

Optimización de curvas de crecimiento a través de análisis univariado

Optimization Process of Growth Curves Through Univariate Analysis

SARA CRISTINA GUERRERO^{1,a}, ÓSCAR ORLANDO MELO^{2,b}

¹DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, TUNJA, COLOMBIA

²DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ, COLOMBIA

Resumen

En este artículo se propone una metodología de modelamiento conjunto de tratamientos con niveles cuantitativos medidos en el tiempo, a través de la combinación de las metodologías de superficie de respuesta y curvas de crecimiento. Se realiza la estimación de los parámetros del modelo propuesto, los cuales miden el efecto de los factores asociados al modelo de superficie de respuesta de segundo orden a lo largo del tiempo. Se plantea y juzga las diferentes hipótesis de interés y, finalmente, con el modelo ajustado se encuentran las condiciones de optimización de un conjunto de tratamientos a través del tiempo. Por último, se presenta una aplicación, analizada mediante curvas de crecimiento por Kshirsagar & Boyce (1995).

Palabras clave: superficie de respuesta, curvas de crecimiento, optimización, análisis univariado.

Abstract

A methodology is suggested to jointly model treatments with quantitative levels measured in time, by combining the response surface and the growth curve techniques. The model's parameters are estimated; these measure the effect through time of the factors related to the second order response surface model. The hypothesis of interest are formulated and tested. Additionally, by means of the fitted model, the optimality conditions throughout time are established for a set of specific treatments. As a final step, an application previously analyzed with growth curves by Kshirsagar & Boyce (1995) is now assessed with the proposed model.

Key words: Response surfaces, Growth curves, Optimization, Univariate analysis.

^aProfesora auxiliar. E-mail: cguerreros@unal.edu.co

^bProfesor asociado. E-mail: oomelom@unal.edu.co

1. Introducción

En experimentación, en algunas ocasiones el interés del investigador se centra en analizar datos a través del tiempo para conocer la tendencia de un individuo o grupos de individuos. Para otros investigadores, el objetivo no solo está en la tendencia, sino también en conocer la combinación de tratamientos que optimizan el proceso a través del tiempo. Este último contexto tiene como punto de partida el análisis de curvas de crecimiento y la metodología de superficie de respuesta (MSR).

Rao (1959) realizó un análisis de curvas de crecimiento para un solo grupo, pero el modelo propuesto por Potthoff & Roy (1964) marcó el punto de partida para el estudio de las curvas de crecimiento; Khatri (1966) estudió este problema teniendo en cuenta varios grupos o tratamientos. Rao (1965), Rao (1966) y Grizzle & Allen (1969) discutieron el uso de covariables en el modelo de curvas de crecimiento propuesto por Potthoff & Roy (1964). Khatri (1973) estudió la estructura de la matriz de covarianza para variables que están igualmente correlacionadas y encontró el criterio por máxima verosimilitud para el juzgamiento de las hipótesis sobre la covarianza.

Reinsel (1982) generalizó el modelo de Potthoff & Roy (1964) cuando se miden varias respuestas simultáneamente; estudió el modelo de curva de crecimiento asumiendo la característica de efectos aleatorios. Posteriormente, Reinsel (1984) analizó los efectos fijos y aleatorios en el modelo de curvas de crecimiento multidimensional. También Laird & Ware (1982) discutieron la ventaja de trabajar con modelos de efectos aleatorios a dos vías para datos longitudinales, incluidos los modelos de curvas de crecimiento y medidas repetidas como casos especiales. Verbyla & Venables (1988) trabajaron el modelo de curvas de crecimiento como un modelo de suma de perfiles. Además, Lee (1991) propuso un criterio para la selección adecuada de la estructura de covarianza. Srivastava (2001) estudió el modelo de curvas de crecimiento anidado propuesto por Srivastava & Katri (1979).

Por otra parte, si el interés del investigador es modelar y analizar situaciones con el propósito de determinar las condiciones operativas óptimas para un proceso particular, este análisis se realiza por medio de la MSR, ampliamente aplicable en ciencias biológicas, químicas, sociales, experimentación en agricultura, ingeniería, ciencias relacionadas con los alimentos, control de calidad y economía, entre otras.

La MSR fue desarrollada principalmente en experimentación industrial y en producción por Box & Wilson (1951). Hill & Hunter (1966) estudiaron la MSR aplicándola en procesos químicos e industriales, cuyos datos fueron medidos a través del tiempo. Thompson (1973) estudió otros criterios para superficies de respuesta que incluyen estimaciones con mínimo sesgo. Mead & Pike (1975) aplicaron la MSR en experimentación en agricultura y el modelamiento de datos biológicos. Lucas (1976) estudió la eficiencia D y G. Box & Draper (1982) analizaron el poder de las transformaciones de las variables para reducir los grados en busca del buen ajuste en el modelo.

Box & Jones (1992) ilustraron la MSR considerando tres arreglos de diseños experimentales con parámetros robustos usando parcelas divididas; posteriormente,

Box & Jones (2001) consideraron las parcelas divididas como experimentos ideales en estudios de robustez debido a que la interacción entre los factores del diseño y los factores ambientales se estiman minimizando el error al establecer subparcelas.

Draper & Ying (1994) mostraron que los grados de libertad de un diseño de Box-Behnken, empleado para el ajuste de modelos, pueden ser utilizados para chequear la adecuación de estos, construyendo contrastes en términos de alto orden. Cuando hay ortogonalidad de los contrastes, estos pueden ser ilustrados en un gráfico de probabilidad normal para determinar el tipo de sesgo. Myers et al. (2002) ilustraron cómo usar la gráfica de dispersión de la varianza (GDV) para comparar y evaluar diseños de superficie de respuesta.

Por otro lado, si en un proceso de producción desarrollado en cierta industria se puede estar interesado en conocer la combinación de niveles de los tratamientos y el tiempo en el cual se lleva a cabo dicha optimización, o si en un tratamiento clínico conformado por la combinación de varios factores (tratamientos) se tiene interés en conocer la combinación de estos y el momento en el tiempo en el cual se observa un resultado óptimo sobre la salud de un paciente, en ambos casos surge la necesidad de modelar conjuntamente los tratamientos y los tiempos, haciendo una mezcla entre MSR y curvas de crecimiento. Por tal razón, en este artículo se propone una metodología que permite modelar conjuntamente la superficie de respuesta (tratamientos) y la curva de crecimiento (tiempos) de manera univariada, donde las condiciones de estudio para los tiempos y niveles de tratamientos deben ser igualmente espaciados. Los tratamientos representan una variable de tipo cuantitativo y son aplicados a la misma unidad experimental a través del tiempo.

Para tratar este problema, en primer lugar se establece el modelo de manera univariada y, debido a la similaridad que presentan los parámetros con la forma multivariada, se recurre al modelo multivariado propuesto por Potthoff & Roy (1964) y al estudio del caso multivariado realizado por Rivera et al. (2006). La solución se logra mediante una transformación que permite trabajar matricialmente el modelo de forma univariada, con lo cual se estiman los parámetros por medio de mínimos cuadrados generalizados, para luego plantear las hipótesis de interés y, finalmente, construir la expresión para generar el óptimo en la situación de interés.

En la segunda sección se describe la superficie de respuesta a través del tiempo, se presenta el modelo, el planteamiento de las hipótesis y la caracterización del óptimo. En la tercera, se presenta una aplicación de la metodología planteada y, finalmente, las conclusiones de este artículo.

2. Curvas de crecimiento en superficies de respuesta

En esta sección, se implementa la metodología que permite la optimización de procesos a través del tiempo. El tipo de análisis propuesto es univariado, con un enfoque clásico, aplicado bajo el supuesto de normalidad. Se modela un caso multivariado de manera univariada, donde los grupos corresponden a cada uno de los niveles de los tratamientos, los cuales son medidos en los diferentes individuos y evaluados a través del tiempo.

El tipo de datos para este estudio corresponde a datos longitudinales. Por las características de estos, Davis (2002) plantea que se tiene específicamente un estudio de medidas repetidas porque el individuo es evaluado en diversas ocasiones a través del tiempo. Además, plantea que, por la naturaleza de la información evaluada en periodos cortos, se está en un análisis de datos longitudinales y, más específicamente, en análisis de medidas repetidas.

Según Davidian (2005), es razonable asumir que todas las fuentes de variación actúan similarmente en cada población y podría asumirse que $\Sigma_h = \Sigma$, una varianza común para todas las poblaciones. Así se cumple el supuesto de homocedasticidad multivariada.

A continuación se presenta un modelo que relaciona las combinaciones de los niveles de los diferentes factores (tratamientos) a través del tiempo, modelando conjuntamente los tratamientos a través de la MSR y el tiempo con la curva de crecimiento de la siguiente forma:

$$Y_{hit} = \sum_{m=0}^q \left[\left(\xi_{0hm} + \sum_{j=1}^{k'} \xi_{jhm} x_{ij} + \sum_{j=1}^{k'} \sum_{z=1}^{k'} \xi_{jzhm} x_{ij} x_{iz} \right) P_m(t_i) \right] + e_{hit} \quad (1)$$

donde Y_{hit} corresponde al vector de observaciones que mide la respuesta del h -ésimo tratamiento en el i -ésimo individuo a través del t -ésimo tiempo. El modelo planteado es de segundo orden; cada ξ_{jhm} es el parámetro asociado a la superficie de respuesta a través del tiempo, para cada una de las n unidades experimentales, con $i = 1, \dots, n_h$, $t = t_1, \dots, t_p$, $h = 1, \dots, k$ y $m = 0, 1, \dots, q$; corresponde al grado de la curva de crecimiento, que a lo sumo es q ; ξ_{0hm} es el intercepto de la superficie de respuesta a través del tiempo; ξ_{jzhm} es el parámetro que asocia la interacción de los factores a través del tiempo; x_{ij} y x_{iz} son las variables explicativas codificadas (Montgomery 2003) asociadas a los factores j -ésimo y z -ésimo, respectivamente ($j, z = 1, \dots, k'$). $P_m(t_i)$ está asociado a los polinomios ortogonales. Según el número de tiempos, estos polinomios pueden ser los coeficientes de los tiempos codificados si el ajuste del modelo es de segundo orden para el caso de la curva de crecimiento (tiempos), pero si el ajuste es de orden superior y los tiempos están igualmente espaciados, se puede recurrir a los polinomios dados en Hinkelman & Kempthorne (1994) o, si se quieren construir, estos deben satisfacer las condiciones presentadas en Kshirsagar & Boyce (1995). Los errores e_{hit} se distribuyen normal con media 0 y varianza σ_{hit}^2 , es decir,

$$e_{hit} \sim N(0, \sigma_{hit}^2)$$

Para este caso, es de notar que

$$\begin{aligned} \text{Cov}(e_{hit}, e_{hit'}) &\neq 0, \\ \text{Cov}(e_{hit}, e_{h'v't'}) &= 0, \quad \forall i' \neq i \end{aligned}$$

Por tanto, el vector Y_{hit} respuesta se distribuye de la siguiente manera:

$$Y_{hit} \sim N \left(\sum_{m=0}^q \left[\left(\xi_{0hm} + \sum_{j=1}^{k'} \xi_{jhm} x_{ij} + \sum_{j=1}^{k'} \sum_{z=1}^{k'} \xi_{jzhm} x_{ij} x_{iz} \right) P_m(t_i) \right]; \sigma_{hit}^2 \right) \quad (2)$$

El vector de observaciones sobre el mismo individuo puede ser expresado como $Y_{hit} = (y_{hi1}, y_{hi2}, \dots, y_{hit_p})$, el cual sigue una distribución normal cuyas medidas están correlacionadas dentro de cada individuo.

Una manera alternativa de presentar el modelo (1) es expresándolo matricialmente a partir de una adaptación del modelo clásico multivariado propuesto por Potthoff & Roy (1964), y realizada por Rivera et al. (2006):

$$Y = X\xi G + e \quad (3)$$

donde Y es una matriz de respuesta de orden $N \times p$ con $N = \sum n_k$; X es la matriz cuyas columnas representan los niveles de los tratamientos asociada a los parámetros de la superficie de respuesta de orden $N \times s$, con s el número de parámetros asociado a la superficie de respuesta; ξ es la matriz de parámetros de la superficie a través del tiempo de orden $s \times q$; G es la matriz de diseño que relaciona los tiempos con el grado de la curva de crecimiento de orden $q \times t$; q es el grado del polinomio respecto al tiempo y $q = p - 1$, donde p es el número de tiempos. Además, se asume que la matriz de respuestas Y se distribuye $N(X\xi G, I_n \otimes \Sigma)$.

El modelo (3) se lleva a la forma univariada aplicando el operador Vec , el cual transforma una matriz en un vector colocando cada columna debajo de la otra hasta formar una sola (Magnus 1994). De este modo, se obtiene

$$Vec(Y^t) = Vec(G^t \xi^t X^t) + Vec(e^t) \quad (4)$$

donde cada una de las componentes del vector de parámetros (ξ) corresponde a una combinación lineal, observándose el comportamiento de la superficie de respuesta a través de la curva de crecimiento. Así, el modelo univariado (4) se presenta como una estrategia alternativa de análisis donde no se pierde la correlación existente entre las mediciones a través del tiempo, es decir, sigue teniéndose en cuenta el comportamiento de correlación presente en la curva de crecimiento.

2.1. Estimación de parámetros e inferencia

La estimación de los parámetros se obtiene a partir del modelo (4) y, a diferencia del análisis multivariado trabajado por Rivera et al. (2006), no es necesario aplicar la transformación planteada por Potthoff & Roy (1964) para contrastar las hipótesis en el caso multivariado. Ellos propusieron posmultiplicar por $T^{-1}G^t(GT^{-1}G^t)^{-1}$ en ambos lados del modelo (3), donde T es de tamaño $t \times t$, simétrica y definida positiva. En esta transformación, se debe garantizar que $GT^{-1}G^t$ sea de rango t y T conocida para que los estadísticos de prueba sean exactos, lo cual no se cumple fácilmente. Entonces es necesaria una estimación de dicha matriz recurriendo a la matriz de covarianzas, lo cual lleva a realizar una estimación de la estimación de los parámetros, produciendo estadísticos inexactos en el caso de muestras pequeñas. Para el modelo propuesto en este artículo, no se presenta esta dificultad en la estimación de los parámetros (Guerrero 2006).

Debido a la naturaleza de la información, cada individuo se encuentra correlacionado a través del tiempo; no se conoce la estructura de la matriz de covarianza.

Por esta razón, los parámetros se estiman por mínimos cuadrados generalizados. En primer lugar, se expresa el vector de errores como:

$$Vec(e^t) = Vec(Y^t) - (X \otimes G^t)Vec(\xi^t)$$

con $(X \otimes G^t)Vec(\xi^t) = Vec(G^t \xi^t X^t)$ (veáse Magnus 1994).

Efectuando el producto matricial $(Vec(e^t))^t V^{-1} (Vec(e^t))$ y derivando parcialmente con respecto a $Vec(\xi^t)$ e igualando a cero, y como $(X \otimes G^t)^t V^{-1} (X \otimes G^t)$ es no singular, se encuentra el estimador de los parámetros, dado por

$$Vec(\hat{\xi}^t) = \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} (X \otimes G^t)^t V^{-1} Vec(Y^t) \quad (5)$$

donde $V = I_n \otimes \Sigma_t$.

A continuación, se obtiene la esperanza y la varianza del estimador presentado en (5) con la finalidad de hacer inferencia sobre la relación de la superficie a través del tiempo con respecto a la variable respuesta de interés.

El valor esperado del vector estimado de parámetros presentado en (5) es

$$\begin{aligned} E\left(Vec(\hat{\xi}^t) \right) &= \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} (X \otimes G^t)^t V^{-1} E\left(Vec(Y^t) \right) \\ &= \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} (X \otimes G^t)^t V^{-1} (X \otimes G^t) Vec(\xi^t) \\ &= Vec(\xi^t) \end{aligned}$$

el cual es un estimador insesgado de $Vec(\xi^t)$. La varianza del vector estimado de parámetros es

$$\begin{aligned} V\left(Vec(\hat{\xi}^t) \right) &= \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} (X \otimes G^t)^t V^{-1} V \\ &\quad V^{-1} (X \otimes G^t) \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} \\ &= \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} \end{aligned}$$

Por tanto, el estimador presentado en (5) tiene la siguiente distribución:

$$Vec(\hat{\xi}^t) \sim N\left(Vec(\xi^t); \left[(X \otimes G^t)^t V^{-1} (X \otimes G^t) \right]^{-1} \right) \quad (6)$$

Además, según Davis (2002), un estimador insesgado de la varianza está dado por

$$\hat{V} = S = \frac{\left[Vec(Y^t) - (X \otimes G^t)Vec(\hat{\xi}^t) \right] \left[Vec(Y^t) - (X \otimes G^t)Vec(\hat{\xi}^t) \right]^t}{Np - sq} \quad (7)$$

La inferencia del modelo planteado en (3) se realiza a partir de la comprobación de las hipótesis de interés. Estas están dadas para determinar el efecto del

tratamiento a través del tiempo, efecto de los tratamientos y el tiempo. Dichas hipótesis deben analizar el ajuste de la curva de crecimiento y la superficie de respuesta, teniendo en cuenta la hipótesis lineal general dada por

$$H_0 : LVec(\xi^t) = d \tag{8}$$

donde $L = A \otimes C^t$ y d es un vector de constantes, con $A_{(h-1) \times h}$ la matriz que mide el comportamiento de los tratamientos y $C_{(t-1) \times t}$ de los tiempos. Las hipótesis de interés se presentan en la tabla 1.

TABLA 1: Hipótesis más comunes sobre tratamientos y tiempos.

Hipótesis	Tiempo	Tratamiento
$H_0^{(1)}$: los tratamientos a través del tiempo afectan la respuesta.	$A_{(s-1) \times (s-1)} = [0_{(s-1) \times 1} I_{(s-1)}]$	$C_{(q-1) \times (q-1)} = [0_{(q-1) \times 1} I_{(q-1)}]$
$H_0^{(2)}$: hay interacción entre tiempo y tratamiento.	$A_{(h-1) \times h} = [1_{(h-1)} -I_{(h-1)}]$	$C_{t \times (t-1)} = [I_{(t-1) \times (t-1)} -1^t]$
$H_0^{(3)}$: no hay efecto de tratamiento.	$A_{(h-1) \times h} = [1_{(h-1)} -I_{(h-1)}]$	$C_{t \times 1} = 1_t$
$H_0^{(4)}$: no hay diferencia entre las condiciones de evaluación.	$A_{h \times 1} = 1^t$	$C_{t \times (t-1)} = [I_{(t-1) \times (t-1)} -1^t]$
$H_0^{(5)}$: el m -ésimo grado del polinomio de la curva de crecimiento no es significativo.	$A = I_s$	$C_{t \times 1} = [0, \dots, 1, \dots, 0]^t$ ↑ m - ésimo grado
$H_0^{(6)}$: el l -ésimo grado del polinomio de la superficie de respuesta no es significativo.	$A_{1 \times h} = [0, \dots, 1, \dots, 0]$ ↑ l - ésimo grado	$C = I_q$
$H_0^{(7)}$: la lm -ésima interacción entre la curva de crecimiento y la superficie de respuesta es significativa.	$A_{1 \times h} = [0, \dots, 1, \dots, 0]$ ↑ l - ésimo grado	$C_{t \times 1} = [0, \dots, 1, \dots, 0]^t$ ↑ m - ésimo grado

Al igual que en el modelo de curvas de crecimiento tradicional, en la aplicación de esta metodología se espera que la interacción entre tratamientos (visto por medio de la superficie de respuesta) y el tiempo (curva de crecimiento) sean significativas ($H_0^{(2)}$); si no lo es, entonces se procede a verificar la significancia o no de los efectos marginales de tratamiento ($H_0^{(3)}$) y tiempo ($H_0^{(4)}$).

El estadístico de prueba para juzgar las diferentes hipótesis presentadas en la tabla 1 es

$$F = \frac{SCH_0/gl_{H_0}}{CME} \sim F_{(ran(L), Np-sq)} \quad (9)$$

donde gl_{H_0} son los grados de libertad de H_0 (lo cual equivale al rango de la matriz L), $SCH_0 = [LVec(\hat{\xi}^t) - d]^t (LM^{-1}L^t)^{-1} [LVec(\hat{\xi}^t) - d]$ y $CME = \frac{SCE}{Np-sq}$, con $SCE = (Vec Y^t)^t V^{-1} [I - (X \otimes G^t)M^{-1}(X \otimes G^t)^t V^{-1}] Vec Y^t$ y $M = (X \otimes G^t)^t V^{-1} (X \otimes G^t)$.

El estadístico presentado en (9) tiene una distribución exacta si la estructura de V es conocida; si esto no se tiene, deben seguirse algunas consideraciones dadas por Singer & Andrade (1994), para garantizar una correcta inferencia sobre el modelo.

La selección apropiada de V facilita la interpretación de las hipótesis y la elección de los términos del error y, en modelos donde la estructura de la matriz de covarianzas satisfaga al menos la restricción de esfericidad, permite que el estadístico F sea exacto para algunas hipótesis. En muchos casos, cuando el supuesto de esfericidad no se mantiene, el estadístico de prueba para las hipótesis se aproxima a una F ; pero si el análisis realizado es de tipo multivariado, no es necesario tener restricciones sobre la elección de V (Singer & Andrade 1994).

La región de confianza de $(1 - \alpha) \%$ para cada una de las hipótesis del tipo (8) está dada por

$$\frac{[Vec(\hat{\xi}^t) - Vec(\xi^t)]^t L^t (LM^{-1}L^t)^{-1} L [Vec(\hat{\xi}^t) - Vec(\xi^t)]}{ran(L)(CME)} \leq F_{(ran(L), Np-sq, \alpha)}$$

Una vez ajustado el modelo, se procede a encontrar la expresión que caracterice el óptimo.

2.2. Caracterización del óptimo

Assumiendo el ajuste del modelo (1), se busca la combinación de tratamientos que optimice el proceso, teniendo presente que ese óptimo depende de la naturaleza de la información y que puede ser una respuesta máxima o mínima o un punto de silla.

La caracterización parte de la respuesta predicha en el modelo (1), donde esta puede escribirse como

$$\hat{Y}_{ht} = \sum_{m=0}^q \left(\hat{\xi}_{ohm} + \sum_{j=1}^{k'} \hat{\xi}_{jhm} x_j + \sum_{j=1}^{k'} \sum_{z=1}^{k'} \hat{\xi}_{jzhm} x_j x_z \right) P_m(t) \quad (10)$$

Si se deriva parcialmente \widehat{Y}_{ht} con respecto a x_j y x_z , las expresiones que generan el óptimo son:

$$x_{j_0} = \frac{-2E \left[\sum_{m=0}^q \widehat{\xi}_{jhm} P_m^2(t) \right] + \sum_{m=0}^q \left[\sum_{z=1}^{k'} \widehat{\xi}_{jzhm} \sum_{m=0}^q \widehat{\xi}_{zhm} P_m^2(t) \right] P_m(t)}{4DE - \sum_{m=0}^q \left[\sum_{z=1}^{k'} \widehat{\xi}_{jzhm} \sum_{m=0}^q \sum_{\substack{j=1 \\ j \neq j_0}}^k \widehat{\xi}_{jzhm} x_j P_m(t) \right] P_m(t)}$$

$$x_{z_0} = \frac{-2D \left[\sum_{m=0}^q \widehat{\xi}_{zhm} P_m^2(t) \right] + \sum_{m=0}^q \left[\sum_{j=1}^k \widehat{\xi}_{jzhm} \sum_{m=0}^q \widehat{\xi}_{jhm} P_m^2(t) \right] P_m(t)}{4DE - \sum_{m=0}^q \left[\sum_{z=1}^{k'} \widehat{\xi}_{jzhm} \sum_{m=0}^q \sum_{\substack{z=1 \\ z \neq z_0}}^k \widehat{\xi}_{jzhm} x_j P_m(t) \right] P_m(t)}$$

para $j_0, z_0 = 1, \dots, k$ con $j_0 \neq z_0$, donde $D = \sum_{m=0}^q \widehat{\xi}_{jjhm} P_m(t)$ y $E = \sum_{m=0}^q \widehat{\xi}_{zzhm} P_m(t)$.

Las anteriores expresiones están dadas en función de la(s) otra(s) variable(s), de manera que se estarían generando tantas ecuaciones como factores intervengan en el análisis. Esto último dificulta la estimación del punto estacionario x_s . Por tanto, reescribiendo (10), se obtiene

$$\widehat{Y}_{ht} = \sum_{m=0}^q \widehat{\xi}_{ohm} P_m(t) + \sum_{j=1}^{k'} \left(x_j \sum_{m=0}^q P_m(t) \widehat{\xi}_{jhm} \right) + \sum_{j=1}^{k'} \sum_{z=1}^{k'} \left(x_j x_z \sum_{m=0}^q P_m(t) \widehat{\xi}_{jzhm} \right) \quad (11)$$

Esta expresión refleja en cada término la estructura de la superficie de respuesta afectada por la curva de crecimiento.

El modelo de segundo orden de la ecuación (11), expresado en forma matricial, corresponde a

$$\widehat{Y} = b_0 + x^t b + x^t B x \quad (12)$$

donde $b_0 = \sum_{m=0}^q \widehat{\xi}_{ohm} P_m(t)$, $x^t = [x_1, x_2, \dots, x_{k'}]$, $b^t = [b_1(t), b_2(t), \dots, b_{k'}(t)]$ y

$$B = \begin{bmatrix} a_{11} & a_{12}/2 & \cdots & a_{1k'}/2 \\ a_{12}/2 & a_{22} & \cdots & a_{2k'}/2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k'}/2 & a_{2k'}/2 & \cdots & a_{k'k'} \end{bmatrix}$$

con cada $b_j(t) = \sum_{m=0}^q P_m(t) \widehat{\xi}_{jhm}$ y $a_{jz} = \sum_{m=0}^q P_m(t) \widehat{\xi}_{jzhm}$ ($j, z = 1, 2 \dots k'$).

El punto estacionario de segundo orden se obtiene derivando parcialmente \widehat{Y} con respecto a x e igualando a cero, es decir,

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

coincidiendo la solución de este para modelos de segundo orden, la cual se puede consultar en Myers & Montgomery (1995).

Para determinar si hay máximo o mínimo, se construye la matriz Hessiana (H) formada con las segundas derivadas, donde las componentes de la diagonal corresponden a la derivadas parciales de segundo orden de cada uno de los niveles de los factores x_j y x_z :

$$\frac{\partial^2(\widehat{Y}_{ht})}{\partial x_j \partial x_j} = 2 \sum_{m=0}^q \widehat{\xi}_{jjhm} P_m(t)$$

De igual manera, se encuentra la expresión para el factor x_z . Las componentes fuera de la diagonal corresponden a la forma generada por

$$\frac{\partial^2(\widehat{Y}_{ht})}{\partial x_j \partial x_z} = \sum_{m=0}^q \widehat{\xi}_{jzhm} P_m(t)$$

Los signos de los valores propios de H determinan la naturaleza del punto estacionario (x_s) de forma que si $\lambda_1, \lambda_2, \dots, \lambda_k$ son todos negativos, es un *máximo*; si todos son positivos, es un *punto de mínima respuesta*; y si tienen signos alternos, es un *punto de silla*.

Al estimar el punto estacionario o la combinación de los niveles que optimizan un proceso, se recurre a la metodología de pendiente en ascenso o descenso cuando se ajusta un modelo de primer orden; o a un análisis canónico, si se ha ajustado un modelo de segundo orden (Myers & Montgomery 1995).

3. Aplicación

En esta sección se presenta un estudio reportado por Frey et al. (1992), en el cual se experimenta el efecto de Sodio Zeolite A (SZA) en el crecimiento y fisiología de 60 caballos. La dieta alimenticia individual se aplicó de manera aleatoria; el tratamiento consta de 4 niveles (0, 0.66, 1.32 y 2%). A los 84 días de haber ingresado a la dieta, se midió la concentración de silicio en el plasma mediante muestras de sangre tomadas a las 0, 1, 3, 6 y 9 horas después de la ingestión de la dieta. Para el caso de aplicación, se hace una adaptación de este estudio eliminando el tiempo 1, con la finalidad de garantizar tiempos igualmente espaciados, aunque la metodología presentada en este artículo puede ser aplicada a tiempos que no sean igualmente espaciados.

La matriz de varianzas y covarianzas se estima, vía PROC MIXED del SAS (Davidian 2005); mediante criterios bayesianos y Akaike se elige la estructura que mejor describa la variación entre las observaciones (Peña 2002). La estimación se

realiza teniendo en cuenta una forma no estructurada (NE), compuesta simétrica (CS) y una autorregresiva de orden 1 ($AR(1)$). La matriz NE resulta ser el mejor modelo, como lo indican los valores resaltados en la tabla 2.

TABLA 2: Análisis de estructura de la matriz Σ .

	NE	SC	$AR(1)$
AIC	807.9	844.5	848.8
$AICC$	808.9	844.6	848.8
BIC	828.8	848.7	853.0

Por tanto, la estimación de $\hat{\Sigma}_{NE}$ es

$$\hat{\Sigma}_{NE} = \begin{bmatrix} 2.8898 & 1.6126 & 1.3595 & 1.5509 \\ 1.6126 & 3.9854 & 2.6601 & 0.8890 \\ 1.3595 & 2.6601 & 2.9863 & 1.1126 \\ 1.5509 & 0.8890 & 1.1126 & 1.6151 \end{bmatrix}$$

La metodología se desarrolla bajo el supuesto de normalidad. Se verifica la calidad del modelo ajustado, analizando la gráfica de probabilidad normal de los residuales y el estadístico de prueba Shapiro y Wilk, detectando que las observaciones de la variable respuesta (concentración de silicio en el plasma) no evidencian normalidad, puesto que el valor p del estadístico es 0.0002; por tal razón se realiza una transformación a las observaciones. Por máxima verosimilitud se estima la transformación apropiada de las observaciones con $\eta = 0.4$. Transformadas las observaciones, se procedió nuevamente a estimar Σ y se obtuvo que la estructura $AR(1)$ describe la variación de las observaciones así:

$$\hat{\Sigma}_{AR(1)} = \begin{bmatrix} 174.3600 & -38.6408 & 8.5635 & -1.8978 \\ -38.6408 & 174.3600 & -38.6408 & 8.5635 \\ 8.5635 & -38.6408 & 174.3600 & -38.6408 \\ -1.8978 & 8.5635 & -38.6408 & 174.3600 \end{bmatrix} \quad (13)$$

Posteriormente, se analiza el conjunto de hipótesis presentado en la tabla 1, las cuales indican el mejor ajuste del modelo.

3.1. Hipótesis de interés y ajuste del modelo

La tabla 3 presenta el primer conjunto de hipótesis dadas en la tabla 1. Las variables tiempo y tratamiento explican la respuesta; así mismo, los tratamientos y el tiempo están interactuando, y el crecimiento y la fisiología de los caballos depende de la concentración SZA a través del tiempo. Este primer conjunto de hipótesis, se rechazan al 5% de significancia; el estadístico de prueba F calculado (F_c) y el tabulado (F_T), para todos los casos, es $F_c > F_T$ (valor $p = 0.000$).

Continuando con el análisis de las hipótesis contenidas en la tabla 1, se presenta inicialmente un análisis marginal para ver el grado de ajuste del polinomio en la curva, luego en la superficie; posteriormente, el ajuste conjunto para las dos metodologías.

TABLA 3: Sumas de cuadrados asociadas a las hipótesis planteadas en la tabla 1.

Hipótesis	gl_{H_0}	SCH_0	F_c	F_T	valor p
$H_0^{(1)}$	16	8.531	1441.43	1.69	0.0000
$H_0^{(2)}$	9	6.523	1959.39	1.92	0.0000
$H_0^{(3)}$	3	2.039	1837.47	2.64	0.0000
$H_0^{(4)}$	3	3.059	2756.88	2.64	0.0000

En la tabla 4, se muestra el conjunto de hipótesis que indican el grado de ajuste para la curva de crecimiento. Con un nivel de significancia de 5 %, existe evidencia estadística que valida el ajuste de un polinomio de primer o de segundo orden para el tiempo (*valor $p < 0.05$*).

TABLA 4: Suma de cuadrados para el ajuste de la curva de crecimiento.

H_{05}	$ran(H_0)$	SCH	F_c	valor p
3^{er}	4	0.0024	1.62	0.1697
2^{do}	4	0.0036	2.43	0.0483
1^{ro}	4	0.0249	16.83	0.0000

Posteriormente, se valida el ajuste del grado del polinomio para la superficie de respuesta de manera marginal, teniendo en cuenta que a lo sumo puede ser un polinomio de tercer orden. Según la tabla 5, existe evidencia estadística suficiente para ajustar una superficie de primer o segundo orden en el tratamiento (*valor $p < 0.05$*).

TABLA 5: Suma de cuadrados para el ajuste del grado del tratamiento.

H_{06}	$ran(H_0)$	SCH	F_c	valor p
3^{er}	4	0.0007	0.45	0.5023
2^{do}	4	0.0325	21.96	0.0000
1^{ro}	4	0.1518	102.59	0.0000

Una vez verificado este conjunto de hipótesis, se analiza el ajuste conjunto del tratamiento y tiempo; teniendo en cuenta las hipótesis que se presentan en la tabla 6, hay evidencia estadística suficiente para ajustar un modelo conjunto de primer orden en el tiempo y el tratamiento (*valor $p = 0.000209$*).

El modelo finalmente se obtiene a partir de la estimación de Σ en (13) y del conjunto de hipótesis de la tabla 6, el cual sugiere que el mejor ajuste corresponde a un modelo de primer orden, que en términos de las variables codificadas es

$$\begin{aligned}
 \hat{Y}_{ht}^* &= \sum_{m=0}^1 \left(\hat{\xi}_{0hm} + \hat{\xi}_{1hm}x \right) P_m(t) \\
 &= \left(\hat{\xi}_{0h0} + \hat{\xi}_{1h0}x \right) P_0(t) + \left(\hat{\xi}_{0h1} + \hat{\xi}_{1h1}x \right) P_1(t) \\
 &= (2.0599 + 0.3647x)P_0(t) + (0.0507 + 0.0352x)P_1(t) \quad (14)
 \end{aligned}$$

donde $\hat{Y}^* = \hat{Y}^{0.4}$ es la variable respuesta predicha transformada, x representa el efecto de tratamiento codificado y los polinomios ortogonales están dados por

TABLA 6: Suma de cuadrados para el ajuste conjunto de la superficie y curva de crecimiento.

H_{06}	$ran(H_0)$	SCH	F_c	valor p
3 ^{er} y 3 ^{er}	1	0.00014	0.00014	0.5346
3 ^{er} y 2 ^{do}	1	0.00018	0.00019	0.4783
3 ^{er} y 1 ^{ro}	1	0.00000	0.00000	0.9825
2 ^{do} y 3 ^{er}	1	0.00000	0.00000	0.9272
2 ^{do} y 2 ^{do}	1	0.00001	0.02520	0.8740
2 ^{do} y 1 ^{ro}	1	0.00130	3.51437	0.0621
1 ^{ro} y 3 ^{er}	1	0.00011	0.30980	0.5783
1 ^{ro} y 2 ^{do}	1	0.00004	1.07937	0.2999
1 ^{ro} y 1 ^{ro}	1	0.00530	14.20940	0.0002

$P_0(t) = 1$, $P_1(t) = 2t - 9$, $P_2(t) = t^2 - 9t + 9$ y $P_3(t) = 10t^3 - 135t^2 + 423t - 324$, donde t representa el efecto del tiempo codificado. Sustituyendo estos polinomios en (14), el modelo ajustado es

$$\widehat{Y}_{ht}^* = 1.6036 + 0.1014t + (0.0479 + 0.0704t)x \quad (15)$$

El modelo ajustado presenta la estructura de los modelos empleados en la MSR; en este sentido, el intercepto (dos primeros términos) y el coeficiente de la variable regresora (último término) corresponden a curvas de crecimiento de primer orden.

Al decodificar las variables en el modelo (15) y dejarlas en las variables naturales, es decir, haciendo $t = \frac{T-4.5}{3}$ y $x = \tau - 1$, se obtiene:

$$\widehat{Y}_{ht}^* = 1.5092 + 0.010334T + (0.023466T - 0.0577)\tau$$

donde T y τ son los efectos de tiempo y tratamiento, respectivamente.

3.2. Optimización del proceso

La caracterización del óptimo generada por la metodología propuesta es un polinomio de primer orden en el tiempo. Observando la figura 1, la superficie presenta una curvatura debido a la interacción tiempo y tratamiento, la cual indica que la concentración de SZA que maximiza la concentración de silicio en el plasma depende del tiempo, pero no se encuentra un máximo absoluto. Esto se ratifica al revisar el gráfico de contornos.

Como el modelo ajustado (15) corresponde a un modelo de primer orden, el óptimo se obtiene siguiendo la metodología de pendiente en máximo ascenso o descenso (Myers & Montgomery 1995). Los coeficientes del modelo ajustado presentan signos iguales, indicando un análisis de pendiente en máximo ascenso.

Teniendo en cuenta (15), la trayectoria del ascenso más pronunciado pasa por el punto $x = 0$ y su pendiente es

$$0.0479 + 0.0704t \quad (16)$$

que está en función del tiempo.

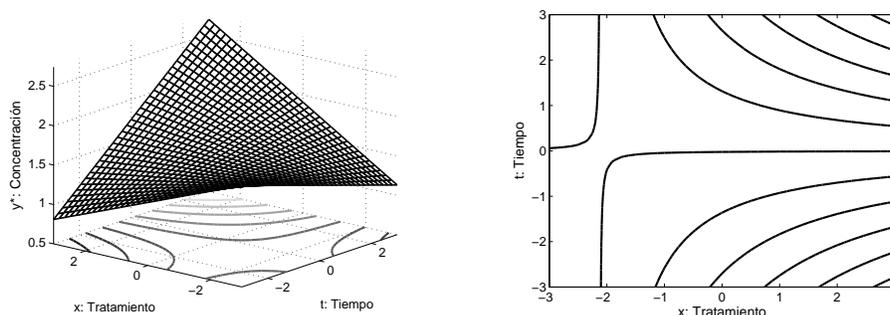


FIGURA 1: Superficie de respuesta y gráfico de contorno para las variables codificadas.

Como existe un solo factor (tratamiento), las unidades de movimiento en la trayectoria de máximo ascenso dependen únicamente del efecto de Sodio Zeolite y, según Myers & Montgomery (1995), no se puede determinar el máximo puesto que la forma general es

$$x = \frac{0.0479 + 0.0704t}{2\xi}$$

y como esta es independiente del valor que tome ξ , no se obtiene un punto estacionario, pues siempre se estaría adicionando una constante, la cual se determina de acuerdo con el conocimiento del experimentador y las condiciones experimentales.

En esta aplicación no es recomendable modelar la variable respuesta utilizando la MSR clásica, pues no se tendría en cuenta la correlación presente entre las mediciones a través del tiempo del mismo individuo, lo cual se considera en la metodología planteada en este artículo.

4. Conclusiones

La metodología presenta una alternativa de análisis que facilita la manera de modelar de forma univariada la superficie de respuesta (tratamientos) y la curva de crecimiento (tiempos), permitiendo encontrar la combinación óptima de tratamientos a través del tiempo. Esta metodología tiene la ventaja de proporcionar estimaciones sencillas debido a que no se recurre a la transformación planteada por Potthoff & Roy (1964) para el trabajo del modelo clásico multivariado.

En el modelamiento de superficies de respuesta a través del tiempo, se recurre a cada una de las alternativas propuestas para el manejo de curvas de crecimiento, empleado para modelar el caso de estudio. El análisis de perfiles fue tenido en cuenta al plantear las hipótesis; el análisis multivariado, debido a la semejanza en el comportamiento de los parámetros, fue el punto de partida para el planteamiento del caso univariado en forma matricial.

La caracterización del óptimo está dada en función de un polinomio en el tiempo y este refleja el grado de la curva de crecimiento.

Agradecimientos

Agradecemos a los evaluadores por sus valiosas y oportunas observaciones que permitieron mejorar el artículo. Este trabajo está enmarcado dentro del proyecto de investigación Estadística aplicada a la investigación experimental, industria y biotecnología, de la Universidad Nacional de Colombia.

[Recibido: febrero de 2008 — Aceptado: septiembre de 2008]

Referencias

- Box, G. E. P. & Draper, N. R. (1982), 'Measures of Lack of Fit for Response Surface Designs and Predictor Variable Transformations', *Tecnometrics* **24**, 1–8.
- Box, G. E. P. & Jones, S. (1992), 'Split Plot Designs for Robust Product Experimentation', *Journal Applied Statistics* **19**, 3–26.
- Box, G. E. P. & Jones, S. (2001), 'Split Plot Designs for Robust Process Experimentation', *Quality Engineering* **13**, 124–127.
- Box, G. E. P. & Wilson, K. B. (1951), 'On the Experimental Attainment of the Optimum Conditions', *Journal Royal of the Statistical Society* **13**, 1–45.
- Davidian, M. (2005), *Applied Longitudinal Data Analysis*, Chapman and Hall, North Carolina, United States.
- Davis, C. (2002), *Statistical Methods for the Analysis of Repeated Measurements*, Springer, New York, United States.
- Draper, N. & Ying, L. H. (1994), 'A Note on Slope Rotatability Over all Directions', *Journal of Statistical Planning and Inference* **41**, 113–119.
- Frey, K. S., Potter, G. D., Odom, T. W., Senior, M. A., Reagan, V. D., Weir, V. H., Elsslander, R. V. T., Webb, M. S., Morris, E. L., Smith, W. B. & Weigand, K. E. (1992), 'Plasma Silicon and Radiographic Bone Density on Weanling Quarter Horses Fed Sodium Zeolite A. J.', *Equine Vet. Sci.* **12**, 292–296.
- Grizzle, J. E. & Allen, D. M. (1969), 'Analysis of Growth an Dose Response Curves', *Biometrics* **25**, 357–381.
- Guerrero, S. C. (2006), Optimización de un proceso a través de superficies de respuesta en curvas de crecimiento, Tesis de maestría, Departamento de Estadística, Universidad Nacional de Colombia.
- Hill, W. J. & Hunter, W. G. (1966), 'A Review of Response Surface Methodology: A Literature Review', *Technometrics* **8**, 571–590.
- Hinkelmann, K. & Kempthorne, O. (1994), *Design and Analysis of Experiments. Vol I. Introduction to Experimental Design*, John Wiley & Sons, New York, United States.

- Khatri, C. A. (1966), 'A Note on a Manova Model Applied to Problems in Growth Curves', *Ann. Inst. Stat. Math* **18**, 75–86.
- Khatri, C. A. (1973), 'Testing Some Covariance Structures Under Growth Curve Model', *Journal Multivariate Analysis* **3**, 102–116.
- Kshirsagar, A. M. & Boyce, S. (1995), *Growth Curves*, Marcel Dekker, New York, United States.
- Laird, N. & Ware, J. (1982), 'Random-Effects Models for Longitudinal Data', *Biometrics* **38**, 963–974.
- Lee, J. C. (1991), 'Growth Curve Models and Technological Forecasting', *Statistical Theory and Data Analysis* **II**, 369–379.
- Lucas, J. M. (1976), 'Which Response Surfaces Is Best?', *Technometrics* **18**, 411–417.
- Magnus, J. R. (1994), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, New York, United States.
- Mead, R. & Pike, D. J. (1975), 'A Review of Responses Surface Methodology from a Biometric Viewpoint', *Biometrics* **31**, 830–851.
- Montgomery, D. C. (2003), *Diseño y análisis de experimentos*, segunda edn, Grupo Editorial Limusa, S. A., México.
- Myers, R. H. & Montgomery, D. C. (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, New York, United States.
- Myers, R. H., Montgomery, D. C. & Vinning, G. G. (2002), *Generalized Linear Models. With Applications in Engineering and the Sciences*, John Wiley & Sons, New York, United States.
- Peña, D. (2002), *Análisis de datos multivariantes*, McGraw-Hill, España.
- Potthoff, R. & Roy, S. N. (1964), 'A Generalized Multivariate Analysis of Variance Models Useful Specially for Growth Curves Problems', *Biometrika* **51**, 313–326.
- Rao, C. R. (1959), 'Some Problems Involving Linear Hypothesis in Multivariate Analysis', *Biometrika* **46**, 49–58.
- Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, United States.
- Rao, C. R. (1966), Covariance Adjustment and Related Problems in Multivariate Analysis, in Proceedings of an International Symposium on Multivariate Analysis, pp. 87–103.

- Reinsel, G. (1982), 'Multivariate Repeated Measurements for Growth Curve Models with Multivariate Random Effects Covariance Structure', *Journal of the American Statistical Association* **77**, 190–210.
- Reinsel, G. (1984), 'Estimation and Prediction in a Multivariate Random Effects Generalized Linear Model', *Journal of the American Statistical Association* **77**, 190–210.
- Rivera, J. C., Ortiz, A. F. & Melo, O. O. (2006), Un enfoque multivariado de curvas de crecimiento y superficies de respuesta, Trabajo de grado, Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de Estadística.
- Singer, J. & Andrade, D. (1994), 'Uso de transformaciones en análisis de varianza y análisis de regresión', *Royal Statistical Society* **2**, 259–256.
- Srivastava, M. (2001), *Nested Growth Curves Models*, Special Issue.
- Srivastava, M. & Katri, C. A. (1979), *An Introduction to Multivariate Statistics*, North-Holland, New York, United States.
- Thompson, W. O. (1973), 'Secondary Criteria in the Selection of Minimum Bias Design in Two Variables', *Technometrics* **2**, 319–328.
- Verbyla, A. P. & Venables, W. N. (1988), 'An Extension of the Growth Curve Models', *Biometrika* **75**, 129–138.