

REGRESIÓN FUNCIONAL MEDIANTE BASES OBTENIDAS POR DESCOMPOSICIÓN ESPECTRAL DEL OPERADOR COVARIANZA

Ramírez John¹

Resumen. En este artículo se estudia el modelo de regresión lineal funcional en el cual las variables explicativas son puntos muestrales de un proceso continuo en el tiempo es decir funciones aleatorias observadas sobre un subconjunto compacto y Y es la variable de respuesta escalar. Se proponen dos estimadores para la regresión: por medio del método de componentes principales funcionales (PCA) y por medio del método de mínimos cuadrados parciales (PLS). El objetivo es comparar los coeficientes de determinación (R^2) obtenidos por los métodos anteriormente mencionados.

Palabras claves: Datos funcionales, modelo lineal. Análisis de componentes principales funcionales.

Abstract. In this paper, we study a functional linear regression model in which explanatory variables are sampling points of a continuous-time process namely random function observed on a compact subset of \mathbb{R} , and Y is scalar response variable. We propose two estimators of regression: by means of a functional PCA method and by functional PLS method. Our aim is to compare coefficients of determination (R^2) obtained by methods mentioned above.

Keywords: functional data, linear model, functional principal component analysis.

Recibido: Agosto 2014.

Aceptado: Septiembre 2014.

1. INTRODUCCIÓN

El rápido avance en las tecnologías de cálculo numérico, tanto en velocidad como en capacidad de almacenamiento, permiten manejar grandes bases de datos y procesarlas rápidamente. En muchos casos los datos consisten de observaciones de una variable aleatoria tomadas a lo largo de un intervalo o en particiones cada vez más finas de este intervalo que puede ser temporal o espacial. Si la partición es lo suficientemente fina podemos considerar que la muestra corresponde a una observación de una variable aleatoria definida en un determinado espacio funcional. Si analizamos esta clase de datos con las técnicas estándar del análisis multivariante se presentan problemas tales como multicolinealidad y alta dimensionalidad. Es necesario un cambio de paradigma: analizar los datos desde un enfoque funcional.

Este hecho ha convertido al Análisis de Datos Funcionales (ADF) en uno de los más activos campos de la estadística en los últimos años. Desde los trabajos iniciales de Ramsay y Silverman en los ochenta y Bosq en los noventa el ADF se ha desarrollado en forma exponencial, tanto teóricamente como en aplicaciones en diferentes campos como medio ambiente y climatología, reconocimiento de imágenes, economía, genética, etc., etc.

Debido a que el ADF es una disciplina emergente, existe un amplio rango de líneas de investigación a ser exploradas que van desde la extensión de métodos multivariados clásicos al contexto fun-

cional hasta la creación de técnicas estadísticas dedicadas exclusivamente al manejo de datos funcionales.

A pesar de la naturaleza funcional de los datos, los objetivos perseguidos son los mismos que pretende el análisis multivariante:

Pre proceso de datos: registro y alineación, técnicas de alisamiento, etc.

Representación y exploración: resaltar las características más importantes, medidas de centralización y dispersión, detección de datos atípicos.

Clasificación: construcción de modelos que expliquen la relación entre las variables funcionales.

Inferencia: intervalos de confianza, prueba de hipótesis.

2. DATOS FUNCIONALES

¿Qué son los datos funcionales?

En muchos campos de las ciencias aplicadas, el desarrollo de nuevas tecnologías ha posibilitado el monitoreo de muchas magnitudes de forma continua o a cortos intervalos produciendo datos que directamente se pueden considerar como curvas y no escalares o vectores, por lo que el tratamiento funcional es crucial.

El enfoque funcional permite estudiar e interpretar la conducta global del proceso estocástico, algo que no se puede alcanzar considerándolo como la realización de un vector aleatorio. Sin embargo, no debe olvidarse la naturaleza discreta de los conjuntos de datos.

Aunque la partición donde los datos son tomados es cada vez más fina, el efecto de la discretización no se puede ignorar, [8].

¹ Ramírez Jhon, M.Sc., Profesor, Departamento de Matemáticas, Facultad de Ciencias Naturales y Matemáticas, ESPOL. (Email: jramirez@espol.edu.ec).

La idea principal en ADF es considerar que una variable aleatoria X que es observada en una partición $\{t_i, i = 1, \dots, L\}$ del conjunto T , es decir $\{X(t_i)\}_{i=1}^L$ es una observación de la variable aleatoria funcional:

$$X = \{X(t) / t \in T\} \quad (2.1)$$

A menudo se considera que $T = [T_0, T_1] \subset \mathbb{R}$ y por tanto cada observación de X es una curva. Sin pérdida de generalidad supondremos que T es un intervalo temporal, aunque T puede ser un subconjunto de \mathbb{R}^n .

Es decir, dado un espacio probabilístico (Ω, \mathcal{A}, P) la variable aleatoria X es un proceso estocástico y por tanto una función medible de la forma:

$$X: \Omega \rightarrow S$$

$$\omega \mapsto X(\omega, t) = X_\omega(t) = X(t) \quad (2.2)$$

Donde S es un espacio funcional. Se puede decir que X es una variable aleatoria S -valuada

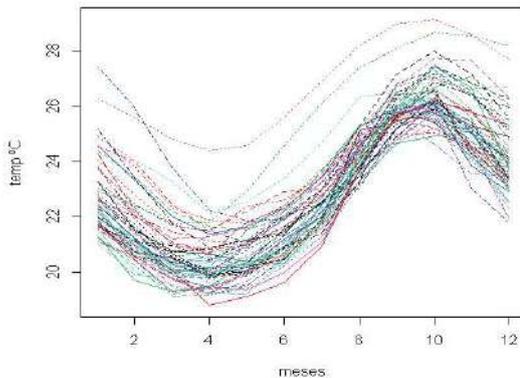
Dada una variable aleatoria funcional X y $n \in \mathbb{N}$, una muestra aleatoria funcional (o una muestra) de X es un conjunto $\{X_i(t) / t \in T\}_{i=1}^n$ (o $\{X_i\}_{i=1}^n$) de variables aleatorias funcionales independientes e idénticamente distribuidas con la misma distribución de X . A cada observación se le denomina dato funcional.

A lo largo de este documento vamos a suponer que $S = (\mathcal{H}, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert real y separable, esto último implica que existe una base ortonormal numerable.

Ejemplos:
Temperatura de la superficie del mar.
 Temperatura promedio mensual de la superficie del mar, desde junio 1950 a mayo de 2004.

Fuente: <http://www.cpc.ncep.noaa.gov/data/indices/>

Figura 1
 Regresión funcional mediante bases obtenidas por descomposición espectral del operador covarianza
 Niño 3-4

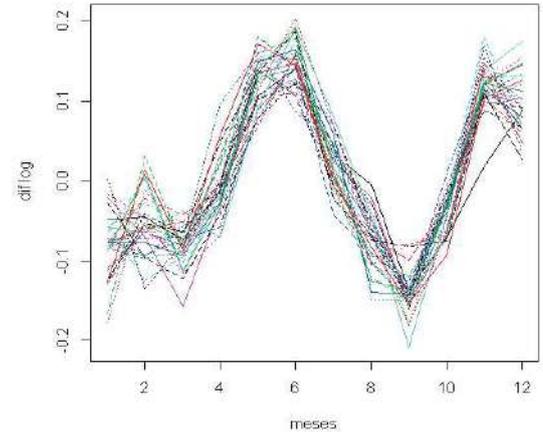


Consumo eléctrico.

Consumo de electricidad en el sector residencial y comercial de los E.E.U.U de enero de 1973 a diciembre 2001.

Fuente <http://www.economagic.com>

Figura 2
 Regresión funcional mediante bases obtenidas por descomposición espectral del operador covarianza
 Consumo Eléctrico

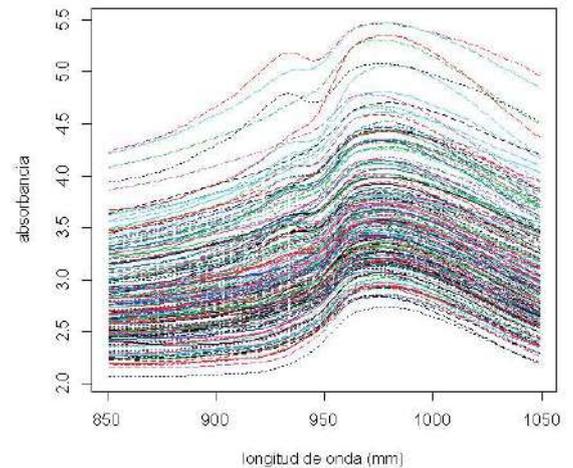


Espectrometría de absorción.

Para cada unidad de una muestra de 250 trozos de carne se observa la denominada curva espectrométrica que corresponde a la absorbancia medida en cien distintas longitudes de onda.

Fuente: <http://lib.stat.cmu.edu/datasets/tecat>

Figura 3
 Regresión funcional mediante bases obtenidas por descomposición espectral del operador covarianza
 Curvas Espectrométricas



Pre procesamiento de datos.

Desde un punto de vista práctico sólo podemos observar versiones discretas de los datos funcionales. Razón por la cual debemos utilizar algún procedimiento de interpolación/ajuste para obtener datos funcionales continuos. Estos procedimientos están ideados tanto para suprimir

o minimizar el ruido inherente a los datos como para expresarlos en forma de funciones continuas.

3. SUAVIZADO DE DATOS

Los procedimientos de suavizado incluyen técnicas no paramétricas como representación por bases (Fourier, splines, wavelets), núcleos y técnicas paramétricas como los modelos lineales. Muchas de estas técnicas se basan en algunas condiciones de suavidad, como la existencia de una o más derivadas.

3.1. Suavizado Lineal: Kernels

Suavizado por funciones base

Este enfoque de suavizado consiste en la aproximación de X por medio de su representación como combinación lineal de funciones base. Un sistema de funciones bases es un conjunto de funciones conocidas $\{\phi_j\}_{j=1}^{\infty}$ linealmente independientes tal que cualquier función X se puede aproximar tanto como se desee tomando combinaciones lineales de una longitud J lo suficientemente grande de funciones ϕ_j . Un sistema de funciones base permite trasladar la información de X desde un espacio infinito dimensional a un espacio de dimensión finita J generado por $\{\phi_j\}_{j=1}^J$. De aquí que cualquier X se puede aproximar por:

$$X^{[J]}(t) = \sum_{j=1}^J c_j \phi_j(t) \quad (3.1)$$

Es decir: $X^{[J]} = \Phi' \mathbf{C}$ donde $\mathbf{C} = (c_1, \dots, c_J)'$ y $\Phi = (\phi_1, \dots, \phi_L)'$

Sin perder generalidad vamos a considerar que la variable aleatoria funcional toma valores en $L^2[T_0, T_1]$. También vamos a suponer que todas las realizaciones del proceso estocástico son observadas en los mismos instantes t_j con $j = 1, \dots, L$. Las observaciones disponibles son: $\{X_i(t_j) / i = 1, \dots, I; j = 1, \dots, L\}$ que se pueden almacenar en una matriz de tamaño $I \times L$. Como dijimos más arriba, las observaciones están dadas en forma discreta y a partir de ellas debemos construir el objeto funcional, de tal forma que se lo pueda evaluar en cualquier instante de tiempo entre T_0 y T_1 .

Esta construcción se la hace por interpolación o por alisamiento, eligiendo un conjunto de funciones base apropiadas para el problema que

se tiene en mano. Las funciones muestrales pueden expresarse como:

$$X_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t) \quad (3.2)$$

donde $\{\phi_k(t)\}$ son las funciones base. La conversión de los datos discretos en datos funcionales reside en la estimación de los coeficientes o ponderadores c_{ik} .

Las bases más utilizadas son: Fourier, B-spline, wavelet.

Se prefiere la base de Fourier cuando el proceso es periódico o cuasi periódico. Mientras que la base de B-splines se la prefiere cuando no existe periodicidad manifiesta en los datos. La base de wavelets se prefiere cuando las curvas son discontinuas. El desarrollo teórico de esta base y otras lo encontramos en [5, 9].

Una base se especifica mediante cuatro elementos: El tipo de base, el rango de valores del argumento $[T_0, T_1]$, el número de funciones bases a tomarse en cuenta y los parámetros de la base a utilizarse.

La base de Fourier para datos periódicos consiste de las funciones:

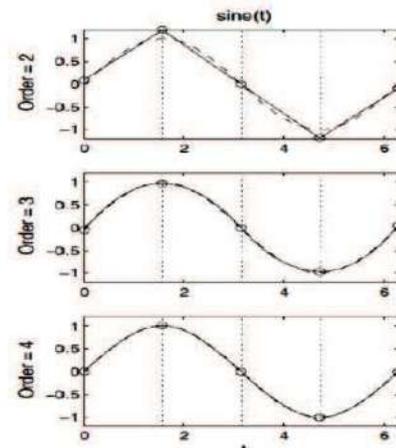
$$1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \dots$$

Donde $\omega = 2\pi/S$ y S es el periodo.

Por ejemplo, tenemos que $S=12$ si los datos son observados mensualmente con variación periódica anual.

La base B-spline consiste de funciones polinomiales por tramos y requiere especificar el orden O del spline y los nodos, que a menudo se los escoge como los puntos en los que se evaluaron los datos muestrales. El orden del spline, $O=2$ corresponde a segmentos de recta, $O=3$ a segmentos de parábolas, etc.

Figura 4
Ejemplos de Splines



Sea x_{ij} la observación de la i -ésima función al tiempo t_j . La estimación de los coeficientes en la representación de $x_i(t)$ por medio de una base (cualquiera que sea la elegida) se la realiza por medio de mínimos cuadrados, minimizando para cada i :

$$\sum_{j=1}^L \left(x_{ij} - \sum_{k=1}^K c_{ik} \phi_k(t) \right)^2 \quad (3.3)$$

Con respecto a los coeficientes c_{ik} , con $K \leq L$.

Por otro lado, dado un dato funcional de la forma

$$x_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t), \quad \text{podemos suavizarlo}$$

minimizando, por el método de los mínimos cuadrados, añadiendo un factor que penalice la rugosidad o aspereza. Para cada i hay que minimizar:

$$\int_{T_0}^{T_1} (x_i(t) - \tilde{x}_i(t))^2 dt + \lambda g(\tilde{x}_i(t)) \quad (3.4)$$

Donde g es una medida de la rugosidad y λ es el peso asignado a la suavidad de la curva. A mayor λ más suavidad. Ver [9].

Las medidas de rugosidad se basan en la segunda derivada. Aquí se utiliza la denominada *curvatura total* de x :

$$g(x) = \int_{T_0}^{T_1} (x''(t))^2 dt \quad (3.5)$$

Al representar a un dato como una función podemos calcular sus derivadas, éstas pueden ser erráticas si el dato funcional no es lo suficientemente suave.

4. MÉTRICAS Y SEMI-MÉTRICAS

Hay una pregunta que tarde o temprano aparecen en ADF:

¿Cómo medir la distancia entre dos datos funcionales?

Una medida que permita cuantificar la separación o cercanía de dos datos funcionales es fundamental en el análisis estadístico, en especial para tener una cuantificación de la dispersión de los datos.

La elección de la métrica del espacio funcional debe ser consecuente con el tipo de dato disponible. Por ejemplo, hay que tener en cuenta si la información de los datos funcionales está relacionada con en la escala o con la forma de ellos. Así, si la escala o amplitud de los datos es la importante, la norma usual de $L^2[a,b]$ será suficiente, pero si la información relevante depende de la forma de la función, debemos utilizar una medida que tome en cuenta la segunda derivada de los datos funcionales.

Las medidas de proximidad desempeñan un papel fundamental en estadística. En el caso finito, por ejemplo \mathbb{R}^n , todas las normas son equivalentes, en el sentido de que definen métricas equivalentes y por tanto topologías equivalentes. Esto implica que la elección de una u otra norma no será crítica en cuanto a continuidad, convergencia o separabilidad del espacio. En contraste, si intervendrá en los aspectos computacionales del cálculo de distancias. Por ejemplo, la norma euclídea tiene asociada un producto interno y por tanto una geometría subyacente a éste. En cambio, si la norma no proviene de un producto interno sólo se tendrán asociadas nociones de cercanía más no geométricas.

Sin embargo, en el enfoque funcional no se da la equivalencia entre las métricas, y los problemas deben ser abordados en forma distinta según la métrica elegida. En este sentido, es posible que la elección de cierto espacio métrico sea insuficiente y sea ventajoso utilizar espacios semi-métricos, y por tanto semi-métricas como las que se describen a continuación.

4.1. SEMI-MÉTRICAS BASADAS EN FPCA

El FPCA (Functional Principal Component Analysis) es la adaptación del Análisis de Componentes Principales al contexto funcional, ver [9]. Dada una variable aleatoria funcional el FPCA nos proporciona la expansión:

$$X = \sum_{j=1}^{\infty} \left(\int_{T_0}^{T_1} X(t) v_j(t) dt \right) v_j \quad (4.1)$$

donde $\{v_j\}$ es la familia ortogonal de funciones propias del operador Γ , la función de covarianzas del proceso.

Este desarrollo se puede truncar en un valor k (o nivel de resolución). El valor k se elige en función de la cantidad de información contenida en la k primeras componentes principales funcionales. Si esta información retenida en las primeras k componentes es suficientemente alta (sobre el 75%), la serie truncada representará una buena aproximación de la variable aleatoria X :

$$X^{(k)} = \sum_{j=1}^k \left(\int_{T_0}^{T_1} X(t) v_j(t) dt \right) v_j \quad (4.2)$$

Nota.- $X^{(k)}$ minimiza

$$E \left[\int_{T_0}^{T_1} (X(t) - P_k X(t))^2 dt \right] \text{ sobre todas las}$$

proyecciones P_k de X en el espacio k -dimensional formado por las primeras k componentes principales funcionales.

Ferraty y Vieu (ver [5]) proponen la siguiente semi-métrica, basada en la norma de $L^2[T_0, T_1]$:

$$d_k^{PCA}(X_i, x) = \sqrt{\sum_{j=1}^k \left(\int_{T_0}^{T_1} (X_i(t) - x(t)) v_j(t) dt \right)^2} \quad (4.3)$$

En la práctica, las funciones propias de Γ son desconocidas, así que se las reemplaza por sus estimaciones \hat{v}_j obtenidas de la función de covarianza empírica. Además, las funciones son observadas en una partición $\{t_l\}_{l=1}^L$ y por tanto la integral que aparece en la semi-métrica se puede reemplazar por una aproximación:

$$\tilde{d}_k^{PCA}(X_i, x) = \sqrt{\sum_{j=1}^k \left(\sum_{l=1}^L w_l (X_i(t_l) - x(t_l)) \hat{v}_j(t_l) \right)^2} \quad (4.4)$$

Donde $\{w_l\}_{l=1}^L$ son los pesos de la cuadratura al aproximar la integral.

4.2. SEMI-MÉTRICAS BASADAS EN PLSR

En ciertas ocasiones se observan simultáneamente dos tipos de variables: variables respuesta y variables regresoras o independientes. En este caso es aconsejable utilizar la familia de semi-métricas obtenida por medio de la regresión de mínimos cuadrados parciales (Partial Least Squares Regression, PLSR), ver [8].

PLSR es un método estadístico utilizado cuando la variable respuesta como el regresor son multivariantes. Fue desarrollado para predecir variables respuesta a partir de regresores cuando existe un alto grado de colinealidad entre éstos y/o cuando su número es bastante grande en comparación con el número de observaciones (que es el caso de los datos funcionales discretizados).

PLSR consiste en obtener una descomposición simultánea de las variables dependientes e independientes en componentes ortogonales tal que la covarianza entre los dos conjuntos de componentes obtenidos sea máxima. PLSR nos proporciona p componentes o factores cada uno de éstos correspondiendo a una respuesta. A mayor número de factores tomados en cuenta se aumenta la precisión en la representación de una función. Sin embargo, tomando una gran cantidad de factores aumenta la variabilidad (puesto que aumenta el ruido contenido en cada componente). El número de factores a considerarse debe ser un valor de equilibrio entre la precisión y la variabilidad. El número de factores desempeña un rol parecido a la dimensión k en el análisis de componentes principales

Aunque exista cierta similitud entre ambos métodos, hay que tener presente que las componentes obtenidas por PLS explican la variabilidad entre variables respuesta y regresores, mientras que el análisis de componentes principales sólo toma en cuenta la variabilidad de los regresores.

Al igual que en el análisis de componentes principales, PLSR puede ser útil para diferentes propósitos, en particular para el manejo de datos funcionales. PLSR permite la construcción de una familia de semi-métricas de la siguiente forma:

Sean $\{\hat{v}_j^q\}_{j=1}^p$ las funciones obtenidas por PLSR,

donde q es el número de factores y p es el número de variables respuesta. Se define la familia de semi-métricas:

$$\tilde{d}_q^{PLS}(X_i, x) = \sqrt{\sum_{j=1}^p \left(\sum_{l=1}^L w_l (X_i(t_l) - x(t_l)) \hat{v}_j^q(t_l) \right)^2} \quad (4.5)$$

Estas dos familias de semi-métricas no requieren que los datos funcionales verifiquen condiciones de suavidad, pueden ser aplicadas a datos ásperos y rugosos. Sin embargo, se deben satisfacer dos condiciones: los datos deben estar balanceados y la partición debe ser lo suficientemente fina.

4.3. SEMI-MÉTRICAS BASADAS EN DERIVADAS

Cuando los datos funcionales satisfacen ciertas condiciones de regularidad, podemos utilizar la familia de semi-métricas que utiliza las derivadas de los datos funcionales:

$$d_k^{deriv}(X_i, x) = \sqrt{\int_{T_0}^{T_1} (X_i^{(k)}(t) - x^{(k)}(t))^2 dt} \quad (4.6)$$

Para evitar la inestabilidad numérica en el cálculo de las derivadas, los datos funcionales se aproximan mediante B-splines².

Observación: d_0^{deriv} es la norma usual en $L^2[T_0, T_1]$.

5. ESTADÍSTICAS DESCRIPTIVAS

Muchos métodos utilizados en ADF son generalizaciones de los procedimientos multivariados clásicos, otros son nuevas propuestas nacidas en el análisis funcional. Sin embargo, la falta de una representación para la función de densidad en el ámbito funcional causa

² Ver [3].

problemas cuando se exploran los datos funcionales. Este inconveniente se debe a que aún no existe una medida funcional que desempeñe el mismo rol que la medida de Lebesgue en el contexto multivariado.

5.1. MEDIDAS DE POSICIÓN

La media.

En el análisis de datos clásico la medida de tendencia central por excelencia es la media. En el contexto funcional las variables aleatorias pertenecen al espacio $L^2[T_0, T_1]$ y la media funcional se define por:

$$(E(X))(t) = E(X(t)) \quad \forall t \in [T_0, T_1] \quad (5.1)$$

Su estimador es:

$$\bar{X}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t) \quad \forall t \in [T_0, T_1] \quad (5.2)$$

Que es el promedio punto a punto de los elementos de la muestra. Este estimador también es conocido como estimador natural.

Actualmente hay varias líneas de investigación tendientes a obtener una medida de centralización desde una óptica estrictamente funcional. Por ejemplo, la planteada por Munk y Dette (1998) basada variedades de Hilbert ver [7, 10].

La mediana.

Una medida de tendencia central alternativa es la mediana funcional. En estos últimos años se han propuesto varias definiciones para la mediana de una variable aleatoria que toma valores en un espacio funcional basadas en medidas de profundidad.

Una medida de profundidad funcional nos permite ordenar un conjunto de datos funcionales desde el centro hacia la periferia: desde las curvas más profundas (aquellas que tienen la profundidad más alta) hasta las curvas más exteriores (con profundidad mínima). La mediana funcional es la curva o dato funcional con la mayor profundidad. La noción de profundidad apareció en el contexto multivariado para generar estadísticos de orden y listas ordenadas de datos de alta dimensionalidad. Los estadísticos de orden son necesarios cuando se emplean técnicas no paramétricas, sobre todo cuando la distribución de los datos no es normal. Es importante disponer de procedimientos que permitan la ordenación de puntos en un espacio n-dimensional y que consideren la geometría de la nube de datos. Por ejemplo, si calculamos la mediana coordinada a coordinada, el vector de medianas no toma en cuenta la geometría de los datos. Por otro lado, cuando tenemos datos univariados el orden natural de los mismos nos da una ordenación o ranking de las observaciones. Al tener datos multidimensionales el orden natural

desaparece y se hace necesario como primer paso definir un centro y a partir de éste ordenar las observaciones de dentro hacia afuera.

Fraiman y Muniz (2001)³ diseñaron una medida de profundidad basada en la forma natural en que aparecen los estadísticos de orden. Sin pérdida de generalidad se consideran funciones definidas en el intervalo $[0, 1]$: $\forall t \in [0, 1]$ si $F_{n,t}$ es la función de distribución empírica de $x_1(t), \dots, x_n(t)$, se nota la profundidad de cada dato $x_i(t)$ por:

$$D_i(t) = 1 - \left| \frac{1}{2} - F_{n,t}(x_i(t)) \right| \quad (5.3)$$

y para cada i se define $I_i = \int_0^1 D_i(t) dt$

En función de este índice de profundidad se pueden ordenar los datos.

5.2. MEDIDAS DE DISPERSIÓN

Covarianza.

Se supone que la variable aleatoria funcional X toma valores en $L^2[T_0, T_1]$ y que $E(\|X\|^2) < \infty$. La dependencia entre los diferentes registros de los datos funcionales a lo largo del tiempo se puede capturar por medio de la función covarianza:

$$\begin{aligned} \Gamma_X(t_1, t_2) &= E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))] \quad \forall t_1, t_2 \in [T_0, T_1] \\ &= E[X(t_1)X(t_2)] - \mu_X(t_1)\mu_X(t_2) \end{aligned} \quad (5.4)$$

En la práctica se utiliza la versión empírica:

$$\hat{\Gamma}_X(t_1, t_2) = \frac{1}{n} \sum_{i=1}^n (X_i(t_1)X_i(t_2) - \bar{X}(t_1)\bar{X}(t_2)) \quad \forall t_1, t_2 \in [T_0, T_1] \quad (5.5)$$

6. MODELOS DE REGRESIÓN FUNCIONAL

Los modelos de regresión son modelos matemáticos utilizados para estudiar la relación entre variables. Se pueden distinguir distintos tipos de modelos de regresión en función de las características de las variables regresoras y de la variable respuesta, pero el problema siempre es el mismo: modelar la variable respuesta a partir de las variables regresoras. Estos modelos se adecúan a una amplia variedad de escenarios y son aplicados ampliamente en todas las áreas de las ciencias exactas o no.

³ Fraiman, R. & Muniz, G. (2001), "Trimmed means for functional data".

Un modelo de regresión es de la forma

$$Y = m(X) + \varepsilon \quad (6.1)$$

Donde la variabilidad de la variable aleatoria Y (variable respuesta o dependiente) es explicada por la variable X (regresor o variable independiente). La función m se denomina función de regresión y ε es una variable aleatoria llamada error. En la siguiente tabla se muestran las cuatro posibles situaciones de acuerdo a la naturaleza de los datos:

	$Y \in \mathbb{R}^p$	$Y \in \mathcal{H}$
$X \in \mathbb{R}^p$	Caso A: $m: \mathbb{R}^p \rightarrow \mathbb{R}^p$	Caso B: $m: \mathbb{R}^p \rightarrow \mathcal{H}$
$X \in \mathcal{H}$	Caso C: $m: \mathcal{H} \rightarrow \mathbb{R}^p$	Caso D: $m: \mathcal{H} \rightarrow \mathcal{H}$

En el presente artículo vamos a concretarnos en el caso **C**: modelo de regresión funcional con respuesta escalar, con $p = 1$ y $m: \mathcal{H} \rightarrow \mathbb{R}$

Recordemos que $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert real y separable y que $\|\cdot\|$ es la norma inducida por el producto escalar $\langle \cdot, \cdot \rangle$.

El modelo de regresión funcional con respuesta escalar se puede expresar como:

$$Y = \langle \theta, X \rangle + \varepsilon \quad (6.2)$$

donde Y es una variable aleatoria real y centrada, $m(\cdot) = \langle \theta, \cdot \rangle$ es el denominado operador de regresión lineal con $\theta \in \mathcal{H}$ y $\|\theta\|^2 < \infty$, X es una variable aleatoria centrada \square -valuada tal que $E(\|\theta\|^2) < \infty$, y ε es una variable aleatoria real que satisface $E[\varepsilon] = 0$, $Var[\varepsilon] = \sigma^2 < \infty$, y $E[\varepsilon X] = 0$.

El modelo dado en (6.2) es un modelo paramétrico, la función $\theta(t)$ es el parámetro.

Si las variables aleatorias X y Y no son centradas, procedemos a centrarlas:

$$\tilde{Y} = Y - E[Y], \quad \tilde{X} = X - E[X] \quad (6.3)$$

El modelo con las variables centradas $\tilde{Y} = \langle \theta, \tilde{X} \rangle + \varepsilon$ es equivalente al modelo

$$Y = \theta_0 + \langle \theta, X \rangle + \varepsilon \quad (6.4)$$

con $\theta_0 = E[Y] - \langle \theta, E[X] \rangle$, de aquí que, sin perder generalidad vamos a considerar que las variables están centradas.

Aunque estos modelos pueden ser construidos considerando la estructura $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, los datos

funcionales son valuados en espacios habituales tal como $L^2[T_0, T_1]$, en este caso utilizando el producto interno usual en L^2 , el modelo (6.2) se puede expresar como:

$$Y = \int_{T_0}^{T_1} \theta(t) X(t) dt + \varepsilon \quad (6.5)$$

El problema que surge en este punto es el de como estimar el parámetro del modelo θ a partir de una muestra $\{(X_i, Y_i)\}_{i=1}^n$ de (X, Y) . En la

práctica no se dispone de las funciones X_i sino de sus versiones discretas $(X_i(t_1), X_i(t_2), \dots, X_i(t_L))$, por lo que la muestra es de la forma

$\{(X_i(t_1), X_i(t_2), \dots, X_i(t_L), Y_i)\}_{i=1}^n$ donde

$\{t_l\}_{l=1}^L$ es la partición en la cual se han tomado los valores de cada dato funcional X_i .

Como un primer intento (ingenuo) de estimar θ podemos olvidarnos de la naturaleza funcional de X y considerar que cada X_i es un vector en \mathbb{R}^L .

Esto nos faculta a utilizar el modelo de regresión lineal clásico: $Y = X\theta + \varepsilon$ donde:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_1(t_1) & \dots & X_1(t_L) \\ \vdots & \ddots & \vdots \\ X_n(t_1) & \dots & X_n(t_L) \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta(t_1) \\ \vdots \\ \theta(t_L) \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (6.6)$$

De donde:

$$(X'X)\hat{\theta} = X'Y \quad (6.7)$$

Es decir, tenemos que resolver un sistema de n ecuaciones con L incógnitas. Puesto que $L \gg n$ el sistema tiene solución paramétrica lo que implica que existen infinitas soluciones que ajustan los datos perfectamente. La estimación de $\theta = (\theta(t_1), \theta(t_2), \dots, \theta(t_L))'$ puede resultar en extremo variable, obteniéndose un estimador con poco o ningún significado práctico.

Otro intento (ingenuo): consideremos que tanto X como θ son funciones en $L^2[T_0, T_1]$, podemos

encontrar una función $\hat{\theta}$ que sea continua y produzca un ajuste perfecto de (6.5). Sin embargo, $\hat{\theta}$ sería una función bastante variable sin información útil.

Se tiene que el espacio de funciones que satisfacen (6.5) es de dimensión infinita, razón por la que hay que imponer ciertas condiciones de regularidad para obtener un estimador consistente y útil.

6.1. REGULARIZACIÓN MEDIANTE BASES DE FUNCIONES

Un método de regularización bastante extendido es el de truncar la base del espacio $L^2[T_0, T_1]$ ⁴:

Sea $\{\phi_j\}_{j=1}^{\infty}$ un sistema de funciones base en el espacio $L^2[T_0, T_1]$, el parámetro θ se puede representar como

$$\theta^{[J_\theta]} = \sum_{j=1}^{J_\theta} b_j \phi_j = \Phi' \mathbf{B} \quad (6.8)$$

Con $\mathbf{B} = (b_1, \dots, b_{J_\theta})'$ el vector de coeficientes y $\Phi = (\phi_1, \dots, \phi_{J_\theta})'$ el vector de las funciones base.

La elección de J_θ es determinante, debe ser lo suficientemente grande como para capturar toda la información relevante y lo suficientemente pequeño para que la interpretación del modelo sea fácil.

Podemos repetir el procedimiento para los regresores utilizando un sistema de funciones base $\{\psi_j\}_{j=1}^{\infty}$ ($\{\psi_j\}_{j=1}^{\infty}$ no necesariamente es

igual a $\{\phi_j\}_{j=1}^{\infty}$):

$$X_i^{[J_x]} = \sum_{j=1}^{J_x} c_{ij} \psi_j = \Psi' \mathbf{C}_i, \quad i=1, \dots, n \quad (6.9)$$

Donde $\mathbf{C} = (c_{i1}, \dots, c_{iJ_x})'$ es el vector de coeficiente para X_i y $\Psi = (\psi_1, \dots, \psi_{J_x})'$ es el vector de las funciones base.

En consecuencia, el modelo $Y = \langle \theta, X \rangle + \varepsilon$ se puede aproximar por:

$$Y_i = \langle \theta^{[J_\theta]}, X_i^{[J_x]} \rangle + \varepsilon_i = \sum_{j_1=1}^{J_\theta} \sum_{j_2=1}^{J_x} b_{j_1} c_{i j_2} \langle \phi_{j_1}, \psi_{j_2} \rangle + \varepsilon_i \quad (6.10)$$

$$Y_i = \mathbf{C}_i' \mathbf{M}_{\Psi\Phi} \mathbf{B} + \varepsilon_i, \quad i=1, \dots, n \quad (6.11)$$

Donde $\mathbf{M}_{\Psi\Phi}$ es la matriz de tamaño $J_x \times J_\theta$ dada por $(\mathbf{M}_{\Psi\Phi})_{j_2 j_1} = \langle \psi_{j_2}, \phi_{j_1} \rangle$

Sea \mathbf{C} la matriz cuyas columnas son los vectores \mathbf{C}_i , (4.11) se puede escribir:

$$\mathbf{Y} = \mathbf{C} \mathbf{M}_{\Psi\Phi} \mathbf{B} + \mathbf{e} = \Xi \mathbf{B} + \mathbf{e} \quad (6.12)$$

Donde \mathbf{Y} es el vector de las respuestas observadas, $\Xi = \mathbf{C} \mathbf{M}_{\Psi\Phi}$ es una matriz de tamaño $n \times J_\theta$ y \mathbf{e} es el vector de errores $n \times 1$.

El vector \mathbf{B} se puede estimar mediante la minimización de la suma de cuadrados residuales.

$$\min_{\mathbf{B}} (\mathbf{Y} - \Xi \mathbf{B})' (\mathbf{Y} - \Xi \mathbf{B}) \quad (6.13)$$

De aquí \mathbf{B} se puede estimar por:

$$\hat{\mathbf{B}} = (\Xi' \Xi)^{-1} \Xi' \mathbf{Y} \quad (6.14)$$

Y las predicciones se calculan con:

$$\hat{\mathbf{Y}} = \Xi \hat{\mathbf{B}} \quad (6.15)$$

Introduciendo una penalización en los mínimos cuadrados podemos reducir la excesiva variabilidad del estimador y aumentar la regularidad. Tenemos los mínimos cuadrados penalizados:

$$\min_{\mathbf{B}} [(\mathbf{Y} - \Xi \mathbf{B})' (\mathbf{Y} - \Xi \mathbf{B}) + \rho \text{PEN}_q(\theta)] \quad (6.16)$$

Donde ρ es el parámetro de suavidad y $\text{PEN}_q(\cdot)$ es la función de penalización.

$$\text{PEN}_q(\theta) = \int_{T_0}^{T_1} (\theta^{(q)}(t))^2 dt \quad (6.17)$$

En este caso

$$\text{PEN}_q(\theta) = \int_{T_0}^{T_1} ((\Phi^{(q)}(t))' \mathbf{B})^2 dt = \mathbf{B}' \mathbf{R} \mathbf{B} \quad (6.18)$$

Donde

$\Phi^{(q)}(t) = (\phi_1^{(q)}(t), \dots, \phi_{J_\theta}^{(q)}(t))'$ es el vector de las derivadas

\mathbf{R} es la matriz de tamaño $J_\theta \times J_\theta$ dada por

$$\mathbf{R} = \int_{T_0}^{T_1} (\Phi^{(q)}(t) (\Phi^{(q)}(t))')^2 dt$$

En consecuencia, el problema se reduce a

$$\min_{\mathbf{B}} [(\mathbf{Y} - \Xi \mathbf{B})' (\mathbf{Y} - \Xi \mathbf{B}) + \rho \mathbf{B}' \mathbf{R} \mathbf{B}] \quad (6.19)$$

Y para ello hay que determinar el vector \mathbf{B} que satisface $2\Xi' \Xi \mathbf{B} - 2\Xi' \mathbf{Y} + 2\rho \mathbf{R} \mathbf{B} = 0$, de aquí que:

$$\hat{\mathbf{B}} = (\Xi' \Xi + \rho \mathbf{R})^{-1} \Xi' \mathbf{Y} \quad (6.20)$$

y el modelo ajustado:

$$\hat{\mathbf{Y}} = \Xi \hat{\mathbf{B}} = \Xi (\Xi' \Xi + \rho \mathbf{R})^{-1} \Xi' \mathbf{Y} \quad (6.21)$$

El balance entre el sesgo y la varianza es controlado por el parámetro de suavidad. Su elección puede realizarse subjetivamente o utilizando algún método de comprobación en función de los datos como la validación cruzada.

Sea la matriz $\mathbf{H} = \Xi (\Xi' \Xi + \rho \mathbf{R})^{-1} \Xi'$, $\text{tr}(\mathbf{H})$ representa los grados de libertad efectivos de la regresión.

La varianza residual

$$S_R^2 = \sum \frac{(y_i - \hat{y}_i)^2}{n - df} \quad (6.22)$$

Es un estimador insesgado de σ^2 y $R^2 = 1 - \frac{SCR}{SCT}$

con $R^2 \leq 1$

⁴ Ver [1], [2]

6.2. ESTIMACIONES BASADAS EN ACPF

El principal objetivo del ACPF (Análisis de Componentes Principales Funcionales) es el de extraer de un conjunto de datos funcionales (curvas) sus características más sobresalientes, que revelarán la complejidad de los datos, y permitirán entender la estructura de sus modos dominantes de variación. Además, conocer cuántos de éstos son necesarios para obtener una aproximación satisfactoria de los datos originales. El ACPF es una técnica de reducción de la dimensión que permite representar a los datos funcionales (de dimensión infinita) en función de unas pocas funciones base, lo que a su vez posibilita su representación en espacios de dimensión finita y por ende como vectores coordenada en \mathbb{R}^n .

El ACPF se fundamenta en dos pilares: el teorema de Mercer y la representación de Karhunen-Loeve.

Sea un proceso estocástico estacionario de segundo orden $X = \{X(t) / t \in [T_0, T_1]\}$, i.e X es una variable aleatoria funcional, definido sobre un espacio probabilístico (Ω, \mathcal{A}, P) , continuo en media cuadrática y con función de covarianza Γ .

Teorema de Mercer:

$$\Gamma(t, s) = \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(s) \quad t, s \in [T_0, T_1] \quad (6.23)$$

La convergencia de la serie es uniforme en $[T_0, T_1]^2$. (4.23) es la descomposición espectral de Γ , el operador de covarianza.

Las constantes λ_i y las funciones ψ_i son los valores y funciones propias de Γ , es decir:

$$\int_{T_0}^{T_1} \Gamma(t, s) \psi_i(s) ds = \lambda_i \psi_i(t) \quad (6.24)$$

Representación de Karhunen-Loeve;

$$X(t) = \sum_{i=1}^{\infty} b_i \psi_i(t) \quad t \in [T_0, T_1] \quad (6.25)$$

La serie converge uniformemente en $[T_0, T_1]$, donde

$$b_i = \int_{T_0}^{T_1} \psi_i(t) X(t) dt \quad (6.26)$$

6.3 APLICACIÓN

En el presente trabajo vamos a utilizar la librería de R `fds.usc`, en particular la función `fregre.pc` que realiza la regresión funcional mediante componentes principales funcionales.

Esta función calcula la base de componentes ortonormales $\{\psi_j\}_{j=1}^{\infty}$ para representar a los datos funcionales como:

$$X_i = \sum_{j=1}^{\infty} c_{ij} \psi_j, \quad i = 1, \dots, n \text{ y al parámetro funcional}$$

$$\theta = \sum_{i=1}^{\infty} \beta_i \psi_i$$

A continuación aplicamos el modelo de regresión funcional mediante componentes principales a los datos de espectrometría. Las 215 curvas espectrométricas constituyen la muestra de la variable aleatoria funcional. Para cada curva se han medido tres cantidades: el contenido de grasa, de agua y de proteína de cada muestra de carne. Estas variables serán consideradas como las variables escalares. Comenzamos por la variable proteína:

Contenido proteico

```
> data(tecator)
> absorp=tecator$absorp.fdata
> ind=1:215
> x=absorp[ind,]
> y=tecator$y$Protein[ind]
> res=fregre.pc(x,y)
> summary(res)
*** Summary Functional Data Regression with
Principal Components ***
Call:
fregre.pc(fdataobj = x, y = y)
Residuals:
    Min       1Q   Median       3Q      Max
-5.6404 -1.5808  0.1239  1.5399  5.5479
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)  17.6828      0.1594    110.959 <2e-16***
PC1          -0.1359      0.0224    -6.068 5.91 <e-09 ***
PC2           1.6466      0.2326     7.080 2.10 <e-11 ***
PC3           3.5700      0.4117     8.671 1.15 <e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

Residual standard error: 2.337 on 211 degrees of freedom
Multiple R-squared: **0.4084**, Adjusted R-squared: 0.4
F-statistic: 48.55 on 3 and 211 DF, p-value: < 2.2e-16

-With 3 Principal Components is explained 99.46 %
of the variability of explicative variables.
-Variability for each principal components -PC- (%):
PC1 PC2 PC3
98.27 0.89 0.29
-Names of possible atypical curves: No atypical curves
-Names of possible influence curves: 34 35 44
99 139 140 185 204

Contenido de grasa

```
data(tecator)
absorp=tecator$absorp.fdata
ind=1:215
x=absorp[ind,]
y=tecator$y$Fat[ind]
res=fregre.pc(x,y)
summary(res)
*** Summary Functional Data Regression with
Principal Components ***
Call:
fregre.pc(fdataobj = x, y = y)
Residuals:
  Min      1Q   Median      3Q      Max
-25.1086 -4.3943  0.3084  5.9803 14.2086
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.14233	0.55658	32.596	< 2e-16 ***
PC1	0.93994	0.07824	12.013	< 2e-16 ***
PC2	-3.04530	0.81221	-3.749	0.000229 ***
PC3	-20.17869	1.43793	-14.033	< 2e-16 ***

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
Residual standard error: 8.161 on 211 degrees of
freedom
Multiple R-squared: 0.5954, Adjusted R-
squared: 0.5897
F-statistic: 103.5 on 3 and 211 DF, p-value: <
2.2e-16
```

-With 3 Principal Components is explained 99.46 %
of the variability of explicative variables.
-Variability for each principal components -PC-
(%):
PC1 PC2 PC3
98.27 0.89 0.29
-Names of possible atypical curves: 14 15 51
144
-Names of possible influence curves: 34 35 44
99 139 140 185 204
>

Humedad

```
data(tecator)
absorp=tecator$absorp.fdata
ind=1:215
x=absorp[ind,]
y=tecator$y$Water[ind]
res=fregre.pc(x,y)
summary(res)
*** Summary Functional Data Regression with
Principal Components ***
Call:
fregre.pc(fdataobj = x, y = y)
```

```
Residuals:
  Min      1Q   Median      3Q      Max
-10.9702 -4.3547 -0.3774  3.5476 18.6047
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	63.20442	0.39934	158.27	< 2e-16 ***
PC1	-0.78163	0.05614	-13.92	< 2e-16 ***
PC2	2.24365	0.58274	3.85	0.000156 ***
PC3	16.32870	1.03169	15.83	< 2e-16 ***

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

```
Residual standard error: 5.855 on 211 degrees of
freedom
Multiple R-squared: 0.6553,
Adjusted R-squared: 0.6504
F-statistic: 133.7 on 3 and 211 DF, p-value: <
2.2e-16
```

-With 3 Principal Components is explained 99.46 %
of the variability of explicative variables.
-Variability for each principal components -PC-
(%):
PC1 PC2 PC3
98.27 0.89 0.29
-Names of possible atypical curves: 144
-Names of possible influence curves: 34 35 44 99
139 140 185 204
>

Figura 5
*Regresión funcional mediante bases obtenidas por descomposición
espectral del operador covarianza*
Primera Componente Principal

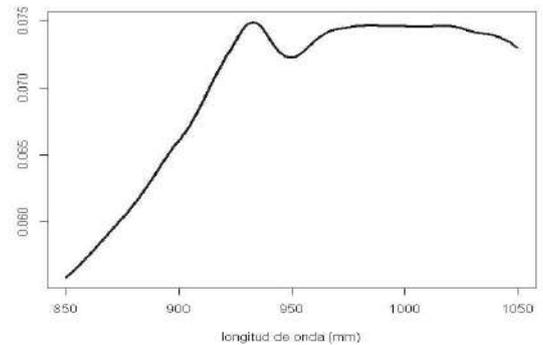


Figura 6

Regresión funcional mediante bases obtenidas por descomposición espectral del operador covarianza
Primera y Segunda Componentes Principales

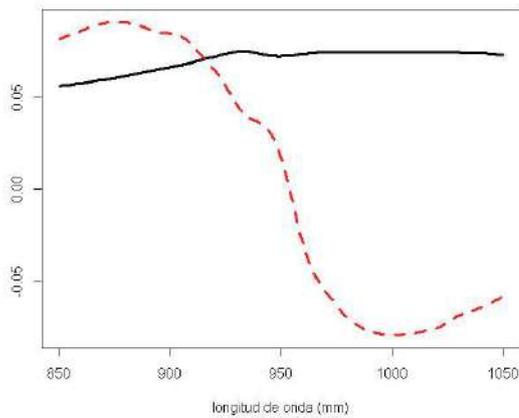


Figura 7

Regresión funcional mediante bases obtenidas por descomposición espectral del operador covarianza
Primeras Tres Componentes Principales

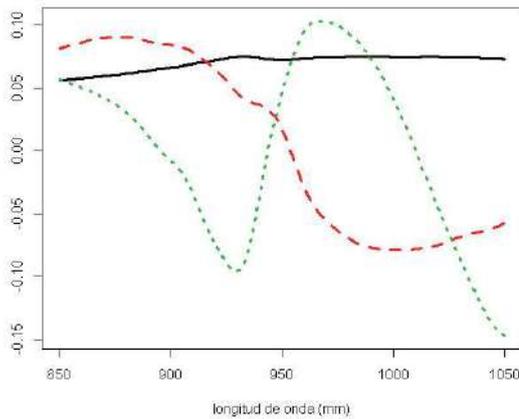
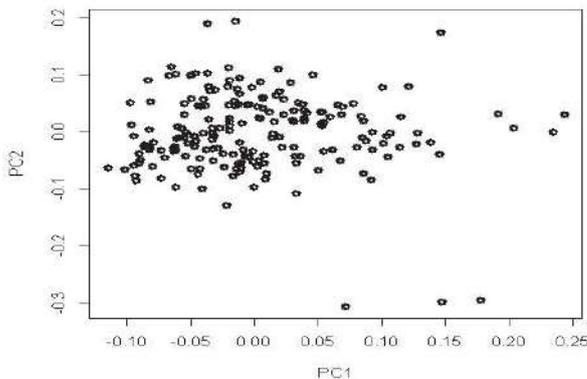


Figura 8

Regresión funcional mediante bases obtenidas por descomposición espectral del operador covarianza
Individuos en el Plano Principal 1-2



En todos los casos el coeficiente de determinación R^2 es inferior a 0.7, lo que significa que no hay un buen ajuste del modelo. Esto podría deberse a que hay muchas variables (en este caso el número de variable es igual al número de medidas en cada curva) y pocos individuos (cada curva es un individuo). Por esta razón vamos a utilizar la regresión PLS:

Utilizando PLSR

```
> data(tecator)
> absorp=tecator$absorp.fdata
> ind=1:215
> x=absorp[ind,]
> y=tecator$y$Fat[ind]
> res=fregre.pls(x,y)
> summary(res)
```

*** Summary Functional Regression with Partial Least Squares***

-Call: fregre.pls(fdataobj = x, y = y)

	Estimate	Std. Error	t value	Pr(> t)
(Inter cept)	18.1423256	0.36640865	49.513912	1.587881e-117
PLS1	-0.3016332	0.03924106	-7.686673	5.776922e-13
PLS2	10.0242344	0.42986925	23.319264	4.156364e-60
PLS3	5.9757045	0.43631267	13.695923	6.620177e-31

-R squared: **0.8260812**

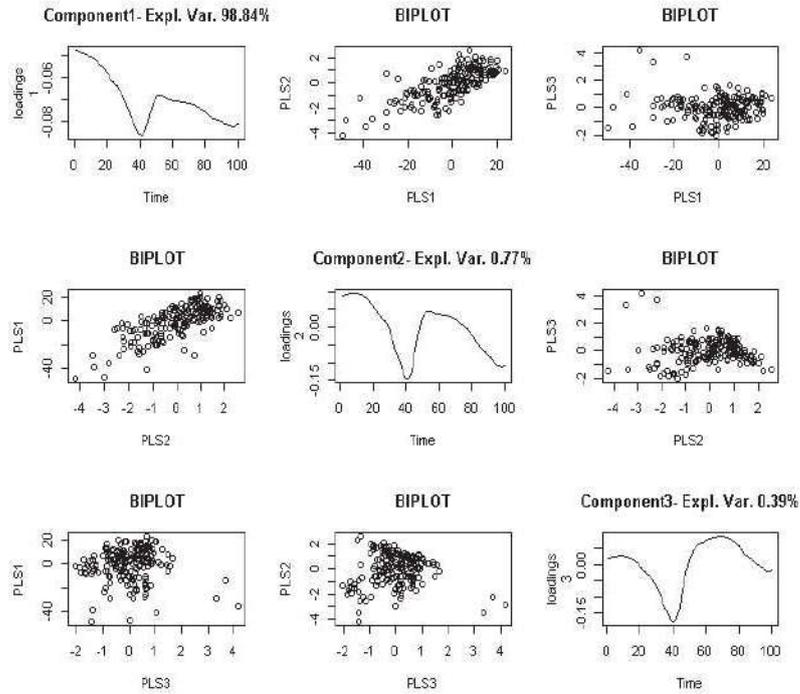
-Residual variance: 28.86489 on 209.2904 degrees of freedom

-Names of possible atypical curves: 13 14 15 51 144 188

-Names of possible influence curves: 34 35 43 44 99 139 140 185 204

Como se aprecia el coeficiente de determinación aumenta considerablemente, lo que indica que la regresión funcional PLS ajusta mejor los datos funcionales que los modelos convencionales.

Figura 9
Componentes principales PLS



7 CONCLUSIONES

El coeficiente de determinación R^2 del modelo conseguido mediante PLS es mucho mayor (**0.8260812**), lo que indica que la regresión funcional mediante componentes principales obtenidos por PLS ajusta mejor los datos funcionales que los modelos basados en ACP convencional.

Todos los regresores o predictores son significativos, su p-value es menor al 0.01%. Debido a que las tres primeras componentes principales explican el 100% de la varianza (cantidad de información de los datos), la dimensionalidad se reduce drásticamente a tres. Es más, el porcentaje de varianza explicado por la primera componente principal es de 98.84%, lo que significa que es suficiente considerar sólo a esta componente en el modelo de regresión, con una mínima pérdida de información

REFERENCIAS BIBLIOGRÁFICAS Y ELECTRÓNICAS

- [1]. **CARDOT, H., FERRATY, F., SARDA, P. (1999).** “*Functional linear model. Statistics & Probability Letters*”. Vol. 45, p.11 – 22.
- [2]. **CARDOT, H., FERRATY, F., SARDA, P. (2003).** “*Spline estimators for the functional linear model. Statistica Sinica*”. Vol. 13, p. 571-591.
- [3]. **DABO-NIANG, S., FERRATY, F. (2008).** “*Functional and Operatorial Statistics*”. Physica-Verlag Heidelberg.
- [4]. **ESCABIAS, M., VALDERRAMA, M., AGUILERA, M. (2012).** “*Functional Data Analysis in Biometrics and Biostatistics*”. Journal of Biometrics & Biostatistics. Vol. 3 No. 8, p. 1-3.
- [5]. **FERRATY, F., VIEU, P. (2006).** “*Nonparametric Functional Data Analysis: Theory and Practice*”. Springer Series in Statistics. New York.
- [6]. **GOIA, A. (2003).** “*Selection Model in Functional Linear Regression Models for Scalar Response*”. Developments in Applied Statistics. Vol. 19, p. 51-68.
- [7]. **MULLER, H., STADTMULLER, U. (2005).** “*Generalized functional linear models*”. The Annals of Statistics. Vol. 33, No. 2, p. 774–805.
- [8]. **OVIEDO M., (2011).** “*Utilities for Statistical Computing in Functional Data Analysis: The R Package fda.usc*”. Master thesis, Universidad de Santiago de Compostela.
- [9]. **RAMSAY, J., SILVERMAN, B. W. (2005).** “*Functional Data Analysis*” (2nd Ed.). Springer Series in Statistics. New York
- [10]. **REEDER R. (2011).** “*Limit Theorems in Functional Data Analysis with Applications*”. PhD. thesis. Department of Mathematics. University of Utah.