



Universidad Autónoma de Querétaro
Facultad de Ingeniería

MODELOS LINEALES FUNCIONALES

APLICACIONES EN LA TECNOLOGÍA DE
ALIMENTOS

TESIS

Que para obtener el título de

Licenciada en Matemáticas Aplicadas

Presentan

Cristina Gorrostieta Hurtado

Jacquelin Chamorro Jiménez

Dirigida por

Dr. Eduardo Castaño Tostado

Santiago de Querétaro, Qro. 2004

No. Adq. C.H.69506

No. Título _____

Clas. 15

511.8

G 673m



Universidad Autónoma de Querétaro
 Facultad de Ingeniería
 Licenciatura en Matemáticas Aplicadas

MODELOS LINEALES FUNCIONALES Y SUS APLICACIONES EN LA
 TECNOLOGIA DE ALIMENTOS

TESIS

Que para obtener el grado de:

Licenciada en Matemáticas Aplicadas

Presentan:

Cristina Gorrostieta Hurtado
 Jacquelin Chamorro Jiménez

Dirigida por:

Dr. Eduardo Castaño Tostado

SINODALES

Dr. Eduardo Castaño Tostado
 Presidente

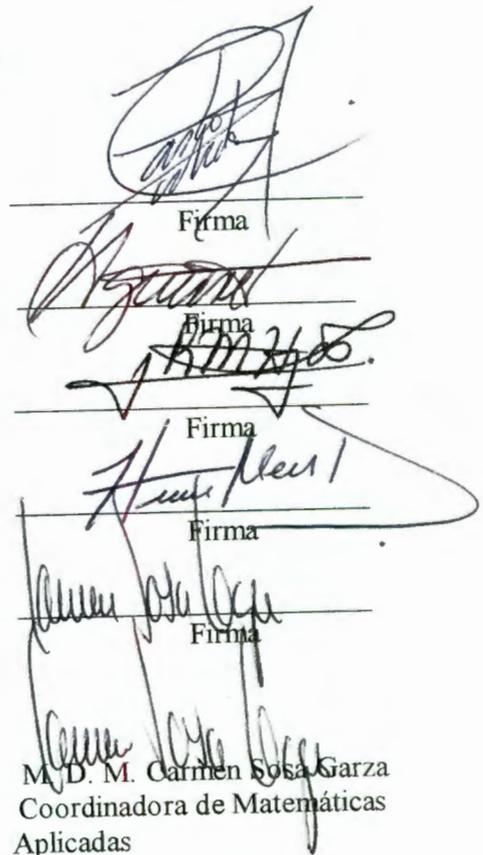
Dr. Víctor Manuel Armando Aguirre Torres
 Secretario

Mat. Rocío Minerva Hidalgo Flores
 Vocal

M. Héctor Mendieta Martínez
 Suplente

M. D. M. Carmen Sosa Garza
 Suplente

M. en I. Gerardo René Serrano Gutiérrez
 Director de la Facultad de Ingeniería



Firma
 Firma
 Firma
 Firma
 Firma

M. D. M. Carmen Sosa Garza
 Coordinadora de Matemáticas
 Aplicadas

Handwritten scribble or signature



Acuerdo **170/04**
25 de agosto del 2004

C. CRISTINA GORROSTIETA HURTADO Y

C. JACQUELIN CHAMORRO JIMÉNEZ

Pasantes de la Licenciatura de Matemáticas Aplicadas

Presente,

Con relación a su solicitud enviada al H. Consejo Académico de la Facultad en la que solicitan autorización para titularse por la opción de **Tesis Colectiva**, me permito informarles que en sesión ordinaria del 25 de agosto del año en curso, este Cuerpo Colegiado, acordó autorizar la opción de titulación y deberán trabajar con el tema : **“Modelos Lineales Funcionales y sus aplicaciones en la Tecnología de Alimentos”**, bajo la dirección del Dr. Eduardo Castaño Tostado.

El contenido aceptado por el H. Consejo Académico es el siguiente:

CONTENIDO:

1. Introducción

- 1.1 ¿Qué son los datos funcionales?
- 1.2 Algunos ejemplos de datos funcionales
- 1.3 Objetivos de Análisis Funcional de datos
- 1.4 Primeros pasos en el Análisis Funcional
- 1.5 El uso de derivadas en Análisis Funcional

2. Modelos lineales

- 2.1 Introducción a diseño de experimentos
- 2.2 Modelo lineal
- 2.3 Diseños estadísticos

3. Representación funcional de datos

- 3.1 Introducción
- 3.2 Métodos de funciones base
- 3.3 Splines
- 3.4 Suavizamiento de datos funcionales
- 3.5 Penalización de rugosidad

4. Modelos lineales funcionales

- 4.1 Introducción
- 4.2 Modelos funcionales lineales y ANOVA funcional
- 4.3 Modelo lineal funcional. Caso 1: variable dependiente funcional y variables independientes
- 4.4 multivariadas.
- 4.5 Modelo lineal funcional. Caso 2: variable dependiente escalar y variables independientes funcionales.
- 4.6 Conclusiones

5. Aplicación al modelo lineal funcional. Caso 1

- 5.1 Introducción
- 5.2 Aplicación a un experimento con ajos
- 5.3 Aplicación a un experimento con yogurt de GERBER®
- 5.4 Conclusiones

6. Aplicación al modelo lineal funcional. Caso 2

- 6.1 Introducción



- 6.2 Aplicación a un experimento con yogurt de GERBER®
- 6.3 Conclusiones
- 7. Aspectos computacionales**
- 7.1 Instrucciones en S-Plus para Análisis Funcional de datos
- 7.2 Ajuste del modelo lineal funcional en S-Plus
- 7.3 Programas realizados en S-Plus para análisis de aplicaciones
- 8. Conclusiones**

REFERENCIAS BIBLIOGRÁFICAS

- Ramsay J. O., Silverman B. W. (1997)
Applied Functional Data Analysis
Springer Verlag, New York
- Ramsay J. O., Silverman B. W.
Applied Functional Data Analysis
Methods and Case Studies
Springer Series in Statistics
- Trevor Hastie, Robert Tibshirani, Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
- Wu C. F. J. & Hamada M., (2000).
Experiments: Planning, Analysis, and Parameter Design Optimization
Jonh Wiley & Sons. INC.

También hago de su conocimiento las disposiciones de nuestra Facultad, en el sentido de que antes del Examen Profesional deberán cumplir con los requisitos de nuestra Legislación y deberán imprimir el presente oficio en todos los ejemplares de su tesis.

Sin más por el momento, quedo de Ustedes.

Atentamente,

“ EL INGENIO PARA CREAR, NO PARA DESTRUIR ”

M. en I. Gerardo René Serrano Gutiérrez
Director de la Facultad

c.c.p. archivo.
*GRSG/RRPV/img.

Agradecimientos

Al Dr. Eduardo Castaño Tostado por su dirección y apoyo.

A Conacyt por su colaboración en este proyecto mediante el fondo 36616-E.

Contenido

Resumen	1
1 INTRODUCCIÓN	2
1.1 ¿QUÉ SON LOS DATOS FUNCIONALES?	2
1.2 ALGUNOS EJEMPLOS DE DATOS FUNCIONALES	3
1.2.1 Mediciones de altura durante el crecimiento	3
1.2.2 Mediciones de temperatura en 35 ciudades de Canadá	5
1.2.3 Datos tomados durante el ciclo de gateo de niños	8
1.3 OBJETIVOS DE ANÁLISIS FUNCIONAL DE DATOS	10
1.4 PRIMEROS PASOS EN ANÁLISIS FUNCIONAL DE DATOS	11
1.4.1 Representación de datos: suavizamiento e interpolación	11
1.4.2 Resumen de estadísticas para datos funcionales	11
1.5 USO DE DERIVADAS EN ANÁLISIS FUNCIONAL DE DATOS	12
2 MODELOS LINEALES	14
2.1 INTRODUCCIÓN A DISEÑO DE EXPERIMENTOS	14
2.2 MODELO LINEAL	21
2.3 DISEÑOS EXPERIMENTALES	27
2.3.1 Diseño con un solo factor	28
2.3.2 Diseños factoriales con dos niveles	30

3 REPRESENTACIÓN FUNCIONAL DE DATOS	35
3.1 INTRODUCCIÓN	35
3.1.1 Muestreo y error observacional suavizamiento	36
3.1.2 Suavizamiento lineal	37
3.2 MÉTODOS DE FUNCIONES BASE	38
3.2.1 Determinación de c_k por mínimos cuadrados	39
3.2.2 Eligiendo una buena base	41
3.2.3 Series de Fourier	42
3.2.4 Base polinomial	44
3.2.5 Bases "wavelets"	45
3.2.6 Recomendaciones finales	47
3.3 SPLINES	49
3.4 PENALIZACIÓN DE RUGOSIDAD	52
3.4.1 Introducción	52
3.4.2 Spline suavizado	53
3.4.3 Estimación de derivadas mediante spline suavizado	55
3.4.4 Penalización de rugosidad generalizada	55
3.4.5 Enfoque de bases regularizadas	56
3.4.6 Algunas propiedades de los estimados	58
3.4.7 Relación con el enfoque de penalización de rugosidad	59
4 MODELOS LINEALES FUNCIONALES	60
4.1 INTRODUCCIÓN	60
4.2 MODELO FUNCIONAL LINEAL CASO 1	61

4.2.1	Estimación del vector de parámetros β	63
4.2.2	Aspectos computacionales	64
4.3	MODELO LINEAL FUNCIONAL CASO 2	68
4.3.1	Introducción	69
4.3.2	Regularización usando funciones base	70
4.3.3	Regularización con penalización de rugosidad	70
4.3.4	Aspectos computacionales de regularización	71
4.3.5	Método directo de penalización para calcular β	73
4.3.6	Validación cruzada	77
5	APLICACIONES AL MODELO LINEAL FUNCIONAL	79
5.1	INTRODUCCIÓN	79
5.2	APLICACIÓN UNO AL MODELO LINEAL FUNCIONAL CASO 1	79
5.3	PRUEBAS DE PERMUTACIÓN	94
5.3.1	Introducción	94
5.3.2	Pruebas de permutación	98
5.3.3	Aplicación	98
5.4	APLICACIÓN DOS AL MODELO LINEAL FUNCIONAL CASO 1	101
5.5	APLICACIÓN AL MODELO LINEAL FUNCIONAL CASO 2	121
6	ASPECTOS COMPUTACIONALES	127
6.1	INTRODUCCIÓN	127
6.2	INSTRUCCIONES EN S-PLUS PARA ANÁLISIS FUNCIONAL DE DATOS	127
6.2.1	create.bspline.basis	128

6.2.2	create.fd	129
6.2.3	data2fd	64.U.130
6.2.4	smooth.fd	131
6.2.5	plot.fd	132
6.2.6	par	133
6.2.7	key	134
6.2.8	eval.fd	134
6.2.9	apply	135
6.2.10	deriv	135
6.2.11	inprod	136
6.2.12	mean.fd	136
6.2.13	linmod.fd	137
6.3	PROGRAMAS REALIZADOS EN S-PLUS	138
6.3.1	Programas utilizados en las aplicaciones al caso 1	138
6.3.2	Programa utilizado en la aplicación al caso 2	148
6.4	AJUSTE DEL MODELO FUNCIONAL LINEAL EN S-PLUS	149
7	Conclusiones	159
	Bibliografía	161

Índice de Figuras

1.1	Patrón de crecimiento de 10 niños	4
1.2	Aceleración estimada de las funciones de altura	5
1.3	Patrones de temperatura de 35 ciudades de Canadá	6
1.4	Funciones ajustadas de los patrones de temperatura	7
1.5	Ángulos formados por cadera y rodilla durante el ciclo de gateo	8
1.6	Ciclo de gateo de un niño	9
2.1	Pasos para la planeación de un experimento	17
3.1	Primeras cinco funciones base de la serie de Fourier	43
3.2	Un ejemplo de wavelet de la familia Haar	45
3.3	Wavelets en diferente traslación y escala	46
3.4	Funciones base B-spline de grado 3.....	49
3.5	Splines de orden lineal, cuadrático y cuártico	50
5.1	Pérdida de peso con el tratamiento 1 de la réplica 1	81
5.2	Funciones ajustadas de pérdida de peso	86
5.3	Funciones modeladas de pérdida de peso	87
5.4	Funciones R^2 y cociente F del experimento de pérdida de peso en ajos	88
5.5	Funciones de regresión para el experimento de pérdida de peso en ajos	89
5.6	Funciones ajustadas de velocidad de pérdida de peso	90
5.7	Funciones R^2 y cociente F del experimento de velocidad de pérdida de peso en ajos	91

5.8	Funciones de regresión para el experimento de velocidad de pérdida de peso en ajos	92
5.9	Funciones modeladas para el experimento de velocidad de pérdida de peso en ajos	93
5.10	Jugo con yogurt dividido en fases	101
5.11	Máquina utilizada para medir los tamaños de partícula	104
5.12	Funciones ajustadas de volumen en el experimento de yogurt	110
5.13	Funciones ajustadas de volumen en el experimento de yogurt (segundo intervalo)	111
5.14	Funciones modeladas en el experimento de yogurt	112
5.15	Funciones modeladas en el experimento de yogurt (segundo intervalo)	113
5.16	Funciones de R^2 y cociente F en el experimento de yogurt	114
5.17	Funciones de R^2 y cociente F en el experimento de yogurt (segundo intervalo)	115
5.18	Función de regresión de la media en el experimento de yogurt	116
5.19	Función de regresión de la media en el experimento de yogurt (segundo intervalo)	117
5.20	Funciones de regresión con banda de confianza al 95% en el experimento de yogurt	118
5.21	Funciones de regresión con banda de confianza al 95% en el experimento de yogurt (segundo intervalo)	119
5.22	Funciones de regresión para tamaños de partículas pequeños	120
5.23	Promedio de perfiles de pérdida de peso	124
5.24	Función de ponderación $\beta(t)$	125
5.25	Producto de la función $\beta(t)$ con cada uno de los perfiles	126

Resumen

El presente trabajo de tesis trata de dos aplicaciones de Análisis Funcional a la tecnología de alimentos.

Los primeros capítulos pretenden dar una explicación general de la teoría desarrollada en Análisis Funcional, en ellos se utilizó principalmente el material de Ramsay y Silverman mencionado en la bibliografía, debido a que son los precursores más importantes de esta teoría.

En el capítulo 4 se presenta el modelo estadístico funcional, el capítulo 5 contiene las aplicaciones a este modelo. La primera de ellas se dió en la industria del ajo, el objetivo es buscar condiciones para no perder la calidad del ajo durante su almacenamiento. La segunda aplicación se realizó con datos proporcionados por la empresa GERBER a través del Posgrado de Alimentos de la Facultad de Química, en la cual se quiere mejorar la calidad de un jugo de yogurt con frutas.

En una de las aplicaciones se realizan pruebas de hipótesis usando pruebas por aleatorización adecuadas en el contexto de análisis funcional de datos.

Finalmente, se explican los programas realizados en S-Plus para el análisis de las aplicaciones.

1 INTRODUCCIÓN

1.1 ¿QUÉ SON LOS DATOS FUNCIONALES?

Los datos en muchos campos surgen de manera natural así como la necesidad de hacer uso de ellos de forma conveniente para que proporcionen información útil, es en ello donde el análisis estadístico juega un papel fundamental. La forma de manejar estos datos ha sido motivo de estudio de nuevos métodos entre los que destaca el Análisis Funcional de Datos el cuál se estudiará en esta tesis.

La mayoría de análisis estadísticos involucran una o más observaciones tomadas en cada uno de los individuos en una muestra, con el afán de hacer inferencias sobre la población general de la cuál la muestra es tomada. En un creciente número de campos estas observaciones son curvas, que son ejemplos de funciones, por esta razón llamamos a las curvas observadas datos funcionales y al análisis estadístico para analizar estos datos Análisis Funcional de Datos. Análisis Funcional de Datos es también una manera general de pensar donde la unidad básica de información es una función observada más que una sucesión de números.

El Análisis Funcional de Datos es una propuesta relativamente reciente por lo que no es muy utilizada en México, sin embargo muchos datos generados actualmente pueden pensarse como expresiones discretas de funciones suaves subyacentes y es de interés analizar

estas curvas como respuesta ante condiciones cambiantes o como condicionantes de respuestas, es por ello que es importante estudiar esta propuesta.

La estadística multivariada tiene muchos métodos desarrollados utilizados específicamente por el Análisis Funcional de Datos que lo han precedido, pero los cuales no han explotado específicamente la suavidad de los datos funcionales o no han hecho uso de sus derivadas. Por la suavidad de una observación funcional tenemos la posibilidad de que una o más derivadas como velocidad o aceleración puedan ser examinadas o usadas en el proceso de modelado lo cual justifica su estudio.

Además contar con interpretaciones funcionales es más eficiente en el uso de la información que interpretar los datos de manera tradicional (estadística multivariada).

1.2 ALGUNOS EJEMPLOS DE DATOS FUNCIONALES

Los datos funcionales surgen en muchas formas pero su característica principal obviamente es que consisten de funciones, es decir la i -ésima observación es una función real $x_i(t)$, $i = 1, \dots, N$, $t \in \mathcal{T}$, donde \mathcal{T} es un intervalo real. Esta sección presenta ejemplos de este tipo de datos, tomados de [1].

1.2.1 Mediciones de altura durante el crecimiento

La Figura 1.1 muestra el patrón de crecimiento de 10 niños suizos que participaron en un estudio de crecimiento, la altura de cada niño fue medida en 29 etapas, pero es más útil pensar en estos datos como una muestra de 10 curvas o funciones que como un conjunto de 290 observaciones. Las edades en las cuales se midió la altura no son igualmente espa-

ciadas, como puede verse en los puntos que indican las edades en los cuales se realizaron las mediciones.

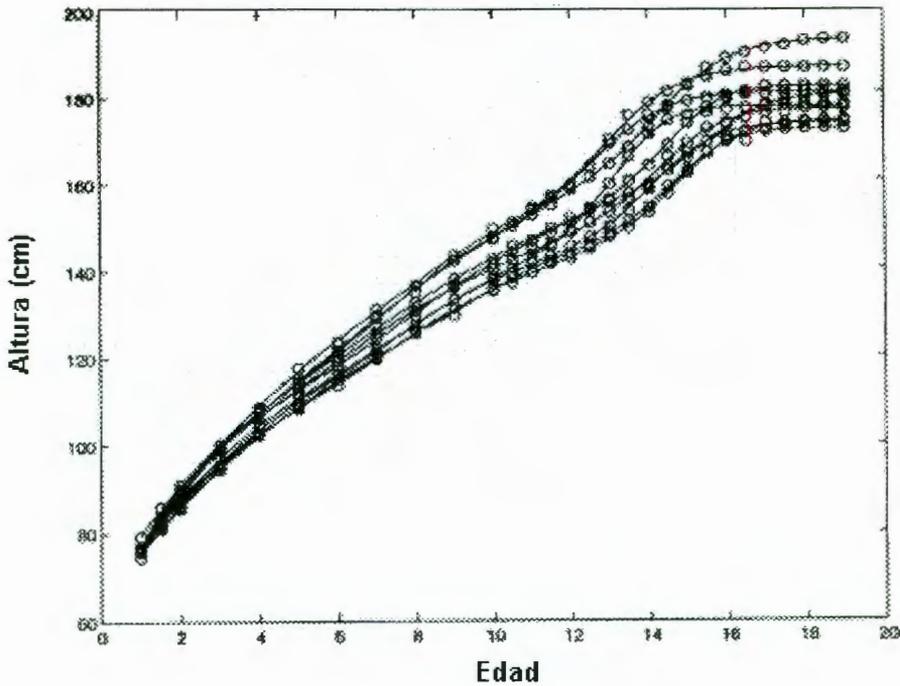


Figura 1.1. Patrón de crecimiento de 10 niños

El estudio del crecimiento humano es importante para establecer un patrón de crecimiento normal y poder detectar posibles problemas en el crecimiento cuando éstos existan, es de particular interés la velocidad y aceleración de crecimiento a lo largo del tiempo (años), pensar en los datos como funciones nos permite estudiar fácilmente estos factores por medio de las derivadas, primera derivada para velocidad y segunda derivada para aceleración de crecimiento. La Figura 1.2 muestra la aceleración estimada de estas 10 funciones de altura, la línea punteada indica el promedio de las 10 curvas. Se observa que alrededor de la pubertad (12-16 años) hay un fuerte pulso en la aceleración, este cambio se puede

apreciar muy poco en la Figura 1.1 pero es muy obvio en la Figura 1.2. Un análisis más profundo de estos datos a través de análisis funcional revelará mayor información.

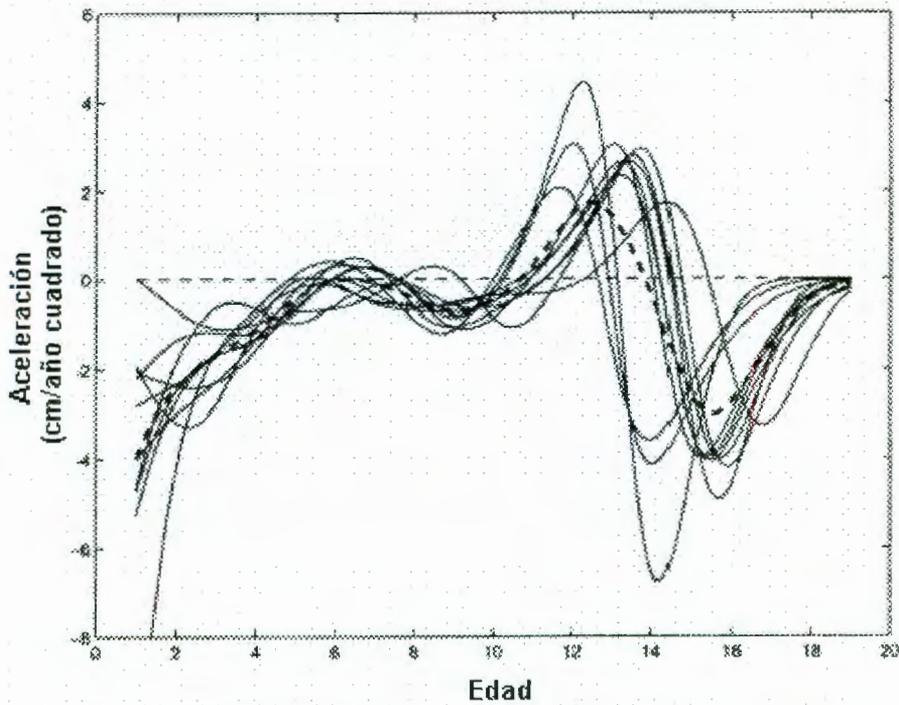


Figura 1.2. Aceleración estimada de las funciones de altura

1.2.2 Mediciones de temperatura en 35 ciudades de Canadá

A continuación se observa, en la Figura 1.3, el cambio de temperatura en el transcurso de un año, en 35 ciudades de Canadá, cada observación marcada con una "x" representa la temperatura media registrada en todo el mes, recolectada en 30 años, los colores indican la zona climática a la que pertenece cada ciudad, ciudades graficadas en rojo corresponden a la zona Atlántico, Continental en azul, Pacífico en verde y Ártico en negro. En este caso puede verse que los puntos de observación si son igualmente espaciados.

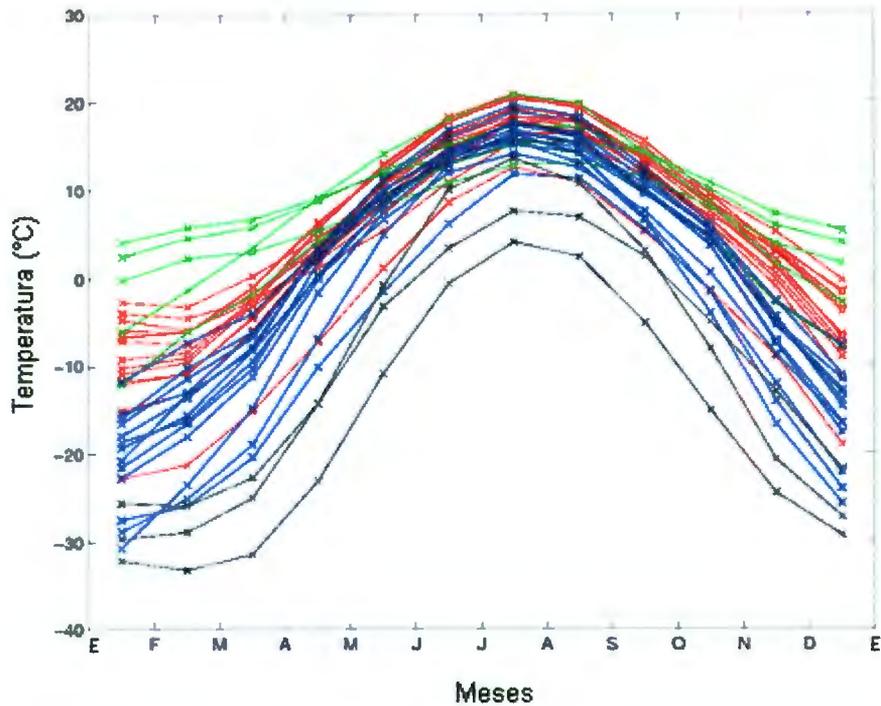


Figura 1.3. Patrones de temperatura de 35 ciudades de Canadá

A cada línea en la gráfica anterior, correspondiente a cada ciudad, es posible ajustarle una función suave de manera que contaremos con 35 funciones ajustadas para utilizar análisis funcional; éstas se muestran en la Figura 1.4.

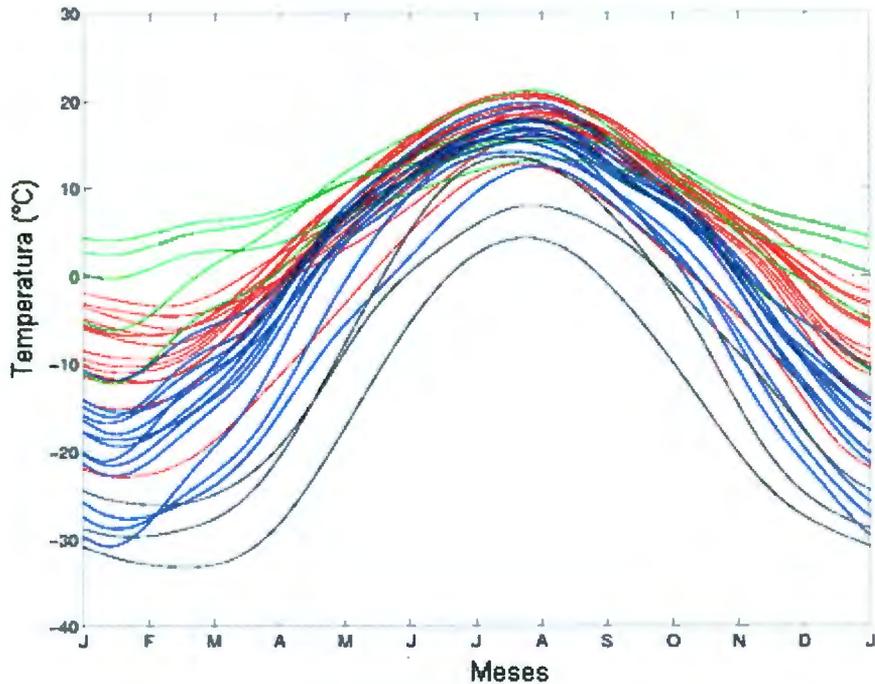


Figura 1.4. Funciones ajustadas de los patrones de temperatura

Mediante el análisis de este tipo de datos podemos ayudar a los climatólogos a responder las siguientes preguntas:

¿Cuáles son los meses en los que el clima es más variable en todo Canadá y cuáles son los meses en el que es más similar?

¿Existe un patrón de temperatura para una ciudad que se encuentre en determinada zona climática?

¿Qué tan influenciado está el clima de determinada ciudad por pertenecer a cierta zona climática?

1.2.3 Datos tomados durante el ciclo de gateo de niños

Los datos funcionales son a menudo multivariados, como en este ejemplo. El laboratorio de análisis de movimiento en los niños en el hospital de San Diego, recolectó los datos que se presentan en la Figura 1.5 que consisten en los ángulos formados por la cadera y la rodilla de 39 niños durante el ciclo de gateo, que comienza y termina cuando la rodilla bajo observación toca el piso, como el tiempo es medido en términos del ciclo de gateo entonces cada curva es dada en términos del argumento $t \in [0, 1]$ (0 inicia ciclo de gateo y 1 termina). Ambos conjuntos de funciones son periódicos.

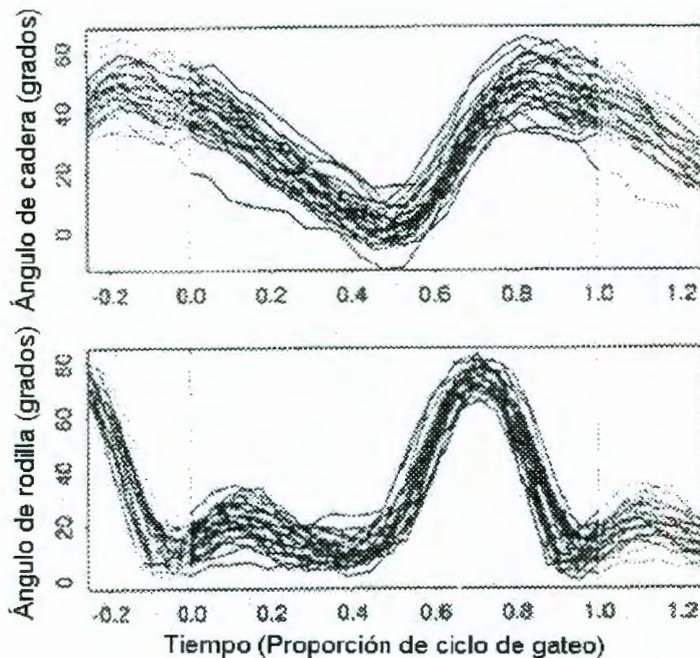


Figura 1.5. Ángulos formados por cadera y rodilla durante el ciclo de gateo

Podemos observar que el ángulo de la rodilla muestra un proceso de dos fases (los valores del ángulo de rodilla se aproximan a cero después aumentan su valor y se vuelven a acercar a cero, este comportamiento se presenta dos veces) mientras que el de la cadera de

una sola fase (los valores del ángulo de cadera decrecen aproximándose a cero y después vuelven a aumentar su valor). Lo que es más difícil observar es la interacción de los dos ángulos. La Figura 1.6 muestra el ciclo de gateo para un solo niño, la naturaleza periódica del proceso ocasiona que se forme una curva cerrada. También se muestra en la figura para propósitos de referencia la curva formada al tomar el promedio sobre los 39 niños (línea punteada). Los puntos indican 20 puntos igualmente espaciados en el ciclo de gateo. Las letras están colocadas en intervalos de un quinto del ciclo, de manera que el ciclo comienza y termina en la letra A.

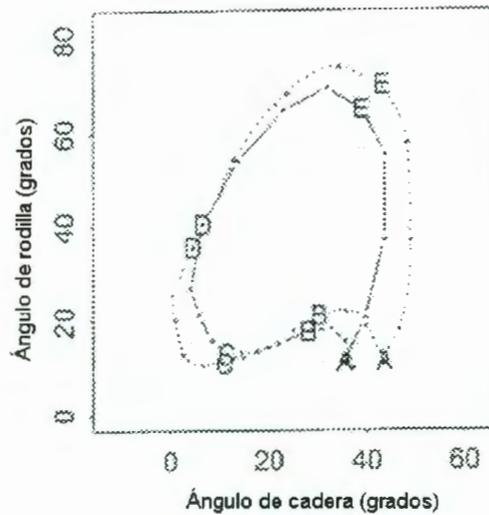


Figura 1.6. Ciclo de gateo de un niño

En la figura puede verse como el gateo de un niño en particular representado por la línea continua difiere del promedio.

1.3 OBJETIVOS DE ANÁLISIS FUNCIONAL DE DATOS

Según [1] las metas de análisis funcional de datos son esencialmente las mismas que la de cualquier otra rama de la estadística, éstas incluyen lo siguiente

- Representar los datos de tal forma que se facilite un análisis más profundo.
- Presentar los datos de tal forma que se puedan apreciar sus características.
- Estudiar las fuentes importantes de variación y patrones presentados en los datos.
- Explicar la variación en la variable dependiente usando la información de las variables independientes.
- Comparar dos o más conjuntos de datos con respecto a ciertos tipos de variación, donde dos conjuntos de datos pueden contener diferentes conjuntos de réplicas de la misma función o diferentes funciones para un conjunto común de réplicas.

El análisis de datos puede verse como exploratorio, confirmatorio o predictivo.

En el modo exploratorio se espera mediante las técnicas revelar aspectos nuevos e interesantes de los datos, este tipo de análisis utiliza sólo la información o los datos que se tienen disponibles.

El análisis confirmatorio tiende a ser inferencial y a ser determinado por preguntas específicas en los datos, en él se supone un tipo de estructura presente en los datos y se desea saber si ciertas afirmaciones o hipótesis se pueden confirmar mediante éstos.

Los estudios predictivos son menos comunes y se centran principalmente en utilizar los datos a la mano y hacer afirmaciones acerca de estados no observados como el futuro.

1.4 PRIMEROS PASOS EN ANÁLISIS FUNCIONAL DE DATOS

1.4.1 Representación de datos: suavizamiento e interpolación

Supongamos que un dato funcional observado de la unidad i es un conjunto de valores de medidas discretas, $y_{i1}, y_{i2}, \dots, y_{in}$; el primer paso es convertir dichos valores a una función x_i con valores $x_i(t)$ calculable para cualquier valor del argumento t . Si los valores discretos se suponen sin error, entonces el proceso es *interpolación*, pero si las observaciones presentan algún error, el método de conversión de los datos discretos a funciones es *suavizamiento*. Los detalles de esta representación se presentan en el capítulo 3.

1.4.2 Resumen de estadísticas para datos funcionales

Supongamos que contamos con N datos funcionales $\{x_i(t)\}_{i=1}^N$. En [1] se proponen las siguientes estadísticas base.

La función de la media es el promedio de las réplicas de las funciones puntualmente.

$$\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t).$$

Similarmente, la función de varianza es

$$\text{var}_X(t) = (N - 1)^{-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)]^2$$

y la función de la desviación estándar es la raíz cuadrada de la varianza.

La función de covarianza resume la dependencia entre diferentes valores de argumentos, y se calcula para todo t_1 y t_2 como

$$cov_X(t_1, t_2) = (N - 1)^{-1} \sum_{i=1}^N \{x_i(t_1) - \bar{x}(t_1)\} \{x_i(t_2) - \bar{x}(t_2)\}.$$

La función de correlación es

$$corr_X(t_1, t_2) = \frac{cov_X(t_1, t_2)}{\sqrt{var_X(t_1)var_X(t_2)}}.$$

En general, si tenemos pares de observaciones funcionales (x, z) la manera en que dependen una de la otra puede ser cuantificada por medio de la función de covarianza cruzada

$$cov_{X,Z}(t_1, t_2) = (N - 1)^{-1} \sum_{i=1}^N \{x_i(t_1) - \bar{x}(t_1)\} \{z_i(t_2) - \bar{z}(t_2)\},$$

o por la función de correlación cruzada

$$corr_{X,Y}(t_1, t_2) = \frac{cov_{X,Y}(t_1, t_2)}{\sqrt{var_X(t_1)var_Y(t_2)}}.$$

1.5 USO DE DERIVADAS EN ANÁLISIS FUNCIONAL DE DATOS

En la sección 1.2 presentamos un ejemplo en que las derivadas y en general los operadores lineales diferenciales son útiles en análisis de datos. El uso de derivadas es importante para extender el rango de métodos gráficos exploratorios simples y en el desarrollo de metodologías más detalladas.

Indicios útiles en los procesos que generan datos funcionales se pueden encontrar en la relación entre derivadas. Graficando la derivada más alta contra la menor es informativo,

ya que podemos encontrar desviaciones de linealidad y porque la diferenciación puede exponer efectos no vistos fácilmente en las funciones originales.

2

MODELOS LINEALES

El propósito de este capítulo es presentar de manera muy general los conceptos básicos de diseño de experimentos y modelos lineales en su manera convencional para después en capítulos siguientes extender algunos de estos conceptos a la forma funcional.

2.1 INTRODUCCIÓN A DISEÑO DE EXPERIMENTOS

Una de las herramientas más usadas para entender y mejorar el funcionamiento de un sistema es la experimentación ya que ésta ayuda a comprender al investigador los cambios que ocurren en las salidas o respuestas de un sistema cuando los valores de las variables de entrada se cambian, en estos casos el análisis estadístico es de gran utilidad para analizar resultados.

Un sistema puede ser un producto o un proceso, por ejemplo los procesos de manufactura, el desarrollo de un producto, la realización de un experimento en un laboratorio de cualquier rama Biología, Ingeniería o Ciencias Físicas.

Las variables de entrada, comunmente llamadas *factores*, son las variables que son estudiadas en el experimento para conocer su efecto en la *variable respuesta*, que es la variable de interés en el experimento que se mide o se registra ante cada cambio y que generalmente se desea maximizar, minimizar o acercar a un valor preestablecido.

Es importante definir antes de la realización de un experimento cuál es la variable respuesta y en qué unidades se medirá. Naturalmente esta variable está condicionada a

los objetivos del experimento y debe tener la posibilidad de que refleje las condiciones de interés en el sistema experimental estudiado. Por ejemplo si se quiere detectar una diferencia de 0.005 gramos en la respuesta no es apropiado tomar medidas con precisión cercana a un gramo.

La elección de los factores en la planeación del experimento es de gran importancia para el éxito del experimento. Es importante también identificar los factores que sean difíciles de cambiar y los factores de ruido ya que estos nos permitirán tener un mayor control sobre el experimento y reducir la variabilidad no planificada.

A los diferentes valores de un factor que se tendrán en cuenta en la realización del experimento se les llama *niveles*, y a la combinación de niveles de factores *tratamientos*.

Diremos que un factor es cuantitativo cuando éste toma valores sobre un rango continuo y cualitativo cuando el factor toma valores sobre un rango discreto.

Ejemplos de factores cualitativos y sus niveles respectivos son: tipo de máquina (diferentes tipos o marcas de máquinas), trabajador (los trabajadores encargados de hacer una tarea), un aditivo químico (diferentes tipos de aditivos químicos), el género (hombre, mujer). Un ejemplo de factor cuantitativo es la temperatura, supongamos que ésta puede tomar valores en un rango continuo de 100 a 200 grados centígrados en cierto experimento, para elegir dos niveles de temperatura se podría elegir cualquier par de números en el intervalo (100, 200), por ejemplo 125 y 175 ó 100 y 200. Los niveles del factor deben elegirse suficientemente alejados para que el efecto que produce cambiarlos se detecte pero no tan lejanos para que se involucren diferentes mecanismos físicos. Cuando solamente se espera

un efecto lineal en un factor dos niveles son suficientes pero cuando se espera curvatura en el efecto del factor es mejor elegir al menos tres niveles.

El siguiente diagrama muestra siete pasos que el investigador debe seguir para la planeación e implementación de un experimento.

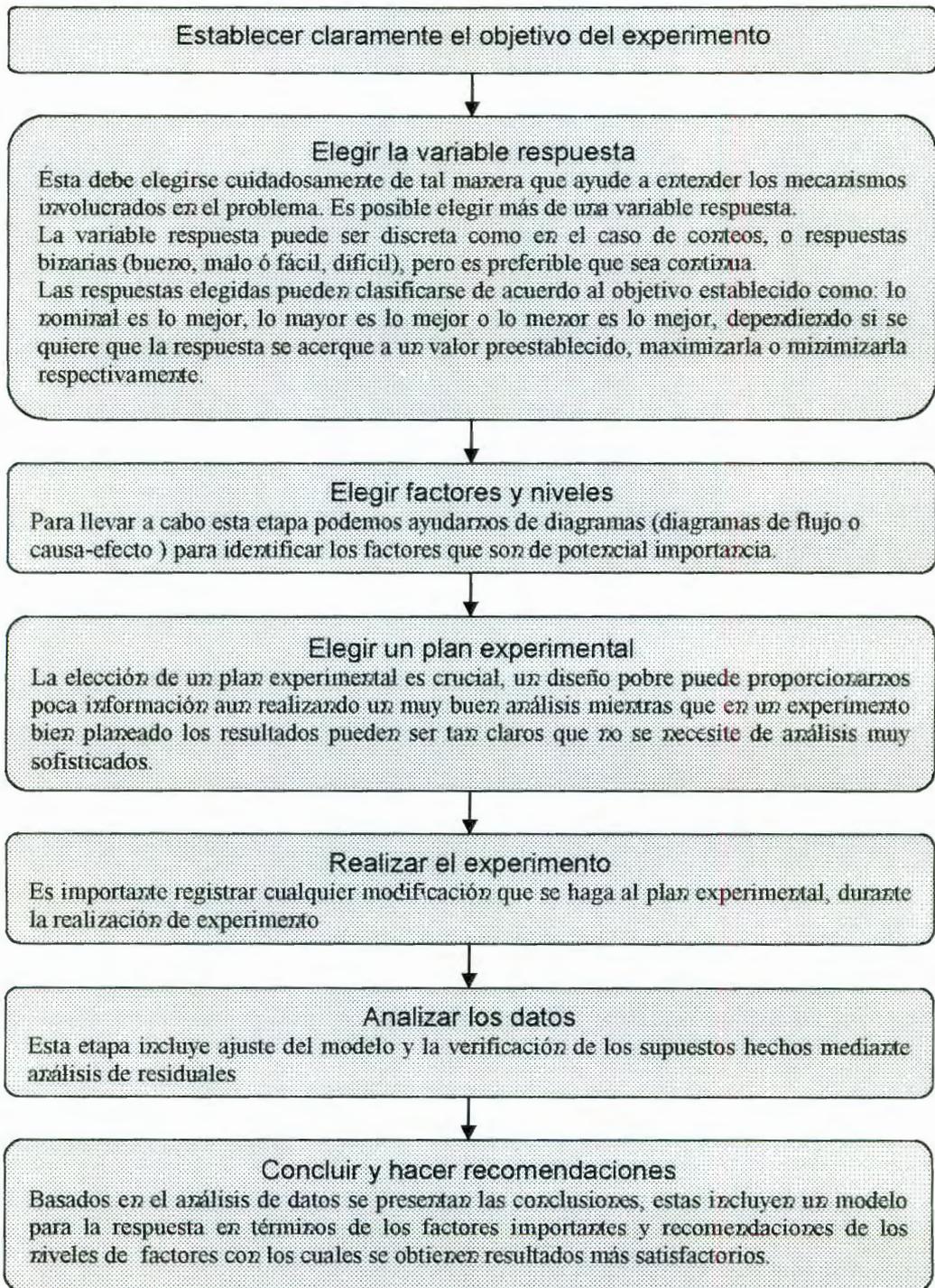


Figura 2.1. Pasos para la planeación de un experimento

Hay tres principios fundamentales que deben de conocerse en la realización del experimento: replicación, aleatorización y manejo de ruido, los cuales se explican a continuación; antes es necesario definir varios conceptos:

- **Unidad experimental**

Es un término genérico que se refiere a una unidad básica (material, animal, persona máquina o periodo de tiempo) donde se evalúa la variable respuesta a la cual se aplican los tratamientos.

- **Replicación**

Por replicación entenderemos cuando cada tratamiento es aplicado a varias unidades experimentales.

La replicación permite la estimación de la magnitud del error experimental contra la cual se comparan las diferencias entre tratamientos. Aumentando el número de réplicas, decrece la varianza de la estimación del efecto de tratamiento por lo que se tiene más poder para detectar diferencias entre tratamientos.

Es importante hacer notar la diferencia entre réplicas y repeticiones ya que a menudo estos conceptos son confundidos; tres lecturas de la misma unidad experimental a la que se le aplicó tres veces el mismo tratamiento son repeticiones, mientras que las lecturas que provengan de unidades experimentales diferentes a las que se aplicó el mismo tratamiento son réplicas. La varianza del error en el primer caso es menor que la obtenida con las réplicas, porque lecturas repetidas sobre la misma unidad experimental sólo miden la variación

del error en la lectura, mientras que en el otro caso también se mide la variación al cambiar de unidad a unidad.

- **Aleatorización de tratamientos**

La aleatorización debe aplicarse a la asignación de tratamientos a unidades experimentales, al orden en el cual los tratamientos son aplicados en la realización del experimento y al orden en el que las respuestas son medidas. Es conveniente aplicarla porque nos proporciona protección contra sesgos debidos a variables no controladas (como factores de ruido) en el experimento que puedan afectar las respuesta, además asegura la validez del error experimental y proporciona una base para la inferencia estadística en el análisis de experimento.

- **Manejo de factores de ruido**

Además de los factores planificados cuyo efecto en la variable respuesta se quiere estudiar, existen otros factores que no fueron considerados y que sin embargo su efecto en la respuesta puede ocasionar un efecto significativo en la variabilidad no planificada, llamados comúnmente factores de ruido. Para controlar la influencia de estos factores en la respuesta en ocasiones se incluyen en el diseño del experimento factores que pueden ser principalmente de dos tipos: factor bloque o covariable, los cuales se explicarán a continuación; otra opción ante los factores de ruido es dejarlos sin controlar, dando así la posibilidad de un mayor grado de generalidad de los resultados estadísticos obtenidos; de esta manera si se está interesado en la variabilidad de la respuesta ante variadas condiciones experimen-

tales, los factores de ruido incluidos deliberadamente ayudan a dar mayor generalidad a los resultados obtenidos.

– Bloqueo

Se conoce como bloque a un grupo de unidades homogéneas, por ejemplo días, semanas, lotes, pares de riñones.

Para realizar bloqueo las unidades deben de ser acomodadas de tal forma que la variación dentro del bloque sea menor que la variación entre los bloques. Al comparar tratamientos dentro de mismo bloque el efecto de bloque es eliminado en la comparación de efectos de tratamientos, por lo que el experimento resulta más eficiente. Por ejemplo puede saberse de antemano en un experimento que la respuesta se verá afectada al cambiar el día, si se realizaran todos los tratamientos el mismo día la variación día a día sería eliminada.

En caso de bloqueo se incluye un factor en el experimento que indica a cuál bloque pertenece cada unidad experimental, a este factor se le llama factor bloque y su tratamiento dentro del análisis del experimento es muy parecido al que se le da a los otros factores, pero el papel de cada uno de ellos es diferente un factor bloque se incluye en el experimento para disminuir la variabilidad del modelo mientras que los otros factores se incluyen por que se está interesado en sus efectos en la respuesta.

– Covariables

Si el factor de ruido es una propiedad cuantitativa de las unidades experimentales que puede ser medida antes de realizar el experimento (el tamaño de un fichero informático, la presión sanguínea de un paciente en un experimento médico o la acidez de una parcela de tierra en un experimento agrícola). El factor se denomina covariable y juega un papel importante en el análisis estadístico.

El modelo matemático especificado para representar el modelo del experimento, debe indicar la relación que se supone que existe entre la variable respuesta y las principales fuentes de variación, es fundamental que este modelo se ajuste a la realidad con la mayor precisión posible. El modelo más usado es el lineal que se explicará en la sección siguiente.

2.2 MODELO LINEAL

Un modelo es una abstracción de la realidad que nos da una aproximación a un fenómeno relativamente complejo. Los modelos pueden ser clasificados en determinísticos o probabilísticos. En los modelos determinísticos los resultados del sistema y las respuestas son exactamente definidos generalmente por un conjunto de ecuaciones. En los modelos probabilísticos las entradas o respuestas de un sistema presentan variabilidad, el modelo contiene elementos aleatorios o es afectado de alguna manera aleatoria.

Los datos experimentales son a menudo modelados mediante un modelo estadístico general lineal (también llamado modelo de regresión múltiple), en el que la respuesta y está relacionada con q covariables (también llamados regresores, predictores, variables explica-

torias o factores) x_1, x_2, \dots, x_q , de la siguiente manera

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q + \varepsilon \quad (2.1)$$

donde $\beta_0, \beta_1, \dots, \beta_q$ son parámetros desconocidos y ε es la parte aleatoria del modelo la que generalmente se supone se distribuye normalmente con media cero y varianza σ^2 , $\varepsilon \sim N(0, \sigma^2)$, entonces en valor esperado de la respuesta y en nuestro modelo es

$$\begin{aligned} E(y | x_1, \dots, x_p) &= E(\beta_0 + \beta_1 x_1 + \dots + \beta_q x_q) + E(\varepsilon) \\ &= \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q \end{aligned}$$

El objetivo es estimar los β_i 's, en ese sentido el modelo (2.1) es lineal ya que la respuesta y es una función lineal de los parámetros β_i .

Si N observaciones o corridas son registradas en un experimento, el modelo lineal es el siguiente

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq} + \varepsilon_i, \quad i = 1, \dots, N \quad (2.2)$$

donde y_i es el i -ésimo valor de la respuesta y x_{i1}, \dots, x_{iq} son los correspondientes valores de las q covariables; estas N ecuaciones pueden escribirse en notación matricial de la siguiente manera:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

donde $\mathbf{y} = (y_1, y_2, \dots, y_N)'$ es el vector de respuestas obtenidas en todas las corridas experimentales incluyendo las réplicas de tratamientos en el experimento, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_q)$ es el vector de parámetros, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$ y

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nq} \end{pmatrix}$$

es llamada la matriz de diseño, construida de tal forma que en la i – *ésima* corrida experimental los factores tomen los valores $(x_{i1}, x_{i2}, \dots, x_{iq})$. en el i – *ésimo* renglón de la matriz y los unos en la primera columna indican la presencia en el modelo del parámetro β_0 .

Los parámetros desconocidos en el modelo son los coeficientes de regresión β y la varianza del error σ^2 .

Una vez realizado el experimento y supuesto un modelo, los objetivos que pueden plantearse son:

1. Estimar los parámetros en el modelo.
2. Evaluar si el modelo representa de manera significativa la variabilidad observada en la variable respuesta.
3. Evaluar si existen desviaciones en los datos que invaliden algunos de los supuestos del modelo.
4. Realizar predicciones de interpolación y/o extrapolación del comportamiento de y como función de valores no experimentados de los factores.
5. Encontrar niveles en los factores en que se optimice el valor de la variable respuesta.

Una de las razones de la popularidad de modelos de regresión lineal es que se pueden determinar los parámetros de una forma muy directa utilizando la técnica de mínimos cuadrados, esta técnica de estimación es introducida por Gauss a principios del siglo XIX. Cuando este método es aplicado al modelo lineal el analista sólo tiene que resolver un conjunto de $q + 1$ ecuaciones lineales simultáneas con las incógnitas $\beta_0, \beta_1, \dots, \beta_q$. Más concre-

tamente para estimar β , el criterio de mínimos cuadrados minimiza la siguiente cantidad

$$\sum_{i=1}^N (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq}))^2 \quad (2.3)$$

que es la suma de cuadrados del error, o en términos matriciales

$$(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta).$$

En otras palabras la distancia entre el vector de respuestas y y el vector de valores ajustados $\mathbf{X}\beta$ es minimizada. Para obtener el mínimo se deriva la expresión anterior para cada parámetro en β , lo que genera un sistema de ecuaciones, la solución a este sistema es el mínimo, el cual representa el vector de los valores estimados de los parámetros

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.4)$$

Una vez estimados los parámetros se prueba si éstos son iguales a cero estadísticamente, bajo la hipótesis

$$H_0 : \beta_1 = \dots = \beta_q = 0$$

Ningún factor tiene poder explicatorio en la respuesta si se satisface H_0 . Para probar esta hipótesis nula, es necesario cuantificar qué parte de la variación total en la respuesta puede ser explicada por el modelo. Para ello expresaremos la suma de cuadrados de las respuestas, llamada suma de cuadrados totales, como la suma de cuadrados de la regresión más la suma de cuadrados de residuales (RSS), en notación matricial

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\ &= (\mathbf{X}\hat{\beta})'(\mathbf{X}\hat{\beta}) + (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} + (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \end{aligned}$$

Para probar la hipótesis nula también es necesario eliminar la contribución del término β_0 , restando su contribución $N\bar{y}^2$ (\bar{y} el promedio de las N observaciones) se obtiene:

$$\begin{aligned} CTSS &= \mathbf{y}'\mathbf{y} - N\bar{y}^2 = RegSS + RSS \\ &= (\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - N\bar{y}^2) + (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \end{aligned} \quad (2.5)$$

donde $CTSS$ es llamada suma de cuadrados corregida y es igual a $\sum_{i=1}^N (y_i - \bar{y})^2$, que mide la variación en los datos. Entonces la variación en los datos es descompuesta en la variación explicada por el modelo de regresión más la variación de residuales. Esta relación esta dada en la tabla de análisis de varianza (ANOVA)

Fuente	Grados de libertad	Suma de cuadrados	Cuadrados medios
regresión	q	$\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - N\bar{y}^2$	$\frac{\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - N\bar{y}^2}{q}$
residual	$N - q - 1$	$(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$	$\frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{N - q - 1}$
total	$N - 1$	$\mathbf{y}'\mathbf{y} - N\bar{y}^2$	

De la relación (2.5) definimos

$$R^2 = \frac{RegSS}{CTSS} = 1 - \frac{RSS}{CTSS}$$

R^2 mide la proporción de la variación total explicada por el ajuste del modelo de regresión, una R^2 alta indica un mejor ajuste del modelo de regresión. R es llamada coeficiente de correlación múltiple puesto que es la correlación entre $\mathbf{y} = (y_i)_{i=1}^N$ y $\hat{\mathbf{y}} = (\hat{y}_i)_{i=1}^N$

Los grados de libertad son asociados con cada suma de cuadrados. El cuadrado medio es la suma de cuadrados dividida por los correspondientes grados de libertad. El cuadrado medio de residuales, es llamado comunmente cuadrado medio del error (denotado MSE

por siglas en inglés) y es una estimación $\hat{\sigma}^2$, de σ^2

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{N - q - 1}.$$

Si la hipótesis H_0 se satisface, la estadística F

$$\frac{(\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} - N\bar{y}^2)/q}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/N - q - 1}$$

tiene una distribución F con parámetros p y $N - p - 1$, que son los grados de libertad del numerador y denominador respectivamente. El valor p de H_0 es calculado al evaluar:

$$\text{Prob}(F_{q, N-q-1} > F_{obs})$$

donde Prob denota la probabilidad del evento, $F_{q, N-q-1}$ es una variable aleatoria que tiene una distribución F con parámetros q y $N - q - 1$ y F_{obs} es el valor observado de la estadística F .

El valor p da la probabilidad bajo la hipótesis nula de que el valor de la estadística F para un experimento realizado en condiciones comparables exceda el valor observado F_{obs} , entonces mientras más pequeño sea el valor p , mayor es la evidencia de que la hipótesis nula no se satisface. Este valor p provee una medida de significancia de los efectos en el experimento. La misma interpretación puede ser aplicada a otras estadísticas de prueba e hipótesis nulas consideradas.

Puede mostrarse que el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ tiene una distribución normal multivariada con media el vector $\boldsymbol{\beta}$ y matriz de varianzas y covarianzas $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

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

La entrada (i, j) de la matriz de varianzas y covarianzas es $Cov(\hat{\beta}_i, \hat{\beta}_j)$ y el j -ésimo elemento en la diagonal es $Cov(\hat{\beta}_j, \hat{\beta}_j) = Var(\hat{\beta}_j)$, entonces la distribución para $\hat{\beta}_j$ es $N(\beta_j, \sigma^2(\mathbf{X}'\mathbf{X})_{jj}^{-1})$, lo que sugiere que para la hipótesis nula

$$H_0 : \beta_j = 0$$

la siguiente estadística t es usada:

$$\frac{\hat{\beta}_j}{\sqrt{\sigma^2(\mathbf{X}'\mathbf{X})_{jj}^{-1}}}$$

Bajo H_0 , esta tiene una distribución t con $N - q - 1$ grados de libertad. Esto también puede usarse para construir intervalos de confianza puesto que el denominador de la estadística t es el error estándar de su numerador $\hat{\beta}_j$.

$$\hat{\beta}_j \pm t_{N-q-1, \frac{\alpha}{2}} \sqrt{\hat{\sigma}^2(\mathbf{X}'\mathbf{X})_{jj}^{-1}}$$

donde $t_{N-q-1, \frac{\alpha}{2}}$ es el cuantil superior de una distribución t con $N - q - 1$ grados de libertad.

Si alguno de estos intervalos contiene el valor cero, puede decirse que hay evidencia de que el factor correspondiente no influye sobre la respuesta y .

2.3 DISEÑOS EXPERIMENTALES

Los modelos de diseños de experimentos son modelos estadísticos clásicos cuyo objetivo es averiguar si unos factores determinados influyen en la variable de interés, y si existe influencia de los factores, cuantificarla.

El objetivo del diseño de experimentos es estudiar si utilizar un determinado tratamiento produce una mejora en el sistema o no, de la manera más eficiente posible.

Un diseño experimental es una regla que determina la asignación de las unidades experimentales a los tratamientos. Aunque los experimentos difieren unos de otros en muchos aspectos, existen diseños estándar que se usan con mucha frecuencia. Aquí se presentaran los diseños que más adelante serán empleados para ilustrar las aplicaciones de análisis funcional.

2.3.1 Diseño con un solo factor

En un experimento con un factor de q niveles, (q tratamientos correspondientes a cada nivel) el modelo lineal es el siguiente

$$y_{ij} = \beta_0 + \beta_i + \varepsilon_{ij}, \quad i = 1, \dots, q; \quad j = 1, \dots, n_i \quad (2.6)$$

donde n_i es el número de observaciones para el tratamiento i (en caso de usar replicación), y_{ij} es la j -ésima observación con el tratamiento i , β_i es el efecto del tratamiento i y $\varepsilon_{ij} \sim N(0, \sigma^2)$.

En este caso la matriz \mathbf{X} es la siguiente:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & \vdots & \vdots & 0 & & 0 \\ 1 & 1 & 0 & 0 & & 0 \\ 1 & 0 & 1 & 0 & & 0 \\ 1 & 0 & \vdots & \vdots & & 0 \\ 1 & 0 & 1 & 0 & & \vdots \\ 1 & 0 & 0 & 1 & & 0 \\ 1 & 0 & 0 & \vdots & & 1 \\ \vdots & \vdots & \vdots & 1 & & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

y para estimar los efectos de tratamientos es necesario imponer restricciones en los parámetros β_i , porque hay q tipos de observaciones pero $q + 1$ parámetros de regresión, el modelo está sobreparametrizado, al intentar calcular el estimador por mínimos cuadrados $\hat{\beta}$ en (2.4), la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ no existe. La matriz \mathbf{X} no es de rango completo, es decir sus columnas no son linealmente independientes, entonces la matriz $\mathbf{X}'\mathbf{X}$ es singular. Para hacer de la matriz $\mathbf{X}'\mathbf{X}$ una matriz singular es necesario imponer restricciones en los parámetros.

La restricción comúnmente usada es la suma cero

$$\sum_{i=1}^q \beta_i = 0. \quad (2.7)$$

De (2.7) se tiene que $\beta_k = -\sum_{i=1}^{q-1} \beta_i$, sustituyendo esta restricción en la expresión (2.6) el número de parámetros es reducido en 1, es decir $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{q-1})$ y los parámetros restantes tienen diferentes significados:

$$\frac{1}{q} \sum_{i=1}^q E(y_{ij}) = \frac{1}{q} \sum_{i=1}^q (\beta_0 + \beta_i) = \beta_0 + 0 = \beta_0$$

$$E(y_{ij}) - \beta_0 = \beta_0 + \beta_i - \beta_0 = \beta_i$$

β_0 representa la media global y β_i es la diferencia de la media global de y con el valor esperado de y cuando recibe el tratamiento β_i .

Otra restricción que puede imponerse es:

$$\beta_1 = 0.$$

Con esta restricción se elimina el término β_1 del vector de parámetros β , así como su columna correspondiente en la matriz de diseño \mathbf{X} , de esta manera es posible calcular

$(\mathbf{X}'\mathbf{X})^{-1}$ y el vector $\hat{\beta}$. Pero la manera de interpretar a los parámetros restantes cambia, en este caso se tendrá,

$$\beta_0 = E(y_{1j})$$

así β_0 representa el valor esperado para la respuesta del tratamiento 1 y

$$\beta_i = E(y_{ij}) - E(y_{1j})$$

entonces β_i $i \geq 2$, representa la comparación entre el primer tratamiento y el tratamiento i .

Es natural imponer esta restricción cuando el tratamiento 1 es un tratamiento estándar, de control o un tratamiento ya existente mientras que los otros $q - 1$ tratamientos son nuevos, ya que así el funcionamiento o efecto de los nuevos tratamientos será medido en comparación con el estándar.

2.3.2 Diseños factoriales con dos niveles

En muchos casos es adecuado elegir dos niveles para cada uno de los factores involucrados en el experimento. Si se cuenta con k factores con dos niveles cada uno (usualmente denotados como 1 y -1), entonces se tienen 2^k tratamientos. El diseño es factorial completo si se requieren observaciones para todos los posibles tratamientos, es decir, si se requieren 2^k observaciones; es por ello que este tipo de diseños son llamados 2^k .

Cuando se tienen muchos factores que pueden influir en la variable de interés la utilización de experimentos completos tiene la desventaja de presentar un número muy elevado de observaciones, pues al crecer k evidentemente se incrementa drásticamente 2^k , en estos casos es recomendable utilizar diseños factoriales fraccionados los cuales utilizan una fracción del total de observaciones en el diseño completo 2^k , estos se denotan como

2^{k-p} , donde $2^{-p} = \frac{1}{2^p}$ es la fracción tomada, k el número de factores y 2^{k-p} el número de observaciones.

Los diseños fraccionales factoriales 2^{k-p} mantienen la propiedad de ortogonalidad de los factores, pero para reducir el número de observaciones se suponen no significativas interacciones de orden generalmente alto (éstas se alían con otros efectos), de manera que cuanto mayor sea p se requerirá un menor número de observaciones pero se tendrán más alias en los efectos.

Para ejemplificar utilizaremos un diseño con $k = 4$ factores denotados como A, B, C y D . El diseño completo es el siguiente:

	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
1	----	-1	-1	-1	-1	1	1	1	1	1	-1	-1	-1	-1	1
2	---+	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	1	-1
3	--+-	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	1	-1
4	---++	-1	-1	1	1	1	-1	-1	-1	-1	1	1	-1	-1	1
5	-+--	-1	1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
6	-+-+	-1	1	-1	1	1	-1	-1	1	-1	1	-1	1	-1	1
7	-++-	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
8	-+++	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1
9	+---	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	-1
10	+--+	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1
11	++-	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
12	+---+	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1
13	++--	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1
14	++-+	1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	-1
15	+++-	1	1	1	-1	1	1	-1	1	-1	1	-1	-1	-1	-1
16	++++	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Tabla 2.1. Diseño factorial completo 2^4

Para tomar un medio del total de observaciones, $p = 1$, ya que $\frac{1}{2} = 2^{-1}$ y se tendrán $2^{4-1} = 8$ corridas, la elección de las corridas la haremos tomando solo los 1's ó sólo los -1's de las columnas correspondiente al los p efectos que se consideren poco significativos, usualmente se eligen los de mayor interacción, en este caso $p = 1$ y elegimos la interacción

$ABCD$ (que es el producto de las columnas A , B , C y D). Aunque esta elección depende de los efectos o factores que se juzguen importantes de acuerdo a cada experimento, existen criterios para definirlo.

Al tomar los 1's de la columna $ABCD$ se obtiene el diseño factorial fraccionado 2^{4-1} siguiente

	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
1	----	-1	-1	-1	-1	1	1	1	1	1	-1	-1	-1	-1	1
4	--++	-1	-1	1	1	1	-1	-1	-1	-1	1	1	-1	-1	1
6	-++-	-1	1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	1
7	+-+-	-1	1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	1
10	+--+	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1
11	+-+-	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
13	++--	1	1	-1	-1	1	-1	-1	-1	1	-1	-1	1	1	1
16	++++	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Tabla 2.2. Diseño factorial fraccionado 2^{4-1}

A la columna $ABCD$ la denotamos como I , $I = ABCD$, a esta relación la llamaremos relación de definición (o subgrupo de contraste de definición), se dice que el diseño es de resolución IV por que la relación de definición consiste de la palabra $ABCD$ que tiene longitud 4, por lo que el diseño se denota como 2^{4-1}_{IV} .

A partir de esta relación de definición $I = ABCD$, se obtienen las relaciones que indican cuáles efectos están confundidos con cuáles, a éstas se les llama relaciones alias para este ejemplo se presentan a continuación.

$$A = BCD \quad B = ACD \quad C = ABD$$

$$D = ABC \quad AB = CD \quad AC = BD \quad AD = BC$$

Cada uno de los efectos principales (A , B , C , D) es estimable si las interacciones triples con que estan aliados (BCD , ACD , ABD , ABC respectivamente) se suponen no significativas, lo cual es generalmente un supuesto razonable especialmente cuando las

interacciones son de orden alto. Las interacciones dobles AB , AC , y AD serán estimables solamente si se tiene un conocimiento previo de que las interacciones dobles con que estan aliadas no lo sean.

Como en [4], llamaremos a un efecto principal o a una interacción doble clara si ninguno de sus alias son efectos principales o interacciones dobles y fuertemente claros si ninguno de sus alias son efectos principales, interacciones dobles o triples. En el ejemplo sólo los efectos principales son claros.

Para un diseño 2^{k-p} , sea w_i que denote el número de palabras de longitud i en su relación de definición, con estos elementos formamos el vector $\mathbf{w} = (w_3, \dots, w_k)$. La definición de un diseño 2^{k-p} es el r más pequeño tal que $w_r \geq 1$, es decir la longitud de la palabra más corta en el subgrupo de contraste de definición

De acuerdo a la elección de la relación de definición se pueden tener varios diseños de un experimento, para elegir uno de éstos se tienen principalmente dos criterios. El *criterio de máxima resolución* establecido por Box y Hunter, elige el diseño 2^{k-p} con máxima resolución, este criterio puede ser justificado ya que los diseños con resolución baja contienen palabras de longitud corta en la relación de definición, lo que implica que muchos efectos estarán confundidos con efectos de bajo orden que usualmente son los que se tiene interés en estimar claramente.

Un diseño no puede ser elegido solamente buscando tener la resolución más alta ya que se pueden tener varios diseños con la misma resolución, por ejemplo, consideremos

los siguientes dos diseños 2^{7-2} :

$$d_1 : I = 4567 = 12346 = 12357$$

$$d_2 : I = 1236 = 1457 = 234567$$

los dos diseños son de resolución IV pero presentan diferentes patrones en el vector w

$$w(d_1) = (0, 1, 2, 0, 0)$$

$$w(d_2) = (0, 2, 0, 1, 0)$$

por lo que se tiene también el criterio de aberración mínima descrito en [4], que elige el diseño con mínima aberrancia.

Criterio de mínima aberrancia:

Dados dos diseños 2^{k-p} , d_1 y d_2 , sea r el entero más pequeño tal que $w_r(d_1) \neq w_r(d_2)$, se dice que d_1 tiene menor aberrancia que d_2 si $w_r(d_1) < w_r(d_2)$. Si no hay diseño con menor aberrancia que d_1 , se dice que d_1 es el diseño de mínima aberrancia.

Este criterio se puede justificar de manera similar al anterior, ya que diseños con menor aberrancia tendrán un número mayor de efectos estimables de manera clara.

3

REPRESENTACIÓN FUNCIONAL DE DATOS

3.1 INTRODUCCIÓN

La idea básica del análisis funcional es pensar en los datos observados como funciones, a diferencia del enfoque tradicional en el que estos se piensan como una secuencia de datos discretos. El término funcional empleado se refiere a las estructura intrínseca de los datos más que a su forma explícita, es decir, se cree que hay una función subyacente a los datos que obtenemos en la práctica, usualmente observados y registrados de manera discreta.

Un registro de una observación funcional x , consiste de n parejas (t_j, y_j)

$$((t_1, y_1), (t_2, y_2), \dots, (t_n, y_n))$$

donde cada y_j es una observación de una función desconocida $x(t)$ evaluada en t_j .

En este capítulo consideraremos algunas técnicas para determinar una aproximación $\hat{x}(t)$, a la función $x(t)$, una vez encontrada tendremos la ventaja de estimar las derivadas $D^m x(t)$ que existan de ella en cualquier valor del argumento t . Como sólo se cuenta con observaciones de la función en un número finito de puntos, utilizaremos interpolación o suavizamiento de éstos para encontrar la forma funcional de la observación.

En general estaremos interesados en una colección de datos funcionales, más que en una sola función, específicamente el registro u observación de la función x_i consiste de n_i

parejas (t_{ij}, y_{ij}) , $j = 1, \dots, n_i$, donde los valores del argumento t_{ij} pueden variar de una observación funcional a otra:

$$\begin{aligned} & ((t_{11}, y_{11}), (t_{12}, y_{12}), \dots, (t_{1n_1}, y_{1n_1})), \\ & ((t_{21}, y_{21}), (t_{22}, y_{22}), \dots, (t_{2n_2}, y_{2n_2})), \\ & \quad \vdots \\ & ((t_{i1}, y_{i1}), (t_{i2}, y_{i2}), \dots, (t_{in_i}, y_{in_i})). \end{aligned}$$

se considera el problema como funcional por que se supone que existe una función detrás de cada una de las observaciones $((t_{i1}, y_{i1}), (t_{i2}, y_{i2}), \dots, (t_{in_i}, y_{in_i}))$ con que contamos.

Se supondrá que el rango de valores de interés para t es un intervalo acotado \mathcal{T} y que x satisface ciertas condiciones de continuidad o suavizamiento en ese intervalo.

3.1.1 Muestreo y error observacional suavizamiento

Suavidad (en el sentido de que existan cierto número de derivadas) es una propiedad de una función x , que puede no ser evidente en el vector $\mathbf{y} = (y_1, \dots, y_n)$ observado, debido al error observacional o ruido, por ejemplo a malas mediciones. Por lo que se tiene

$$y_j = x(t_j) + \varepsilon_j, \tag{3.1}$$

donde el error ε_j contribuye a la rugosidad de los datos.

El modelo estadístico estándar para los errores ε_j que se supone es que son independientemente distribuidos con media cero y varianza finita.

Uno de los objetivos al representar de manera funcional, es filtrar este error tan eficientemente como sea posible. Otra alternativa puede ser dejar el error presente en los datos observados y suavizar los resultados del análisis realizado en lugar de suavizar los datos observados.

El modelo (3.1) puede remplazarse por el modelo siguiente

$$y_j = x(t_j) + \varepsilon(t_j) \quad (3.2)$$

en éste el ruido es remplazado por una función ε , que podemos suponer tiene las características de ruido blanco, media cero, varianza constante y covarianza cero para distintos valores del argumento t . Suponer que el tipo de ruido que se presenta es discreto (3.1) o funcional (3.2), depende del experimento.

Aunque se analizarán principalmente las técnicas para reducir el error observacional en las funciones por sí mismas, este error ya sea discreto o funcional, no es el único modelo posible para la variación de una observación funcional a otra, ya que el ruido también puede afectar a una o más derivadas. Un método en particular destacado para el tratamiento del ruido observacional de manera directa es el método de penalización por rugosidad, éste juega un papel importante dentro del análisis funcional de datos, es por ello que lo describimos de manera separada en la sección 4. Por el momento explicaremos los suavizadores lineales ya que estos son usados para el desarrollo de técnicas más complejas.

3.1.2 Suavizamiento lineal

Un suavizador lineal estima el valor de la función $x(t)$ por una combinación lineal de las observaciones discretas

$$\hat{x}(t) = \sum_{j=1}^n S_j(t)y_j,$$

$$\hat{x}(t) = S_1(t)y_1 + \dots + S_n(t)y_n.$$

El comportamiento del suavizador en t es determinado por los pesos $S_j(t)$

Suavizadores lineales también pueden ser representados en forma matricial:

Sean s_1, s_2, \dots, s_m una sucesión de puntos de evaluación en \mathcal{T} (no necesariamente los valores observados t_j) en los cuales se desea evaluar x

\hat{x} \longrightarrow el vector de los m valores $x(s_i)$

y \longrightarrow el vector de los datos observados y_j

Entonces podemos escribir:

$$\hat{x} = \mathbf{S}y \quad (3.3)$$

donde \mathbf{S} es una matriz de $m \times n$ tal que $s_{ij} = S_j(s_i)$,

$$\mathbf{S} = \begin{pmatrix} S_1(s_1) & \dots & S_n(s_1) \\ \vdots & \ddots & \vdots \\ S_1(s_m) & \dots & S_n(s_m) \end{pmatrix} \text{ y } \mathbf{S}(ay + bz) = a\mathbf{S}y + b\mathbf{S}z.$$

Muchos suavizadores muy usados son lineales. La linealidad de un suavizador es deseable ya que es importante para trabajar varias propiedades de la representación suave y la simplicidad del suavizador implica relativa rapidez en los cálculos. Por otro lado algunos suavizadores no lineales pueden ser más adaptables a diferentes comportamientos en diferentes partes del rango de observación y pueden ser robustos ante observaciones atípicas.

3.2 MÉTODOS DE FUNCIONES BASE

Uno de los procesos más familiares de suavizamiento es representar la función por una combinación lineal de K funciones base conocidas ϕ_k

$$x(t) = \sum_{k=1}^K c_k \phi_k(t)$$

La construcción de esta función $x(t)$, se centra entonces en encontrar los valores de las K constantes c_k .

El grado en el cual los datos y_j son suavizados es determinado por el número de funciones base K utilizados, una representación exacta (interpolación) es generalmente posible cuando $K = n$ en el sentido que podemos elegir los coeficientes c_k , para que $x(t_j) = y_j$, para cada j .

Hay muchas formas de elegir un conjunto de funciones base $\{\phi_1, \phi_2, \dots, \phi_k\}$, se presentarán algunas más adelante. Primero consideraremos un método para determinar los coeficientes c_k .

3.2.1 Determinación de c_k por mínimos cuadrados

Supongamos que se tiene una observación funcional:

$$((t_1, y_1), (t_2, y_2), \dots, (t_n, y_n))$$

de la cual se desea encontrar una función que la aproxime mediante el método de las funciones base.

La manera más simple de encontrar los coeficientes c_k , es mediante el criterio de mínimos cuadrados, el cual determina los coeficientes c_k de tal forma que se minimize la siguiente suma:

$$SMSSE(\mathbf{y}/\mathbf{c}) = \sum_{j=1}^n [y_j - x(t_j)]^2 = \sum_{j=1}^n \left[y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2. \quad (3.4)$$

Expresada en términos matriciales:

$$SMSSE(\mathbf{y}/\mathbf{c}) = (\mathbf{y} - \Phi\mathbf{c})'(\mathbf{y} - \Phi\mathbf{c}) = \|\mathbf{y} - \Phi\mathbf{c}\|^2 \quad (3.5)$$

donde \mathbf{c} representa el vector de dimensión K que contiene los elementos c_k y la matriz Φ es

$$\Phi_{n \times K} = \begin{pmatrix} \phi_1(t_1) & \dots & \phi_K(t_1) \\ \vdots & \ddots & \vdots \\ \phi_1(t_n) & \dots & \phi_K(t_n) \end{pmatrix},$$

la cual se supone que es de rango completo.

Con este criterio se obtiene la solución:

$$\mathbf{c} = (\Phi' \Phi)^{-1} \Phi' \mathbf{y} \quad (3.6)$$

Si los puntos de evaluación son los puntos observados, la matriz de suavizamiento \mathbf{S} en 3.6 está dada por:

$$\mathbf{S} = \Phi (\Phi' \Phi)^{-1} \Phi' \quad (3.7)$$

La matriz de suavizamiento en este caso es una proyección ortogonal puesto que es simétrica y satisface la relación $\mathbf{S}^2 = \mathbf{S}$, esta propiedad se conoce como idempotencia.

Si los puntos de evaluación (s_1, \dots, s_m) no son los mismos que los observados (t_1, \dots, t_n) entonces la matriz de suavizamiento está dada por:

$$\mathbf{S} = \tilde{\Phi} (\Phi' \Phi)^{-1} \Phi' \quad (3.8)$$

donde $\tilde{\Phi}_{m \times n} = \{\phi_{ij} = \phi_j(s_i)\}$.

Podemos extender el criterio de mínimos cuadrados de manera más general a la forma

$$SMSSE(\mathbf{y}/\mathbf{c}) = (\mathbf{y} - \Phi \mathbf{c})' \mathbf{W} (\mathbf{y} - \Phi \mathbf{c}) = \|\mathbf{y} - \Phi \mathbf{c}\|_{\mathbf{W}}^2 \quad (3.9)$$

en el cual \mathbf{W} es una matriz simétrica positiva que permite asignar un peso específico a los residuales, esta representación puede ser importante si se sabe que las varianzas de ε_j no son necesariamente constantes o no son independientemente distribuidas de alguna forma conocida. Las estimaciones de los coeficientes c_k están dados en este caso por la

expresión:

$$\mathbf{c} = (\Phi'W\Phi)^{-1}\Phi'W\mathbf{y} \quad (3.10)$$

En caso de que los puntos de evaluación y los argumentos en los que se registraron las observaciones t_j sean los mismos, la matriz de suavizamiento correspondiente es

$$\mathbf{S} = \Phi(\Phi'W\Phi)^{-1}\Phi'W \quad (3.11)$$

y esta es aún un operador de proyección ortogonal dado que WS es simétrica.

3.2.2 Eligiendo una buena base

Una característica deseable de las funciones base es que tengan características similares a las de las funciones que se quieren estimar, idealmente una base debe elegirse de manera que se logre una muy buena aproximación usando relativamente un número pequeño de funciones base K , ya que además de reducirse los cálculos, se puede proporcionar información de interés en los coeficientes calculados; en consecuencia hay bases que se adaptan bien a cierto tipo de datos mientras que otras no, por lo que no se tiene una base que funcione bien en todos los casos.

La elección de la base es particularmente de interés para la estimación de las derivadas,

$$D\hat{x}(t) = \sum_{k=1}^K c_k D\phi_k(t);$$

hay bases que funcionan bien para la estimación de la función pero no para sus derivadas, esto sucede debido a que una representación de $\hat{x}(t)$ muy apegada a las observaciones puede forzar a que \hat{x} tenga altas oscilaciones lo cual afecta a las derivadas, es por ello que uno de

los criterios para elegir una base puede ser si la primera o más derivadas de la aproximación $\hat{x}(t)$, se comportan de manera razonable.

En seguida se presentaran algunas de las funciones base que son más usadas en la práctica.

3.2.3 Series de Fourier

Este sistema fue desarrollado a principios del siglo XIX en Francia, y consiste en una serie de senos y cosenos que van aumentando su frecuencia

$$\hat{x}(t) = c_0 + c_1 \sin (wt) + c_2 \cos (wt) + c_3 \sin (2wt) + c_4 \cos (2wt) + \dots$$

Definimos las funciones base

$$\begin{aligned}\phi_0(t) &= 1 \\ \phi_{2r-1}(t) &= \sin(rwt) \\ \phi_{2r}(t) &= \cos(rwt) \\ r &= 1, 2, 3, \dots\end{aligned}$$

Esta base es periódica y el parámetro w determina el periodo $2\pi/w$, que es igual a la longitud del intervalo \mathcal{T} en el que estamos trabajando. Si los valores t_j son igualmente espaciados en \mathcal{T} , entonces la base es ortogonal en el sentido de que la matriz producto cruz $\Phi\Phi'$ es diagonal y puede ser igual a la identidad al dividir las funciones base por una constante adecuada. La Figura 3.1 muestra las cinco primeras funciones base de la serie Fourier

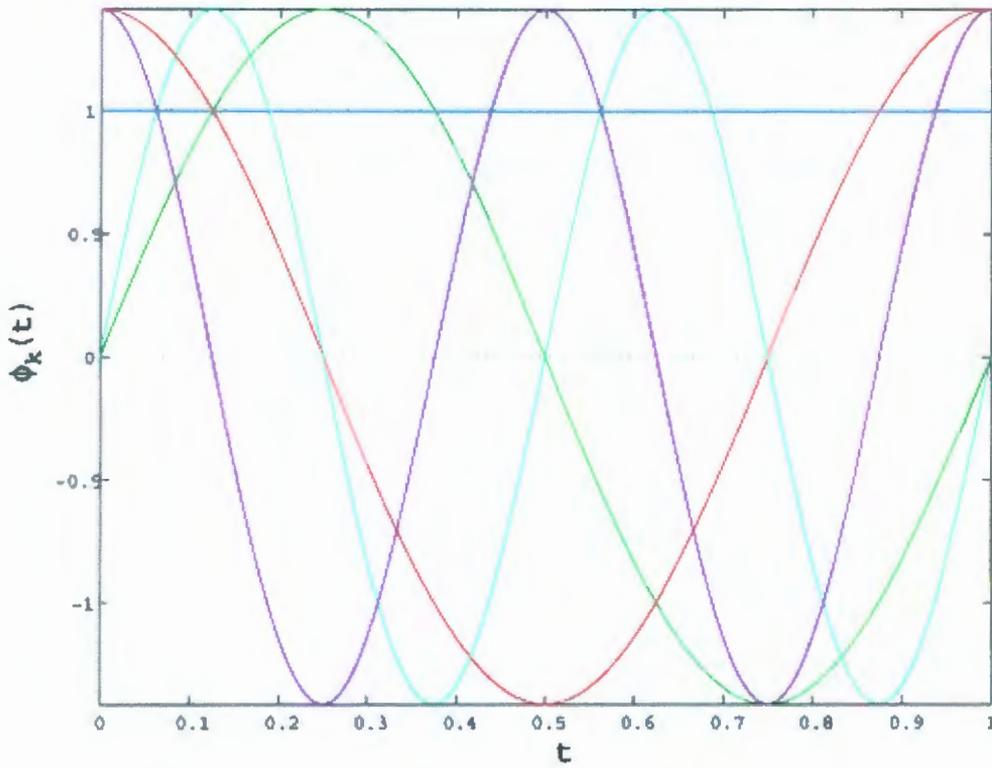


Figura 3.1. Primeras cinco funciones base de la serie de Fourier

La estimación de las derivadas es simple ya que

$$D \sin(rwt) = rw \cos(rwt)$$

$$D \cos(rwt) = -rw \sin(rwt)$$

Esto implica que la expansión de Dx en serie de Fourier tiene los coeficientes $(0, -wc_2, wc_1, -2wc_4, 2wc_3, \dots)$ y D^2x tiene los coeficientes $(0, -w^2c_1, -w^2c_2, -4w^2c_3, -4w^2c_4, \dots)$, similarmente podemos encontrar las expansiones de Fourier para derivadas de orden más alto.

Las series de Fourier son especialmente útiles para funciones muy estables, es decir que no presentan cambios locales muy fuertes y que su curvatura es la misma o del mismo orden en todo el intervalo. Idealmente la periodicidad de las series de Fourier debería verse

reflejada en los datos y las series de Fourier resultan inapropiadas para datos en los que se sabe o se sospecha presentan discontinuidades en la función o en alguna de sus derivadas.

3.2.4 Base polinomial

Este tipo de bases

$$\phi_k(t) = (t - w)^k, \quad k = 0, \dots, K$$

son también clásicas, pero desafortunadamente la matriz del producto cruz $\Phi' \Phi$ es singular en muchas ocasiones y debe elegirse con mucho cuidado el parámetro de cambio w . Sin embargo si los valores del argumento t_j son igualmente espaciados, la expansión en polinomios ortogonales pueden ser obtenida en $O[(m + n)K]$ operaciones para todos los valores suaves, de otra manera necesitaremos $O[(m + n)K^2]$ operaciones.

Como en el caso de las funciones de Fourier, los polinomios no presentan muchos cambios locales sin utilizar un número de funciones base K relativamente grande, más aún los polinomios tienden a ajustarse relativamente bien en el centro de los datos pero no en las colas. Son generalmente poco útiles para extrapolación.

Aunque las derivadas de las expansiones polinomiales son fáciles de calcular, son rara vez satisfactorios como estimadores de las derivadas.

Debido a las desventajas que presenta la base polinomial ésta ha sido prácticamente remplazada por otro tipo de base llamado spline que aunque se relaciona con el polinomial presenta más ventajas y es ampliamente usado, por lo que se explica más a detalle en la sección.3.3

3.2.5 Bases "wavelets"

Los wavelets son ondas que oscilan y satisfacen $\int_{-\infty}^{\infty} \psi(t) dt = 0$, donde $\psi(t)$ es el wavelet, estos nos proporcionan un conjunto conveniente de funciones base para espacios de funciones como $L^2(\mathbb{R})$ en $(-\infty, \infty)$.

El conjunto de funciones base depende de la elección del wavelet madre, esta elección da lugar a una variedad de familias de funciones wavelets.

Para construir wavelets comenzamos eligiendo un wavelet madre adecuado ψ y después consideramos todas las traslaciones y dilataciones de la forma

$$\psi_{jl}(t) = 2^{\frac{j}{2}} \psi(2^j t - l)$$

para enteros j y l .

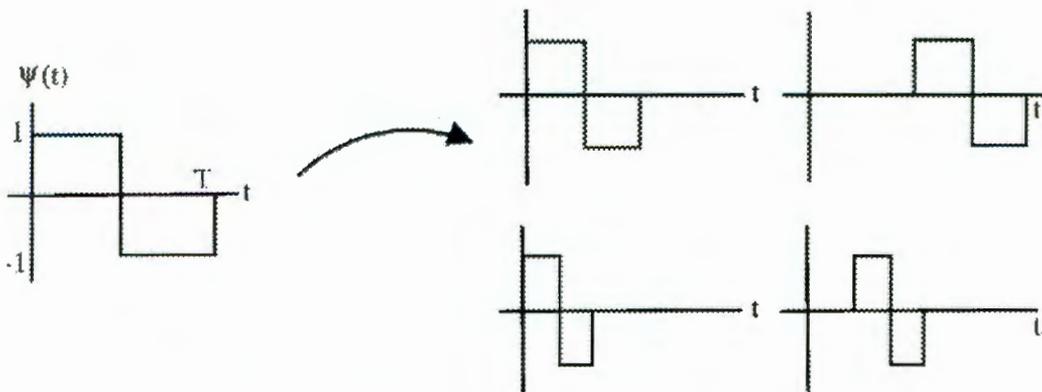


Figura 3.2. Un ejemplo de wavelet de la familia Haar

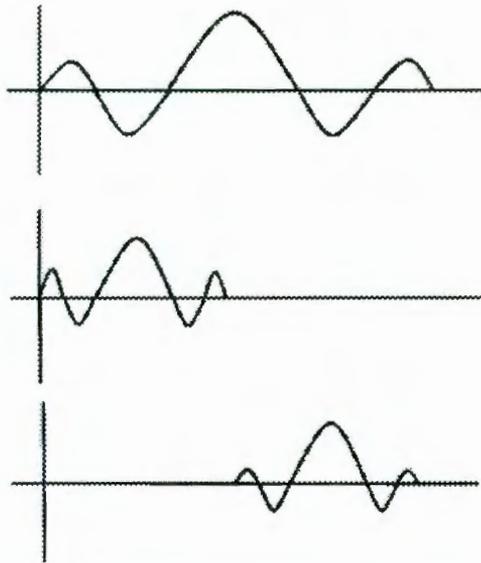


Figura 3.3. Wavelets en diferente traslación y escala

Las Figuras 3.2 y 3.3 muestran algunos wavelets en diferente traslación y escala para dos familias, la primera es llamada Haar.

Construimos el wavelet madre de tal forma que se asegure que la base sea ortonormal, en el sentido de que la integral del producto de cualquier par de funciones base distintas sea cero.

En contraste con la base de Fourier, la expansión con wavelets lidia muy bien con discontinuidades y rápidos cambios de comportamiento; esta propiedad así como otros resultados matemáticos un poco más técnicos, indican que es razonable suponer que una función observada está bien aproximada con una expansión wavelet con pocos coeficientes distintos de cero, aún si ésta presenta picos o muchos cambios locales.

El hecho de que los wavelets estén localizados en tiempo y frecuencia les permite representar funciones que estén igualmente localizadas en tiempo y frecuencia. Los wavelets

son principalmente útiles en representar funciones que presentan muchas ondas o picos y cambian mucho localmente

3.2.6 Recomendaciones finales

La mayoría de las veces se utilizan funciones base Fourier para ajustar curvas a observaciones que muestran un comportamiento periódico, como por ejemplo datos climatológicos o económicos, si no se observa periodicidad se utilizan splines. La siguiente lista muestra algunas características que deben de tomarse en cuenta para la elección de funciones base.

– *Cálculo rápido de funciones base individuales*

En algunas situaciones podemos necesitar miles de funciones base, por lo que ser capaces de evaluar $\phi_k(t)$ de manera rápida es crucial, es por ello que los polinomios eran tan populares cuando los cálculos se realizaban a mano, ahora los métodos modernos de cálculo permiten también la evaluación rápida de senos, cosenos y otras funciones trascendentes.

– *Flexibilidad*

Muchas curvas presentan aspectos que involucran muchos picos en puntos específicos, los polinomios pueden ajustarse bien a estas curvas pero solamente utilizando un número muy grande de funciones base, los splines son un poco mejores en este caso, las bases de Fourier no son tan buenas para ajustarse a las partes en las que se presenten muchos picos; la base que es ideal para este tipo de curvas son los wavelets.

– *Cálculo rápido de los coeficientes c_k*

Los coeficientes son calculados al resolver un sistema de ecuaciones lineales. No es muy difícil calcular estos coeficientes cuando se tienen a lo más cientos de ellos, pero cuando se tienen miles, sólo servirán algunos sistemas de funciones bases que involucran estructuras especiales para estas situaciones. De las funciones base vistas la base splines resulta muy buena, mientras que la polinomial, Fourier y wavelets sólo funcionan en ocasiones especiales, como cuando los valores del argumento t son equiespaciados.

– *Diferenciabilidad*

Las derivadas muchas veces son importantes en el análisis de los datos, por lo que se requiere que las derivadas de las funciones base se comporten de manera satisfactoria sin importar el orden que se requiera; tanto la base spline como Fourier funcionan adecuadamente en estos casos pero la base polinomial no, porque sus derivadas tienden a comportarse de manera más simple al incrementar el orden, lo cual no sucede de manera común en la práctica.

– *Restricciones especiales*

Algunas funciones pueden ser sólo positivas, otras estrictamente crecientes, este tipo de restricciones necesitan ser consideradas desde la elección de las funciones base. Un ejemplo común es la periodicidad, para la que ya hemos visto que tanto la base de Fourier como la wavelet son muy buenas opciones.

3.3 SPLINES

Supongamos que t es unidimensional. Una función $x(t)$ formada por trozos de funciones polinómicas se obtiene dividiendo el dominio de t en intervalos contiguos y representando x mediante polinomios separados en cada intervalo.

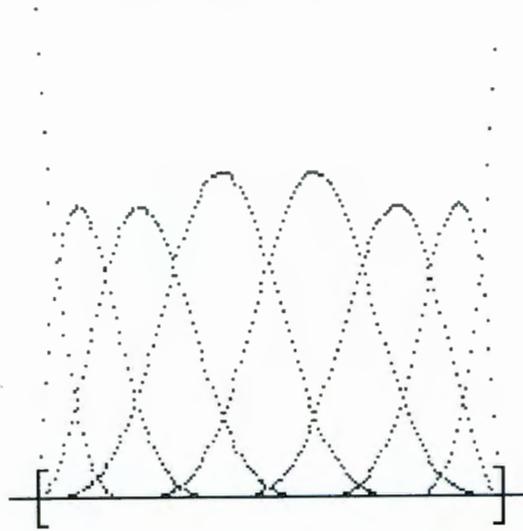


Figura 3.4. Funciones base B-spline de grado tres.

La siguiente base representa un spline cúbico con nodos en τ_1 y τ_2 :

$$\phi_1(t) = 1, \phi_2(t) = t, \phi_3(t) = t^2, \phi_4(t) = t^3,$$

$$\phi_5(t) = (t - \tau_1)_+^3, \phi_6(t) = (t - \tau_2)_+^3$$

Hay seis funciones base correspondientes al espacio lineal de funciones de dimensión seis. Una forma rápida de confirmarlo es mediante la siguiente fórmula:

$$\begin{aligned} \text{Núm. de funciones base} &= (\text{Núm. de regiones}) \times (\text{Núm. de parámetros por región}) - \\ &(\text{Núm. de nodos}) \times (\text{Núm. de restricciones por nodo}) \end{aligned}$$

En general, un spline de orden M con nodos $\tau_j, j = 1, \dots, r$ son trozos de polinomios de orden M y tiene derivadas continuas hasta el orden de $M - 2$. Un spline cúbico tiene $M = 4$. La forma general, de acuerdo a [3], de la base de potencias trunca es:

$$\begin{aligned}\phi_j(t) &= t^{j-1}, j = 1, \dots, M \\ \phi_{M+l}(t) &= (t - \tau_l)_+^{M-1}, l = 1, \dots, r\end{aligned}$$

Se afirma que el spline cúbico es el spline de menor orden para el cuál la discontinuidad de los nodos no es visible para el ojo humano. En la práctica los más usados son los de orden 1, 2 y 4.

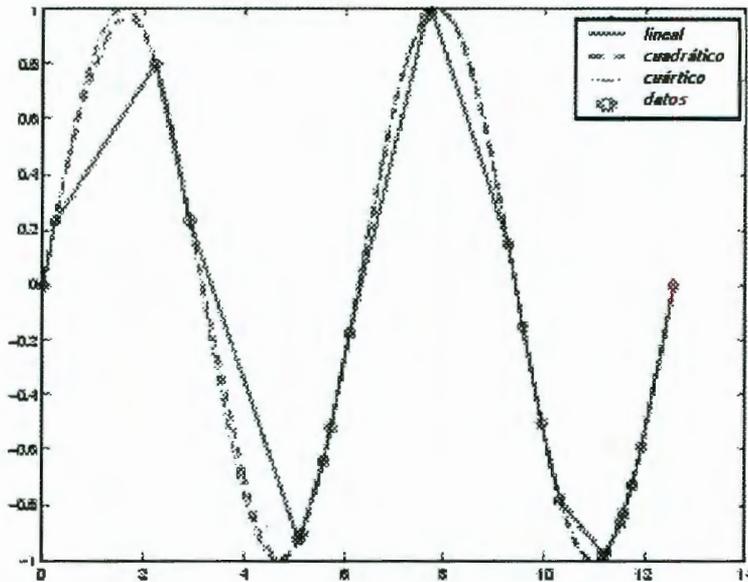


Figura 3.5. Splines de orden lineal, cuadrático y cúbico

Necesitamos seleccionar el orden del spline, el número de nodos y su posición. Un enfoque simple es parametrizar una familia de splines por el número de funciones base o grados de libertad, y que la posición de los nodos sea determinada por las observaciones x_i .

Dado que el espacio de funciones spline de un orden particular y secuencia de nodos es un espacio vectorial, existen muchas bases equivalentes para representarlos (incluso por polinomios ordinarios). Aunque la base de potencias es conceptualmente simple, no es muy atractiva numéricamente ya que potencias de números muy grandes pueden ocasionar graves problemas de series. La base *B-spline* permite cálculos eficientes aunque el número de nodos r sea grande.

El comportamiento de los polinomios que ajustan los datos tiende a tener error cerca de la frontera, y la extrapolación puede ser peligrosa. Los polinomios ajustados fuera de los nodos frontera tienen un comportamiento menos controlado que el correspondiente a los polinomios globales en la región.

Un spline cúbico agrega la restricción adicional de que la función sea lineal fuera de los nodos frontera, con lo que tenemos cuatro grados de libertad libres (dos restricciones en las dos fronteras), que pueden ser utilizados para agregar más nodos interiores en la región. Suponer que la función es lineal cerca de la frontera (que es donde menos información tenemos) es causa de error cerca de la frontera, pero es considerado razonable.

Un spline cúbico con r nodos es representado por K funciones base, donde:

$$\phi_1(t) = 1, \phi_2(t) = t, \phi_{k+2}(t) = d_k(t) + d_{K-1}(t),$$

donde

$$d_k(t) = \frac{(t - \tau_k)_+^3 - (t - \tau_K)_+^3}{\tau_K - \tau_k}.$$

Cada una de estas funciones base tienen segunda y tercera derivada igual a cero para

$$t \geq \tau_K.$$

3.4 PENALIZACIÓN DE RUGOSIDAD

3.4.1 Introducción

En esta sección introduciremos una tercera opción para aproximar datos discretos a una función.

Hemos visto que la expansión de bases trabaja bien si las funciones base tienen la misma característica esencial que el proceso que generó los datos. Por ejemplo, una base de Fourier es útil si las funciones observadas son periódicas y no exhiben fluctuaciones en algún intervalo particular. Sin embargo la expansión en bases no tiene control sobre el grado de suavizamiento y puede ser costoso calcularlo si la base no exhibe ortogonalidad o soporte local.

Técnicas de ajuste local por polinomios están basadas en algoritmos atractivos, eficientes y fáciles de entender que son simples modificaciones de técnicas clásicas de estadística, ellas ofrecen control continuo de la suavidad de la aproximación, pero rara vez son soluciones óptimas de un problema estadístico explícito.

Como en el enfoque de expansión en bases, los métodos de penalización de rugosidad están basados en una declaración explícita de la representación suave de los datos, pero la necesidad de tener una representación suave se expresa explícitamente en el nivel en el que el criterio sea optimizado, más aún, se pueden aplicar a un rango mayor de problemas de suavizamiento que simplemente estimar una curva x de observaciones $x(t_i)$ para ciertos puntos t_i .

3.4.2 Spline suavizado

Consideremos como trabaja el método de regularización en el caso funcional cuando el objetivo es estimar una función x sobre la base de observaciones discretas y ruidosas de un vector y .

El método de spline suavizado estima una curva x de observaciones $y_j = x(t_j) + \varepsilon_j$ haciendo explícito dos posibles objetivos en la estimación de la curva. Por un lado, queremos garantizar que la curva estimada sea un buen ajuste de los datos, por ejemplo en términos de la suma de cuadrados de residuales $\sum_j \{y_j - x(t_j)\}^2$. Por otro lado, no queremos que el ajuste sea muy bueno si resulta una curva x que sea localmente variable. Estos dos objetivos corresponden a los dos elementos del

$$\text{Error cuadrático medio} = \text{Sesgo}^2 + \text{Varianza muestral}$$

De acuerdo al modelo $y_j = x(t_j) + \varepsilon_j$, una estimación completamente insesgada del valor de la función $x(t_j)$ puede ser producida por una curva que ajuste a y_j exactamente, pero dicha curva debe tener una varianza alta, manifestada en la rápida variación local de la curva. En el spline suavizado, como en otros métodos suavizantes, el error cuadrático medio puede ser reducido dramáticamente sacrificando algún sesgo que reduzca la varianza muestral, lo cual es una razón importante para imponer suavidad a la curva estimada.

En el enfoque de función base restringimos x para que sea una combinación lineal de un número pequeño K de funciones base y minimizamos la suma de cuadrados de residuales sujeto a dicha restricción. Por otro lado, en el método de penalización de rugosidad se permite una gran clase de funciones x para cuantificar la rapidez de la variación local de

x y hacer un equilibrio entre regularidad y bondad de ajuste de los datos, lo que corresponde a un balance entre varianza y sesgo.

Una medida de la rugosidad de una función presentada en [1], se obtiene integrando su segunda derivada al cuadrado,

$$PEN_2(x) = \int \{D^2x(t)\}^2 dt = \|D^2x\|^2 \quad (3.12)$$

la cual evalúa la curvatura total en x . Por lo que para funciones altamente variables se espera tener un valor alto de $PEN_2(x)$.

Se define la suma de cuadrados de residuales penalizada en [1] como

$$PENSSSE_\lambda(x|\mathbf{y}) = \sum_j \{y_j - x(t_j)\}^2 + \lambda \times PEN_2(x) \quad (3.13)$$

La función estimada se obtiene encontrando la función x que minimice $PENSSSE_\lambda(x)$ sobre el espacio de funciones de x para la cual $PEN_2(x)$ está definida.

El parámetro λ es un parámetro de suavizamiento que mide el cambio entre el ajuste de los datos, y se obtiene por la suma de cuadrados de residuales, y la variabilidad de la función x se cuantifica por medio de $PEN_2(x)$. Si λ es grande, entonces funciones que no son lineales deben tener alta penalización de rugosidad en $PENSSSE_\lambda(x)$. Por lo que cuando $\lambda \rightarrow \infty$ la curva ajustada x aproxima la regresión lineal estándar a los datos observados. Si λ es pequeño la curva tiende a ser más y más variable dado que es menor la penalización de rugosidad, cuando $\lambda \rightarrow 0$ la curva x aproxima un interpolador de los datos, satisfaciendo $x(t_j) = y_j$ para todo j .

La curva x resultante del uso de penalización de rugosidad PEN_2 puede ser mostrada mediante un spline cúbico con nodos en los puntos t_j .

3.4.3 Estimación de derivadas mediante spline suavizado

Muchos análisis funcionales de datos estiman derivadas porque son de interés directo o bien porque juegan un papel importante en alguna parte del análisis. La penalización (3.12) puede no ser conveniente dado que ella controla la curvatura de x y sólo la pendiente en la derivada Dx .

Si la derivada de orden m es la mayor que se requiere debemos penalizar probablemente dos derivadas más altas. Por ejemplo, el estimado de la aceleración es mejor si usamos

$$PEN_4(x) = \int_T \{D^4x(s)\}^2 ds = \|D^4x\|^2 \quad (3.14)$$

en (3.13) dado que éste controla la curvatura en D^2x .

3.4.4 Penalización de rugosidad generalizada

Una extensión del método de penalización de rugosidad es usar más medidas de rugosidad además de $\|D^2x\|^2$. Existen dos razones para considerar penalización de rugosidad generalizada. Por un lado, queremos que la clase de funciones con rugosidad cero sea la mayor, o diferente de aquellas que son de la forma $a + bt$. Por otro, que al menos localmente las curvas x satisfagan una ecuación diferencial particular y penalizar su desviación.

Podemos lograr estos dos objetivos reemplazando la segunda derivada D^2 por un operador lineal diferencial más general L , definido como

$$Lx = w_0x + w_1Dx + \dots + w_{m-1}D^{m-1}x + D^m x,$$

donde los pesos w_j pueden ser considerados como constantes o como funciones $w_j(t)$.

Entonces [1] define

$$PEN_L(x) = \|Lx\|^2$$

la integral del cuadrado de Lx .

3.4.5 Enfoque de bases regularizadas

En esta sección permitiremos que la función tenga una expansión en bases de dimensión mayor y usaremos penalización de rugosidad en el ajuste de la función a los datos observados.

Supongamos que tenemos dos conjuntos de funciones base, ϕ_j , $j = 1, \dots, J$ y ψ_k , $k = 1, \dots, K$, que se complementan una a otra. Sean las funciones ϕ_j las de menor número y seleccionadas para dar una razonable descripción a gran escala de la característica de los datos. Las funciones base complementarias ψ_k son generalmente mayores en número, y son diseñadas para captar características locales y otras no representadas por ϕ_j . Supongamos que alguna función x de interés puede ser expresada en términos de las dos bases como

$$x(t) = \sum_{j=1}^J d_j \phi_j(t) + \sum_{k=1}^K c_k \psi_k(t) \quad (3.15)$$

Se desarrolló una penalización de rugosidad para x (en [1]) de tal forma que las combinaciones lineales de ϕ_j tienen un efecto completamente suave y no contribuyen a la penalización de rugosidad. Entonces la penalización de rugosidad debe depender sólo de los coeficientes de ψ_i . Podemos pensar x como la suma de dos partes, una función ultrasuave $x_S = \sum_j d_j \phi_j$ y una función $x_R = \sum_k c_k \psi_k$. Buscaremos una medida de rugosidad $PEN(x_R)$ y consideraremos la función x_R expresada únicamente en términos de ψ_k . Se

define

$$PEN_0(x_R) = \int_T x_R(t)^2 dt = \int (\mathbf{c}'\boldsymbol{\psi})^2 = \int_T \left[\sum_{k=1}^K c_k \psi_k(t) \right]^2 dt.$$

Otra posibilidad es tomar un cierto orden de derivadas al cuadrado e integrar. Por ejemplo, podemos utilizar

$$PEN_2(x_R) = \int (D^2 x_R)^2 = \int_T \left[\sum_{i=1}^I c_i D^2 \psi_i(t) \right]^2 dt.$$

En general, podemos usar algún operador lineal diferencial L , definiendo

$$PEN_L(x_R) = \int (Lx_R)^2 = \int_T \left[\sum_{k=1}^K c_k L\psi_k(t) \right]^2 dt.$$

Haciendo L el operador identidad o la segunda derivada obtenemos PEN_0 y PEN_2 como casos especiales.

Podemos expresar estas penalizaciones en términos de matrices como

$$PEN_L(x_R) = \mathbf{c}'\mathbf{R}\mathbf{c}$$

donde la matriz simétrica \mathbf{R} de orden K contiene los elementos

$$r_{kl} = \int_T L\psi_k(t)L\psi_l(s)ds = \langle L\psi_k, L\psi_l \rangle.$$

Consideremos una función general x de la forma (3.15) y definamos la rugosidad de x como

$$PEN(x) = \mathbf{c}'\mathbf{R}\mathbf{c}$$

Para expresar la suma de cuadrados penalizada necesitamos expresar la suma de cuadrados de residuales en términos de los coeficientes de los vectores \mathbf{d} y \mathbf{c} .

$$\sum_i \{y_i - x(t_i)\}^2 = \|\mathbf{y} - \Phi\mathbf{d} - \Psi\mathbf{c}\|^2$$

donde la matriz Ψ de $n \times K$ tiene elementos $\psi_{ik} = \psi_k(t_i)$. De [1] tomamos el siguiente criterio de suavizamiento

$$PENSSE_\lambda(x|y) = \|y - \Phi d - \Psi c\|^2 + \lambda c' R c. \quad (3.16)$$

Podemos minimizar ésta forma cuadrática en d y c para encontrar la curva ajustada x en términos de su expansión (3.15). La solución para d en cualquier valor ajustado de c es

$$d = (\Phi' \Phi)^{-1} \Phi' (y - \Psi c), \quad (3.17)$$

y

$$\Phi d = P_\phi (y - \Psi c)$$

donde la matriz proyección $P_\phi = \Phi(\Phi' \Phi)^{-1} \Phi'$. Sustituyendo (3.17) en $PENSSE_\lambda$ y definiendo la proyección complementaria $Q_\phi = I - P_\phi$ y recordando que $Q_\phi Q_\phi = Q_\phi$, obtenemos

$$c = (\Psi' Q_\phi \Psi + \lambda R)^{-1} \Psi' Q_\phi y.$$

3.4.6 Algunas propiedades de los estimados

El tamaño de penalización de la componente ψ está controlado por el parámetro de suavizamiento λ . En el caso límite, cuando $\lambda \rightarrow 0$, no se aplica penalización y los estimados se obtienen minimizando el criterio $PENSSE_\lambda$. Cuando $\lambda \rightarrow \infty$, la penalización es tan severa que la contribución de ψ a la penalización de rugosidad es forzada a ser cero.

3.4.7 Relación con el enfoque de penalización de rugosidad

Para minimizar la suma de cuadrados de residuales penalizada por $\|Lx\|^2$ no necesitamos especificar las funciones ϕ_j , sólo asegurarnos que $\{\psi_k\}$ es una base conveniente para las funciones de interés. En el contexto de suavizamiento de spline, con $L = D^2$ podemos tomar los $\{\psi_k\}$ como base *B-spline* con los nodos en los datos observados.

4

MODELOS LINEALES FUNCIONALES

4.1 INTRODUCCIÓN

En esta sección se presentan algunas ideas de modelos lineales funcionales, basándonos en los conceptos de modelo lineal clásico.

El modelo funcional lineal es el siguiente

$$y = Z\beta + \epsilon \quad (4.1)$$

donde y es el vector de observaciones, β es un vector de parámetros, Z es una matriz que define una transformación lineal del espacio de parámetros al espacio de las observaciones y ϵ es un vector de errores con media cero.

Se pueden tener los siguientes casos:

1. Los parámetros y la variable respuesta son funciones.
2. La variable independiente o covariable es funcional y la variable dependiente escalar.
3. Tanto los parámetros como las variables independientes y la variable respuesta son funcionales.

Las siguientes secciones explican los dos primeros casos, que son los de especial interés en esta tesis.

4.2 MODELO FUNCIONAL LINEAL CASO 1

VARIABLE DEPENDIENTE FUNCIONAL Y VARIABLES INDEPENDIENTES NO FUNCIONALES

Hemos mencionado que los modelos lineales pueden ser funcionales en términos de la variable dependiente, independiente, o de ambas, siendo en todos los casos los coeficientes de regresión (o parámetros) funcionales. En esta sección presentamos algunas ideas del modelo funcional lineal en particular para el caso donde sólo los parámetros y la variable dependiente (observaciones) son funcionales y la matriz de diseño permanece como en un modelo general lineal clásico.

Este modelo es aplicable cuando los datos de la variable que deseamos modelar como respuesta, obtenidos mediante el experimento, pueden presentarse de la siguiente manera:

$$\begin{aligned} y_1 &: ((t_{11}, y_{11}), (t_{12}, y_{12}), \dots, (t_{1n_1}, y_{1n_1})) \\ y_2 &: ((t_{21}, y_{21}), (t_{22}, y_{22}), \dots, (t_{2n_2}, y_{2n_2})) \\ &\vdots \\ y_N &: ((t_{N1}, y_{N1}), (t_{N2}, y_{N2}), \dots, (t_{Nn_N}, y_{Nn_N})) \end{aligned}$$

donde N es el número de corridas en el experimento, y_i , es el conjunto de valores obtenidos en la i -ésima corrida registrados en los diferentes argumentos de t $(t_{i1}, t_{i2}, \dots, t_{in_i})$.

Cada uno de estos conjuntos de observaciones y_i pueden pensarse como un conjunto discreto de valores que una función subyacente $y_i(t)$ toma en los argumentos $t_{i1}, t_{i2}, \dots, t_{in_i}$, por lo que el primer objetivo es representarlos mediante una función $y_i(t)$ $t \in [t_{i1}, t_{in_i}]$, esto se hace mediante las técnicas estudiadas en el capítulo 3.

Una vez representadas cada una de las N observaciones de manera funcional se obtiene el vector de funciones \mathbf{y} , en el modelo (4.1) para el caso de interés.

$$\mathbf{y} = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{pmatrix}.$$

La matriz de diseño \mathbf{Z} , en (4.1), define la transformación lineal del espacio de parámetros, digamos \mathcal{B} , al espacio de observaciones, digamos \mathcal{Y} , en este caso la matriz \mathbf{Z} permanece igual que en el caso tradicional (con la misma estructura obtenida al modelar el experimento como multivariado o univariado); los elementos del espacio \mathcal{Y} son vectores de funciones definidos por $\mathbf{Z}\beta$, y los elementos del espacio \mathcal{B} son vectores β de funciones que tienen como argumento t .

Si la matriz \mathbf{Z} no es de rango completo, se pueden imponer ciertas restricciones sobre los parámetros, para lograr identificarlos de manera única, dichas restricciones se presentaron en el capítulo 2 en el caso de un modelo lineal clásico.

Para desarrollar el análisis del modelo en el caso 1 consideremos la siguiente notación: \mathbf{y} variable respuesta es un vector de N observaciones, cada una en forma funcional; β es un vector de dimensión q que tiene elementos con la misma forma funcional que los elementos de \mathbf{y} , y μ es la media general.

Entonces en el modelo $\mathbf{y} = \mathbf{Z}\beta + \varepsilon$, se tiene que:

$$\mathbf{y} = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{pmatrix},$$

\mathbf{Z} es la matriz de diseño convencional de dimensiones $N \times q$,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 = \mu(t) \\ \beta_1(t) \\ \vdots \\ \beta_{q-1}(t) \end{pmatrix}$$

y $\boldsymbol{\varepsilon}$ es un vector con N funciones de error con media cero $\varepsilon(t) \sim N(0, \sigma_t^2)$.

El valor esperado de $y_i(t)$ para cada t , es modelado como $z'_i \boldsymbol{\beta}(t)$.

4.2.1 Estimación del vector de parámetros $\boldsymbol{\beta}$

Para estimar el vector de funciones $\boldsymbol{\beta}$ extendemos el criterio de mínimos cuadrados al caso funcional, la cantidad $\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}$ es ahora un vector de funciones, donde los renglones de $\mathbf{Z}\boldsymbol{\beta}$ corresponden a $\langle z'_i, \boldsymbol{\beta} \rangle$ (el producto interno del i -ésimo renglón de \mathbf{Z} con el vector $\boldsymbol{\beta}$).

El criterio de ajuste por mínimos cuadrados es el siguiente

$$LMSSE(\boldsymbol{\beta}) = \int \|y(t) - \mathbf{Z}\boldsymbol{\beta}(t)\|^2 dt, \quad (4.2)$$

donde $\|\cdot\|$ representa la norma euclidiana.

Si es necesario imponer alguna restricción sobre los parámetros $\boldsymbol{\beta}$, ésta la podemos expresar como

$$\mathbf{L}\boldsymbol{\beta} = 0$$

para alguna matriz adecuada \mathbf{L} , por ejemplo a la restricción suma cero $\sum_{i=1}^{q-1} \beta_i(t) = 0$, para toda t , equivale la matriz $\mathbf{L} = (0, 1, 1, 1 \dots 1)$ de dimensión $1 \times q$ y a la restricción $\beta_1(t) = 0$, para toda t , la matriz $\mathbf{L} = (0, 1, 0, 0 \dots 0)$ de $1 \times q$.

Por lo tanto, minimizando el criterio (4.2), sujeto a la restricción $\mathbf{L}\beta(t) = 0$, para toda t , se obtiene la estimación por mínimos cuadrados $\hat{\beta}$ de los parámetros funcionales μ , $\beta_i(t)$.

Para llevar a cabo el análisis del modelo, utilizamos las mismas herramientas que en el análisis de varianza clásico con la única diferencia de que ahora estas cantidades dependen del argumento t , es decir tendremos un *ANOVA* para cada valor específico t . La suma de cuadrados del error, la suma de cuadrados alrededor de la media, el coeficiente de determinación, la suma de cuadrados medio del error y el cociente F para el caso funcional se presentan a continuación respectivamente.

$$\begin{aligned}
 SSE(t) &= \sum_j^N [y_j(t) - (\mathbf{Z}\beta)_j(t)]^2 \\
 SSY(t) &= \sum_j^N [y_j(t) - \hat{\mu}(t)]^2 \\
 RSQ(t) &= \frac{SSY(t) - SSE(t)}{SSY(t)} \\
 MSE(t) &= \frac{SSE(t)}{\text{df}(\text{error})} \\
 MSR(t) &= \frac{SSY(t) - SSE(t)}{\text{df}(\text{modelo})} \\
 \text{Cociente } F(t) &= \frac{MSR(t)}{MSE(t)}.
 \end{aligned}$$

4.2.2 Aspectos computacionales

Se ha presentado de manera general la manera de estimar el vector $\hat{\beta}$, lo que ahora se pretende es mostrar de una manera más práctica el cálculo de este vector.

Podemos minimizar el criterio (4.2), minimizando $\|y(t) - \mathbf{Z}\beta(t)\|^2$, para cada t , y minimizar $LMSSSE(\beta)$, sujeto a la restricción $\mathbf{L}\beta = 0$ minimizando la siguiente suma de cuadrados regularizada

$$\|y(t) - \mathbf{Z}\beta(t)\|^2 + \lambda \|\mathbf{L}\beta(t)\|^2 \quad (4.3)$$

para alguna constante $\lambda > 0$, conveniente. Esto nos conduce a encontrar $\hat{\beta}$ mediante la ecuación

$$(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{L}'\mathbf{L})\hat{\beta}(t) = \mathbf{Z}'y(t). \quad (4.4)$$

El vector $\hat{\beta}$, encontrado mediante esta ecuación, no depende del valor asignado a λ , puesto que al asignar cualquier valor positivo a λ solamente se toma en cuenta la restricción, pero no afecta el ajuste. El valor $\lambda = 1$ normalmente funciona pero este puede variar si se encuentran inestabilidades numéricas.

La manera más directa de encontrar $\hat{\beta}$ de la expresión (4.4) es mediante una malla de valores para el argumento t , que usualmente se elige de tal forma que corresponda con la discretización de los datos originales, es decir los valores $t_{i1}, t_{i2}, \dots, t_{in_i}$ correspondientes a cada observación y_i . Una vez que se han encontrado un conjunto discreto de valores para cada una de las funciones que componen el vector $\hat{\beta}$, se interpolan para determinar la función respectiva.

A menudo es más apropiado encontrar el vector de parámetros estimados $\hat{\beta}$, en términos de las funciones base que utilizamos al representar las observaciones y_i como funciones $y_i(t)$.

Sea ϕ el vector de funciones base linealmente independientes, de dimensión K y \mathbf{Y} la matriz de dimensión $N \times K$, con los coeficientes del vector de funciones observadas

\mathbf{y} , tal que $\mathbf{y} = \mathbf{Y}\phi$, el j -ésimo renglón de \mathbf{Y} contiene los coeficientes correspondientes a (ϕ_1, \dots, ϕ_k) en la expansión de la función observada $y_j(t)$.

Expresando el vector $\hat{\beta}$ en términos de la misma base ϕ , se tiene $\hat{\beta} = \mathbf{B}\phi$, para alguna matriz \mathbf{B} , de dimensión $q \times K$, que esta por determinar, al sustituir en (4.4) observamos que \mathbf{B} satisface el siguiente sistema de ecuaciones

$$(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{L}'\mathbf{L})\mathbf{B} = \mathbf{Z}'\mathbf{Y}$$

resolviendolo para \mathbf{B} , encontramos $\hat{\beta}$ mediante $\hat{\beta} = \mathbf{B}\phi$, en términos de las funciones base ϕ usadas.

Hasta ahora nos hemos preocupado solamente por determinar $\hat{\beta}$, sin prestar atención a la suavidad de las funciones que la componen, estas funciones serán razonablemente suaves cuando las funciones observadas que forman el vector \mathbf{y} lo sean, pero en ocasiones no sucede así, como resultado se obtienen funciones de regresión muy rugosas que varían mucho en intervalos pequeños y generalmente son difíciles de interpretar, en estos casos es deseable incorporar restricciones de suavidad en el ajuste del modelo funcional, comentaremos brevemente dos posibles enfoques para lograrlo.

El método más simple es incorporar un número de funciones base relativamente pequeño (K pequeño) en el ajuste, de esta manera se obtendrán funciones observadas más suaves y automáticamente funciones de regresión más suaves.

Otra manera es penalizar la rugosidad en el criterio $LMSSE(\beta)$ (4.2). Si definimos

$$PEN_2(\beta) = \sum_j PEN_2(\beta_j) = \sum_j \int (D^2\beta_j)^2$$

el criterio penalizado se expresa como

$$LMSSE_{\lambda}(\beta) = LMSSE(\beta) + \lambda \times PEN_2(\beta) \quad (4.5)$$

donde $\lambda \geq 0$, es una constante llamada parámetro de suavizamiento, que determina que tanto se tomara en cuenta la rugosidad (mientras más pequeños sea λ se tendrán funciones más rugosas en $\hat{\beta}$).

Al minimizar la expresión (4.5), sujeta a la restricción $L\beta = 0$, obtenemos los parámetros estimados $\hat{\beta}$, suavizados.

Computacionalmente es probablemente más directo implementar este procedimiento de suavizamiento utilizando funciones base. Dada una base ϕ , defina $\mathbf{R} = \int (D^2\phi)(D^2\phi)'$ y $\mathbf{J} = \langle \phi, \phi' \rangle$ la matriz de producto interior, entonces podemos escribir $\beta = \mathbf{B}\phi$ y $\mathbf{y} = \mathbf{Y}\phi$ para matrices adecuadas de coeficientes \mathbf{B} y \mathbf{Y} , y el criterio (4.5) como

$$LMSSE_{\lambda}(\beta) = \text{traza}(\mathbf{Y} - \mathbf{ZB})\mathbf{J}(\mathbf{Y} - \mathbf{ZB})' + \lambda \text{traza}(\mathbf{BRB}') \quad (4.6)$$

Si no se tienen restricciones sobre β entonces encontramos \mathbf{B} mediante la ecuación

$$(\mathbf{Z}'\mathbf{JZ} + \lambda\mathbf{R})\mathbf{B} = \mathbf{Z}'\mathbf{JY}.$$

Si se tienen restricciones sobre el vector de parámetros, expresadas como $L\beta = 0$, utilizamos la técnica de descomposición \mathbf{QR} que se presenta a continuación para escribir $\beta = \mathbf{C}\alpha$ para ciertas matrices \mathbf{C} y α

La descomposición \mathbf{QR}

La técnica de descomposición \mathbf{QR} de una matriz \mathbf{Z} de $m \times n$, nos permite escribir \mathbf{Z} como el producto de dos matrices \mathbf{Q} y \mathbf{R}

$$\mathbf{Z} = \mathbf{QR}$$

donde \mathbf{Q} es una matriz de $m \times m$ ortogonal ($\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$) y \mathbf{R} es una matriz de dimensión $m \times n$ triangular superior ($r_{ij} = 0$ si $i > j$).

En el caso de que $m > n$, los últimos $m - n$ renglones de \mathbf{R} son cero y cada una de las últimas $m - n$ columnas de \mathbf{Q} , denotadas cada una de éstas como \mathbf{x} , satisfacen $\mathbf{x}'\mathbf{Z} = 0$, omitiendo estos renglones y columnas se tiene que $\mathbf{Z} = \mathbf{Q}_1\mathbf{R}_1$, donde \mathbf{R}_1 es una matriz triangular superior de $n \times n$ y \mathbf{Q}_1 de dimensiones $m \times n$, es una matriz con columnas ortogonales.

Para descomponer β de $q \times 1$, aplicando la técnica, se tiene $\beta = \mathbf{C}\alpha$ donde \mathbf{C} es una matriz de $q \times 1$ y α un escalar (matriz de 1×1)

Dado que sabemos que $\beta = \mathbf{B}\phi$ y $\beta = \mathbf{C}\alpha$, entonces $\mathbf{C}\alpha = \mathbf{B}\phi$, de donde $\mathbf{B} = \mathbf{C}\mathbf{A}$ para alguna matriz \mathbf{A} ($\alpha\phi^{-1}$), al sustituir $\mathbf{B} = \mathbf{C}\mathbf{A}$ en (4.6), obtenemos

$$(\mathbf{C}'\mathbf{Z}'\mathbf{J}\mathbf{Z}\mathbf{C} + \lambda\mathbf{C}'\mathbf{R}\mathbf{C})\mathbf{A} = \mathbf{C}'\mathbf{Z}'\mathbf{J}\mathbf{Y}$$

de donde podemos encontrar la matriz \mathbf{C} que nos permitirá encontrar \mathbf{B} y $\hat{\beta}$ con los parámetros suavizados.

4.3 MODELO LINEAL FUNCIONAL CASO 2

VARIABLE DEPENDIENTE ESCALAR Y VARIABLES INDEPENDIENTES FUNCIONALES

4.3.1 Introducción

Consideremos un modelo donde la predicción de valores escalares y_i esta basada en funciones x_i y sea β una función, entonces un modelo factible es:

$$y_i = \alpha + \int_0^T x_i(t)\beta(t)dt + \varepsilon_i = \alpha + \langle x_i, \beta \rangle + \varepsilon_i. \quad (4.7)$$

Podemos estimar los parámetros en (4.7) minimizando la suma de cuadrados de residuales

$$LMSSE(\alpha, \beta) = \sum_{i=1}^N (y_i - \alpha - \langle x_i, \beta \rangle)^2 = \|\mathbf{y} - \alpha - \langle \mathbf{x}, \beta \rangle\|^2, \quad (4.8)$$

donde \mathbf{x} es el vector de longitud N de las funciones covariables $(x_1, \dots, x_N)'$.

Podemos ver que la función de coeficientes de regresión β está indeterminada sobre la base de cualquier muestra finita (x_i, y_i) , porque tenemos un número infinito de parámetros $\beta(t)$ y sólo un número finito de condiciones $y_i = \alpha + \langle x_i, \beta \rangle$ para satisfacer.

Dado que el espacio de funciones que satisfacen (4.7) es de dimensión infinita no importa cuan grande sea la muestra de tamaño N , minimizando la suma de cuadrados de residuales no podemos producir un estimador consistente de β en (4.7). Por lo que para obtener un buen estimador o identificar un único $\hat{\beta}$, debemos usar algún metodo de regularización.

4.3.2 Regularización usando funciones base

Para alguna K dada, supongamos que $\phi = (\phi_1, \dots, \phi_K)'$ es una base y que la función de regresión β tiene la expansión

$$\beta = \sum_{v=1}^K b_v \phi_v = \mathbf{b}'\phi. \quad (4.9)$$

Expandemos las covariables como

$$x_i = \sum_{v=1}^K c_{iv} \phi_v = \mathbf{c}'_i \phi. \quad (4.10)$$

Para tener en cuenta bases que no son necesariamente ortonormales definimos a \mathbf{J} como la matriz $\mathbf{J} = \int \phi \phi'$ con entradas

$$\mathbf{J}_{jk} = \int \phi_j(s) \phi_k(s) ds = \langle \phi_j, \phi_k \rangle. \quad (4.11)$$

Ahora

$$\langle x_i, \beta \rangle = \sum_{j=1}^K \sum_{v=1}^K c_{ij} \mathbf{J}_{jv} b_v \quad (4.12)$$

y la matriz de coeficientes \mathbf{Z} se define ahora como $\mathbf{Z} = [1 \ \mathbf{C}\mathbf{J}]$. De igual manera definimos el vector de parámetros aumentado $\zeta = (\alpha, b_0, \dots, b_{M-1})'$ y su estimado por mínimos cuadrados $\hat{\zeta}$ es la solución de la ecuación

$$\mathbf{Z}'\mathbf{Z}\hat{\zeta} = \mathbf{Z}'\mathbf{y} \quad (4.13)$$

si son ortonormales, como Fourier, $\mathbf{J} = \mathbf{I}$.

4.3.3 Regularización con penalización de rugosidad

El objetivo de la penalización de rugosidad es evadir excesiva fluctuación local en la función estimada.

Dada una función periódica dos veces diferenciable β , definimos la suma de cuadrados de residuales penalizada

$$\begin{aligned} PENSSE_{\lambda}(\alpha, \beta) &= \sum_{i=1}^N [y_i - \alpha - \langle x_i, \beta \rangle]^2 + \lambda \int_{\tau} [D^2\beta(s)]^2 ds \quad (4.14) \\ &= \|\mathbf{y} - \alpha - \langle x_i, \beta \rangle\|^2 + \lambda \|D^2\beta\|^2 \end{aligned}$$

donde $\int (D^2\beta)^2$ cuantifica la variación en β y el parámetro de suavizamiento $\lambda > 0$ controla la variación entre rugosidad e infidelidad de los datos observados.

Podemos escoger el parámetro de suavizamiento λ subjetivamente o por validación cruzada. Para aplicar validación cruzada, sean $\alpha_{\lambda}^{(-j)}$ y $\beta_{\lambda}^{(-j)}$ los estimados de α y β obtenidos minimizando la suma de cuadrados de residuales penalizada basados en todos los datos excepto (x_j, y_j) . Podemos definir validación cruzada como:

$$CV(\lambda) = \sum_{j=1}^N \left(y_j - \alpha_{\lambda}^{(-j)} - \langle x_j, \beta_{\lambda}^{(-j)} \rangle \right)^2 \quad (4.15)$$

y minimizando $CV(\lambda)$ sobre λ obtenemos la elección automática de λ .

4.3.4 Aspectos computacionales de regularización

Ahora consideraremos varios métodos para calcular regresión funcional y validación cruzada. La elección de cuál método es mejor, a veces depende del tamaño del problema.

Discretización directa

Podemos aproximar el mínimo de la suma de cuadrados de residuales penalizada (4.14) discretizando las curvas de covariables x_i y la función parámetro β en una malla fina. También, podemos aproximar la penalización de rugosidad $\lambda \|D^2\beta\|^2$ por un múltiplo de la suma de cuadrados de segundas diferencias, y las integrales $\int x_i\beta$ por sumas. La

minimización es entonces una forma cuadrática de alta dimensión, y puede ser resuelta mediante métodos numéricos.

Expansión completa de bases

El método de funciones base es útil porque es simple de aplicar, y algunos problemas sugieren una elección particular de la base. El primer paso es representar el problema regularizado ajustado en términos de expansión de funciones base, y entonces aplicar el concepto de regularización a dicha representación.

Supongamos que expandemos las funciones covariables x_i y las funciones de regresión β en M términos de las funciones base ϕ_v , como en (4.9) y (4.10). Definimos la matriz \mathbf{K} que tiene entradas

$$k_{ij} = \int D^2 \phi_j(s) D^2 \phi_k(s) ds = \langle D^2 \phi_j, D^2 \phi_k \rangle. \quad (4.16)$$

En el caso de Fourier, \mathbf{K} es diagonal, con elementos w_k^4 , donde $w_{2j-1} = w_{2j} = 2\pi j$.

En general, la suma de cuadrados de residuales penalizada es

$$\begin{aligned} PENSSE_\lambda(\alpha, \beta) &= \sum_i (y_i - \alpha - \sum_v c_{iv} b_v \phi_v)^2 + \lambda \int \left\{ \sum_v b_v D^2 \phi_v(t) \right\}^2 dt \\ &= \|\mathbf{y} - \alpha - \mathbf{C}\mathbf{b}\|^2 + \lambda \mathbf{b}\mathbf{K}\mathbf{b} \end{aligned} \quad (4.17)$$

Definamos el vector $\zeta = (\alpha, \mathbf{b})'$, y $\mathbf{Z} = [1 \ \mathbf{C}\mathbf{J}]$ la matriz de coeficientes de $N \times (M + 1)$, con J definida como en (4.11).

Sea \mathbf{K}_0 la matriz de penalización \mathbf{K} aumentada, agregando un renglón y una columna de $M + 1$ ceros, por lo que la expresión (4.17) simplificada es:

$$PENSSE_\lambda(\zeta) = \|\mathbf{y} - \mathbf{Z}\zeta\|^2 + \lambda \zeta' \mathbf{K}_0 \zeta. \quad (4.18)$$

Por lo que el valor que minimiza $\hat{\zeta}$ satisface

$$(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K}_0)\hat{\zeta} = \mathbf{Z}'\mathbf{y}. \quad (4.19)$$

Un método útil es trabajar dentro de un problema de dimensión pequeña usando un número moderado de K funciones base. Esto reducirá el tamaño del sistema de ecuaciones (4.19).

En contraste con la sección 4.3.2, la idea no es usar K como parámetro de suavizamiento. La regularización es controlada principalmente por el parámetro λ y la reducción de dimensión es una idea numérica para reducir los cálculos sin alterar sustancialmente el resultado actual. En la práctica, el valor de K alrededor de 20 ó 30 da buenos resultados, excepto posiblemente para valores muy pequeños del parámetro de suavizamiento y conjuntos muy grandes de datos.

4.3.5 Método directo de penalización para calcular β

Ahora estudiaremos una manera directa de usar el método de penalización de rugosidad que calcula $\hat{\beta}$ sin usar funciones base.

La tarea central es mostrar cómo podemos instaurar este método como un proceso de dos etapas, primero minimizar una expresión cuadrática simple para obtener el vector de valores $\hat{\mathbf{y}}$, y después calcular el interpolador lineal funcional suavizante de estos valores.

Hemos visto que los datos observados pueden ser ajustados exactamente por un número infinito de posibles elecciones de parámetros (α, β) . Puede ser de interés definir un interpolador funcional $(\tilde{\alpha}, \tilde{\beta})$ para los datos dados por la elección del parámetro de suavizamiento. Necesitamos que los estimados $(\tilde{\alpha}, \tilde{\beta})$ minimicen $\|D^2\beta\|^2$, sujeto a las N

restricciones

$$y_i = \tilde{\alpha} + \langle x_i, \tilde{\beta} \rangle \quad (4.20)$$

El interpolador funcional es el caso límite del estimador regularizado de la sección 4.3.3 cuando $\lambda \rightarrow 0$.

Consideraremos el problema (4.20) como una manera de cuantificar la rugosidad o irregularidad del vector respuesta \mathbf{y} relativo a las covariables funcionales observadas x_i . En general, si z_1, \dots, z_N es alguna sucesión de valores, entonces podemos definir la rugosidad de z relativo a las covariables funcionales x_i como la rugosidad de la función suavizante β_z tal que

$$z_i = \alpha_z + \langle x_i, \beta_z \rangle \quad (4.21)$$

para toda i , para alguna constante α_z .

Las dos etapas del proceso de minimización

Supongamos que podemos definir una matriz \mathbf{R} de orden N tal que la rugosidad de la variable z puede ser expresada como la forma cuadrática

$$\int [D^2 \beta(t)]^2 dt = \mathbf{z}' \mathbf{R} \mathbf{z} \quad (4.22)$$

con la cuál podemos conceptualizar la solución del problema de suavizamiento dividiendo la minimización de la suma de cuadrados de residuales penalizada en dos etapas:

Etapa 1: Encontrar $\hat{\mathbf{y}}$ tal que minimice $PENSSE_\lambda(\hat{\mathbf{y}}) = \sum_i (y_i - \hat{y}_i)^2 + \lambda \hat{\mathbf{y}}' \mathbf{R} \hat{\mathbf{y}}$, la solución es $\hat{\mathbf{y}} = (\mathbf{I} + \lambda \mathbf{R})^{-1} \mathbf{y}$

Etapa 2: Encontrar el interpolador funcional lineal más suave (α, β) que satisfaga

$$\hat{y}_i = \alpha + \int x_i(t) \beta(t) dt. \quad (4.23)$$

El siguiente argumento muestra que estas dos etapas inducen la minimización de $PENSSE_\lambda(\alpha, \beta)$.

Escriba el problema de minimización primero minimizando $PENSSE_\lambda(\alpha, \beta)$ como función de (α, β) con $\hat{\mathbf{y}}$ fijo, después minimice el resultado con respecto a $\hat{\mathbf{y}}$. Formalmente, esto es

$$\min_{\hat{\mathbf{y}}}[\min_{\alpha, \beta}\{PENSSE_\lambda(\alpha, \beta)\}] = \min_{\hat{\mathbf{y}}}\left\{\sum (y_i - \hat{y}_i)^2 + \lambda \min_{\beta} \int [D^2\beta(t)]^2 dt\right\} \quad (4.24)$$

pero de acuerdo a la suposición (4.22)

$$PENSSE_\lambda(\alpha, \beta) = \min_{\hat{\mathbf{y}}}\left\{\sum (y_i - \hat{y}_i)^2 + \lambda \hat{\mathbf{y}}' \mathbf{R} \hat{\mathbf{y}}\right\} \quad (4.25)$$

Ahora presentaremos una solución al problema del interpolador funcional lineal presentado en la etapa 2. El objetivo es encontrar el interpolador funcional lineal más suave $(\tilde{\alpha}, \tilde{\beta})$ para un vector específico $\hat{\mathbf{y}}$ de longitud N relativo a las covariables dadas $x_i, i = 1, \dots, N$.

Sea la matriz \mathbf{Z} definida en términos de las covariables funcionales $x_i, \mathbf{Z} = [1 \ \mathbf{C}\mathbf{J}]$.

En términos de expansión de bases, queremos resolver el siguiente problema:

$$\min\{\zeta' \mathbf{R} \zeta\} \text{ sujeto a } \mathbf{Z} \zeta = \hat{\mathbf{y}} \quad (4.26)$$

Supongamos que las primeras M_0 funciones base ϕ_v tienen rugosidad $\int (D^2 f)^2 = 0$.

Recordemos la matriz \mathbf{K} que tiene entradas

$$k_{ij} = \int D^2 \phi_j(s) D^2 \phi_k(s) ds = \langle D^2 \phi_j, D^2 \phi_k \rangle.$$

Sea \mathbf{K}_2 la matriz obtenida removiendo los primeros M_0 renglones y columnas de \mathbf{K} , luego \mathbf{K}_2 es estrictamente definida positiva, y los renglones y columnas removidos son todos cero.

Sea \mathbf{Z}_1 la matriz de las primeras $M_0 + 1$ columnas de \mathbf{Z} , y sea \mathbf{Z}_2 las restantes columnas. Definiendo \mathbf{P} la matriz proyección de $N \times N$, $\mathbf{P} = \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1$ nos permite definir $\mathbf{Z}^* = \mathbf{P}\mathbf{Z}_2$.

Sea ζ_1 el vector de las primeras $M_0 + 1$ componentes de ζ , y sea ζ_2 las componentes restantes de ζ . Multiplicando ambos lados de la restricción

$$\mathbf{Z}\zeta = \mathbf{Z}_1\zeta_1 + \mathbf{Z}_2\zeta_2 = \hat{y}$$

por \mathbf{Z}'_1 , implica que

$$\mathbf{Z}'_1\mathbf{Z}_1\zeta_1 + \mathbf{Z}'_1\mathbf{Z}_2\zeta_2 = \mathbf{Z}'_1\hat{y}. \quad (4.27)$$

Resolviendo para ζ_1 obtenemos

$$\zeta_1 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1(\hat{y} - \mathbf{Z}_2\zeta_2) \text{ y } \mathbf{Z}_1\zeta_1 = (\mathbf{I} - \mathbf{P})(\hat{y} - \mathbf{Z}_2\zeta_2). \quad (4.28)$$

Entonces una vez que ζ_2 ha sido determinado podemos encontrar ζ_1 . Sustituyendo la solución (4.28) de ζ_1 en la restricción (4.27) y reorganizando podemos encontrar ζ_2 resolviendo el problema de minimización

$$\min_{\zeta_2} \{\zeta'_2\mathbf{K}_2\zeta_2\} \text{ sujeto a } \mathbf{Z}^*\zeta = \mathbf{P}\hat{y} \quad (4.29)$$

usando el hecho que $\zeta'\mathbf{K}\zeta = \zeta'_2\mathbf{K}_2\zeta_2$.

Sea \mathbf{R} definida como la g-inversa Moore-Penrose de $\mathbf{Z}^*\mathbf{K}_2^{-1}\mathbf{Z}'^*$, es decir

$$\mathbf{R} = (\mathbf{Z}^*\mathbf{K}_2^{-1}\mathbf{Z}'^*)^+ = \mathbf{Z}^*\mathbf{K}_2\mathbf{Z}'^* \quad (4.30)$$

La solución del problema (4.29) está dado por

$$\zeta_2 = \mathbf{K}_2^{-1} \mathbf{Z}^* \mathbf{R} \hat{\mathbf{y}}, \quad (4.31)$$

y el valor mínimo de la función objetivo $\zeta' \mathbf{R} \zeta$ esta dado por

$$\zeta' \mathbf{R} \zeta = \zeta_2' \mathbf{K}_2 \zeta_2 = \hat{\mathbf{y}}' \mathbf{R} \mathbf{Z}^* \mathbf{K}_2^{-1} \mathbf{K}_2 \mathbf{K}_2^{-1} \mathbf{Z}^* \mathbf{R} \hat{\mathbf{y}} = \hat{\mathbf{y}}' \mathbf{R} \mathbf{R}^+ \mathbf{R} \hat{\mathbf{y}} = \hat{\mathbf{y}}' \mathbf{R} \hat{\mathbf{y}}, \quad (4.32)$$

la cual es la suposición utilizada en el procedimiento de dos etapas.

Finalmente regresando a la técnica de dos etapas para suavizamiento, podemos realizar la primera etapa resolviendo la ecuación

$$(\mathbf{I} + \lambda \mathbf{R}) \hat{\mathbf{y}} = \mathbf{y}.$$

4.3.6 Validación cruzada

Tenemos la posibilidad de escoger el parámetro de suavizamiento λ por validación cruzada.

Sea \mathbf{S} la matriz sombrero del procedimiento de suavizamiento la cual mapea los valores de datos \mathbf{y} a los valores ajustados $\hat{\mathbf{y}}$ para un valor particular λ . El valor de validación cruzada [7] satisface.

$$CV(\lambda) = \sum_{i=1}^N \left(\frac{\mathbf{y} - \hat{\mathbf{y}}_i}{1 - s_{ii}} \right)^2.$$

Si N es de tamaño moderado y usamos el algoritmo en dos etapas, entonces la matriz sombrero satisface

$$\mathbf{S} = (\mathbf{I} + \lambda \mathbf{R})^{-1}.$$

El cálculo de la matriz simétrica \mathbf{R} como g-inversa Moore-Penrose significa que podemos expresar \mathbf{R} como $\mathbf{U} \text{diag}(\rho_1, \rho_2, \dots) \mathbf{U}'$, donde \mathbf{U} es una matriz ortogonal y ρ_i son

los eigenvalores de \mathbf{R} . Se sigue que los elementos de la diagonal de \mathbf{S} estan dados por

$$s_{ii} = \sum_j (1 + \lambda\rho_j)^{-1} u_{ij}^2.$$

Si N es grande y consideramos una expansión en un número moderado K de funciones base, entonces podemos encontrar los elementos de la diagonal de \mathbf{S} mediante

$$\mathbf{S} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R})^{-1}\mathbf{Z}'$$

De \mathbf{S} podemos calcular un indicador de los grados de libertad efectivos usados en la aproximación, para ésto es recomendado utilizar la traza de \mathbf{S} o la traza de \mathbf{S}^2 .

5

APLICACIONES AL MODELO LINEAL FUNCIONAL

5.1 INTRODUCCIÓN

En este capítulo se presentarán dos aplicaciones del modelo funcional lineal cuando la respuesta es funcional y las variables independientes son multivariadas, y una aplicación al modelo funcional cuando la respuesta es escalar y las variables independientes son funcionales.

La primer aplicación fué un experimento en el cual las unidades experimentales son lotes de cinco ajos que se almacenan durante un periodo en el cual pierden peso, el propósito de este experimento es estudiar la pérdida de peso, los perfiles obtenidos en este experimento fueron utilizados para otra aplicación donde la variable respuesta es el índice de brotación en los ajos y la variable independiente es el perfil de pérdida de peso de cada tratamiento. En la segunda aplicación las unidades experimentales son frascos con jugo de yogurt con frutas, en los que se busca aumentar la calidad del producto.

5.2 APLICACIÓN UNO AL MODELO LINEAL FUNCIONAL CASO 1

El ajo es comúnmente almacenado sin control climático en su periodo de baja disponibilidad, es de interés para los productores que durante su almacenamiento no tenga una pérdida

significativa de peso ya que esto representaría pérdidas económicas. El objetivo del experimento es el estudio de la pérdida de peso, a través del tiempo en diferentes condiciones de almacenamiento (mostradas en la Tabla 5.1).

Tratamientos
0°C
20°C
30°C
5°C
0°C 70% humedad relativa (<i>h.r.</i>)
Ambiente

Tabla 5.1 Condiciones de almacenamiento del ajo

En cada condición se tuvo un lote con 360 ajos que fue monitoreado durante 190 días, de cada lote se escogieron tres conjuntos de cinco bulbos cada uno, los cuales representan tres unidades experimentales correspondientes a las tres réplicas en cada tratamiento. Cada unidad experimental fue medida repetidamente durante 190 días cada 10 días

El diseño experimental es completamente aleatorizado con un factor (condición de almacenamiento) de seis niveles y tres réplicas, la variable respuesta es la pérdida de peso de los ajos a lo largo de los 190 días.

Como ya se mencionó, de cada tratamiento se realizaron tres réplicas, por lo tanto se contó con 18 observaciones medidas a través del tiempo. Cada una de ellas consiste en 20 pares (t_j, y_{ij}) donde y_{ij} es la j -ésima pérdida de peso observada hasta el tiempo t_j de la unidad experimental i , para $t_j \in \{0, 10, \dots, 190\}$ días.

Por ejemplo, en la Figura 5.1 se muestra la observación 1 correspondiente a la primera réplica del tratamiento 0°C.

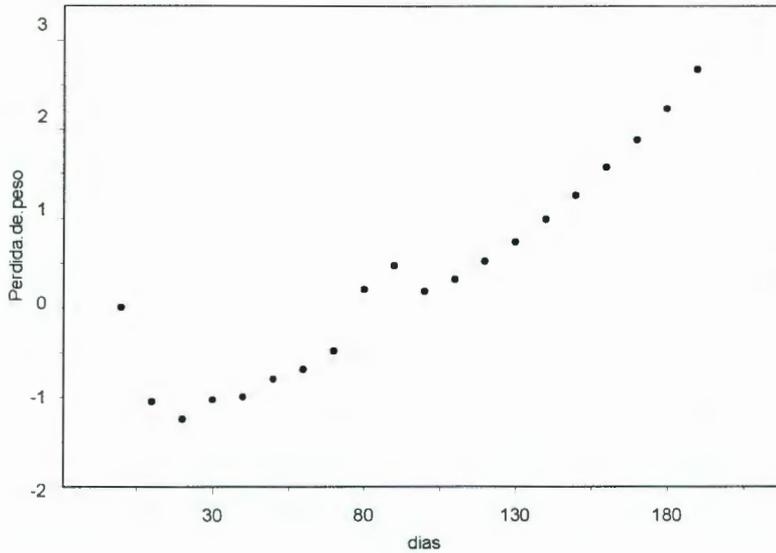


Figura 5.1. Pérdida de peso con el tratamiento 1 de la réplica 1

A cada observación podemos ajustar una función, por lo que el análisis de este experimento puede ser visto como una aplicación de análisis funcional en el que la variable dependiente es funcional y la independiente (temperatura) es multivariada, para llevar a cabo dicho análisis se implementó un programa en S-Plus que se explicará en el capítulo 6.

Las funciones ajustadas a cada una de las observaciones se muestran en la Figura 5.2, para realizar el ajuste utilizamos B-splines de orden 4 con nodos en los días en que se midió la pérdida de peso y 20 funciones base.

Estas muestran que en general la pérdida de peso aumenta conforme transcurre el tiempo. La temperatura de almacenamiento de $5^{\circ}C$ presenta menor pérdida de peso en los primeros 90 días, sin embargo es en el que hay mayor pérdida de peso en los últimos días (150 a 190). Los tratamientos que corresponden a las temperaturas de almacenamiento $20^{\circ}C$, $30^{\circ}C$ y ambiente, muestran un comportamiento de pérdida de peso similar. Los

tratamientos que corresponden a la temperatura de $0^{\circ}C$ y $0^{\circ}C$ 70% *h.r.* son en los que se tiene menor pérdida de peso en los 190 días.

Ajustamos el siguiente modelo funcional:

$$Pp_{kg} = \beta_{amb}(t) + \beta_g(t) + \varepsilon_{kg}(t)$$

donde Pp_{kg} es la función ajustada a la réplica k ($k = 1, 2, 3$) bajo el tratamiento g ($g = 0^{\circ}C, 20^{\circ}C, 30^{\circ}C, 5^{\circ}C, 0^{\circ}C$ 70% *h.r.*), β_{amb} y β_g son efectos de los tratamientos ambiente y g respectivamente.

Con este modelo se compara el tratamiento g contra el tratamiento ambiente, y entonces la matriz de diseño es la siguiente:

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (5.1)$$

Las funciones modeladas se presentan en la Figura 5.3 y las funciones R^2 y el coeficiente F se muestran en la Figura 5.4.

En ella se observa que R^2 es alta (superior a 0.8) en todo el rango y el modelo es significativo en casi todo el rango.

Las funciones de regresión se presentan en la Figura 5.5. Las diferencias de los tratamientos de $0^{\circ}C$ y $0^{\circ}C$ 70% *h.r.* con el de temperatura ambiente es negativa en todo el rango y estas diferencias son mayores conforme aumentan los días, por lo que estos tratamientos presentan menor pérdida de peso comparados con el de temperatura ambiente especialmente hacia el final de los días.

La diferencia del tratamiento $20^{\circ}C$ con el de temperatura ambiente es muy cercana a cero en el intervalo de 0 a 125 días, por lo que podemos decir que en estos días no hay diferencia entre almacenar los ajos a $20^{\circ}C$ y temperatura ambiente; en los días restantes ésta diferencia es negativa pero no tan significativa como la observada en los tratamientos de $0^{\circ}C$ y $0^{\circ}C$ 70% *h.r.*, lo que implica que a partir del día 125 la pérdida de peso en los ajos almacenados a $20^{\circ}C$ es un poco menor que los almacenados a temperatura ambiente.

La diferencia del tratamiento de $30^{\circ}C$ con el tratamiento temperatura ambiente es cercana a cero en los primeros 30 días lo cual indica que al almacenar a $30^{\circ}C$ se tiene prácticamente la misma pérdida de peso que en temperatura ambiente, de 30 a 140 días esta diferencia es positiva, es decir la pérdida de peso obtenida con $30^{\circ}C$ es mayor y en los días de 140 a 190 la diferencia es negativa y se tendrá menor pérdida de peso almacenando a $30^{\circ}C$.

La diferencia del tratamiento $5^{\circ}C$ con temperatura ambiente es negativa hasta el día 150, y presenta un comportamiento similar a la diferencia de los tratamientos $0^{\circ}C$ y $0^{\circ}C$ 70% *h.r.* con temperatura ambiente de 0 a 75 días, pero en los días 150 a 190 la diferencia

es positiva de tal forma que se tiene la mayor pérdida de peso que con cualquier tratamiento comparado con el de temperatura ambiente.

Dado que lo que se quiere es determinar el tratamiento mediante el cual obtengamos la menor pérdida de peso en los 190 días, entonces del análisis anterior podemos decir que los mejores tratamientos son el $0^{\circ}C$ y $0^{\circ}C$ 70% *h.r.*

La velocidad de pérdida de peso es un factor importante para los productores debido a sus consecuencias económicas, es por ello que analizaremos la velocidad de pérdida de peso para cada tratamiento mediante un modelo funcional, tomándola como respuesta y haciendo un análisis similar al anterior.

Las funciones de velocidad de pérdida de peso son las derivadas de las funciones de pérdida de peso, para una mejor interpretación las derivadas fueron suavizadas con parámetro $\lambda = 100$. Éstas se muestran en la Figura 5.6, en ellas se observa que los tratamientos de $0^{\circ}C$, $0^{\circ}C$ 70% *h.r.* y $30^{\circ}C$ presentan una velocidad de pérdida de peso casi constante durante todo el periodo pero la del tratamiento de $30^{\circ}C$ es mayor; los tratamientos $20^{\circ}C$, $5^{\circ}C$ y temperatura ambiente aumentan su velocidad de pérdida de peso a partir de los 75 días y se incrementa conforme transcurren los días, sin embargo el tratamiento $5^{\circ}C$ presenta un aumento de velocidad más drástico a partir del día 100.

Las funciones de R^2 y el cociente F , mostradas en la Figura 5.7 indican que se tiene un buen ajuste a los datos ya que en general la R^2 es superior a 0.8 y el modelo es significativo en todo el intervalo.

Las funciones de regresión obtenidas se presentan en la Figura 5.8, los tratamientos de $0^{\circ}C$, $20^{\circ}C$ y $0^{\circ}C$ 70% *h.r.* presentan una diferencia negativa en todo el periodo, respecto

al tratamiento de temperatura ambiente lo cual indica que la velocidad de pérdida de peso de éste es mayor que la de los tratamientos $0^{\circ}C$, $20^{\circ}C$ y $0^{\circ}C$ 70% *h.r.* La diferencia del tratamiento $30^{\circ}C$ con temperatura ambiente es positiva en el intervalo de 20 a 90 días y negativa en el resto, esto es con el tratamiento de $30^{\circ}C$ se pierde peso más rápido que con el de temperatura ambiente en el intervalo mencionado. La diferencia del tratamiento de $5^{\circ}C$ respecto a temperatura ambiente es negativa en los primeros 90 días y positiva en los días siguientes, la velocidad de pérdida de peso de este tratamiento es menor comparada con temperatura ambiente que la obtenida con los tratamientos restantes, sin embargo en los últimos días esta velocidad se incrementa de tal forma que es mayor a la de los tratamientos restantes. Las funciones modeladas se presentan en la Figura 5.9.

De acuerdo a los resultados se recomienda almacenar el ajo a la temperatura de $0^{\circ}C$, ya que este tratamiento presenta una menor pérdida de peso durante el periodo de almacenamiento observado.

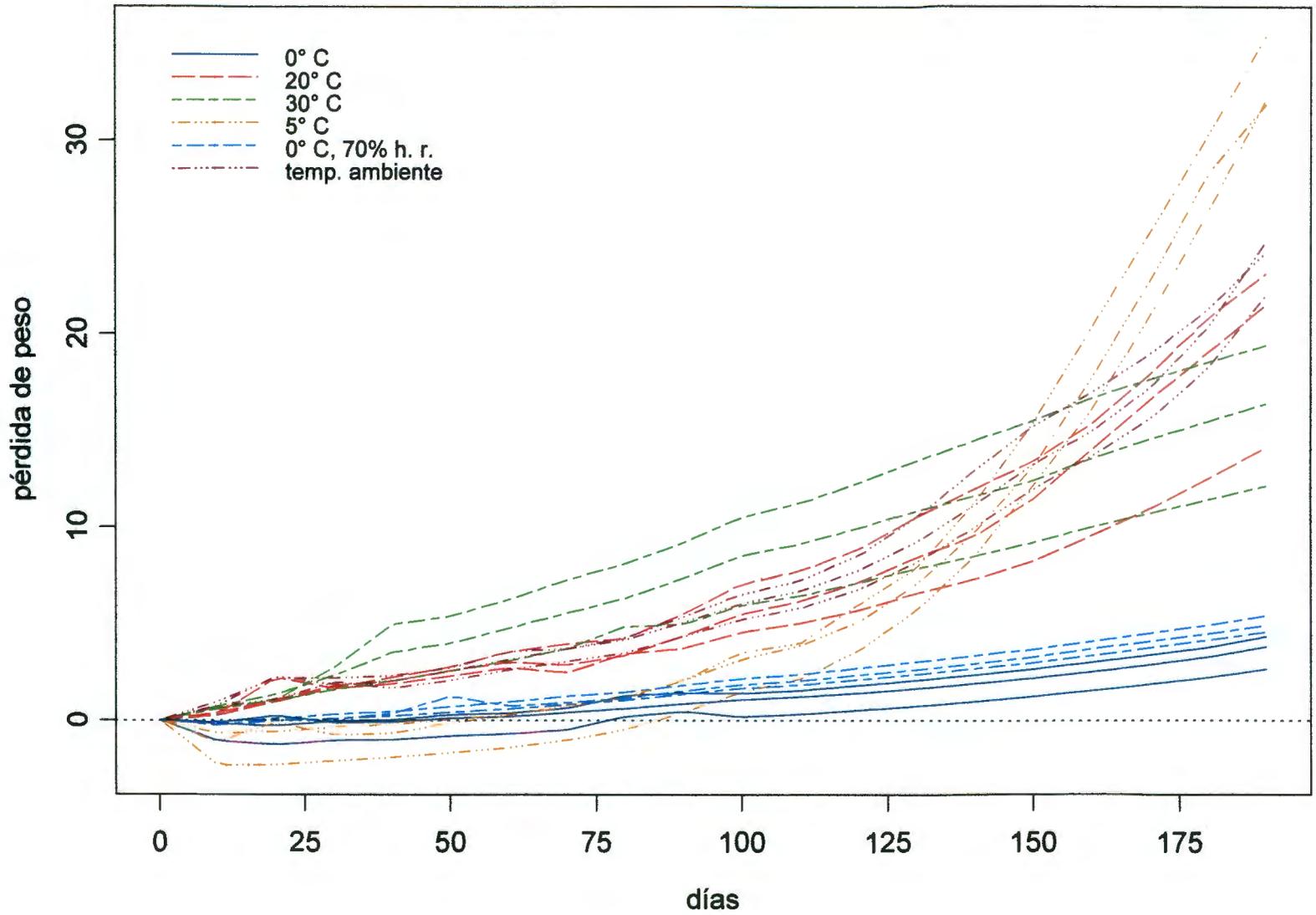


Figura 5.2. Funciones ajustadas de pérdida de peso

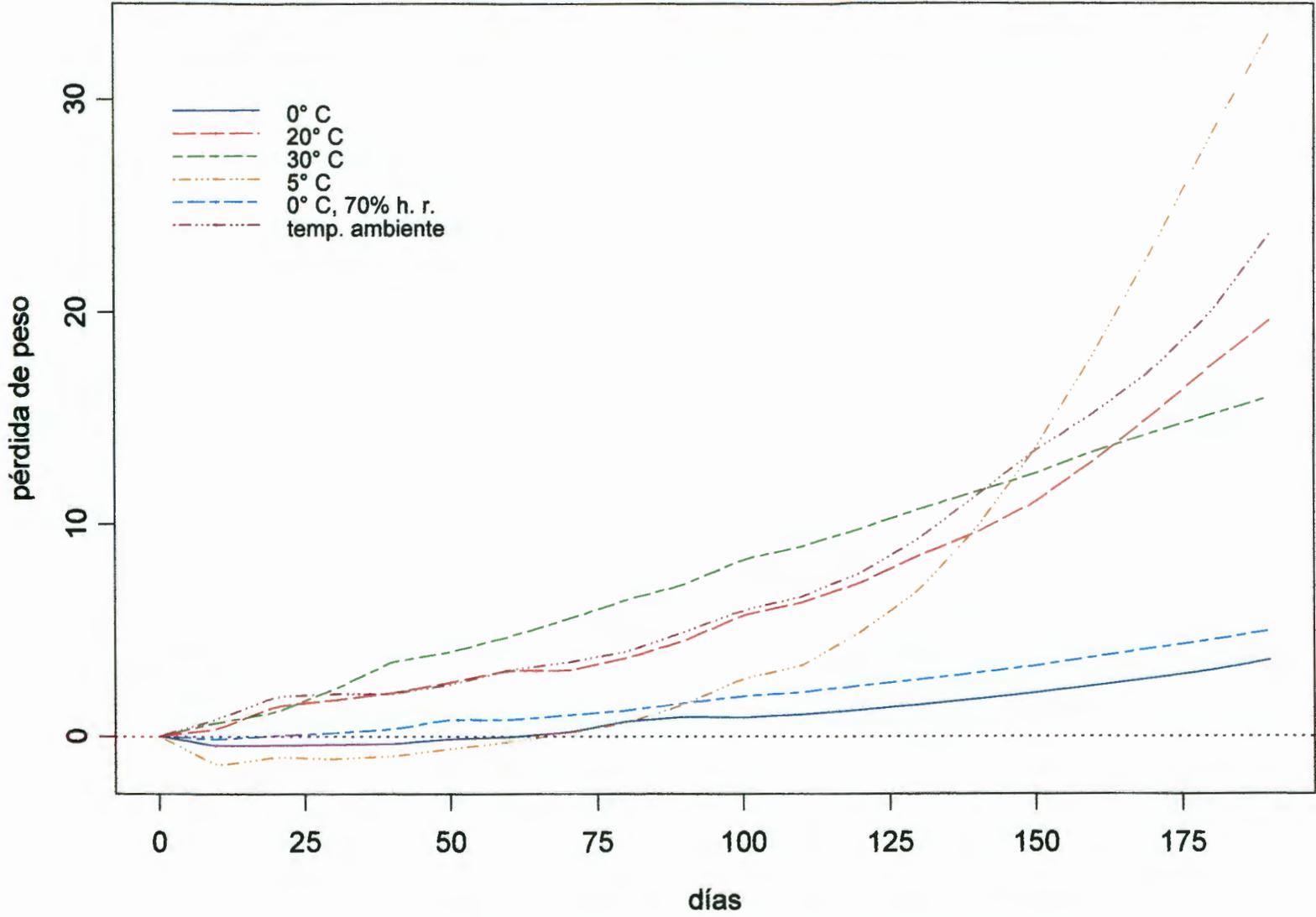


Figura 5.3. Funciones modeladas de pérdida de peso

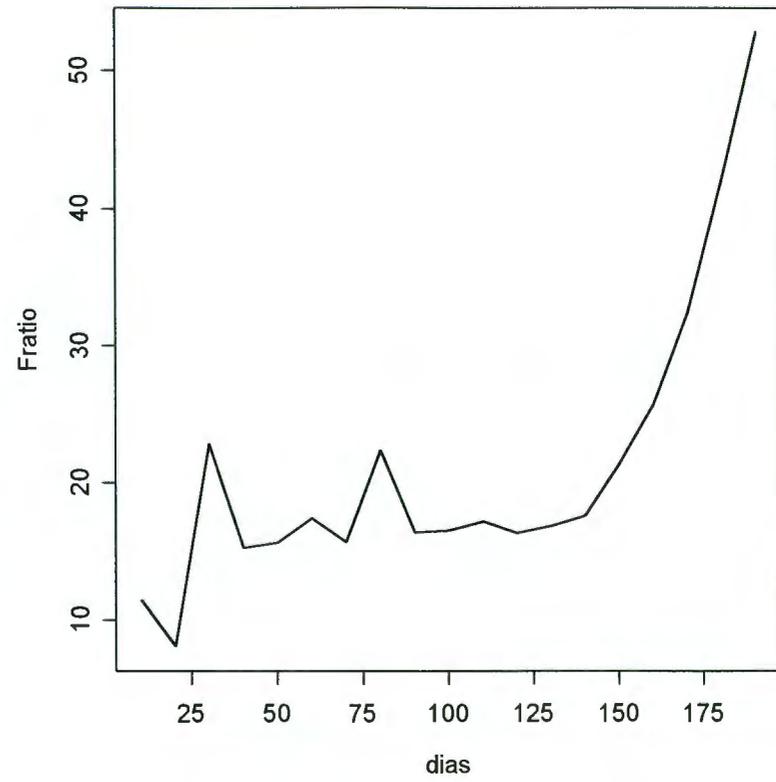
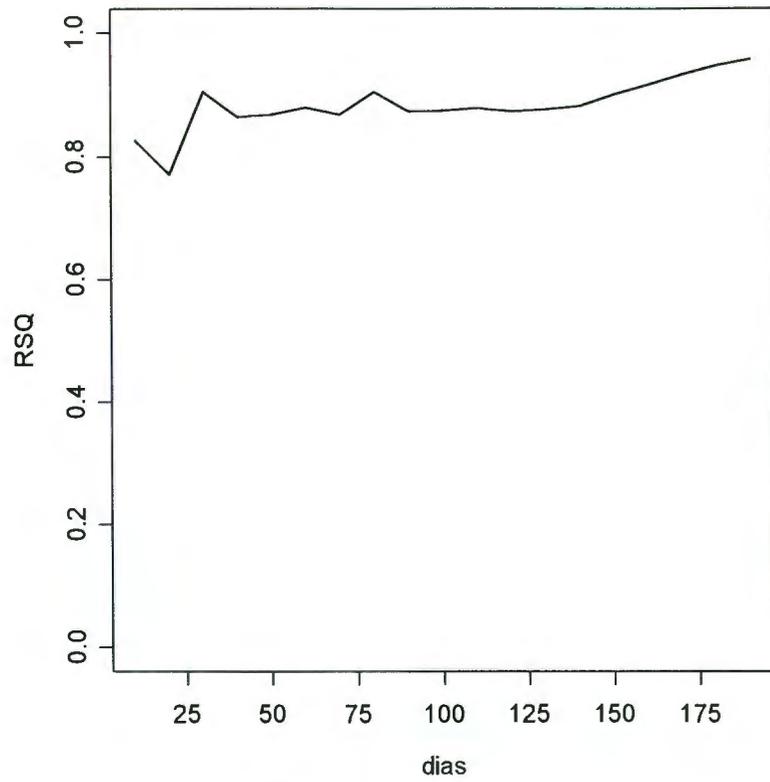


Figura 5.4. Funciones R^2 y cociente F del experimento de pérdida de peso en ajos

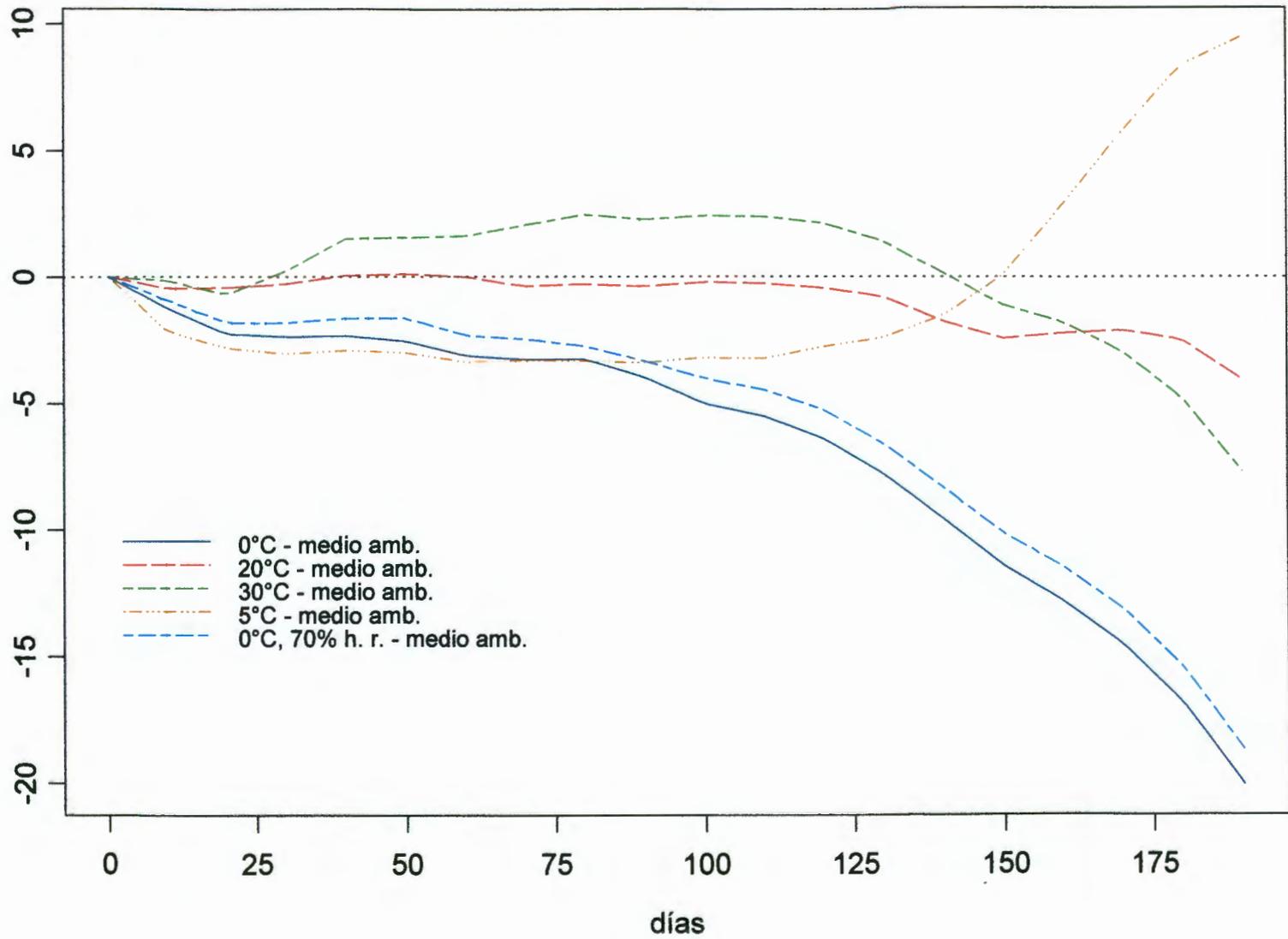


Figura 5.5. Funciones de regresión para el experimento de pérdida de peso en ajos

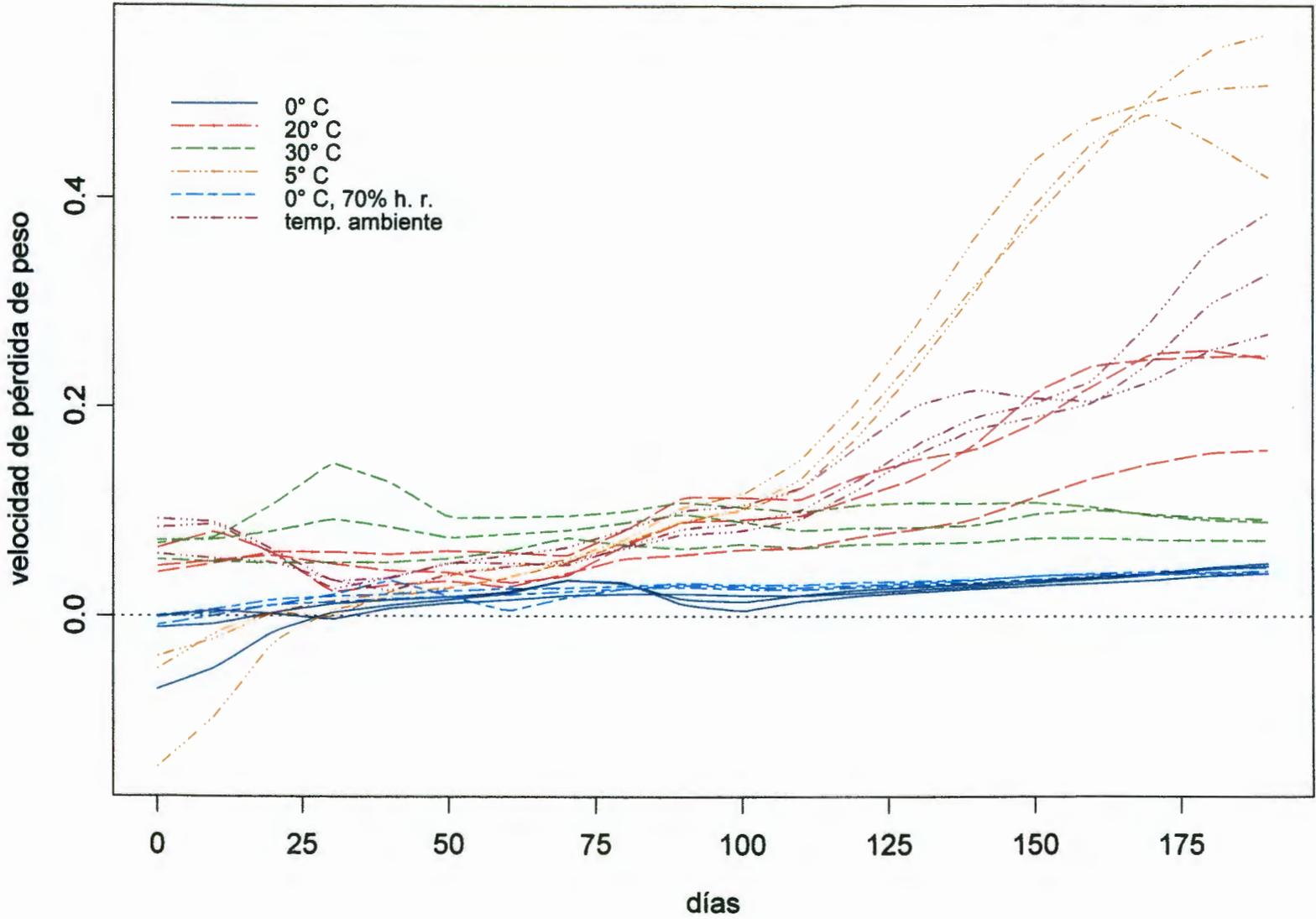


Figura 5.6. Funciones ajustadas de velocidad de pérdida de peso

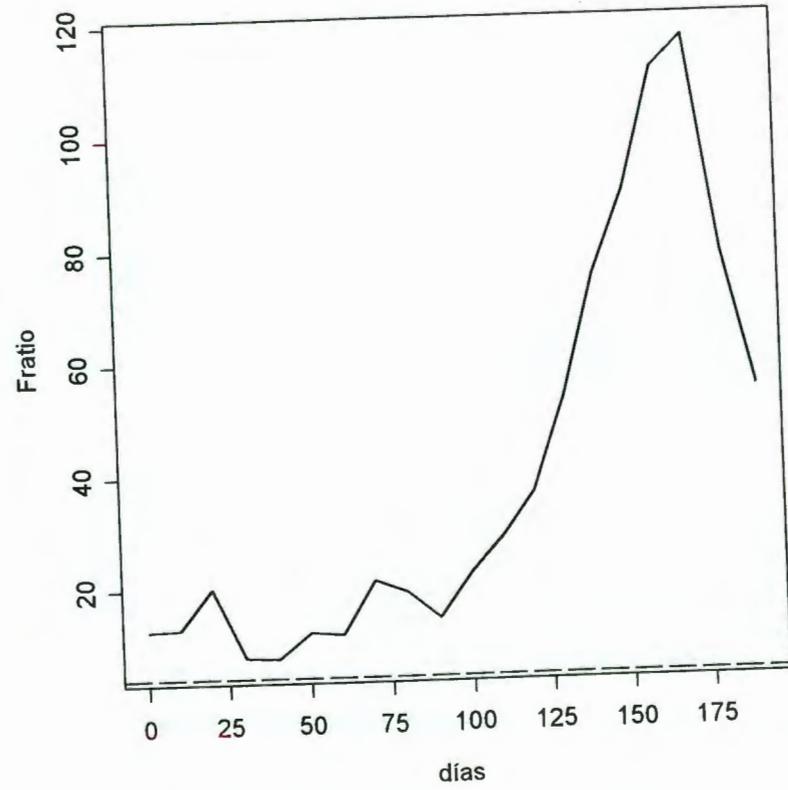
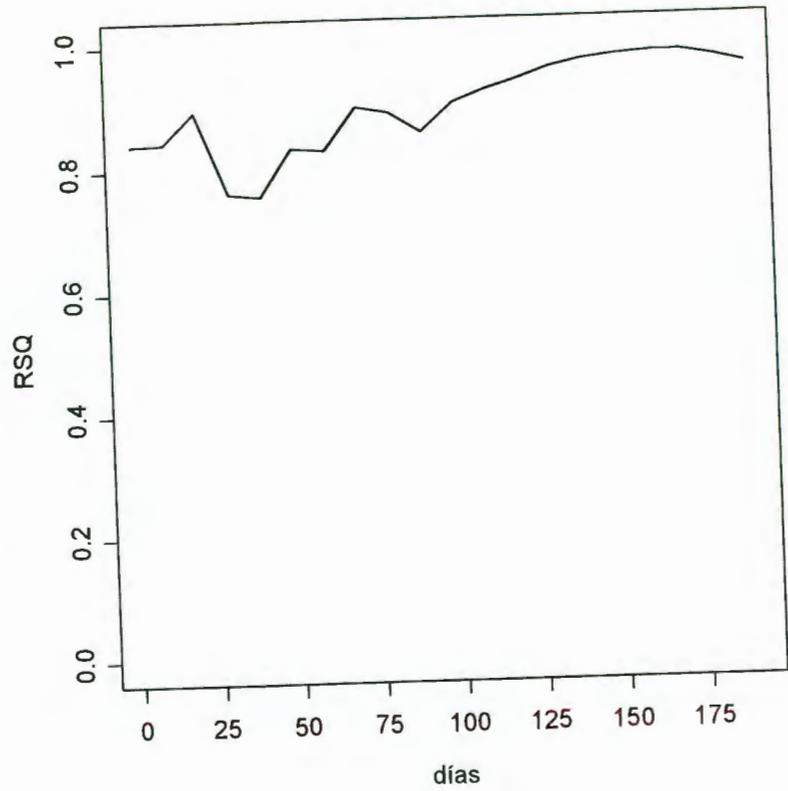


Figura 5.7. Funciones R^2 y cociente F para el experimento de velocidad de pérdida de peso en ajos

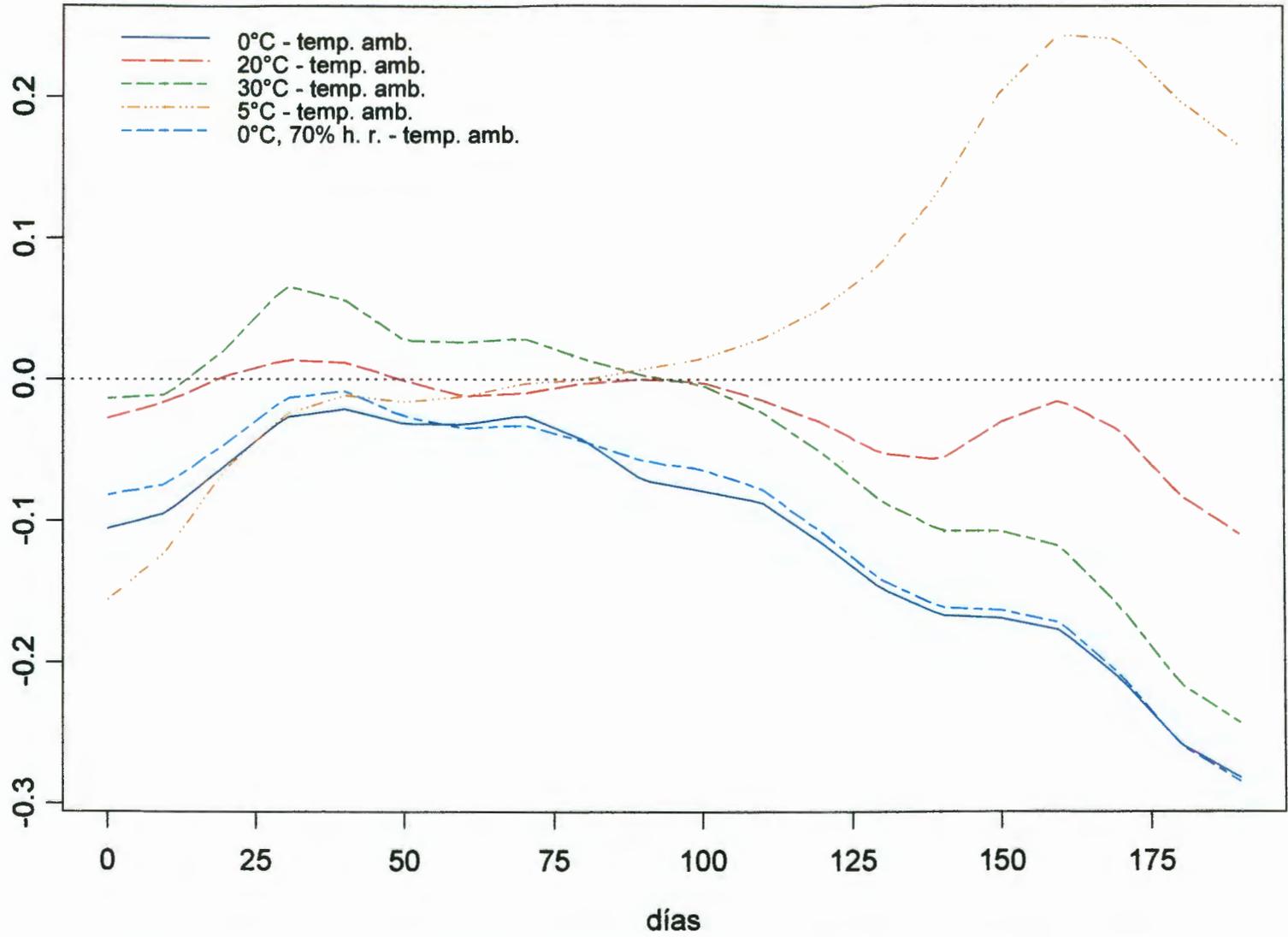


Figura 5.8. Funciones de regresión para el experimento de velocidad de pérdida de peso en ajos

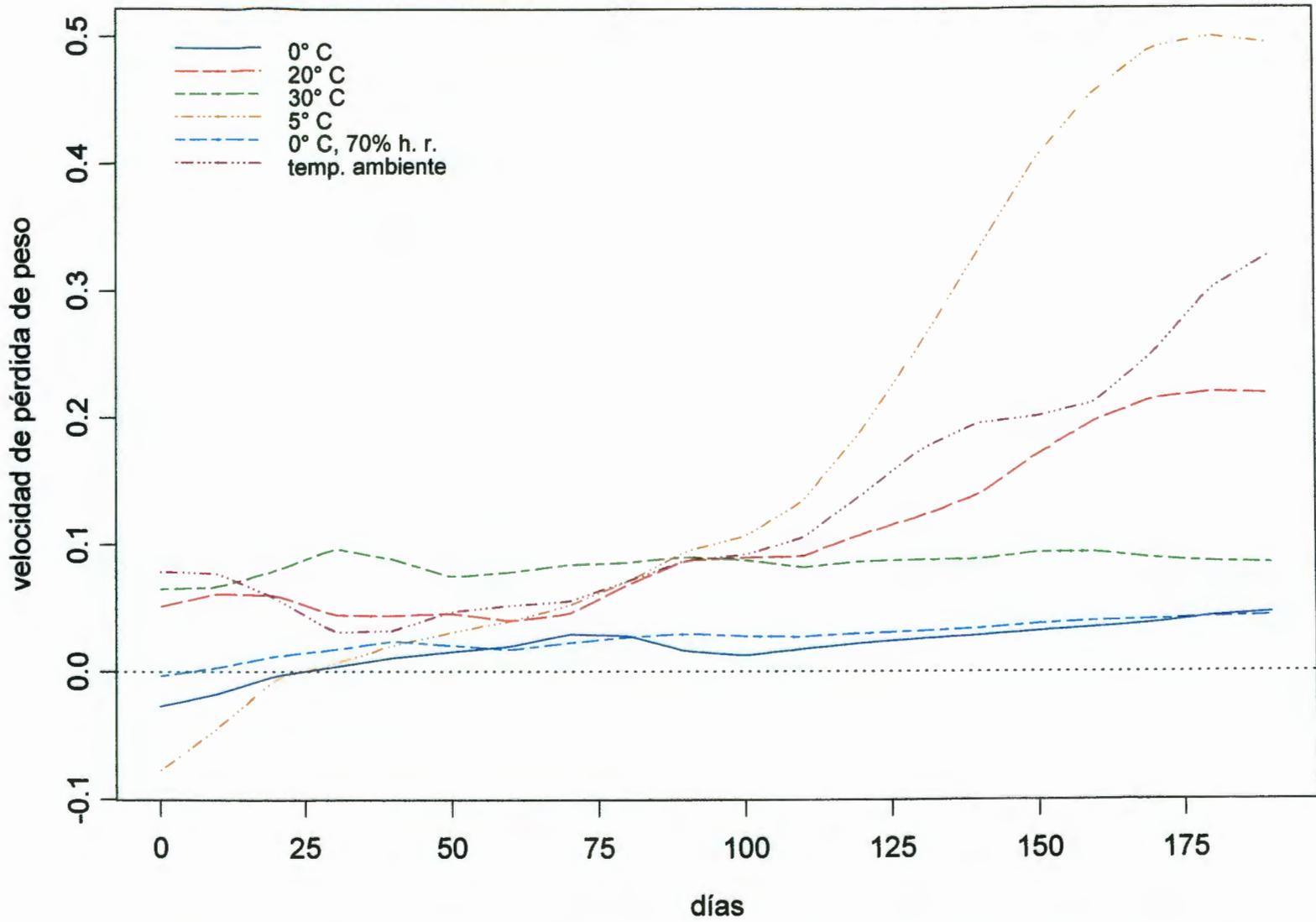


Figura 5.9. Funciones modeladas para el experimento de velocidad de pérdida de peso en ajos

5.3 PRUEBAS DE PERMUTACIÓN

5.3.1 Introducción

Las pruebas de hipótesis estadísticas formulan una pregunta experimental en términos de una hipótesis nula (escrita como \mathcal{H}_0), la cual supone que no hay efecto experimental. La prueba rechaza esta hipótesis nula en favor de una hipótesis alternativa (escrita como \mathcal{H}_A), si es improbable que los datos observados pudieran haber surgido bajo la hipótesis nula, donde los datos son representados por una estadística bajo supuestos convenientes. La probabilidad de rechazar falsamente una hipótesis nula (error Tipo I) es denotada por α . Comúnmente α es elegido como 0.05.

En la práctica es conveniente que calculemos el llamado valor p , la probabilidad de que la estadística como variable aleatoria exceda (o iguale) su valor observado, bajo la hipótesis nula, dados los supuestos.

Las pruebas de hipótesis por permutación son un tipo de pruebas no paramétricas propuestas a principios del siglo XX, sin embargo su popularidad es reciente debido a los cálculos facilitados por la computadora.

El concepto principal de una prueba de permutación es relativamente intuitivo. Por ejemplo, considere un sujeto que es sometido repetidamente a dos condiciones "descanso" y "activación", si realmente no hay diferencia entre las dos condiciones, entonces sería muy sorprendente que la mayoría de las observaciones correspondientes a "activación" fueran más grandes que las de "descanso" y estaríamos inclinados a concluir que hay evidencia de diferencia entre las dos condiciones. La prueba de permutación simplemente provee un

mecanismo formal para cuantificar esta "sorpresa" en términos de probabilidad por medio de pruebas de significancia y valores p .

Si no hay efecto experimental entonces el etiquetado de observaciones correspondiente a la condición experimental es arbitrario, ya que los mismos datos surgirían dada cualquier condición. Dada la hipótesis nula de que el etiquetado es arbitrario, el efecto experimental puede ser establecido al compararlo con la distribución de valores obtenida cuando las etiquetas son permutadas.

La justificación para intercambiar las etiquetas surge de suposiciones de distribución débiles o por el esquema de aleatorización usado en el diseño del experimento. Las pruebas justificadas por la aleatorización inicial de condiciones de unidades experimentales, son llamadas frecuentemente pruebas de aleatorización.

Los conceptos fundamentales en una prueba de aleatorización son, aleatorización experimental, hipótesis nula, intercambiabilidad, y distribución empírica aleatorizada; definidos en [10].

- Aleatorización

La idea básica es asignar aleatoriamente unidades experimentales a tratamientos. Este etiquetado es elegido aleatoriamente de acuerdo a un esquema de aleatorización y cualquier otro posible etiquetado de este esquema tuvo la misma probabilidad de ser elegido.

- Hipótesis nula

En las pruebas de aleatorización, la hipótesis nula se propone de acuerdo a los datos adquiridos. Por ejemplo,

\mathcal{H}_0 : "Cada observación habría sido la misma bajo cualquier tratamiento".

La hipótesis es que las condiciones experimentales no afectan los datos significativamente, si se hubiera realizado el experimento con una asignación diferente se habrían obtenido exactamente los mismos datos. En este sentido pensamos los datos como fijos y el diseño experimental como aleatorio.

Bajo esta hipótesis nula, el etiquetado de las observaciones es arbitrario. Entonces podemos re-aleatorizar las etiquetas de los datos, permutándolos, sujetos a la restricción de que cada permutación pudo surgir de un esquema de aleatorización inicial. Los datos observados tienen la misma probabilidad de resultar de cualquiera de estas permutaciones.

- Intercambiabilidad

Considera la situación antes de que los datos sean recolectados pero después de que las etiquetas de los tratamientos han sido asignados a las unidades experimentales. Formalmente, un conjunto de etiquetas de los datos son intercambiables si la distribución de la estadística es la misma para cualquier etiquetado bajo una hipótesis como se define en [10].

Con una prueba de aleatorización, el esquema de aleatorización provee el posible etiquetado y la hipótesis nula asegura que las etiquetas son intercambiables dentro de las restricciones de este esquema.

- Distribución aleatorizada

Considere alguna estadística que exprese el efecto experimental de interés. El cálculo de esta estadística depende del etiquetado de los datos, es decir permutando las etiquetas obtenemos otro valor de la estadística.

Dada la intercambiabilidad bajo la hipótesis nula, los datos observados tienen la misma probabilidad de haber surgido de cualquier posible etiquetado. Entonces las estadísticas asociadas con cada uno de los etiquetados posibles tienen también la misma probabilidad. Por lo que tenemos la distribución aleatorizada de nuestra estadística, que es la distribución muestral de la estadística bajo la hipótesis nula, dados los datos observados. Bajo la hipótesis nula la estadística observada es elegida aleatoriamente de un conjunto de estadísticas correspondientes a todos los posibles re-etiquetados.

El valor p es en este caso la proporción de valores de la estadística en la distribución aleatorizada mayores o igual que el observado. El etiquetado actual usado en el experimento es uno de todos los posibles etiquetados, entonces si la estadística observada es la mayor de la distribución aleatorizada, el valor p es $\frac{1}{B}$, donde B es el número de etiquetados posibles del esquema inicial de aleatorización.

Sea T que denote el valor de la estadística para el etiquetado original del experimento. Rechazamos la hipótesis nula al nivel de significancia α si el valor p es menor que α . De manera equivalente, T debe ser mayor o igual que el percentil $100(1 - \alpha)$ de la distribución aleatorizada. Entonces el valor crítico es $c + 1$, donde $c = \lfloor \alpha B \rfloor$, (αB redondeado hacia abajo) si T excede este valor crítico entonces la prueba es significativa al nivel α .

5.3.2 Pruebas de permutación

Para hacer una prueba de permutación se requiere seguir los siguientes pasos de [5]:

1. Analizar el problema.
 - a) ¿Cuál es la hipótesis nula y cuáles son las alternativas?
 - b) ¿Cómo están distribuidos los datos?
 - c) ¿Qué pérdidas están asociadas con malas decisiones?
2. Elegir una estadística de prueba y establecer una regla de rechazo que distinga la hipótesis nula de la alternativa.
3. Computar la estadística de prueba para el etiquetado original de las observaciones.
4. Rearreglar (permutar) las etiquetas y recomputar la estadística de prueba para las etiquetas rearrregladas. Repetir hasta que obtenga la distribución de la estadística de prueba para todas las posibles permutaciones
5. Aceptar o rechazar la hipótesis usando esta distribución de permutaciones como guía.

5.3.3 Aplicación

La aplicación de pruebas de permutación dentro de análisis funcional se llevará a cabo para el experimento de pérdida de peso en el ajo explicado en la sección 5.1.

Una vez que $\beta(t)$ es estimada establecemos las hipótesis nulas para cada efecto:

$$H_{0g} : \beta_g(t) = 0, \quad g = 0^\circ C, 20^\circ C, 30^\circ C, 5^\circ C, 0^\circ C \text{ 70\% h.r.} \quad (5.2)$$

donde $\beta_g(t)$ representa la comparación entre el tratamiento g y el de temperatura ambiente. Bajo esta hipótesis si se hubiera realizado el experimento con una asignación diferente entre los tratamientos g y ambiente se habrían obtenido observaciones similares, entonces el etiquetado de las observaciones entre estos dos tratamientos es arbitrario. Por lo que se llevará a cabo una permutación restringida, intercambiando las etiquetas de las observaciones de los tratamientos g y ambiente y manteniendo fijas las correspondientes a los tratamientos restantes, por lo que se tienen 20 permutaciones distintas.

En el contexto de modelos lineales funcionales [1] propone las estadísticas siguientes.

$$S_g = \sup_t \left| \left[\widehat{\beta}_g(t) \right] / \left[a_g \sqrt{MSE(t)} \right] \right|, \quad (5.3)$$

y

$$I_g = \int \left| \widehat{\beta}_g(t) \right| / a_g \sqrt{MSE(t)} dt. \quad (5.4)$$

evadiendo así el problema de multiplicidad de pruebas [10]; donde se define

$$a_g^2 = \mathbf{u}'_g (Z'Z)^{-1} \mathbf{u}_g, \quad (5.5)$$

\mathbf{u} un vector con 1 en la posición correspondiente al tratamiento g y 0 en las restantes y Z es la matriz de diseño (5.1).

En la Figura 5.5 se muestran los contrastes considerados en (5.2)

Para calcular los valores p correspondientes a cada contraste, se realizó un programa en S-Plus que se presentará en el capítulo 6, dicho programa realiza los siguientes pasos:

1. Calcula los valores S_g, I_g para el etiquetado original de las observaciones.
2. Obtiene una permutación del etiquetado original.

3. Realiza el análisis funcional del nuevo ordenamiento para estimar el vector $\beta(\mathbf{t})$.
4. Calcula las estadísticas S_g, I_g del contraste de interés.
5. Compara los nuevos valores de S_g, I_g con los valores originales, en caso de ser mayores los cuantifica.
6. Repite este procedimiento a partir del paso 2, para las $B = 20$ permutaciones distintas.

En la Tabla 5.2 se muestran los valores p estimados para cada uno de los contrastes de interés de las estadísticas S_g, I_g .

Contrastes	S_g	I_g
$0^\circ\text{C-temp. ambiente}$	0	0
$20^\circ\text{C-temp. ambiente}$	0.20	0.50
$30^\circ\text{C-temp. ambiente}$	0.05	0.10
$5^\circ\text{C-temp. ambiente}$	0	0
$0^\circ\text{C } 70\% \text{ humedad relativa (h.r.) - temp. ambiente}$	0.05	0

Tabla 5.2 Valores p estimados

Estos resultados indican una diferencia significativa entre la temperatura ambiente y las temperaturas de almacenamiento de 0°C , $0^\circ\text{C } 70\% \text{ h.r.}$ y 5°C , mientras que las condiciones de 20°C y temperatura ambiente son estadísticamente iguales. 30°C es un tratamiento que no es claramente igual a la temperatura ambiente. Viendo la Figura 5.5, la discrepancia esencial con la temperatura ambiente se da después de los 150 días de almacenamiento. Note que los tratamientos 0°C , $0^\circ\text{C } 70\% \text{ h.r.}$ y 5°C , fueron los que mostraron mayor diferencia respecto a la condición de temperatura ambiente en el análisis anterior, en particular los tratamientos de 0°C y $0^\circ\text{C } 70\% \text{ h.r.}$ resultaron ser los mejores.

5.4 APLICACIÓN DOS AL MODELO LINEAL FUNCIONAL CASO 1

La empresa GERBER® produce un jugo de yogurt con frutas, para alimento infantil, con los siguientes ingredientes:

Pectina, Yogur, Purés, Azúcar y Agua.

Una vez terminado el producto se identificó el problema de separación de fases del jugo de frutas con yogurt durante almacenamiento y cadena de transporte, éste se presentó en todos los sabores del producto, manzana, pera y durazno, mango y pera y con mayor severidad en el de frutas mixtas (manzana, piña, naranja, chabacano y platano).

En la Figura 5.10, se observa la separación en fases de un jugo de frutas con yogurt.



Figura 5.10. Jugo con yogurt dividido en fases

Para el estudio de este problema se formó un equipo de trabajo. El objetivo del proyecto de investigación es reducir la separación de fases de un jugo de frutas con yogurt, para lo cual se requiere identificar los factores de formulación / proceso que influyen en la

separación de fases del producto y establecer las condiciones de formulación y de proceso que reduzcan ésta.

Los factores que se identificaron son:

Pectina (Componente del producto)

Temperatura de mezclado (Factor del proceso)

Tiempo de mezclado (Factor del proceso)

Presión de homogeneizado (Factor del proceso)

Temperatura de almacenamiento (Factor de almacenamiento)

Tiempo de almacenamiento (Factor de almacenamiento)

Para estudiar la influencia de estos factores en la reducción de capas, se realizaron dos diseños estadísticos.

Como primer diseño se propuso un OME($3^1 2^4$) [4], que se muestra en la Tabla 5.3:

Tratamiento	Pectina (P) (%)	Tiempo (t) (min)	Temperatura (T) (°C)	Sólidos Tot. (St) (%)	Presión (pr) (psi)	
1	(0,-1, 1, 1,-1)	0.3	5	75	20	3000
2	(0, 1,-1,-1, 1)	0.3	7	65	18	3500
3	(1,-1,-1, 1, 1)	0.5	5	65	20	3500
4	(1,-1, 1, -1, 1)	0.5	5	75	18	3500
5	(1, 1,-1, 1,-1)	0.5	7	65	20	3000
6	(1, 1, 1,-1,-1)	0.5	7	75	18	3000
7	(2, 1, 1, 1, 1)	0.7	7	75	20	3500
8	(2,-1,-1,-1,-1)	0.7	5	65	18	3000

Tabla 5.3. Diseño OME($3^1 2^4$) para el experimento de yogurt

P = % de Pectina

T = Temperatura de mezclado

t = Tiempo de mezclado

St = Sólidos totales

pr = Presión de homogeneización

El segundo es un diseño factorial fraccionado 2_{IV}^{4-1}

Tratamiento	Pectina (P) (%)	Tiempo (t) (min.)	Temperatura (T) (° C)	Presión (pr) (psi)
1 (-1,-1,-1,-1)	0.5	5	70	2000
2 (-1,-1, 1, 1)	0.5	5	80	3500
3 (-1, 1,-1, 1)	0.5	7	70	3500
4 (-1, 1, 1,-1)	0.5	7	80	2000
5 (1, -1,-1, 1)	0.7	5	70	3500
6 (1, -1, 1,-1)	0.7	5	80	2000
7 (1, 1,-1,-1)	0.7	7	70	2000
8 (1, 1, 1, 1)	0.7	7	80	3500

Tabla 5.4. Diseo 2_{IV}^{4-1} para el experimento de yogurt

con relación de definición $I = (Pectina)(Tiempo)(Temperatura)(Presión)$ y los alias de interacciones dobles son:

$$(Pectina)(Tiempo) = (Temperatura)(Presión)$$

$$(Tiempo)(Temperatura) = (Pectina)(Presión)$$

$$(Pectina)(Temperatura) = (Tiempo)(Presión)$$

En ambos experimentos las variables respuesta son: análisis de tamaño de partícula, viscosidad de las muestras y medición visual del número de capas.

Para una aplicación del modelo en el capítulo 4 se considerará el segundo diseño experimental, con dos réplicas por tratamiento, tomando como respuesta el análisis de tamaño de partícula, de manera que tenemos 16 observaciones.

El análisis de tamaño de partícula consiste en determinar el porcentaje de concentración de volumen para 71 tamaños de partícula en una botella de 8 onzas de jugo de

yogurt con frutas, este se realizó con la máquina Dynamic Light scattering, presentada en la Figura 5.11.

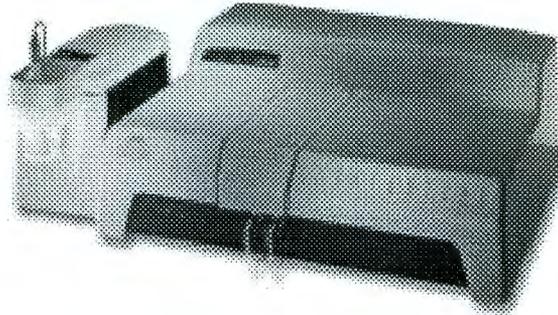


Figura 5.11 Máquina utilizada para medir los tamaños de partícula

Mediante revisiones bibliográficas, el equipo de trabajo llegó a la conclusión que el análisis de tamaño de partícula está relacionado con la separación en fases del yogurt de la siguiente manera, mientras mayor sea la concentración de volumen en partículas pequeñas menor será la separación en fases. Por lo tanto, el objetivo es lograr mayor concentración de volumen en tamaños pequeños de partículas.

Los 71 tamaños de partículas (t) que pertenecen al intervalo $[0.01\mu m, 208.93\mu m]$ se dividieron en dos intervalos, de 0.010 a 4.365 y de 5.012 a 158.489, para un mejor análisis.

Cada respuesta consiste en 71 pares (t_j, y_{ij}) , donde y_{ij} es el volumen que corresponde al tamaño t_j , $j = 1, \dots, 71$, $i = 1, \dots, 16$, de manera que podemos pensar que existe una función subyacente a cada uno de estos perfiles.

Lo que queremos es determinar como afecta cada factor al perfil de distribución de tamaños de partículas; como a éste se le puede ajustar una curva, es natural utilizar análisis funcional en el caso en el que la respuesta es funcional y las variables independientes son

multivariadas. Implementamos un programa en S-Plus para realizar este análisis, programa que se explicará en el capítulo 6.

Lo primero que se busca es expresar cada respuesta de manera funcional, de tal forma que el volumen pueda verse de manera continua en función del tamaño de partículas.

Para cada intervalo se ajustaron los datos utilizando splines de orden 4 con nodos en los tamaños observados de partículas, obteniendo 16 funciones que se muestran en las Figuras 5.12 y 5.13.

Se observa en las figuras de funciones ajustadas de volumen que la mayor concentración de éste corresponde a tamaños de partículas pequeñas. En el primer intervalo el porcentaje de concentración de volumen es mayor a tres para tamaños de partículas de 0 a 1 a diferencia del segundo intervalo en el que se tienen porcentajes de concentración de volumen menores a 2.

También puede apreciarse un efecto de compensación, es decir los tratamientos que tienen mayor concentración en tamaños de partículas pequeñas tendrán menor concentración de volumen en tamaños mayores, como era de esperarse.

Podemos identificar al tratamiento ocho como el mejor, ya que presenta la mayor concentración de volumen en partículas muy pequeñas y la menor concentración en partículas grandes comparado con los otros tratamientos.

Para cada uno de los intervalos se realizó un análisis funcional de varianza mediante el siguiente modelo:

$$\begin{aligned} Vol_k = & \mu(t) + \alpha_1(t)x_P + \alpha_2(t)x_t + \alpha_3(t)x_T + \alpha_4(t)x_{pr} \\ & + \alpha_5(t)x_Tx_{pr} + \alpha_6(t)x_Px_{pr} + \alpha_7(t)x_tx_{pr} + \varepsilon_k \end{aligned}$$

donde la función μ es la media general, $\alpha_i(t)$, el efecto del factor i , y cada función Vol_k corresponde a cada una de las 16 respuestas representadas funcionalmente.

En este caso la matriz de diseño permanece como en el caso multivariado y se muestra a continuación:

	Tratamiento								
$Z =$	1	-1	-1	-1	-1	1	1	1	1
	1	-1	-1	-1	-1	1	-1	-1	1
	1	-1	-1	1	1	-1	-1	1	2
	1	-1	-1	1	1	-1	1	-1	2
	1	-1	1	-1	1	-1	1	-1	3
	1	-1	1	-1	1	-1	-1	1	3
	1	-1	1	1	-1	1	-1	-1	4
	1	-1	1	1	-1	1	1	1	4
	1	1	-1	-1	1	1	1	1	5
	1	1	-1	-1	1	1	-1	-1	5
	1	1	-1	1	-1	-1	-1	1	6
	1	1	-1	1	-1	-1	1	-1	6
	1	1	1	-1	-1	-1	1	-1	7
	1	1	1	-1	-1	-1	-1	1	7
	1	1	1	1	1	1	-1	-1	8
	1	1	1	1	1	1	1	1	8

Las funciones modeladas \widehat{Vol}_k son obtenidas con la siguiente expresión:

$$\begin{aligned} \widehat{Vol}_k = & \widehat{\mu}(t) + \widehat{\alpha}_1(t)x_P + \widehat{\alpha}_2(t)x_t + \widehat{\alpha}_3(t)x_T + \widehat{\alpha}_4(t)x_{pr} \\ & + \widehat{\alpha}_5(t)x_T x_{pr} + \widehat{\alpha}_6(t)x_P x_{pr} + \widehat{\alpha}_7(t)x_t x_{pr} + \varepsilon_k \end{aligned}$$

se presentan en las Figuras 5.14 y 5.15 para ambos intervalos.

Las funciones R^2 y el cociente F se muestran en las Figuras 5.16 y 5.17, tanto para el primero como para el segundo intervalo.

En ellas se observa que para el primer intervalo la R^2 es alta y el modelo es significativo en casi todo el rango, en el segundo intervalo el valor de la R^2 decrece conforme aumenta el tamaño de partícula y el modelo es significativo hasta tamaños de partícula de

75; para tamaños mayores el modelo no es significativo y no se tiene un buen ajuste por lo que no tiene mucho sentido interpretar el modelo en estos tamaños.

Las gráficas de la media $\hat{\mu}(t)$, presentadas en las Figuras 5.18 y 5.19, muestran el comportamiento general de las observaciones, en el primer intervalo se observa mayor concentración en tamaños de partículas pequeñas mientras que en el segundo en general es menor mostrando valores máximos en tamaños de partículas de 40 a 80.

Los efectos estimados para ambos intervalos se presentan en las Figuras 5.20 y 5.21.

Como ya se mencionó, es conveniente para aumentar la calidad del producto concentrar el volumen en tamaños de partículas muy pequeños, a fin de lograr mayor estabilidad por lo que se tendrá un interés principal en estos tamaños de partículas (de 0 a 0.3) al dar las recomendaciones para fijar los niveles de los factores involucrados en el experimento, para una mejor interpretación se presentan los efectos estimados sólo para estos tamaños de partículas en la Figura 5.22.

En estos tamaños de partículas (de 0 a 0.3) se observa que la presión es significativa en el intervalo 0.1 a 0.2 y la temperatura de 0.03 a 0.11 ambas con un efecto positivo. Obsérvese que los efectos restantes, excepto la interacción temperatura - presión, también tienen un efecto positivo aunque no son significativos; como la concentración en tamaños pequeños de partículas es lo deseable, se recomienda fijar todos los factores en nivel alto, siendo los más importantes la presión y la temperatura de esta manera los efectos positivos contribuyen al incremento de la concentración de volumen en dicho intervalo.

En tamaños de partículas más grandes es deseable disminuir la concentración de volumen. Para tal efecto se observarán los efectos que resultaron significativos en tamaños

de partículas mayores a 0.3. La pectina ejerce cierto nivel de significancia con efecto negativo en el intervalo 0.36 a 0.5, y tiene un efecto positivo más importante en tamaños de 75 a 110. El tiempo de mezclado sólo es significativo en tamaños de 0.8 a 1.5, con un efecto negativo. La temperatura de mezclado es significativa en el intervalo 0.5 a 1.4 y en tamaños de 90 a 120 con efecto negativo. La presión es significativa en tamaños de partículas de 25 a 40 con un efecto positivo mientras que ejerce un efecto negativo en tamaños de partículas de 2 a 4.5 y en el intervalo de 50 a 125 donde es notablemente significativa.

La interacción temperatura de mezclado - presión, es significativa en tamaños de partículas de 25 a 50, con efecto negativo y de 100 a 125 con efecto positivo. La interacción de pectina y presión es significativa en el intervalo de 85 a 110 con un efecto negativo. La interacción de tiempo de mezclado con presión es significativa para tamaños de 2.1 a 3.7. Obsérvese que en el intervalo de 0.3 a 4 los efectos que muestran significancia (pectina, temperatura, tiempo, presión y la interacción tiempo-presión) lo hacen de manera negativa, por lo que si los fijamos en nivel alto disminuye la concentración de volumen en estos tamaños que es lo que se busca.

En el intervalo de 20 a 40 la presión y la interacción de temperatura y presión son significativos, con efecto positivo y negativo respectivamente. Para conservar el efecto negativo de la interacción es necesario fijar la temperatura y presión ambos en nivel alto o en nivel bajo, tomando en cuenta que dichos efectos son significativos en tamaños de partículas muy pequeños, donde es prioridad concentrar volumen, se mantendrán en nivel alto.

En el intervalo de 50 a 75 fijando la presión en nivel alto disminuye la concentración de volumen en estos tamaños.

Para tamaños de partículas mayores a 75 se vió que el modelo no es significativo.

La recomendación final es fijar todos los factores en nivel alto, ya que esto aumenta la concentración de volumen en tamaños de partículas muy pequeñas y la disminuye en tamaños grandes de partículas.

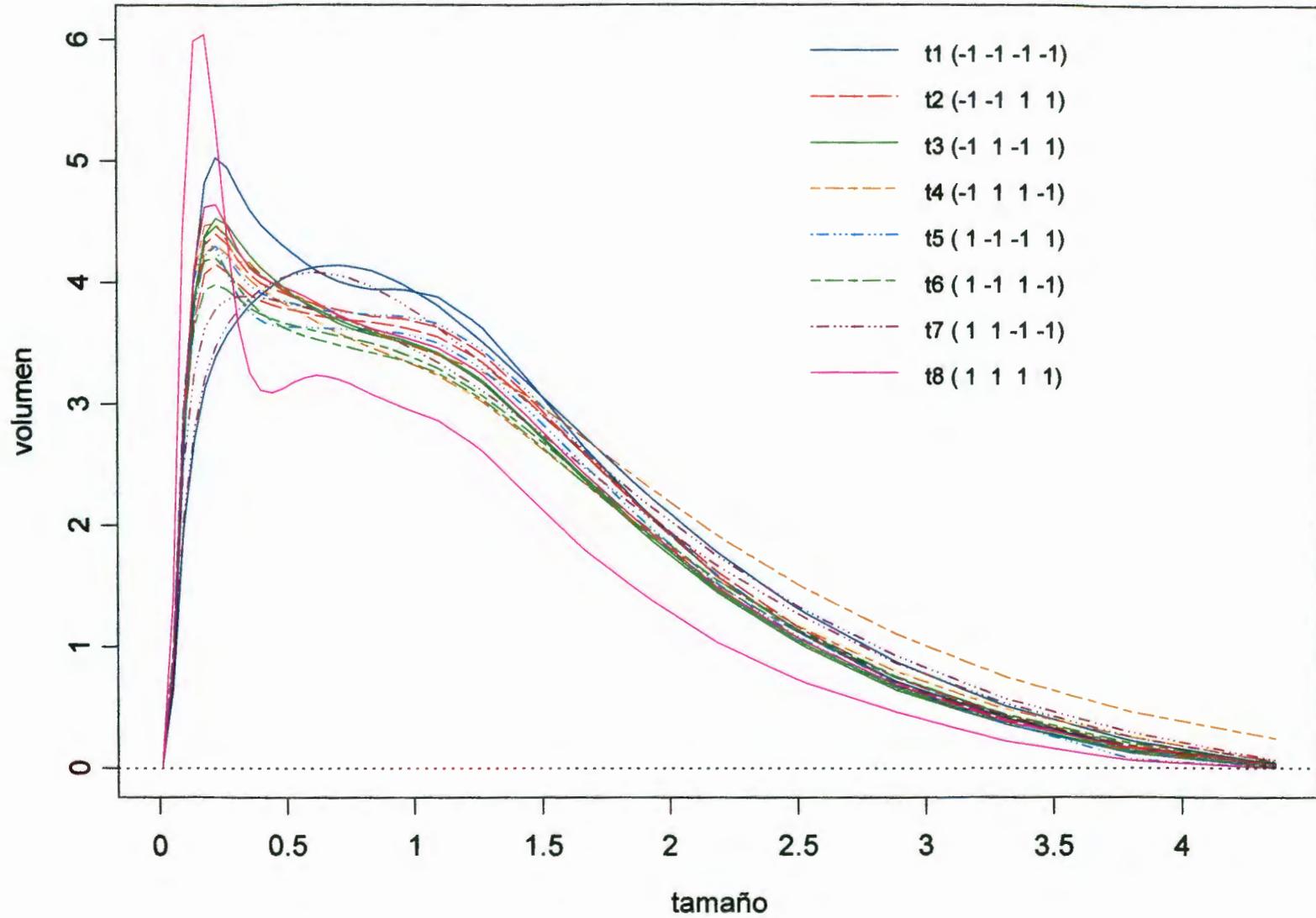


Figura 5.12 Funciones ajustadas de volumen en el experimento de yogurt

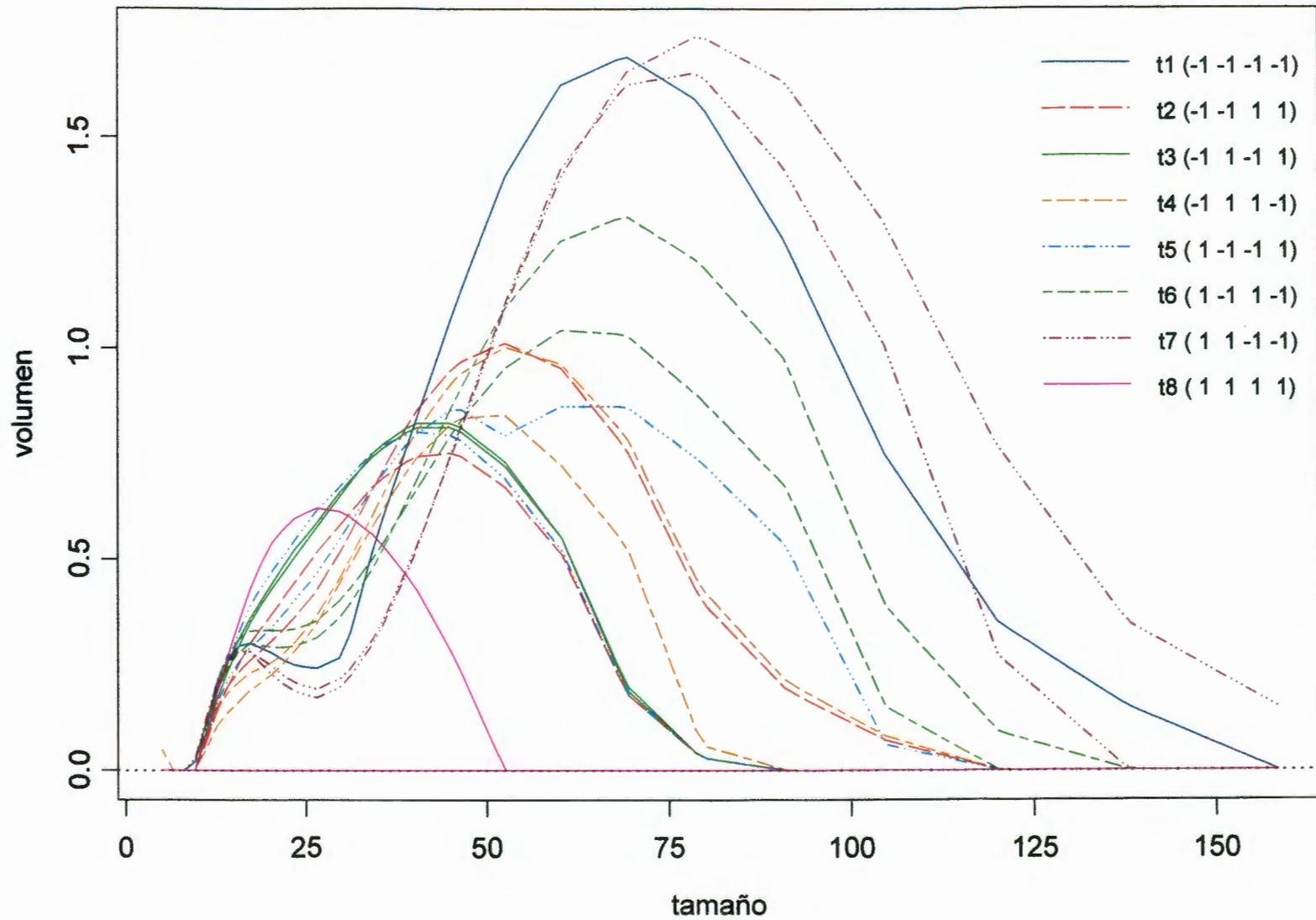


Figura 5.13 Funciones ajustadas de volumen en el experimento de yogurt

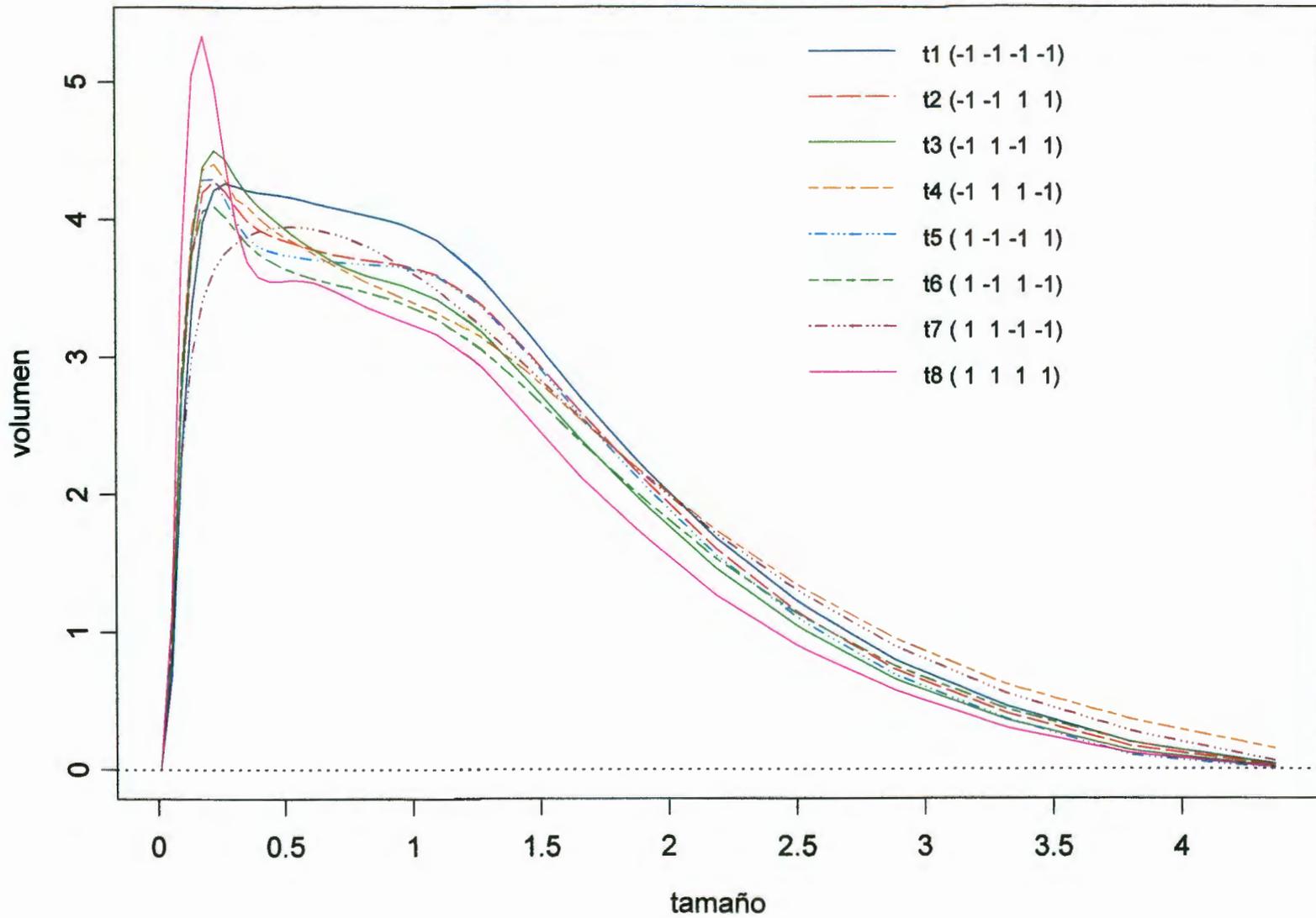


Figura 5.14 Funciones modeladas en el experimento de yogurt

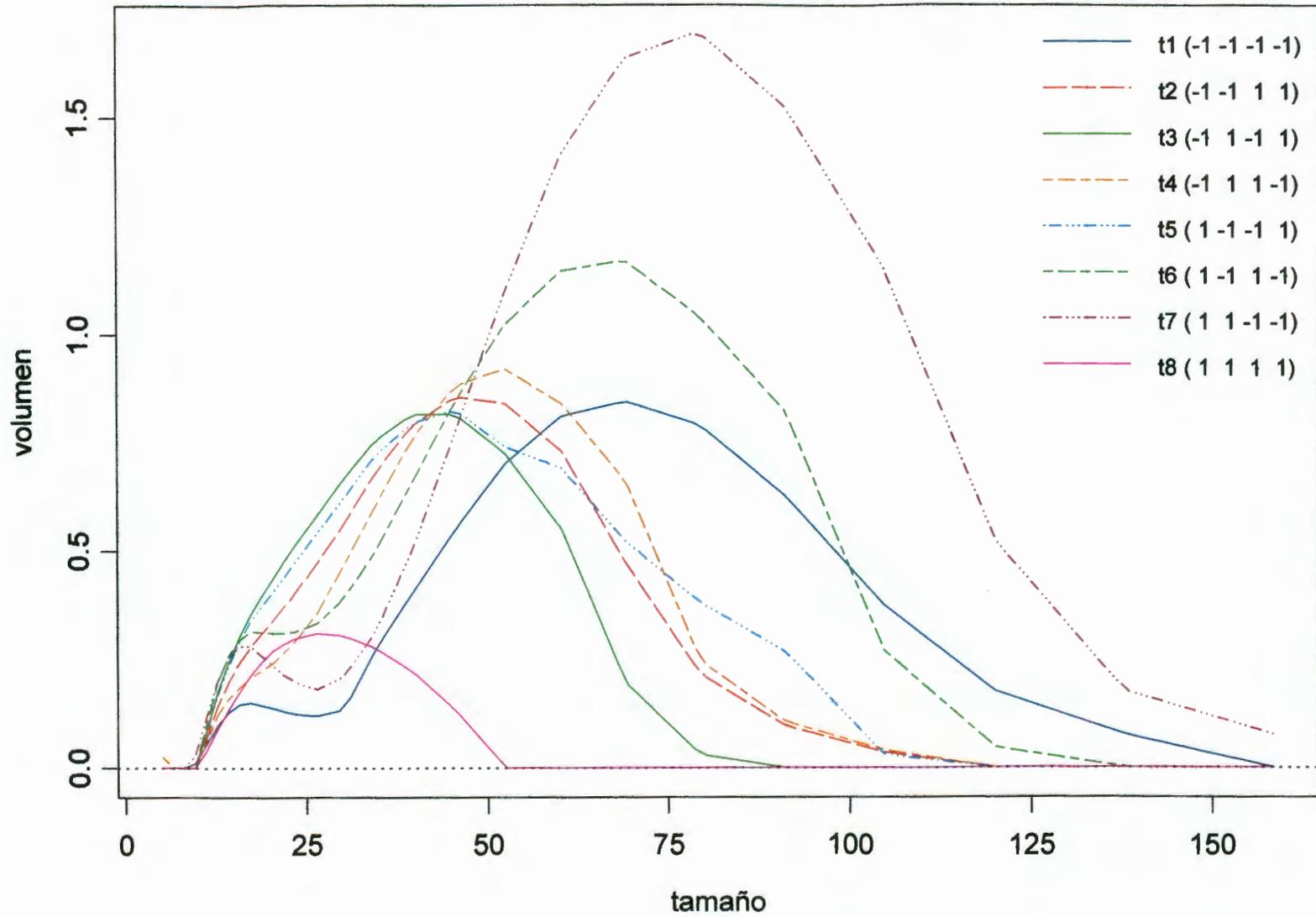


Figura 5.15 Funciones modeladas en el experimento de yogurt

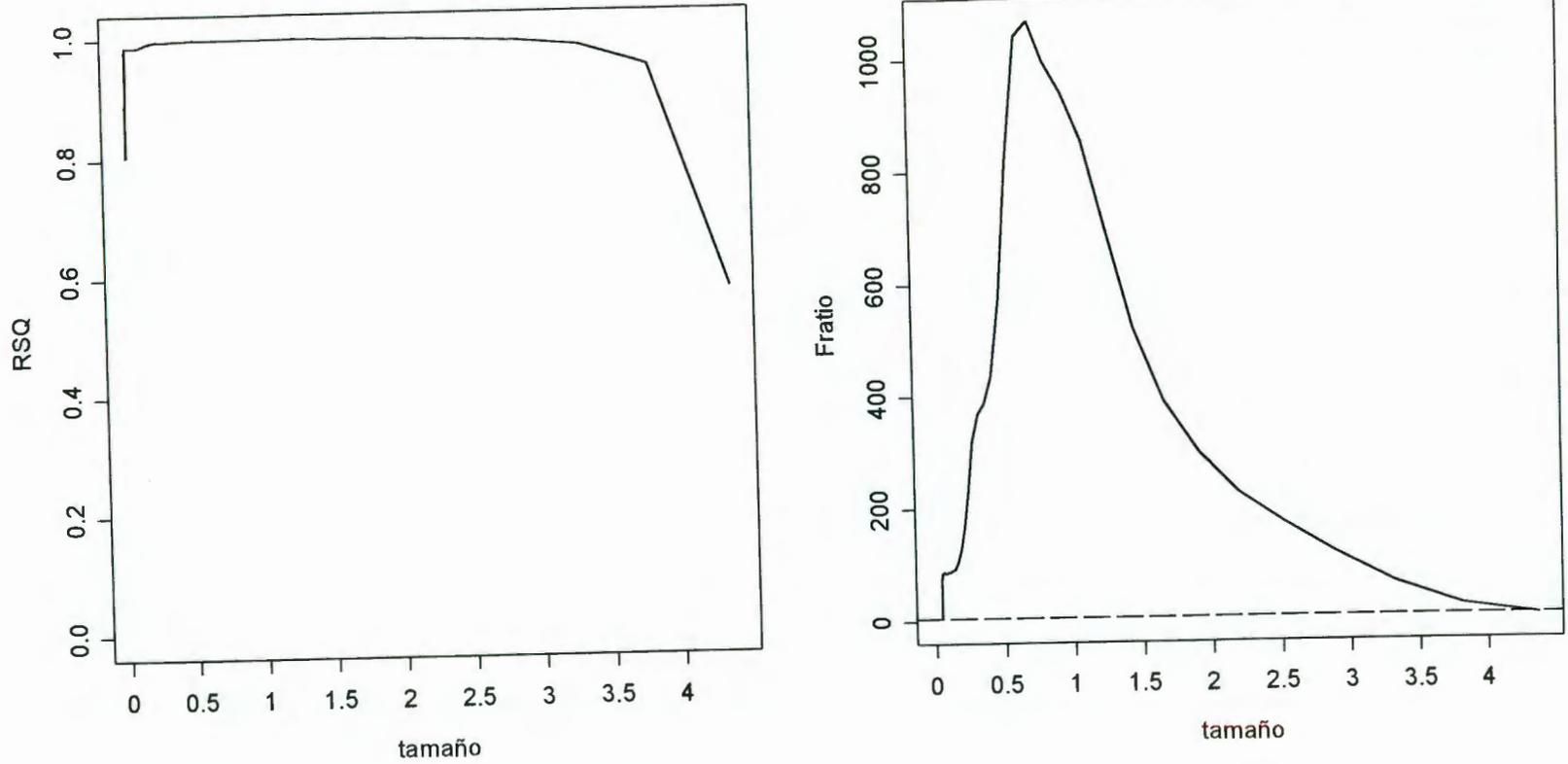


Figura 5.16 Funciones de R^2 y cociente F en el experimento de yogurt

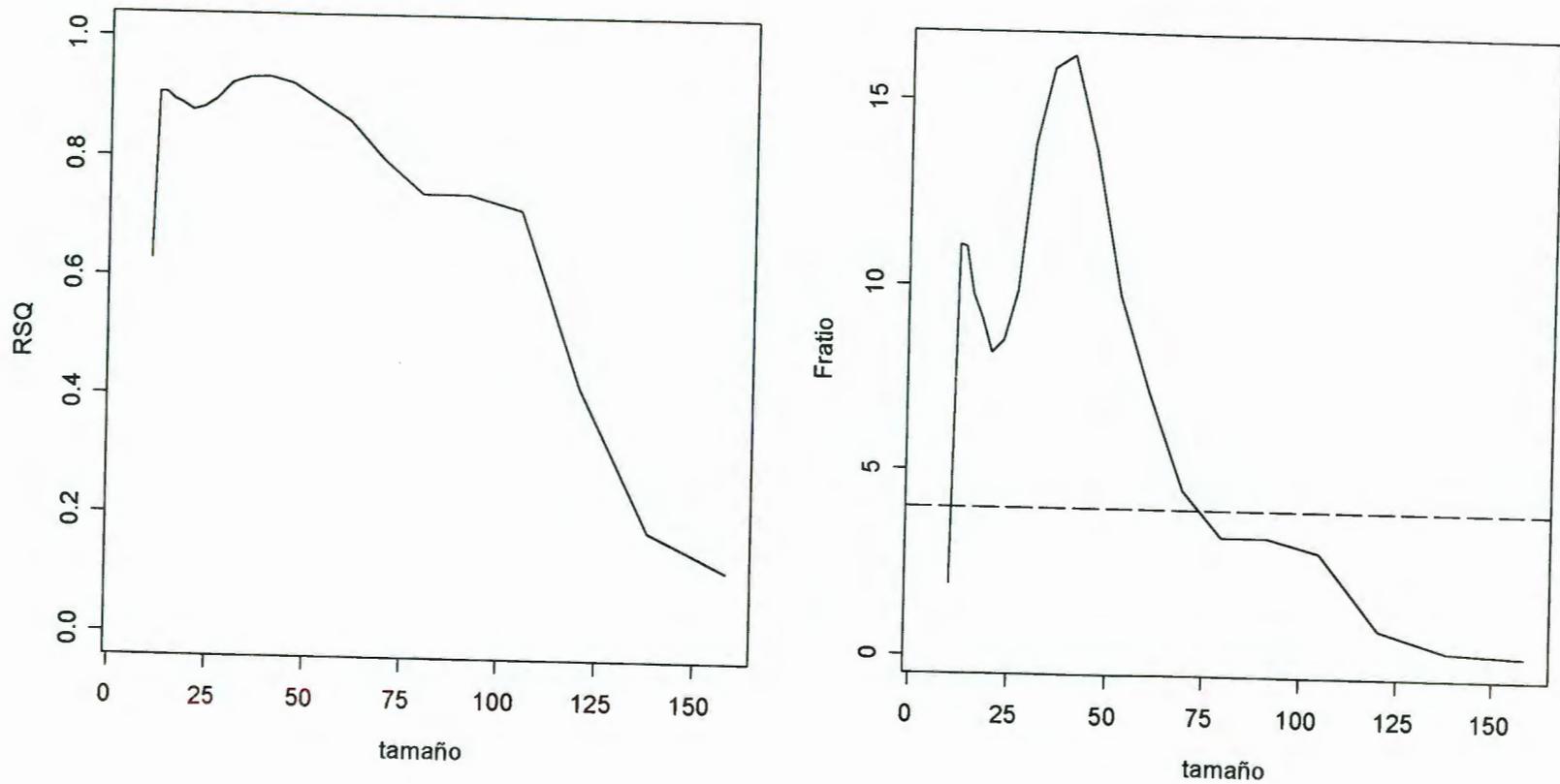


Figura 5.17 Funciones de R^2 y cociente F en el experimento de yogurt

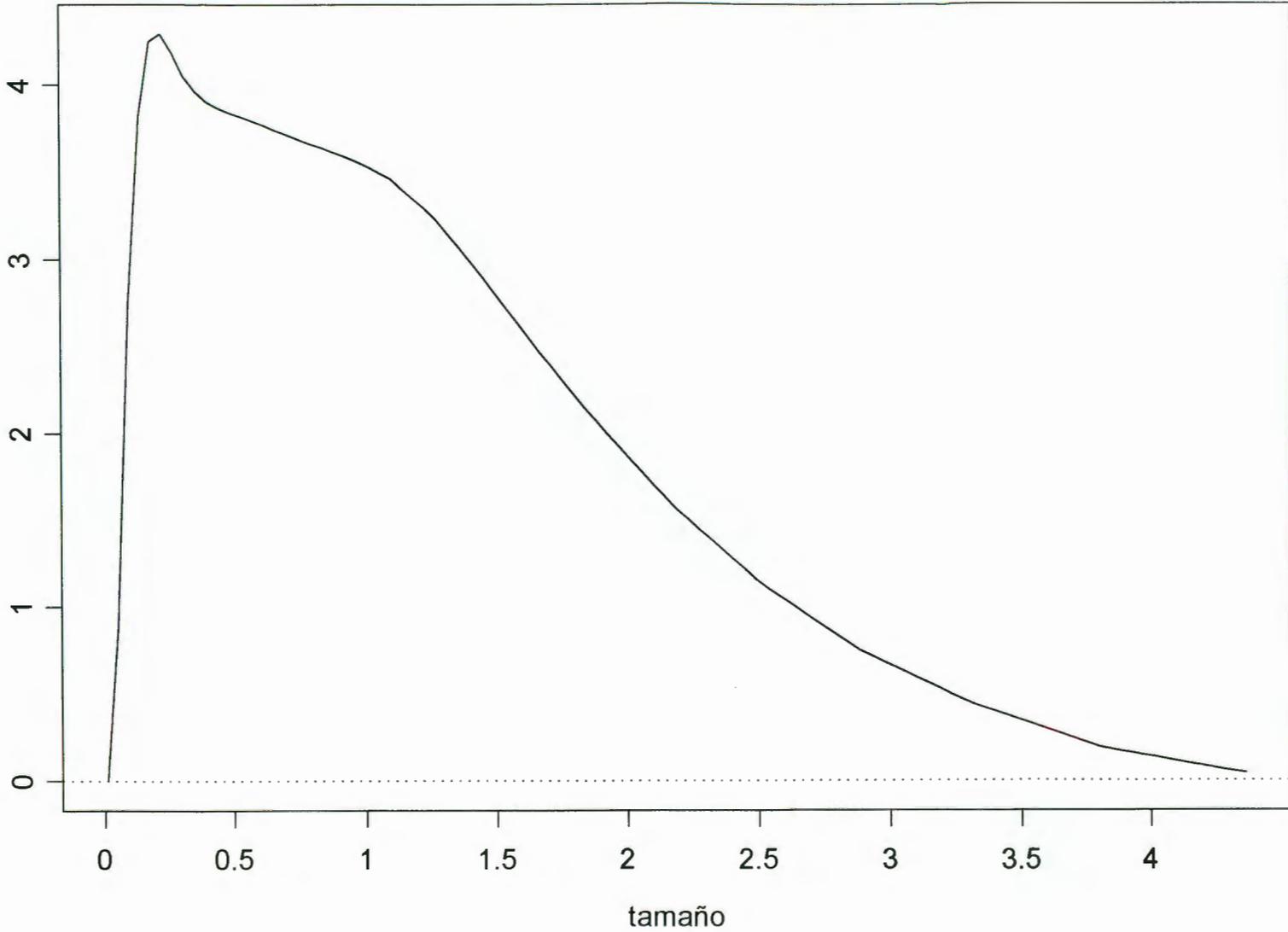


Figura 5.18 Función de regresión de la media en el experimento de yogurt

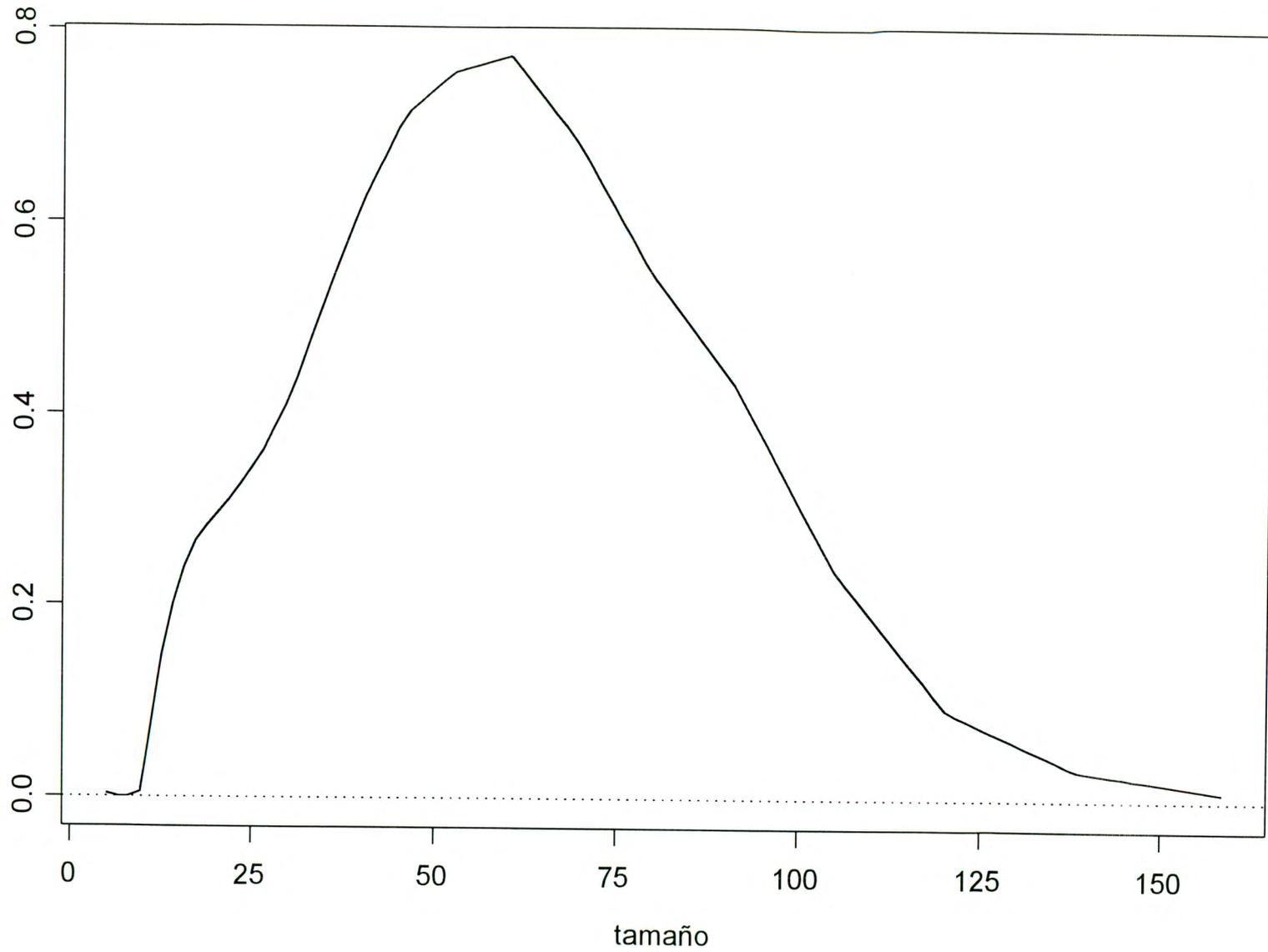


Figura 5.19 Función de regresión de la media en el experimento de yogurt

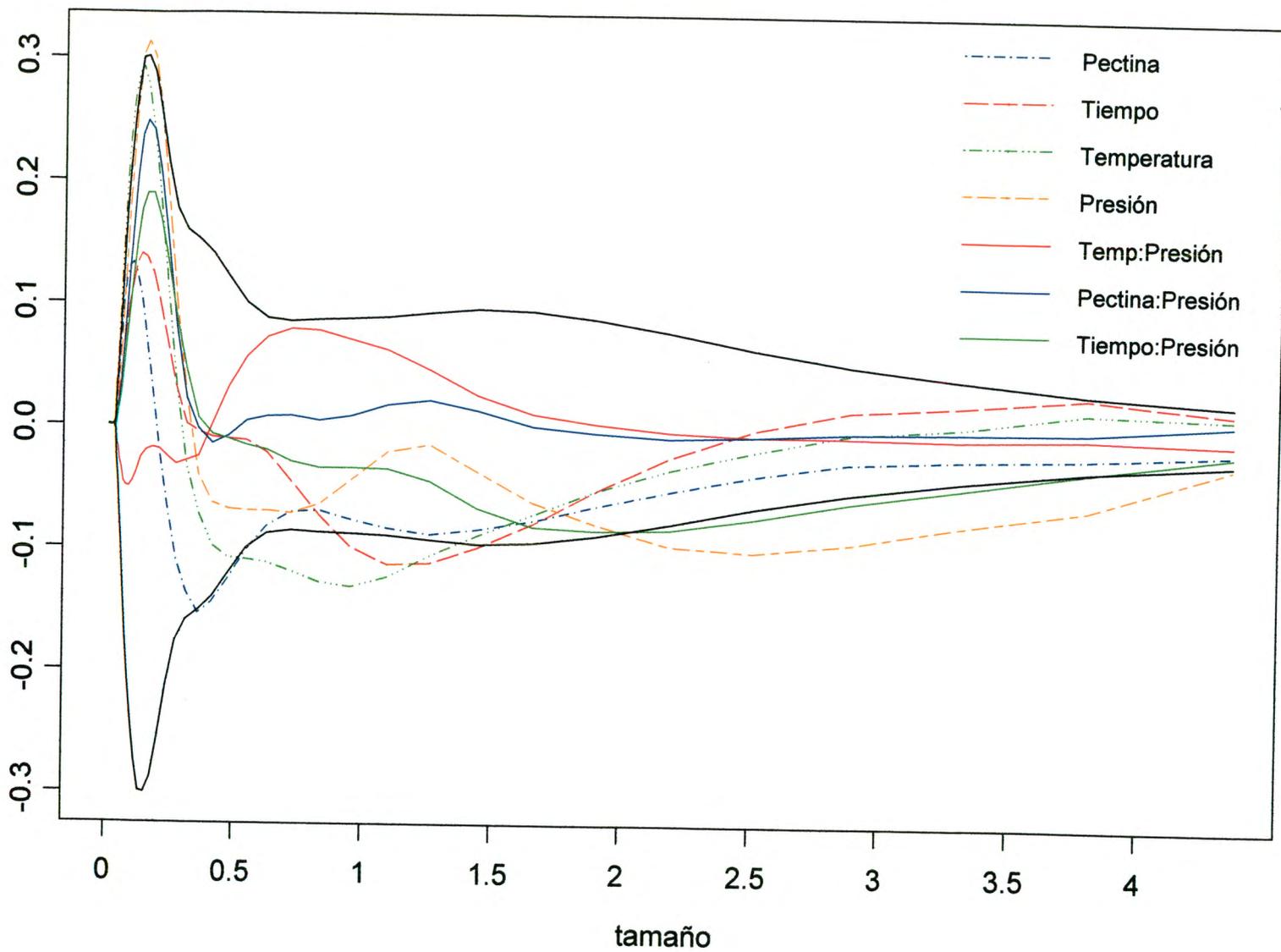


Figura 5.20 Funciones de regresión con bandas de confianza al 95% en el experimento de yogurt

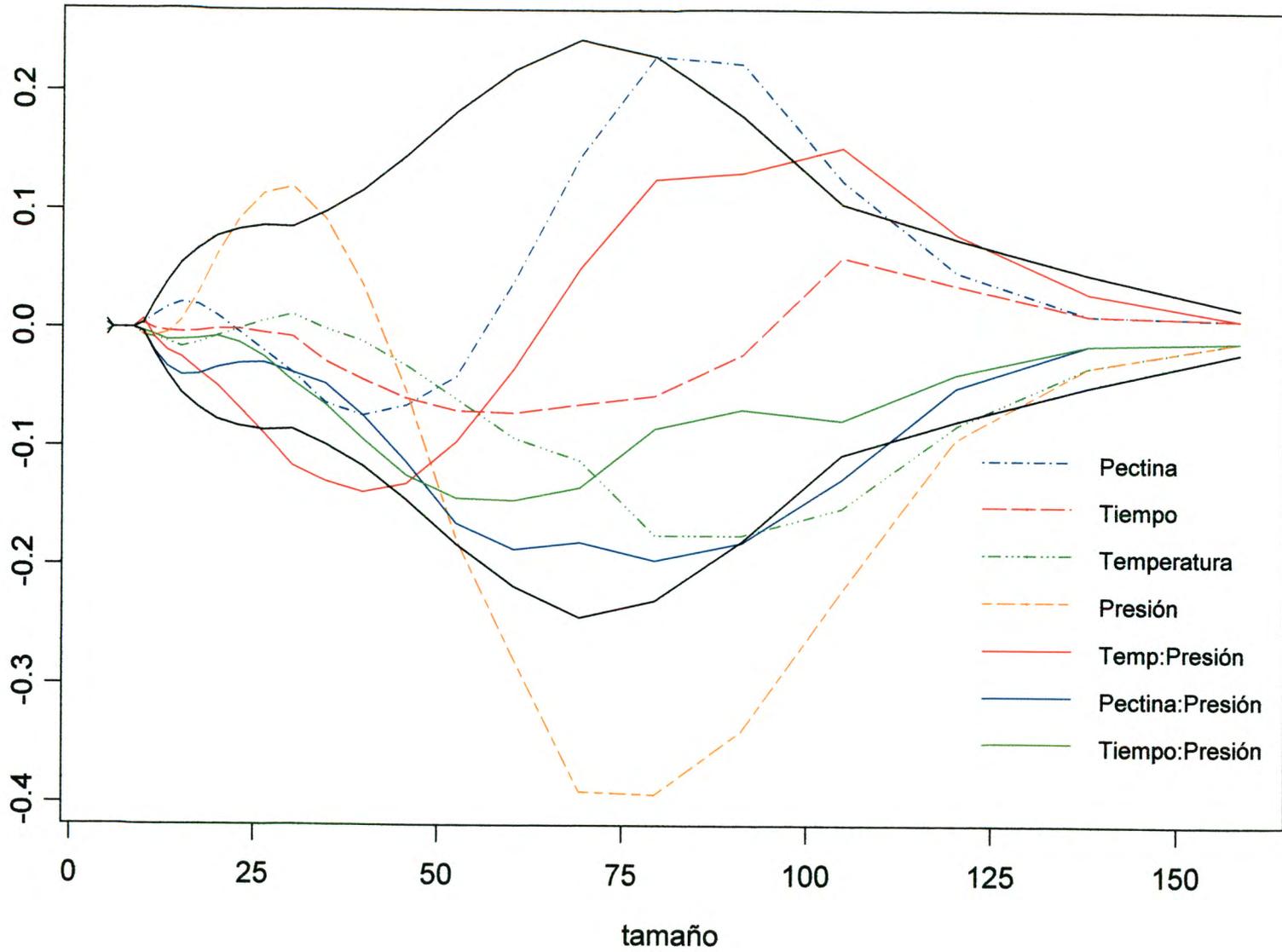


Figura 5.21 Funciones de regresión con bandas de confianza al 95% en el experimento de yogurt

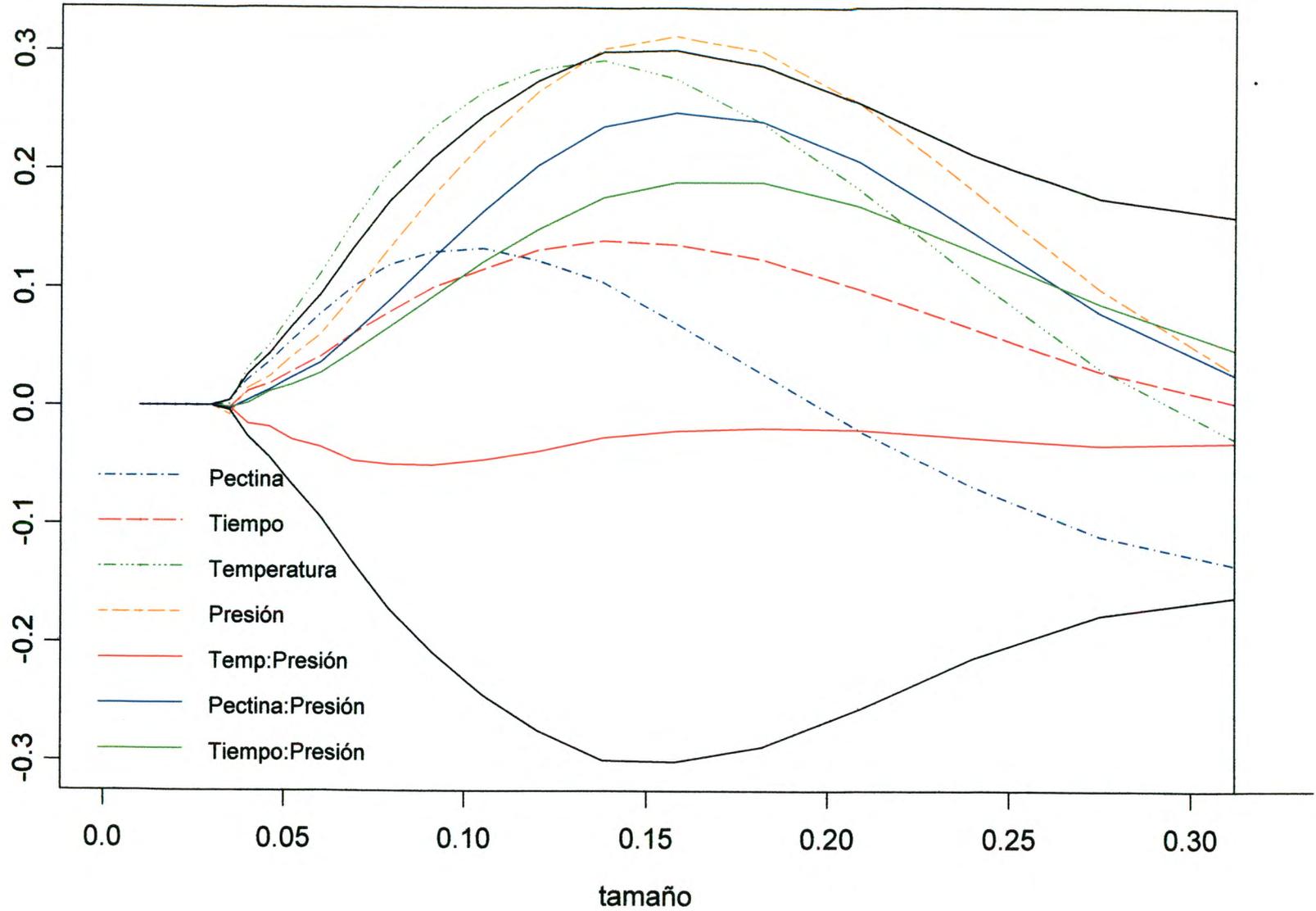


Figura 5.22. Funciones de regresión para tamaños de partícula pequeños

5.5 APLICACIÓN AL MODELO LINEAL FUNCIONAL CASO 2

La brotación es la aparición de raíces en los dientes de ajos, ésta es una propiedad no deseable ante los consumidores, es por ello que se buscan alternativas para retardarla. El siguiente análisis pretende determinar la relación entre el índice de brotación y los perfiles de pérdida de peso, mostrados en el capítulo anterior. Para lo cual se ajustó un modelo funcional en el que la respuesta es el índice de brotación en función de los perfiles de pérdida de peso sujetos a diferentes condiciones de almacenamiento.

Para medir el índice de brotación se almacenaron 5 bulbos de ajos en las mismas condiciones de almacenamiento establecidas para el experimento de la pérdida de peso (Tabla 5.1).

Al cabo de 190 días se seleccionaron tres dientes de ajo de cada uno de los cinco bulbos, que se diseccionaron para medirles la longitud del brote, de manera que se cuenta con 15 mediciones de brotación para cada tratamiento. Para obtener el índice de brotación la longitud del brote se divide entre la longitud del diente y se multiplica por cien. El promedio de los 15 índices se consideró como el valor representativo de cada tratamiento. Los resultados se presentan en la Tabla 5.5. Se promediaron las 3 réplicas de perfiles de pérdida de peso (mostradas en la Figura 5.2) para cada tratamiento, los promedios se presentan en la Figura 5.23, se observa que la pérdida de peso es creciente en general y se tiene la misma interpretación que la de las funciones de pérdida de peso vistas en la sección 5.2.

Tratamiento	Índice de brotación
0°C	77.95
20°C	130.92
30°C	85.15
5°C	145.41
0°C 70% humedad relativa (<i>h.r</i>)	70.32
ambiente	111.90

Tabla 5.5. Índice de brotación para cada tratamiento

Para ajustar el modelo funcional se tomó como la variable independiente estos perfiles promedio y como variable dependiente el índice de brotación, el modelo queda representado de la siguiente manera.

$$y_i = \alpha + \int_0^T x_i(t)\beta(t)dt + \varepsilon_i \quad (5.6)$$

donde:

y_i : Índice de brotación para el tratamiento i . $i = 1, \dots, 6$

α : El valor del intercepto.

$x_i(t)$: Perfil de pérdida de peso para el tratamiento i .

$T = 190$ días.

$\beta(t)$: Función de ponderación.

La función de ponderación $\beta(t)$, se estimó mediante un programa realizado en S-Plus; se presenta en la Figura 5.24, el proceso de estimación resulta en el modelo (5.6) que el intercepto (α) es 78.89, este valor indica un índice de brotación base, que se modificará en función de los perfiles de pérdida de peso. La R^2 obtenida es muy cercana a 1, lo que indica que se tiene un buen ajuste.

La función de ponderación $\hat{\beta}(t)$ muestra que se asigna mayor importancia en el intervalo de 75 a 115 días y los primeros y últimos 15 días, mientras que se tiene menor importancia en los intervalos 35 a 55 y 130 a 170 días.

Obsérvese que en los últimos 15 días es cuando la función de ponderación tiene mayor contribución, y como los perfiles son crecientes entonces aquellos perfiles que presenten un aumento significativo en pérdida de peso en los últimos 15 días aumentarán el índice de brotación.

Este comportamiento se ve reflejado en la velocidad de pérdida de peso para cada perfil mostrados en la Figura 5.9, en la que se ve que los perfiles que aumentan su velocidad en los últimos días, correspondientes a los tratamientos $20^{\circ}C$, $5^{\circ}C$ y temp. ambiente, presentan mayor índice de brotación, mientras que los perfiles que muestran una velocidad casi constante no cambian significativamente el índice de brotación base.

La Figura 5.25 muestra el producto de la función de ponderación con cada uno de los perfiles $(x_i(t)\hat{\beta}(t))$, en ella se confirma lo dicho anteriormente.

Finalmente los tratamientos correspondientes a $0^{\circ}C$ y $0^{\circ}C$ 70% *h.r.* fueron los mejores por que además de presentar un índice de brotación bajo, presentan menor pérdida de peso.

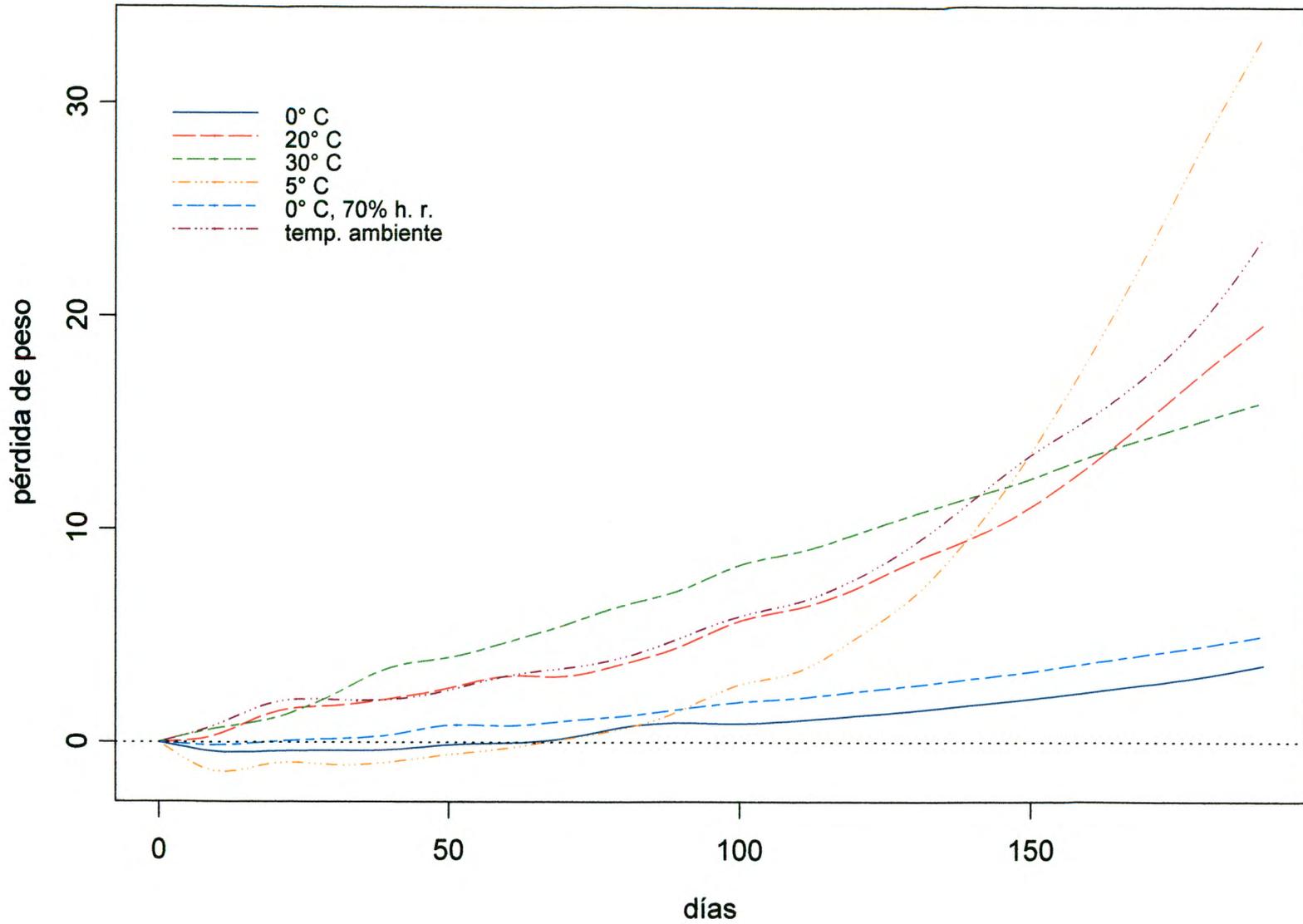
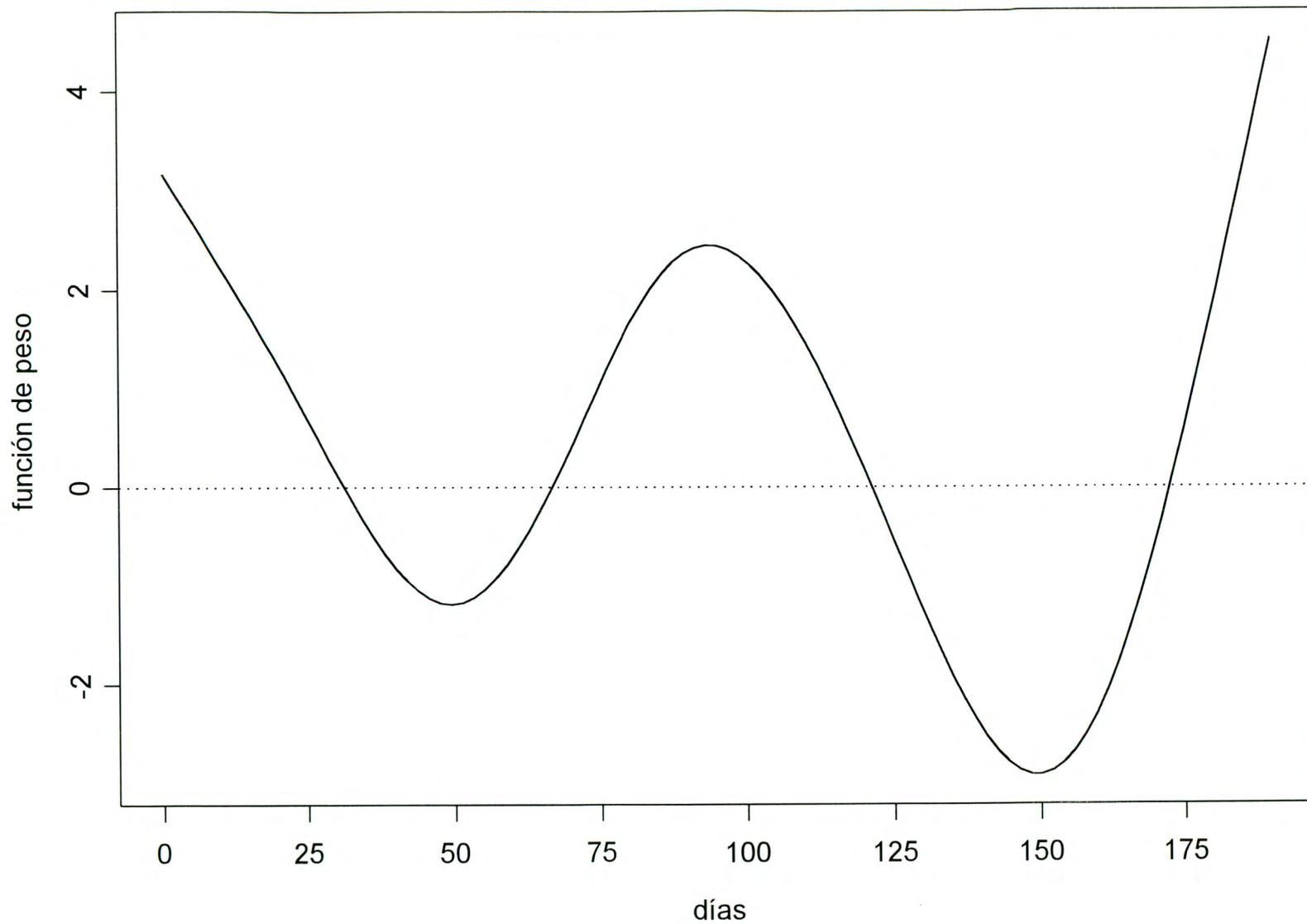
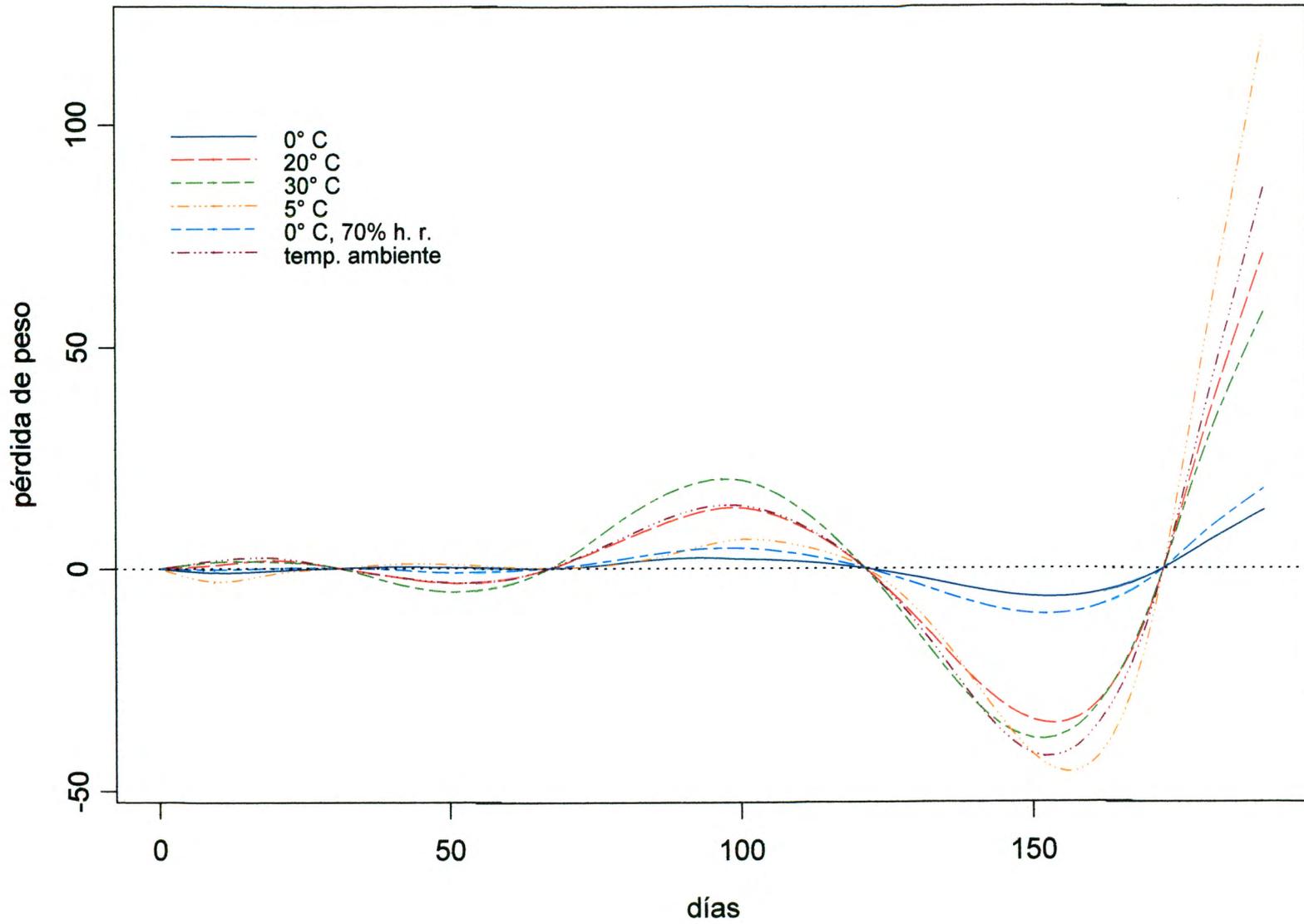


Figura 5.23. Promedios de perfiles de pérdida de peso

Figura 5.24. Función de ponderación $\beta(t)$

Figura 5.25. Producto de la función $\beta(t)$ con cada uno de los perfiles

6

ASPECTOS COMPUTACIONALES

6.1 INTRODUCCIÓN

En este capítulo se explican los programas realizados para llevar a cabo el análisis funcional de las aplicaciones presentadas, dichos programas se implementaron en S-PLUS versión 6.1.

Iniciamos presentando algunas de las instrucciones de S-PLUS más utilizadas en estos programas, una de las más importantes es la instrucción *linmod*, pues con ella se realiza el ajuste del modelo lineal funcional, es por ello que ésta se explica de manera más detallada.

6.2 INSTRUCCIONES EN S-PLUS PARA ANÁLISIS FUNCIONAL DE DATOS

Un objeto en S-PLUS es una lista de datos que pueden ser escalares, vectores, matrices, otras listas, objetos, etc. Por ejemplo un objeto *fd* es una lista que contiene dos elementos, un objeto *base* y una matriz de coeficientes. Podemos extraer partes de esta lista de datos utilizando el nombre del objeto seguido por el signo "\$" y el nombre del elemento que se quiere obtener, o utilizando el nombre del objeto y en dobles corchetes el número del elemento deseado, por ejemplo un objeto funcional llamado *vollin* contiene tres elementos llamados *alpha*, *reg* y *yhat*, para extraer el elemento *yhat* escribimos *vollin\$yhat* o bien *vollin[[3]]*.

Nos referiremos a los objetos funcionales como objetos *fd* un objeto *fd* consiste en una muestra de N observaciones funcionales y contiene tres elementos:

- *coef*: es un arreglo de dimensión 1, 2 ó 3 dependiendo del tipo de función y del número de réplicas. La primera dimensión corresponde al número de funciones base (denotado por K), la segunda dimensión corresponde a las réplicas y la longitud de esta dimensión es N que es el tamaño de la muestra y la tercera se requiere cuando las observaciones funcionales se miden en dos coordenadas o más.
- *basisfd*: especifica una base para la expansión.
- *fdnames*: es una lista en S-PLUS con tres miembros, los cuales son utilizados como etiquetas, el nombre del primer miembro describe el argumento, el segundo se usa para describir las réplicas y el tercero describe los valores de las funciones.

Enseguida describiremos instrucciones en S-PLUS, que son de utilidad para llevar a cabo un análisis funcional.

6.2.1 **create.bspline.basis**

La función *create.basis.fd* es una función para crear cualquier base, pero es más conveniente usar una de las funciones base diseñada para crear una base de un tipo específico, las funciones base que fueron utilizadas en los análisis de las aplicaciones son *splines*, la función para crear este tipo de base es *create.bspline.basis*

```
create.bspline.basis(rangeval, nbasis, norder = 4, breaks = NULL)
```

Argumentos:

rangeval : es un vector de longitud dos que contiene el valor inicial y final del argumento *t*, definiendo el intervalo sobre el cual el objeto funcional puede ser evaluado.

nbasis : es una variable entera que especifica el número de funciones base.

norder (opcional): Un entero que especifica el orden del bspline. Por default se toma como 4.

params (opcional): Es un vector que especifica la posición de los nodos (una sucesión estrictamente creciente), la primera y la última posición deben coincidir con el valor inicial y final de argumento *t* respectivamente.

Salidas:

Regresa una lista con los siguientes nombres: *type*, *rangeval*, *nbasis* y *params* respectivamente.

6.2.2 create.fd

La siguiente instrucción crean un objeto *fd* que contiene observaciones funcionales:

```
create.fd(coef, basisfd, fdnames = defaultnames)
```

Argumentos:

coef : un arreglo de dimensión 2 ó 3, la primera dimensión corresponde a las funciones base, la segunda a las réplicas y la tercera a las funciones

basisfd : un objeto de la clase *basis*.

fdnames (opcional): una lista de longitud 3 utilizadas como etiquetas, donde la primera es para el argumento de valores, por default "time", la segunda para las réplicas por default "reps" y la tercera para funciones por default "values".

Salidas:

Regresa una lista de clase *fd* que contiene un arreglo de coeficientes con el nombre *coefs*, un objeto base llamado *basis* y una lista con el nombre *fdnames*.

6.2.3 data2fd

Ahora detallaremos la técnica principal para crear un objeto funcional a partir de datos discretos, mediante *data2fd*, este paso representa la asociación de datos discretos de cada réplica con una o más funciones definidas por una base, el resultado es un conjunto de funciones que puede ser evaluado por cualquier valor del argumento y puede ser manipulado en varias maneras, como calcular productos internos, tomar derivadas, etc.

$$data2fd(y, argvals, basisfd, fdnames = defaultnames)$$

Argumentos:

y : Es un arreglo que contiene datos discretos de curvas. Si *y* es un vector se supone una réplica y una variable. Si *y* es una matriz los renglones deben corresponder a los valores del argumento y las columnas a las réplicas o casos, se supondrá que hay sólo una variable por observación. Si *y* es un arreglo de tres dimensiones la primera de ellas (renglones) corresponde a los valores del argumento, la segunda (columnas) a las réplicas y la tercera a las variables dentro de las réplicas.

argvals : Conjunto de valores del argumento, si es un vector el mismo conjunto de valores del argumento es usado para las columnas de y , si es una matriz las columnas corresponden a las columnas de y que contienen los valores del argumento para tal réplica o caso.

basisfd : Un objeto de tipo *basis* creado por la función *basis.fd*

fdnames : (opcional) una lista de longitud tres, utilizadas como etiquetas, donde la primera es para el argumento de valores, por default "time"; la segunda para las replicas, por default "reps" y la tercera para funciones, por default "values".

Salidas:

Regresa una lista de la clase *fd* que contiene el arreglo de coeficientes con el nombre *coef*, un objeto *basis* con el nombre *basis*, una lista con el nombre *fdnames*.

6.2.4 smooth.fd

Es una función diseñada para suavizar datos funcionales. Esto es, los datos discretos han sido procesados por *data2fd* para producir un objeto *fd*, y queremos imponer suavidad adicional en los objetos, la versión suavizada de estos objetos puede tener la misma base que los originales o usar una nueva base.

El propósito de esta función es suavizar las funciones en un objeto funcional por el método de regularización o penalización de rugosidad y regresar un objeto funcional que contenga las funciones suaves.

$$smooth.fd(fd, lambda, Lfd, rebase)$$

Argumentos:

fd : objeto funcional a ser suavizado

$lambda$: parámetro de suavizamiento, debe ser un número real positivo, mientras mayor sea este valor se tendrá un mayor suavizamiento.

Lfd (opcional): es un entero que especifica el orden de la derivada o bien un objeto funcional que especifica el operador lineal diferencial cuya valor será penalizado. Los valores de default son tomados de acuerdo a las bases como sigue

base *bspline* : el orden de spline dividido por dos

base *polinomial* : uno

base *constante*: cero

otras bases: dos

rebase (opcional): si es verdadero o distinto de cero, y el tipo de base es poligonal entonces la base es cambiada a *bspline* cúbico antes de suavizamiento.

Salidas:

Regresa un objeto funcional de datos suavizados.

6.2.5 plot.fd

Esta función esta diseñada para graficar objetos funcionales o sus derivadas, réplica por réplica o todas las réplicas simultaneamente.

`plot.fd(fd, Lfd = 0, matplt = T, href = T, nx = 101, xlab, ylab, xlim, ylim, ...)`

Argumentos:

fd : objeto funcional, es decir una lista de la clase fd

Lfd (opcional): un entero de valor cero o mayor o un objeto *fd*. Si es un entero especifica el orden de la derivada que será evaluada.

matplt : (opcional) es una variable lógica. Si es verdadera (T) todas las funciones son graficadas simultaneamente usando la función *matplot*. Si el valor es falso (F) cada función es graficada de manera individual y requiere de un click para avanzar a la siguiente gráfica.

href : es una variable lógica. Si el valor es T una línea horizontal punteada es dibujada a partir de cero

nx : el número de puntos en que las funciones serán evaluadas para graficarse, para funciones suaves 101 valores son suficientes pero para funciones más detalladas será necesario incrementarlo.

xlab : etiqueta del eje *x*

ylab : etiqueta del eje *y*

xlim : un vector que contiene el valor inicial y final del eje *x*

ylim : un vector que contiene el valor inicial y final del eje *y*

...

6.2.6 par

Proporciona control sobre la acción del trazado, color, etc de las gráficas

par(...)

Algunas de las acciones más comunes realizadas por *par* son el tamaño del gráfico, escala de los ejes, tipo y color de línea, etiquetas de los ejes, etc.

6.2.7 key

Utilizamos esta instrucción para colocar etiquetas en las gráficas que se realizan en los programas.

```
key(x, y, ..., title = "", align = T, background = 0, border = 0, between = 2,
corner, divide = 3, transparent = F, cex, cex.title = 1.5 * max(cex), col,
lty, lwd, font, pch, adj = 0, type = "l", size = 5, columns = 1, density = -1,
between.columns = 3, angle = 0, plot = T, space = NULL,
text.width.multiplier = 1)
```

6.2.8 eval.fd

Esta función evalúa un objeto funcional para un conjunto de valores específicos del argumento

$$\text{eval.fd}(\text{evalarg}, \text{fd}, \text{Lfd} = 0)$$

Argumentos:

evalarg : Un vector de valores de argumento en el cual las funciones del objeto funcional serán evaluadas

fd : Objeto funcional, esta es una lista con la clase *fd*

Lfd (opcional): Puede ser un entero de valor cero o mayor o un objeto *fd*. Si es un entero especifica el orden de la derivada que será evaluada.

Salidas:

Regresa un arreglo de dimensiones dos o tres que contiene los valores de la función. La primer dimensión corresponde a los valores del argumento en *evalarg*, la segunda a las réplicas y la tercera a las funciones.

6.2.9 apply

Regresa un vector o arreglo resultado de aplicar una función específica a otro vector o arreglo

$$\mathit{apply}(X, \mathit{MARGIN}, \mathit{FUN}, \dots)$$

Argumentos:

X : Es un arreglo al que se le aplicará la función FUN

$MARGIN$: Indica sobre cuales dimensiones de X se aplicará FUN , por ejemplo, si X es una matriz, $MARGIN = 1$ indica renglones y $MARGIN = 2$ indica columnas.

FUN : Una función que será aplicada al arreglo específico.

...

6.2.10 deriv

Es una función que toma la derivada parcial de una expresión en S-PLUS con respecto a una variable o variables dadas.

$$\mathit{deriv}(\mathit{expr}, \mathit{namevec}, \dots)$$

Argumentos:

expr : expresión que será derivada.

$\mathit{namevec}$: parámetros con respecto a los que se tomará la derivada

...

6.2.11 inprod

Calcula una matriz de productos internos de funciones por integración numérica.

$$\text{inprod}(fd1, fd2, Lfd1 = 0, Lfd2 = 0, JMAX = 15, EPS = 1e^{-4})$$

Argumentos:

$fd1, fd2$: pueden ser datos funcionales o funciones base, en el último caso un objeto funcional es creado de una función base usando la matriz identidad como la matriz de coeficientes. Ambos objetos funcionales deben ser univariados

$Lfd1, Lfd2$: orden de derivadas para el producto interno de $fd1$ y $fd2$ respectivamente

$JMAX$: número máximo de iteraciones

EPS : criterio de convergencia de error relativo.

Salidas:

Una matriz de $NREP1 \times NREP2$ de productos internos para cada posible par de funciones.

6.2.12 mean.fd

Evalua la media puntual de un conjunto de funciones en un objeto funcional

$$\text{mean.fd}(fd)$$

Argumentos:

fd : Objeto funcional.

Salidas:

Un objeto funcional de una réplica que contiene la media de una o varias funciones del objeto *fd*

6.2.13 linmod.fd

La función *linmod* ajusta un modelo funcional lineal en cualquiera de los siguientes casos:

- La variable dependiente es funcional y la variable independiente multivariada.
- La variable dependiente multivariada y la variable independiente funcional.
- La variable dependiente funcional y una sola variable independiente funcional.

$$\text{linmod.fd}(x\text{fd}, y\text{fd}, w\text{vec} = \text{rep}(1, n\text{curves}), xL\text{fd} = 2, yL\text{fd} = 2, x\text{lambda} = 0, \\ y\text{lambda} = 0, z\text{matrnk} = p)$$

Argumentos:

xfd : Si la variable independiente es multivariada, es una matriz de diseño. Si la variable independiente es funcional, es un objeto funcional.

yfd : Si la variable dependiente es multivariada, es una matriz de diseño. Si la variable dependiente es funcional, es un objeto funcional.

wvec : Un vector de pesos para las réplicas. Por default son igual a uno.

xLfd : El orden de la derivada u operador lineal diferencial cuyo valor es penalizado para la variable independiente. 2 por default.

yLfd : El orden de la derivada u operador lineal diferencial cuyo valor es penalizado para la variable dependiente. 2 por default.

xlambda : Un parámetro de suavizamiento para la variable independiente. 0 por default.

ylambda : Un parámetro de suavizamiento para la variable dependiente. 0 por default.

zmatrnk : El rango actual de la matriz de variable independiente para el caso en el que la variable dependiente es funcional y la variable independiente es multivariado.

Salidas:

alpha : Un vector de valores del intercepto en el modelo funcional.

regfd : Un objeto funcional para las funciones de regresión. La estructura de este objeto depende de cada caso.

6.3 PROGRAMAS REALIZADOS EN S-PLUS.

6.3.1 Programas utilizados en las aplicaciones al caso 1

Experimento con ajos

En este caso las variables dependientes son las funciones de pérdida de peso para cada tratamiento (denotadas como *rfd* en el programa) y las variables independientes son los tratamientos (*zmat* en el programa).

```

#-----
#
#                               Experimento con ajos (sección 5.2)
#-----

#----- días en que se midió la pérdida de peso -----
días <-c(0,10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190)

#----- representación funcional de la respuesta -----

knots <- días
norder <- 4
nbasis <- length(knots)
basef <- create.bspline.basis(c(0,190), nbasis, norder, knots)

#----- entrada de datos -----

ajo <- as.matrix(ajo, 20, 18, byrow=T)

#----- creación del objeto fd -----

rfd <-data2fd(ajo,días,basef)

#----- grafica cada curva de las funciones de pérdida de peso -----

plotfit.fd(ajo, knots, rfd)

#----- grafica las funciones de pérdida de peso -----

tipo<-c(1,1,1,4,4,4,6,6,6,13,13,13,14,14,14,37,37,37)
color<-c(2,2,2,8,8,8,8,4,4,4,5,5,5,6,6,6,3,3,3)
par(mfrow=c(1,1),pty="m")
plot(rfd,xlab="días",ylab="pérdida de peso",main="Funciones ajustadas de pérdida
de peso",lty=tipo,col=color)
tiplin<-c(1,4,1,6,13,14,37,1,4)
colr<-c(2,8,4,5,6,4,3,16,5)
axis(1,at=c(25,75,125,175))
key(x=0,y=35,lines = list(lty = c(1,4,6,13,14,37),
col=c(2,8,4,5,6,3)),cex=0.8,
text=list(c("0° C","20° C","30° C","5° C","0° C, 7 h. r.,"temp.
ambiente"),transparent=T))

#----- creación de la matriz de diseño -----

trat1 <- c(1,2,3)
trat2 <- c(4,5,6)
trat3 <- c(7,8,9)
trat4 <- c(10,11,12)
trat5 <- c(13,14,15)
trat6 <- c(16,17,18)

zmat <- matrix(0,18,6)
zmat[trat1,1] <- 1 # columna para el efecto de trat1 - trat6
zmat[trat2,2] <- 1 # columna para el efecto de trat2 - trat6
zmat[trat3,3] <- 1 # columna para el efecto de trat3 - trat6
zmat[trat4,4] <- 1 # columna para el efecto de trat4 - trat6
zmat[trat5,5] <- 1 # columna para el efecto de trat5 - trat6
zmat[c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18),6] <- 1 # efecto de trat6

```

```

#----- ajuste de modelo funcional -----
vollin <- linmod.fd(zmat, rfd)
volregfd <- vollin$reg

#----- grafica cada función de regresión -----

plot.fd(volregfd[1:5], matplt=T, xlab="días", ylab="pérdida de
  peso", lty=c(1, 4, 6, 13, 14),
  col=c(2, 8, 4, 5, 6))
title("Funciones de regresión")
axis(1, at=c(25, 75, 125, 175))
key(x=0, y=-10, lines = list(lty = c(1, 4, 6, 13, 14),
  col=c(2, 8, 4, 5, 6)), cex=0.8,
  text=list(c("0°C - medio amb.", "20°C - medio amb.", "30°C - medio amb.",
  "5°C - medio amb.", "0°C, 7 h. r. - medio amb."), transparent=T))

#----- calcula funciones de residuales -----

resfd <- (rfd - vollin$yhat)

#----- imprime funciones de aproximación -----

plot.fd(vollin$yhat, xlab="días", ylab="pérdida de peso", lty=tipo, col=color)
key(x=0.5, y=30, lines = list(lty = c(1, 4, 6, 13, 14, 37),
  col=c(2, 8, 4, 5, 6, 3)), cex=0.8,
  text=list(c("0° C", "20° C", "30° C", "5° C", "0° C, 7 h. r.", "temp.
  ambiente")), transparent=T)
axis(1, at=c(25, 75, 125, 175))
title("Funciones modeladas")

#----- calcula y grafica media y desviación estándar de pérdida de peso -----

meanfd <- mean.fd(rfd)
stdvfd <- std.fd(rfd)
par(mfrow=c(1, 2), pty="s")
plot(meanfd, main="Media")
plot(stdvfd, main="Desviación Estándar")

#----- calcula la función de suma de cuadrados del error -----

resmat <- eval.fd(resfd, días)
SSE <- apply((resmat)^2, 1, sum)

#----- calcula la suma de cuadrados alrededor de la media -----

resmat0 <- eval.fd(resfd, días) - eval.fd(meanfd, días) %*% matrix(1, 1, 18)
SSY <- apply((resmat0)^2, 1, sum)

#----- calcula RSQ y F-ratio -----

RSQ <- (SSY - SSE) / SSY
Fratio <- ((SSY - SSE) / 5) / (SSE / 12)

```

#----- grafica RSQ y F-ratio con 0.05 valor crítico para F -----

```

par(mfrow=c(1,2),pty="s")
plot(días,RSQ,type="l",ylim=c(0,1),xlab="tamaño",main="R-square")
axis(1,at=c(25,75,125,175))
plot(días,Fratio,type="l",xlab="tamaño",main="F-ratio")
axis(1,at=c(25,75,125,175))
abline(h=4,lty=4)

```

#-----calcula y grafica las derivadas de las funciones de pérdida de peso -----

```

index <- 1:18
Drfd <- deriv(rfd[index],1)
Srfd <- smooth.fd(Drfd,1000)
yfd <- Srfd[index]
par(mfrow=c(1,1),pty="m")
plot.fd(yfd,xlab="días",ylab="velocidad de pérdida de peso",lty=tipo,col=color)
key(x=0.5,y=0.4,lines = list(lty = c(1,4,6,13,14,37),
col=c(2,8,4,5,6,3)),cex=0.8,
text=list(c("0° C","20° C","30° C","5° C","0° C, 7 h. r.,"temp.
ambiente")),transparent=T)
axis(1,at=c(25,75,125,175))

```

Análisis por permutaciones de la aplicación de ajos

```

# -----
#                               Análisis por permutaciones de la aplicación de ajos
# -----

# ----- días en que se midió pérdida de peso -----
días <-c(10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190)

#----- representación funcional de la respuesta -----
knots <- días
norder <- 4
nbasis <- length(knots)
basef <- create.bspline.basis(c(10,190), nbasis, norder, knots)

#----- creación de la matriz de diseño -----
trat1 <- c(1,2,3)
trat2 <- c(4,5,6)
trat3 <- c(7,8,9)
trat4 <- c(10,11,12)
trat5 <- c(13,14,15)
trat6 <- c(16,17,18)

zmat <- matrix(0,18,6)
zmat[trat1,1] <- 1 # columna para el efecto de trat1 - trat6
zmat[trat2,2] <- 1 # columna para el efecto de trat2 - trat6
zmat[trat3,3] <- 1 # columna para el efecto de trat3 - trat6
zmat[trat4,4] <- 1 # columna para el efecto de trat4 - trat6
zmat[trat5,5] <- 1 # columna para el efecto de trat5 - trat6
zmat[,6] <- 1 # columna para el efecto de trat6.

u1<- c(1,0,0,0,0,0)
u2<- c(0,1,0,0,0,0)
u3<- c(0,0,1,0,0,0)
u4<- c(0,0,0,1,0,0)
u5<- c(0,0,0,0,1,0)
u6<- c(0,0,0,0,0,1)

a<-c(0,0,0,0,0,0)
a[1]<- sqrt(t(u1)%*%ginverse(t(zmat)%*%zmat)%*%u1)
a[2]<- sqrt(t(u2)%*%ginverse(t(zmat)%*%zmat)%*%u2)
a[3]<- sqrt(t(u3)%*%ginverse(t(zmat)%*%zmat)%*%u3)
a[4]<- sqrt(t(u4)%*%ginverse(t(zmat)%*%zmat)%*%u4)
a[5]<- sqrt(t(u5)%*%ginverse(t(zmat)%*%zmat)%*%u5)
a[6]<- sqrt(t(u6)%*%ginverse(t(zmat)%*%zmat)%*%u6)

# ----- valores originales de las estadísticas S, I -----
S<-c(9.147537, 1.855406, 3.514401, 7.490903, 8.528764, 15.268318)
I<-c(1009.9906, 126.5583, 329.0874, 657.2252, 841.6099, 1579.3979)
n1<-c(0,0,0,0,0,0)
n2<-c(0,0,0,0,0,0)

# ----- cálculo de permutaciones -----
permutaciones<-matrix(0,19,6)
permutaciones[1, ]<-c(1,2,3,4,5,6)

```

```

permutaciones[2, ]<-c(1,2,5,4,3,6)
permutaciones[3, ]<-c(1,2,6,4,5,3)
permutaciones[4, ]<-c(1,4,3,2,5,6)
permutaciones[5, ]<-c(1,5,3,4,2,6)
permutaciones[6, ]<-c(1,6,3,4,5,2)
permutaciones[7, ]<-c(4,2,3,1,5,6)
permutaciones[8, ]<-c(5,2,3,4,1,6)
permutaciones[9, ]<-c(6,2,3,4,5,1)
permutaciones[10, ]<-c(1,4,5,2,3,6)
permutaciones[11, ]<-c(4,5,3,1,2,6)
permutaciones[12, ]<-c(4,2,5,1,3,6)
permutaciones[13, ]<-c(1,4,6,2,5,3)
permutaciones[14, ]<-c(4,6,3,1,5,2)
permutaciones[15, ]<-c(4,2,3,1,5,3)
permutaciones[16, ]<-c(1,5,6,4,2,3)
permutaciones[17, ]<-c(5,6,3,4,1,2)
permutaciones[18, ]<-c(5,2,6,4,1,3)
permutaciones[19, ]<-c(4,5,6,1,2,3)

```

```
ajop<- ajo
```

```

for(j in 1:19)
{
  nuevap<-c(0,0,0,0,0,0)
  for(i in 1:6)
  {
    if(permutaciones[j,i]<=3)
      nuevap[i]=permutaciones[j,i]
    else
      nuevap[i]=permutaciones[j,i]+12
  }
  ajo[ ,1]=ajo[ ,nuevap[1]]
  ajo[ ,2]=ajo[ ,nuevap[2]]
  ajo[ ,3]=ajo[ ,nuevap[3]]
  ajo[ ,16]=ajo[ ,nuevap[4]]
  ajo[ ,17]=ajo[ ,nuevap[5]]
  ajo[ ,18]=ajo[ ,nuevap[6]]
}

```

```

#----- creación del objeto fd -----
rfd <-data2fd(ajop,días,basef)

```

```

#----- ajuste del modelo funcional -----
vollin <-linmod.fd(zmat,rfd)
volregfd <- vollin$reg

```

```

#----- calcula funciones de residuales -----
resfd<-(rfd-vollin$yhat)

```

```

#----- calcula la función de suma de cuadrados del error -----
resmat<-eval.fd(resfd,días)
SSE<-apply((resmat)^2,1,sum)
rMSE<- sqrt(SSE/12)

```

```

#----- cálculo de estadísticas S (sup) e I (int) -----
coef<-matrix(0,19,6)
beta<-matrix(0,19,6)

```

```

absbeta<-matrix(0,19,6)
sup<-c(0,0,0,0,0,0)
for(i in 1:6)
{
  coef[ ,i]<-getcoef(volregfd[i])
  beta[ ,i]<-coef[ ,i]/(rmse*a[i])
  absbeta[ ,i]<-abs(beta[ ,i])
  sup[i]<-max(absbeta[ ,i],na.rm=T)
}
absbeta[1,]<-0
betaf<-create.fd(absbeta,basef)

aux<-matrix(1,19,6)
auxf<-create.fd(aux,basef)

int<-inprod(betaf,auxf)

```

El siguiente ciclo cuenta el número de veces que las estadísticas S (sup) e I(int) son mayores que los valores de S, I originales

```

for(i in 1:6)
{
  if(int[i]>I[i])
  {
    ni[i]=ni[i]+1
  }
  if(sup[i]>S[i])
  {
    ns[i]=ns[i]+1
  }
}
}

```

Imprime el número de veces que las estadísticas S (sup) e I(int) fueron mayores que los valores de S, I originales

```

print(ni)
print(ns)

```

Experimento con jugo de yogurt con frutas

Las variables dependientes son las funciones de concentración de volumen de tamaños de partículas para cada tratamiento (denotadas como *rfd* en el programa) y las variables independientes son los tratamientos (*dmat* en el programa).

```
#-----
#                               Experimento con yogurt (sección 5.3)
#                               Análisis funcional del primer intervalo
#-----

#----- tamaño de partículas -----

size <-c(0.010,0.011,0.013,0.015,0.017,0.020,0.023,0.026,0.030,0.035,0.040,0.046,
        0.052,0.060,0.069,0.079,0.091,0.105,0.120,0.138,0.158,0.182,0.209,0.240,0.275,
        0.316,0.363,0.417,0.479,0.550,0.631,0.724,0.832,0.955,1.096,1.259,1.445,1.660,
        1.905,2.188,2.512,2.884,3.311,3.802,4.365)

#----- representación funcional de la respuesta -----

knots <- size
norder <- 4
nbasis <- length(knots)
basef <- create.bspline.basis(c(0.01,4.365), nbasis, norder, knots)

#----- creación del objeto fd -----

volumen<-as.matrix(datos5,34,16,byrow=T)
rfd <-data2fd(volumen,size,basef,argnames=c("tamaño","tratamiento","volumen"))

#----- grafica las funciones de volumen -----

tipo<-c(1,1,4,4,1,1,6,6,13,13,14,14,37,37,1,1)
color<-c(2,2,8,8,4,4,5,5,6,6,4,4,3,3,19,19)
par(mfrow=c(1,1),pty="m")
plot(rfd,xlab="tamaño",ylab="volumen",main="Funciones ajustadas de
      volumen",lty=tipo,col=color)
tiplin<-c(1,4,1,6,13,14,37,1)
colr<-c(2,8,4,5,6,4,3,16)
axis(1,at=c(0.5,1.5,2.5,3.5))
key(x=2.5,y=6,lines = list(lty = c(1,1,4,1,1,1,6,1,13,1,14,1,37,1,1),
col=c(2,0,8,0,4,0,5,0,6,0,4,0,3,0,19)),cex=0.8,
text=list(c("t1 (-1 -1 -1 -1)"," ","t2 (-1 -1 1 1)"," ","t3 (-1 1 -1 1)"," ",
"t4 (-1 1 1 -1)"," ","t5 ( 1 -1 -1 1)"," ","t6 ( 1 -1 1 -1)"," ",
"t7 ( 1 1 -
1 -1)",
" ","t8 ( 1 1 1 1)")),transparent=T)
```

```

#-----
#                               Modelo lineal para funciones de volumen
#-----

#----- creación de la matriz de diseño -----

t1<-c(1,2)
t2<-c(3,4)
t3<-c(5,6)
t4<-c(7,8)
t5<-c(9,10)
t6<-c(11,12)
t7<-c(13,14)
t8<-c(15,16)
dmat<-matrix(1,16,8)
dmat[t1,c(2,3,4,5)] <- -1
dmat[t2,c(2,3,7,8)] <- -1
dmat[t3,c(2,4,6,7)] <- -1
dmat[t4,c(2,5,6,8)] <- -1
dmat[t5, c(3,4,6,8)] <- -1
dmat[t6,c(3,5,6,7)] <- -1
dmat[t7,c(4,5,7,8)] <- -1

#----- efectos de factores -----

vollin <-linmod.fd(dmat,rfd)
volregfd <- vollin$reg

#----- grafica cada función de regresión -----

plot.fd(volregfd[1],matplt=T,xlab="tamaño",ylab="volumen")
title("Función de regresión de la media")
axis(1,at=c(0.5,1.5,2.5,3.5))

cor<-c(2,8,4,5,8,2,4)
tip<-c(3,4,5,6,9,1,1)

plot.fd(volregfd[2:8],matplt=T,xlab="tamaño",ylab="volumen",lty=tip,col=cor)
title("Funciones de regresión")
key(x=1.8,y=0.3,lines = list(lty =
  c(3,1,4,1,5,1,6,1,9,1,1,1,1),col=c(2,0,8,0,4,0,5,0,8,0,2,0,4)),cex=0.8,
text=list(c("Pectina"," ","Tiempo"," ","Temperatura"," ","Presión","
  ","Temp:Presión"," ","Pectina:Presión",
  " ","Tiempo:Presión")),transparent=T)
axis(1,at=c(0.5,1.5,2.5,3.5))

#----- calcula funciones de residuales -----

resfd<-(rfd-vollin$yhat)

#----- calcula media de volumen -----

meanfd<-mean.fd(rfd)

#----- calcula la función de suma de cuadrados del error -----

resmat<-eval.fd(resfd,size)
SSE<-apply((resmat)^2,1,sum)

```

```

#----- calcula la suma de cuadrados alrededor de la media -----
resmat0<-eval.fd(resfd,size)-eval.fd(meanfd,size)%*%matrix(1,1,16)
SSY<-apply((resmat0)^2,1,sum)

#----- calcula la correlación cuadrada múltiple y funciones F-radio -----

RSQ<-(SSY-SSE)/SSY
Fratio<-((SSY-SSE)/7)/(SSE/8)

#----- Bandas de confianza -----

varfun<-((SSE/8)^.5)/4
limsup<-2*varfun
liminf<-(-2)*varfun
efectos.eval<-eval.fd(volregfd[2:8],size)
matplot(size,cbind(efectos.eval,limsup,liminf),xlab="tamaño",ylab="volumen",
type="l",lty=c(tip,1,1),col=c(cor,1,1))
key(x=3.25,y=0.32,lines = list(lty =
c(3,1,4,1,5,1,6,1,9,1,1,1,1),col=c(2,0,8,0,4,0,5,0,8,0,2,0,4)),cex=0.8,
text=list(c("Pectina"," ","Tiempo"," ","Temperatura"," ","Presión","
","Temp:Presión"," ","Pectina:Presión",
" ","Tiempo:Presión")),transparent=T)
axis(1,at=c(0.5,1.5,2.5,3.5))
title("Funciones de regresión con bandas de confianza al 95%")

#----- imprime funciones de aproximación -----

plot.fd(vollin$yhat,lty=tipo,col=color)
key(x=2.5,y=5.3,lines = list(lty = c(1,1,4,1,1,1,6,1,13,1,14,1,37,1,1),
col=c(2,0,8,0,4,0,5,0,6,0,4,0,3,0,19)),cex=0.8,
text=list(c("t1 (-1 -1 -1 -1)", " ","t2 (-1 -1 1 1)", " ","t3 (-1 1 -1 1)", " ",
"t4 (-1 1 1 -1)", " ","t5 ( 1 -1 -1 1)", " ","t6 ( 1 -1 1 -1)", " ","t7 ( 1 1 -
1 -1)",
" ","t8 ( 1 1 1 1)")),transparent=T)
axis(1,at=c(0.5,1.5,2.5,3.5))
title("Funciones modeladas")

#----- grafica RSQ y F-ratio con 0.05 valor crítico para F -----

par(mfrow=c(1,2),pty="s")
plot(size,RSQ,type="l",ylim=c(0,1),xlab="tamaño",main="R-square")
axis(1,at=c(0.5,1.5,2.5,3.5))
plot(size,Fratio,type="l",xlab="tamaño",main="F-ratio")
axis(1,at=c(0.5,1.5,2.5,3.5))
abline(h=4,lty=4)

```

6.3.2 Programa utilizado en la aplicación al caso 2

Experimento con ajos

```

#-----
#                               Análisis funcional de índice de brotación de ajos
#-----

#----- tamaño de partículas -----
dias <-c(0,10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190)

#----- representación funcional -----
knots <- dias
norder <- 4
nbasis <- length(knots) + norder- 2
basef <- create.bspline.basis(c(0,190), nbasis, norder, knots)

# -----
#                               Modelo de índice de brotación como función de pérdida de peso
# -----

ibrot <- as.matrix(brotacion,6,1,byrow=T)
ppeso <- as.matrix(brot,20,6,byrow=T)
rfd <-data2fd(ppeso,dias,basef)

#----- grafica las funciones de pérdida de peso -----
tipo<-c(1,4,6,13,14,37)
color<-c(2,8,4,5,6,3)
plot(rfd,xlab="días",ylab="pérdida de peso",main="Funciones de pérdida de peso",
     lty=tipo,col=color)
key(x=0,y=30,lines = list(lty = c(1,4,6,13,14,37),col=c(2,8,4,5,6,3)),cex=0.8,
text=list(c("0° C","20° C","30° C","5° C","0° C, 7h.r.", "temp.ambiente"),
transparent=T))

linmodlist <- linmod.fd(rfd, ibrot, xlambda=0.000005)

intercept <- linmodlist[[1]]
regfd <- linmodlist[[2]]

# -----
#                               grafica la función de regresión
# -----

plot.fd(regfd,matplt=T,xlab="días",ylab="función de peso")
axis(1,at=c(25,75,125,175))
title("Función de peso")
legend(125,6,c("alpha=",intercept))

#----- calcula los valores ajustados y grafica Y contra Y gorro -----
yhat <- inprod(regfd,rfd) + intercept
plot(yhat, ibrot, type="p", xlab="índice de brotación predicho", ylab="índice de
     brotación observado")
abline(0,1,lty=2)
RSQ <- cor(yhat,ibrot)^2
round(RSQ,3)

```

6.4 AJUSTE DEL MODELO FUNCIONAL LINEAL EN S-PLUS

La instrucción *linmod* que vimos en la sección 6.2 es una de las más importantes ya que con ella se ajusta el modelo funcional lineal, como se explicó en el capítulo 4 existen tres casos, las aplicaciones vistas se enfocaron únicamente al caso 1 y al caso 2, por lo que en esta sección se presentarán las subrutinas de *linmod* correspondientes a estos casos.

Para el primer caso del modelo, en el experimento con ajos utilizamos ésta función como:

$$\text{linmod.f}d(\text{zmat}, \text{rfd})$$

y para el experimento con yogurt:

$$\text{linmod.f}d(\text{dmat}, \text{rfd})$$

Mientras que en el segundo caso en el experimento se tiene:

$$\text{linmod.f}d(\text{rfd}, \text{ibrot}, \text{xlamb}d = 0.000005)$$

```

# ++++++
#                               Caso 1: Variable independiente multivariada
#                               y Variable independiente funcional
# ++++++

# Evalúa si el argumento yfd (DV) es un objeto funcional y si xfd (IV) es una matriz
if (inherits(yfd, "fd") && !(inherits(xfd, "fd")))
{
  if (is.matrix(xfd))
  {
    zmat <- as.matrix(xfd)
    fd <- yfd
  }
  else
  {
    stop ("First argument not a vector or a matrix.")
  }
}

# Extrae el arreglo de coeficientes del objeto funcional rfd, lo guarda en coef
coef <- getcoef(fd)

# Asigna a coefd la dimensión de la matriz de coeficientes.
coefd <- dim(coef)

# ndim es el número de elementos en coefd. En el ejemplo igual a 2.
ndim <- length(coefd)

# Si ndim es menor que 2 no se tienen datos suficientes para hacer el análisis, por lo que termina la
# subrutina.
if (ndim < 2)
  stop("Linear modeling impossible with 1 replication")

# nrep contiene el número de renglones de la matriz de diseño
nrep <- nrow(zmat)

# En este caso asigna el valor de 1 a nvar ya que es ndim es 2
if(ndim == 3) nvar <- coefd[3] else nvar <- 1

# Extrae el objeto basis.fd del objeto funcional fd que contiene los siguientes elementos: tipo de base,
# rango de valores sobre la base, parámetros y el número de bases.
basisfd <- getbasis(fd)

# nbasis indica el número de funciones base tomadas del objeto basis.fd
nbasis <- basisfd$nbasis

# rangewt contiene el intervalo en el cual varían los valores de los pesos. Para la aplicación se tomo el

```

vector de pesos wtvec como default y rangewt varía de uno a uno.

```
rangewt <- range(wtvec)
```

Las siguientes instrucciones verifican que el vector de pesos no contenga valores negativos, que sea de la longitud adecuada, que los valores dados a lambda sean no negativos, que el orden de la deriva que será penalizada sea no negativa y que el número de renglones de la matriz de diseño coincida con el número de réplicas.

```
if (rangewt[1] < 0) stop("WTVEC must not contain negative values.")
if (rangewt[1] == rangewt[2]) wtvar <- F else wtvar <- T

if (length(wtvec) != nrep) stop("WTVEC of wrong length")
if (min(wtvec) <= 0) stop("All values of WTVEC must be positive.")
if (ylambda < 0) warning (
  "Value of LAMBDA was negative, and 0 used instead.")
if (ylambda > 0 && yLfd < 0) stop(
  "Order of derivative must be nonnegative.")

if (dim(coef)[2] != nrep) stop(
  "Number of rows of ZMAT must equal number of replications")
```

Asigna a p el número de columnas de la matriz de diseño.

```
p <- ncol(zmat)
```

Construye una matriz bcoef de ceros con renglones igual al número de funciones base y columnas igual al número de parámetros, ya que nvar es igual a 1 en este ejemplo.

```
if (nvar > 1)
{
  bcoef <- array(0,c(nbasis,p,nvar))
}
else
{
  bcoef <- matrix(0,nbasis,p)
}
```

Verifica que el rango de la matriz de diseño no sea mayor que el número de columnas (p).

```
if (zmatrnk > p)
  stop("Specified rank of ZMAT must not be greater than no. columns.")
```

Si el rango es menor que p hace el siguiente análisis

```
if (zmatrnk < p)
{
  rootw <- sqrt(wtvec)
  zmatd <- svd(sweep(zmat,2,rootw,"*"))
  zmatd <- zmatd$d
  if (zmatd[zmatrnk] <= 0) stop("ZMAT is not of specified column rank")
  index <- 1:zmatrnk
  zginvt <- sweep(zmatd$u[,index],1,zmatd[index],"/") %*%
    t(zmatd$v[,index])
  if (nvar == 1)
  {
```

```

      bcoef <- sweep(coef,1,rootw,"*") %**% zginvt
    }
  else
  {
    for (j in 1:nvar)
    {
      bcoef[,j] <- sweep(coef[,j],1,rootw,"*") %**% zginvt
    }
  }
}

```

Para la aplicación el rango de la matriz de diseño es igual a p y $nvar$ es 1, luego se obtiene la estimación de los coeficientes de regresión con la función `lsfit`, que ajusta una regresión por mínimos cuadrados multivariada (ponderada por el vector de pesos `wtvec`), que tiene como variable independiente la matriz de diseño y como variable respuesta la transpuesta de `coef`, dado que `int` es igual a `F` no se consideró el intercepto en cada regresión. De la salida de `lsfit` solo tomamos la lista de coeficientes estimados y a su transpuesta la llamamos `bcoef`.

```

else
{
  if (nvar == 1)
  {
    bcoef <- t(lsfit(zmat,t(coef),wtvec,int=F)$coef)
  }
}

```

En el caso que $nvar$ no sea 1 y el rango de la matriz sea p se tiene lo siguiente

```

else
{
  for (j in 1:nvar)
  {
    bcoef[,j] <- t(lsfit(zmat,t(coef[,j]),wtvec,int=F)$coef)
  }
}
}

```

Para obtener las observaciones estimadas multiplicamos la matriz `bcoef` con la transpuesta de la matriz de diseño.

```

yhatcoef <- bcoef %**% t(zmat)

```

Asigna nombres a los coeficientes estimados de regresión.

```

if (nvar > 1)
{
  dimnames(bcoef) <- list(NULL,dimnames(zmat)[[2]],
                        dimnames(yfd[[1]])[[3]])
}
else
{
  dimnames(bcoef) <- list(NULL,dimnames(zmat)[[2]])
}

fdnames <- getnames(fd)
regfdnames <- fdnames

```

```
regfdnames[[2]] <- paste('Reg. Coef.',1:p)
regfdnames[[3]] <- 'Reg. Coef.'
names(regfdnames)[2] <- 'Reg. Coefficients'

# Crea un objeto funcional a partir del vector de coeficientes estimados con la base especificada
# anteriormente y los nombres dados.

regfd <- create.fd (bcoef, basisfd, regfdnames)

# Crea un objeto funcional a partir de las observaciones estimadas con la base especificada anteriormente
# y los nombres dados.

yhatfd <- create.fd (yhatcoef, basisfd, fdnames)

# Crea una lista que contiene 0, las funciones de regresión, las funciones de observaciones estimadas y los
# coeficientes estimados de las funciones de regresión, a lo que le llama linmodlist

linmodlist <- list(0, regfd, yhatfd, bcoef)

# Asigna nombres a los elementos de esta lista.

names(linmodlist) <- c('alpha', 'regfd', 'yhatfd', 'bcoef')

# Regresa linmodlist

return( linmodlist )
}
```

```

# ++++++
#                               Caso 2: Variable independiente funcional
#                               y variable dependiente multivariada
# ++++++

# Evalúa si el argumento yfd (DV) es una matriz y si xfd (IV) es un objeto funcional

if (inherits(xfd, "fd") && !(inherits(yfd, "fd")))
{
  if (is.numeric(yfd))
  {
    ymat <- as.matrix(yfd)
    fd    <- xfd
  }
  else
  {
    stop (paste("Second argument not a functional data object",
               " and is not numeric",
               "when first argument is functional."))
  }
}

# Extrae el arreglo de coeficientes del objeto funcional pfd

coef  <- getcoef(fd)

# Asigna a coef la dimensión de la matriz de coeficientes. En este caso es (22,6)

coefd <- dim(coef)

# ndim es el número de elementos en coefd. En el ejemplo igual a 2.

ndim  <- length(coefd)

# Si ndim es menor que 2 no se tienen datos suficientes para hacer el análisis, por lo que termina
# la subrutina.

if (ndim < 2)
  stop("Linear modeling impossible with 1 replication")

# Si ndim es tres el análisis no se puede realizar

if (length(dim(coef)) == 3)
  stop("This version cannot accommodate multiple functional IVs")

# Asigna a nrep el segundo elemento de coefd, que es 6

nrep <- coefd[2]

# Asigna a nvar el número de columnas de la matriz de respuestas (índice de brotación), que es 1

nvar <- ncol(ymat)

# rangewt contiene el intervalo en el cual varían los valores de los pesos.
# Para la aplicación se tomó el vector de pesos wtvec como default es decir, este es un vector con
# unos de longitud 6 por lo que rangewt varía de uno a uno.

```

```

rangewt <- range(wtvec)

# Se verifica que el vector de pesos no contenga valores negativos.

if (rangewt[1] < 0) stop("WTVEC must not contain negative values.")

# Si los pesos varían asigna a wtvar un valor verdadero, de lo contrario wtvar es falso

if (rangewt[1] == rangewt[2]) wtvar <- F else wtvar <- T

# Extrae el objeto basis.fd del objeto funcional fd que contiene los siguientes elementos: tipo de
# base, rango de valores sobre la base, parámetros y el número de bases.

basisfd <- getbasis(fd)

# Para esta aplicación basisfd contiene lo siguiente:

> getbasis(rfd)
$type:
[1] "bspline"

$rangeval:
[1] 0 190

$nbasis:
[1] 22

$params:
[1] 10 20 30 40 50 60 70 80 90 100 110 120 130 140
[2] 150 160 170 180

attr(,"class"):
[1] "basis.fd"

# nbasis indica el número de funciones base tomadas del objeto basis.fd que como se vio
# anteriormente son 22.

nbasis <- basisfd$nbasis

# type indica el tipo de base utilizada, en este caso es "bspline"

type <- getbasistype(basisfd)

# Las siguientes instrucciones verifican que el número de renglones de la matriz de respuestas sea igual al
# número de réplicas indicadas en el objeto funcional de la variable independiente. También se verifica que
# la longitud del vector de pesos sea igual al número de réplicas y que el valor de xlambda sea positivo.

if (nrow(yamat) != nrep)
  stop("Number of rows of YMAT must equal number of replications")

# one es un vector de unos de longitud nrep = 6
one <- rep(1,nrep)

if (length(wtvec) != nrep)

```

```

    stop("WTVEC of wrong length")
  if (xlambda < 0)
    warning ("Value of XLAMBDA was negative, and 0 used instead.")

# Asigna a jmat la matriz de dimensión 22*22 J , que tiene las entradas
#  $J_{jk} = \int \phi_j(s)\phi_k(s)ds = \langle \phi_j, \phi_k \rangle$ , donde  $\phi_k$  es el elemento k-ésimo del vector ,
#  $\phi = (\phi_1, \phi_2, \dots, \phi_{45})$  que contiene las 22 funciones base spline.

  jmat <- inprod(basisfd, basisfd)

# zmat contiene la traspuesta de la matriz, adjuntado por renglones el vector one (1*6) a la matriz
# formada de multiplicar jmat (22*22) con coef (22*6) es decir  $Z = [1 \ CJ]$ , donde C es la matriz de
# coeficientes respecto a la base de las observaciones.

  zmat <- t(rbind(one, jmat %*% coef))

# bcoef es una matriz de ceros con nbasis (22) renglones y nvar (1) columnas.

  bcoef <- matrix(0, nbasis, nvar)

# alpha es un vector de longitud nvar (1), que contiene ceros.

  alpha <- rep(0, nvar)

# index es un vector equiespaciado que contiene la secuencia que inicia en 2 y termina en nbasis+1 (23),
# con incrementos de uno.

  index <- 2:(nbasis+1)

# Si el valor de xlambda es negativo o cero y el número de funciones base mas uno excede el
# número de observaciones, no es posible realizar el análisis y la subrutina termina.

  if (xlambda <= 0)
  {
    # no smoothing required, do ordinary least squares
    if (ncol(zmat) > nrow(zmat))
      stop(paste("For XLAMBDA = 0, no. of basis functions exceeds",
                 "number of replications. No fit possible.))

# Si el valor de xlambda es negativo o cero y los pesos varían (wtvar es verdadero), multiplica las
# columnas de zmat por wtvec elemento a elemento y lo guarda en temp, el producto cruzado de la
# matriz temp con zmat lo asigna a la matriz que llama Cmat y el producto cruzado de temp con ymat lo
# asigna a Dmat

    if (wtvar)
    {
      temp <- sweep(zmat, 2, wtvec, "*")
      Cmat <- crossprod(temp, zmat)
      Dmat <- crossprod(temp, ymat)
    }

# Si el valor de xlambda es negativo o cero y los pesos no varían (wtvar es falso). Cmat contiene el
# producto cruzado de zmat con zmat y Dmat el producto cruzado de zmat con ymat

```

```

else
{
.

# Cmat es la matriz  $Z^T Z$ 

Cmat <- crossprod(zmat)

# Dmat es  $Z^T y$ 
Dmat <- crossprod(zmat, ymat)
}
}

# Si xlambda es positivo, se requiere suavizamiento, en el ejemplo xlambda es 0.000005 por lo que este
# es el caso

else
{
# smoothing required

# kmat es una matriz de ceros, en este ejemplo de 23*23 (número de funciones base más uno)

kmat <- matrix(0, nbasis+1, nbasis+1)

# Asigna a la submatriz de kmat, formada de omitir su primer renglón y su primer columna, el producto
# interno de la derivada de orden xLfd de las funciones base, en este caso xLfd se tomo el valor de default
# 2, entonces esta matriz tiene por entradas:  $K_{jk} = \int D^2 \phi_j(s) D^2 \phi_k(s) ds = \langle D^2 \phi_j, D^2 \phi_k \rangle$ .

kmat[index, index] <- inprod(basisfd, basisfd, xLfd, xLfd)

# Si los pesos varían (wtvar verdadero), multiplica las columnas de zmat por wtvec elemento a
# elemento y lo guarda en temp, el producto cruzado de la matriz temp con zmat mas el producto de la
# constante lambda con la matriz kmat lo asigna Cmat y el producto cruzado de temp con ymat lo asigna
# a Dmat

if (wtvar)
{
temp <- sweep(zmat, 2, wtvec, "*")

# Cmat es la matriz  $Z^T Z + \lambda K_0$ , donde  $K_0$  es kmat.

Cmat <- crossprod(temp, zmat) + xlambda*kmat

# Dmat es  $Z^T y$ 

Dmat <- crossprod(temp, ymat)
}

# Si los pesos no varían (wtvar es falso). Cmat contiene el producto cruzado de zmat con zmat mas el
# producto de la constante lambda con la matriz kmat y Dmat el producto cruzado de zmat con ymat.

else
{

# Cmat es la matriz  $Z^T Z + \lambda K_0$ , donde  $K_0$  es kmat.

```

```

Cmat <- crossprod(zmat) + xlambda*kmat

# Dmat es  $Z^T y$ 
Dmat <- crossprod(zmat, ymat)
}

# Resuelve el sistema  $(Z^T Z + \lambda K_0) \hat{\zeta} = Z^T y$  para  $\hat{\zeta} = (\alpha, b)^T$  llamado temp.
temp <- symsolve( Cmat, Dmat )

# Asigna a yhat el producto  $Z \hat{\zeta}$ .
yhat <- zmat %*% temp

# bcoef contiene los coeficientes respecto a la base de b, de temp  $\hat{\zeta} = (\alpha, b)^T$ 
bcoef <- as.matrix(temp[index,])

# alpha contiene el valor  $\alpha$ , del vector  $\hat{\zeta} = (\alpha, b)^T$ 
alpha <- temp[1,]

# obtiene los nombres del objeto funcional xfd (que contiene los perfiles de volumen) dado
# inicialmente
fdnames <- getnames(xfd)

# Asigna nombres a los coeficientes estimados de regresión.
regfdnames <- fdnames
regfdnames[[2]] <- paste('Reg. Coef.', 1:nvar)
regfdnames[[3]] <- 'Reg. Coef.'
names(regfdnames)[2] <- 'Reg. Coefficients'

# Crea un objeto funcional a partir de los coeficientes b y la base usada, con los nombres
# anteriores.
regfd <- create.fd(bcoef, basisfd, regfdnames)

# Crea una lista que contiene el intercepto  $\alpha$ , las funciones de regresión  $\beta$ , las funciones de
# observaciones estimadas  $\hat{y}$ .
linmodlist <- list(alpha, regfd, yhat)

# Asigna nombres a los elementos de esta lista
names(linmodlist) <- c('alpha', 'regfd', 'yhat')
# Regresa la lista linmodlist
return( linmodlist )
}

```

7

Conclusiones

El uso de Análisis Funcional de Datos fué una herramienta de gran utilidad y practicidad en el análisis de las aplicaciones, ya que no requiere de una gran cantidad de supuestos a diferencia de otros enfoques. Un objetivo alcanzado en esta tesis era utilizar los modelos funcionales lineales en problemas reales, debido a sus características, éstos se ajustaron de manera muy natural a las aplicaciones encontradas. En particular se consideraron dos de los tres casos del modelo funcional lineal por el tipo de datos proporcionados, sin embargo consideramos que una aplicación al modelo funcional en el caso 3 mejoraría la presente tesis.

Creemos que es importante estudiar esta teoría ya que una gran cantidad de problemas que podrían ser tratados mediante Análisis Funcional se presentan continuamente, pero debido a que es una herramienta relativamente nueva se usan enfoques que en ocasiones resultan ser más complicados.

La prueba por permutación fué otra herramienta de la que no teníamos conocimiento y nos ayudó al análisis de una de las aplicaciones, además de proporcionarnos una nueva técnica en la prueba de hipótesis.

Respecto al análisis de las aplicaciones, en el caso de GERBER los resultados confirmaron las suposiciones para la mejora del producto hechas por la empresa, pero además se encontraron otros factores que en conjunto contribuyen de manera significativa a dicha mejora. En el caso del experimento con ajo los resultados indicaron de manera clara que al

almacenarlos a $0^{\circ}C$ tendrán un mejor resultado al final del periodo de almacenamiento que al almacenarlos de manera tradicional , es decir a temperatura ambiente.

Bibliografía

- [1] Ramsay, J. O., Silverman, B. W. (1997). *Functional Data Analysis*. Springer Verlag, New York.
- [2] Ramsay, J. O., Silverman, B. W. (2002). *Applied Functional Data Analysis. Methods and Case Studies*. Springer Series in Statistics.
- [3] Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Springer.
- [4] Wu, C. F. J. & Hamada, M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley & Sons. INC.
- [5] Good, P. (2000) *Permutation Test: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer. Second Edition.
- [6] Castaño, E. T., y Domínguez, J. D. (2004). *Experimentos: Estrategia y Análisis en Ciencia y Tecnología*. CIMAT
- [7] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman & Hal/CRC.
- [8] León, F. G. (2004). *Tesis: "Estudio estadístico sobre la calidad del ajo ante diferentes condiciones de almacenamiento"*. Maestría en Ingeniería Industrial ITQ
- [9] Zuñiga, M. C. G. (2004). *Tesis: "Identificación y evaluación de los factores que influyen en la separación de fases en un jugo de frutas con yogurt"*. PROPAC-UAQ
- [10] Nichols, T. E. and Holmes, A. P. (2001). *Nonparametric permutation test for functional neuroimaging: a primer with examples*. *Human Brain Mapping*, 15, pp 1-25.
- [11] Loughin, T. M. and Noble, W. (1997) *A Permutation Test for Effects in an Unreplicated Factorial Design*. *Technometrics*, Vol. 39. N°2
- [12] S-Plus 6 for Windows User's guide, Insightful Corporation, Seattle, WA.
- [13] <http://ego.psych.mcgill.ca/misc/fda>
- [14] http://www.udc.es/dep/mate/estadistica2/estadistica_2.htm