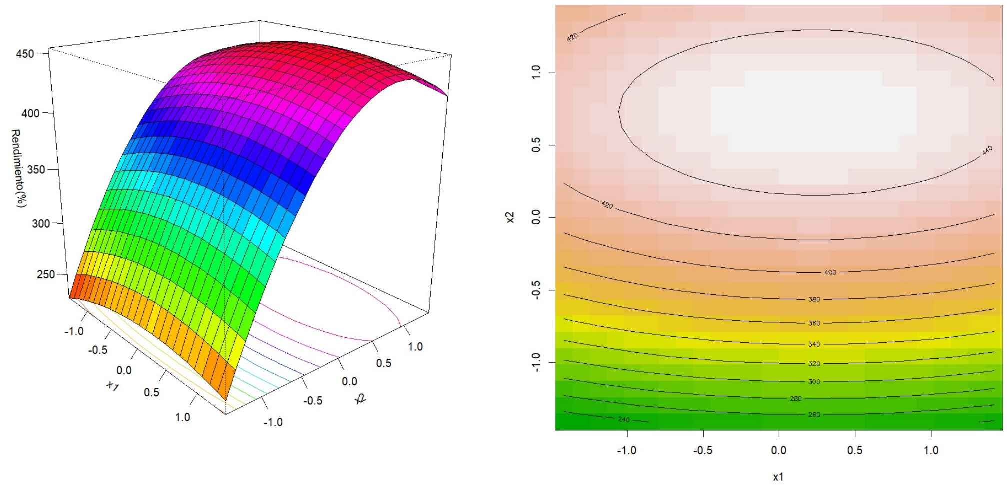

Figura 4: gráfica de superficie y contornos en R.

4. Conclusiones

El modelo ajustado tiene buen ajuste y muestra que $x2$ tiene una influencia significativa en la respuesta. Sin embargo, $x1$ y sus términos cuadráticos no parecen ser tan influyentes en la respuesta en comparación con $x2$. Es posible que la relación con $x1$ sea más compleja y no se pueda modelar de manera lineal o cuadrática en este rango de datos. Además, el análisis de varianza respalda la significancia de los términos cuadráticos y de interacción en el modelo.

Referencias

[1] Figueroa Preciado, G. (2003, diciembre). Optimización de una superficie de respuesta utilizando JMP IN. páginas 17-22.

[2] Box, G. E. P., Wilson, K. G. (1951), On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society*, B 13, 1-45

[3] Cornell, John A. (1984), *How to apply Response Surface Methodology*, American Society for Quality Control, Milwaukee, WI.

[4] Kuehl, Robert O. (2001) *Diseño de Experimentos*, 2a. Edición, Thomson Learning. Montgomery, D. C. (2002), Diseño y Análisis de Experimentos, Editorial Limusa, Segunda Edición.