Departamento de Estadística e Investigación Operativa
Universidad de Granada

TRABAJO DE INVESTIGACIÓN

MÁSTER OFICIAL EN ESTADÍSTICA APLICADA

Análisis de la varianza en el contexto de datos funcionales

Autor: Albert Gabarrús Barri
Tutora: María Dolores Ruiz Medina

Septiembre 2012

TRABAJO DE INVESTIGACIÓN

Análisis de la varianza en el contexto de datos funcionales

Autor:

Fdo.: Albert Gabarrús Barri

Tutora:

Fdo.: María Dolores Ruiz Medina

Septiembre 2012

TABLA DE CONTENIDOS

AGRADECIMIENTOS

Porque el éxito nunca corresponde a una sola persona, y siempre es compartido en la medida en que se ha trasferido el conocimiento, quiero expresar mi agradecimiento sincero a:

- la Dra. María Dolores Ruiz Medina, Tutora de este trabajo, por sus apreciados y relevantes aportes para el desarrollo de esta investigación, además por la confianza depositada, y enseñarme la trascendencia del estudio de esta área.

- todos los extraordinarios profesionales que laboran como profesores en el Master Oficial de Estadística Aplicada de la Universidad de Granada, para mí ha sido un privilegio y un vivo ejemplo de excelencia profesional de los que pude aprender mucho.

Finalmente, quiero expresar mi agradecimiento a mi familia, que siempre está a mi lado.

PRÓLOGO

Este trabajo de investigación se ha llevado a cabo en la línea de investigación en Análisis de Datos Funcionales (ADF) del Departamento de Estadística e Investigación Operativa de la Universidad de Granada.

Las observaciones muestrales de una variable funcional son funciones que en la mayor parte de los casos proceden de la observación temporal de una variable estadística (realizaciones de un proceso estocástico). Los datos funcionales aparecen en campos muy diversos de aplicación de la estadística como la economía, ciencias de la salud y medioambientales, entre otras.

La posibilidad de obtener para un mismo individuo un gran número de registros de la misma característica ha propiciado la aparición de nuevos métodos estadísticos que facilitan el trabajo con grandes volúmenes de información. Entre ellos se encuentran las técnicas no paramétricas de suavizado (Delicado, 2007) y el ADF.

El ADF es una nueva visión que considera el caso en el que el número de medidas repetidas es muy grande. Su reciente crecimiento involucra procedimientos de muchas áreas de la estadística. Cabe destacar las contribuciones en análisis exploratorio y descriptivo de datos (Ramsay, 2005), modelos lineales (Cardot and Sarda, 1999; Malfait and Froda, 2000), modelos lineales generalizados (Escabias and Valderrama, 2004), análisis de la varianza (Cuevas, Febrero and Fraiman, 2004; Delicado, 2007; Shen and Faraway, 2004), métodos no paramétricos (Ferraty and Vieu, 2006), datos longitudinales (Yao and Wang, 2005) y técnicas multivariadas como componentes principales (Pezulli and Silverman, 1993; Silverman, 1995), correlación canónica (Leurgans and Silverman, 1993), análisis discriminante (Ferraty and Vieu, 2003) y análisis clúster (Clarkson and Ramsay, 2005). En la mayoría de los casos se supone que las variables funcionales son independientes, situación que en algunas áreas del conocimiento no se cumple. En muchas situaciones experimentales el supuesto de independencia puede ser inválido, por ejemplo cuando las funciones son evaluadas en diferentes puntos del tiempo (temporalmente correlacionadas) o del espacio (espacialmente correlacionadas). En este caso, algunos de los procedimientos pierden propiedades importantes, razón por la cual es necesario extender dichas metodologías considerando este tipo de correlación. Se destaca que en la literatura para datos funcionales correlacionados temporalmente, Yamanishi y Tanaka (2003) desarrollaron un modelo de regresión que tiene en cuenta la relación entre las variables en el tiempo y en el espacio, combinando metodologías de regresión ponderada geográficamente (Brunsdon, Fotheringham and Charlton, 1998) y de regresión funcional múltiple (Ramsay and Silverman, 2005) y Baladandayuthapani, Mallick, Hong, Lupton, Turner, y Caroll (2008) muestran una alternativa para analizar un diseño experimental con una variable funcional respuesta correlacionada, a través de procedimientos jerárquicos Bayesianos.

En concreto, el análisis de la varianza funcional (FANOVA) proporciona alternativas a los métodos clásicos del ANOVA. Debido al gran número de aplicaciones en las que pueden emplearse los datos funcionales, el interés de los investigadores hacía los modelos FANOVA está en aumento. Para empezar, Ramsay and Silverman (1997) y Stone et al. (1997) proporcionan gran información básica y general sobre los modelos FANOVA. Por otra parte, aunque existe un gran número de trabajos referentes al ajuste de los modelos FANOVA y a la estimación de sus componentes (por ejemplo, Wahba et al., 1995; Stone et al., 1997; Huang, 1998; Lin, 2000; Gu, 2002), no hay demasiados trabajos en desarrollar contrastes de hipótesis en los modelos FANOVA.

Es por ello que el objetivo fundamental de este trabajo es describir algunos resultados recientes sobre el análisis de la varianza en el contexto de datos funcionales, partiendo del artículo desarrollado por Abdelhak Zoglat en 2008.

1. RESUMEN

En este trabajo se extienden algunas de las técnicas del análisis de la varianza multivariante (MANOVA), donde los datos son vectores en $\mathbb{R}^d$, al caso donde las observaciones son funciones. Consideramos el modelo

$$Y(t) = x'\beta(t) + \sigma\epsilon(t), \quad t \in [0,1]$$

donde $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$ es conocida, $\sigma > 0$ es un parámetro real desconocido, y $\beta = (\beta_1, \dots, \beta_p)' \in L_2^p$ es un vector de cuadrado integrable y funciones desconocidas. Se estima β y se describen algunas de las propiedades óptimas del estimador $\hat{\beta}$. También de desarrolla una prueba estadística para probar $H_0\colon K\beta = C$ *versus* $H_1\colon K\beta \neq C$. Los resultados son extensiones naturales de los clásicos.

2.INTRODUCCIÓN

El análisis de la varianza multivariante (MANOVA) pretende contrastar hipótesis lineales sobre la influencia de los distintos niveles de uno o varios factores en el comportamiento de un vector (multidimensional) de variables. En este caso el "parámetro" a estimar es una función. Pero hay situaciones en que los propios datos disponibles son funciones.

El ADF se ocupa de la modelización estadística de variables aleatorias que toman valores en un espacio de funciones (variables funcionales). Por ejemplo, el seguimiento de procesos tecnológicos o industriales, el control de las condiciones atmosféricas [Ramsay J.O. and Dalzell C.J. 1991], la observación del "mercado continuo" en la Bolsa, etcétera, proporcionan observaciones aleatorias que pueden considerarse como funciones. Varias técnicas estadísticas estándares tales como el análisis de regresión, el análisis de la varianza (ANOVA) o el análisis de componentes principales, entre otras, han sido consideradas desde el punto de vista funcional [Parzen E. 1961]. En general, estas metodologías se centran en variables funcionales independientes e idénticamente distribuidas. Sin embargo, a pesar de su importancia, algunos aspectos de la inferencia estadística de datos funcionales no han sido totalmente explorados. Naturalmente este punto de vista "funcional" requiere un cierto proceso de abstracción y de modelización porque, en realidad, las observaciones se obtienen casi siempre en versión "discretizada". Esto podría suponer la pérdida de información potencialmente útil. El objetivo del artículo es desarrollar procedimientos que permitan usar al investigador la totalidad de la información contenida en los datos. Los resultados son una generalización natural de los resultados de la teoría de los modelos lineales clásicos. El espacio más comúnmente usado cuando se habla de datos funcionales es el espacio de Hilbert $L^2(S)$, esto es, las funciones de cuadrado integrable en el intervalo $S = [a,b] \subset \mathbb{R}$. Sin embargo, los principales cálculos están basados en integraciones numéricas simples.

El objetivo fundamental de este trabajo es describir algunos resultados recientes sobre el análisis de la varianza en el contexto de datos funcionales. Más concretamente, y partiendo del artículo desarrollado por Abdelhak Zoglat en 2008, se consideran los siguientes contenidos por secciones:

En el Apartado 3, se presentan algunas definiciones preliminares del ADF. Posteriormente se considera el modelo

$$Y(t) = x'\beta(t) + \sigma\epsilon(t), \quad t \in [0,1]$$

donde $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$ es conocida, $\sigma > 0$ es un parámetro real desconocido, y $\beta = (\beta_1, \dots, \beta_p)' \in L_2^p$ es un vector de cuadrado integrable y funciones desconocidas.

En el Apartado 4, se realiza la estimación de β por mínimos cuadrados y por máxima verosimilitud, y se describen algunas de las propiedades óptimas del estimador $\hat{\beta}$.

Finalmente, en el Apartado 5, se realiza inferencia sobre β mediante el contraste de hipótesis $H_0: K\beta = C$ *versus* $H_1: K\beta \neq C$.

Los resultados son extensiones naturales de las clásicos.

3. ANÁLISIS DE DATOS FUNCIONALES

El análisis de datos funcionales es aquella parte de la estadística que trabaja con muestras de funciones aleatorias. Ramsay y Sikerman (1997), entre otros, han introducido herramientas de análisis para este tipo de datos. Las medidas de tendencia central, de dispersión y de relación entre variables aplicadas a muestras de variables aleatorias se pueden definir de manera análoga para muestras de datos funcionales. Sólo hace falta considerar que ahora estaremos trabajando en un espacio vectorial distinto: el espacio L^2 (funciones de cuadrado integrable). En realidad Ramsay y Silverman definen estas medidas de manera general para cualquier espacio que tenga definido un producto escalar, y luego deduce de ellas las definiciones para muestras de funciones.

A continuación se presentan unas definiciones preliminares sobre medidas clásicas y distancias entre vectores en $\mathbb{R}^p$ en contexto funcional. Se asume que $X(t)$, $Y(t)$, $Z(t)$, $t \in T$ son funciones definidas en algún espacio de funciones.

- *Producto escalar*

$$\langle X(t), Y(t) \rangle = \int_T X(t)\, Y(t)\, dt\, .$$

Propiedades:

1. Simetría: $\langle X(t), Y(t)\rangle = \langle Y(t), X(t)\rangle$.
2. Positividad: $\langle X(t), Y(t)\rangle \geq 0$.
3. Bilinealidad: $\langle aX(t) + bY(t), Z(t)\rangle = a\langle X(t), Z(t)\rangle + b\langle Y(t), Z(t)\rangle$, para cualesquiera números reales *a* y *b*.

- *Norma*

$$\|X(t)\| = \sqrt{\|X(t)\|^2}, \text{ donde}$$

$$\|X(t)\|^2 = \langle X(t), X(t)\rangle = \int_T X(t)\, X(t)\, dt\,.$$

Propiedades:

1. $\|X(t)\| \geq 0$.
2. $\|aX(t)\| = |a|\|X(t)\|$, para cualquier número real *a*.
3. $\|X(t) + Y(t)\| \leq \|X(t)\| + \|Y(t)\|$.

En el ADF la unidad básica de información es la función completa, más que un conjunto de valores [Ramsay J.O. and Dalzell C.J. 1991], es decir, que un dato funcional se puede establecer como la extensión de medidas repetidas cuando la cantidad de mediciones en un determinado individuo es muy grande. En el contexto multivariado los datos provienen de la observación de la familia aleatoria $\{X(s_j)\}_{j=1,\dots,J}$. Por otro lado, en el análisis funcional se asume que estos mismos proceden de una familia continua $X = \{X(s); s \in S\}$

[Ferraty F. and Vieu P. 2006]. Según Ferraty and Vieu dos definiciones importantes para establecer diferencias entre el contexto real y funcional son:

Definición 1. Una variable aleatoria X se llama funcional si toma valores en un espacio de funciones, es decir, un espacio infinito dimensional (espacio funcional). Una observación x de la variable aleatoria X se denomina dato funcional.

Definición 2. Un conjunto de datos funcionales $x_1, x_2, \dots, x_n$ es la observación de n variables funcionales distribuidas como X.

3.1. Consideraciones generales

Sea $(S, \mathcal{S}, \mu)$ un espacio de medida, con μ una medida σ-finita. Normalmente se asume que se tienen elementos del espacio de Hilbert $L_2(S, \mu) = \left\{ f: S \to \mathbb{R};\ \int_S\ f^2(s)\, d\mu(s) < \infty \right\}$.

$L_2(S, \mu)$ está dotado con su producto escalar habitual:

$$\forall f, g \in L_2(S, \mu) \qquad \langle f, g \rangle = \int_S\ f(s) g(s)\, d\mu(s)\ ,$$

y su norma:

$$\|.\|_2^2 = \langle .,. \rangle\ .$$

Veremos que $f(s)$ y $g(s)$ es, en efecto, un producto escalar ya que cumplen lo siguiente:

1. Simetría:

   $\langle f(s), g(s)\rangle = f(s)'g(s) = \int_S \; f(s)g(s)\,d\mu(s) =$

   $\int_S \; g(s)f(s)\,d\mu(s) = \langle g(s), f(s)\rangle.$

2. Positividad:

   $\langle f(s), f(s)\rangle = f(s)'f(s) = \int_S \; f(s)f(s)\,d\mu(s) \geq 0 \; y \; \langle f(s), f(s)\rangle =$

   $\int_S \; f(s)f(s)\,d\mu(s) = 0 \;\Leftrightarrow x = 0.$

3. Bilinealidad:

   $\langle af(s) + bg(s), z(s)\rangle = \int_S \; [af(s) + bg(s)]\, z(s)d\mu(s) =$

   $\int_S \; af(s)\, z(s)d\mu(s) + \int_S \; bg(s)\, z(s)d\mu(s) = a\langle f(s), z(s)\rangle +$

   $b\langle g(s), z(s)\rangle$, para cualesquiera números reales *a* y *b*.

Sea $k: S \times S \longrightarrow \mathbb{R}$ una función simétrica, es decir, para cada $(s,t) \in S \times S$, $k(s,t) = k(t,s)$ y $\mu \times \mu$ de cuadrado integrable. El operador integral T_k, con kernel k, se define en $L_2(S,\mu)$ por: $f \longmapsto T_k f$, donde $T_k f$ es una función real definida en S por

$$\forall t \in S \qquad T_k f(t) = \int_S \; k(s,t)f(s)\,d\mu(s)\,.$$

Se puede ver fácilmente que T_k es lineal y autoadjunto, es decir, $\langle T_k f, g\rangle = \langle f, T_k g\rangle$ para todo $f, g \in L_2(S,\mu)$. Además T_k es continuo y compacto. Por lo tanto, existe una secuencia (λ_k) de números reales y una base ortonormal (φ_k) para $(ker\; T_k)^{\perp}$ tal que para todo $\int \; \in L_2(S,\mu)$

$$T_k\int \; = \sum_{k\geq 1} \lambda_k \langle \int, \varphi_k \rangle \varphi_k \,. \tag{1}$$

En particular, para cada $k \geq 1$, φ_k es una función propia de T_k asociada con el valor propio λ_k.

Sea B_{L2} la σ-álgebra de Borel de $L_2(S,\mu)$, y X una variable aleatoria de media cero de valor $(L_2(S,\mu), B_{L2})$. La función de covarianza de X, $k: S \times S \longrightarrow \mathbb{R}$, se define como: $k(s,t) = \mathbb{E}[X(s)X(t)]$. En este punto siempre vamos a considerar variables aleatorias de media cero de valor $L_2(S,\mu)$ con función de covarianza $\mu \times \mu$ de cuadrado integrable. De este modo T_k, el operador de la covarianza de X satisface (1).

El operador T_k induce en $L_2(S,\mu)$ un productor interior $\langle .,. \rangle_{T_k}$, y de este modo una seminorma $\|.\|_{T_k}$, definida por

$$\forall f, g \in L_2(S,\mu), \quad \langle f, g \rangle_{T_k} = \langle f, T_k g \rangle, \;\; y \;\; \|f\|_{T_k}^2 = \langle f, f \rangle_{T_k}.$$

Gracias a (1) y la definición de $\|.\|_{T_k}$, tenemos que

$$\forall \int \; \in L_2(S,\mu), \quad \left\| \int \right\|_{T_k} = \|h\|_{T_k} \,, \tag{2}$$

donde h es la proyección de f en $(ker\, T_k)^{\perp}$.

Además sabemos que

$$\forall f, g \in L_2(S,\mu), \quad \langle T_k f, g \rangle_{T_k} = \mathbb{E}[\langle X, f \rangle \langle X, g \rangle] \,. \tag{3}$$

Utilizando la definición de T_k y la desigualdad de Schwarz se puede ver fácilmente que la seminorma es continua respecto a la habitual,

$$\forall f \in L_2(S,\mu) \quad \|f\|_{T_k}^2 = \|f\|_2^2 \int_{S\times S} k^2(s,t)\, d\mu(s) d\mu(t) \,. \tag{4}$$

Esto implica en particular que cualquier secuencia que converge en $(L_2(S,\mu), \|.\|_2)$ también converge en $\left(L_2(S,\mu), \|.\|_{T_k}\right)$. Así los cálculos con $\|.\|_{T_k}$ son sencillos. Además, utilizando esta seminorma veremos que los métodos de mínimos cuadrados y máxima verosimilitud conllevan al mismo estimador. Con la normal habitual esto no se establece.

Es conveniente utilizar notación vectorial. Consideramos vectores en el espacio de Hilbert $L_2^r(S,\mu)$ dotado de su producto interno canónico

$$\langle f,g\rangle = \sum_{i=1}^{r} \langle f_i, g_i\rangle, \qquad y \qquad \langle f,g\rangle_{T_k} = \sum_{i=1}^{r} \langle f_i, g_i\rangle_{T_k},$$

donde $f = (f_1, \dots, f_r)'$, y $g = (g_1, \dots, g_r)' \in L_2^r(S,\mu)$. Sea $\xi = (\xi_1, \dots, \xi_r)'$ un elemento de $L_2^r(S,\mu)$. Para cada $i = 1, \dots, r$, y $k \geq 1$, $\xi_{i(k)}$ indicará la proyección de la componente de la proyección de ξ_i en φ_k, es decir, $\xi_{i(k)} = \langle \xi_i, \varphi_k\rangle \varphi_k$. El elemento de $\mathbb{R}^r$, cuyos componentes son $\xi_{1(k)}, \dots, \xi_{r(k)}$ serán indicados por $\xi_{(k)}$, es decir, $\xi_{(k)} = \left(\xi_{1(k)}, \dots, \xi_{r(k)}\right)'$.

Para simplificar la notación y sin perder ningún tipo de generalización supondremos que $(S, S, \mu) = \left([0,1], \mathrm{B}_{[0.1]}, \lambda\right)$, el intervalo $[0,1]$ dotado de la σ-álgebra de Borel y la medida de Lebesgue. El espacio de Hilbert $L_2([0,1], \lambda)$ es indicado con L_2.

3.2. El Modelo

El modelo se basa en el modelo tradicional de Análisis de la Variancia. Las observaciones son funciones del tiempo, y cada observación $Y(t), t \in [0,1]$ es la suma de una función m determinista, y una función aleatoria escalar , es decir, $Y = m + \sigma\epsilon$. Llamamos m a la función del valor medio, ϵ al error, y σ es un escalar positivo desconocido. Asumimos que la función m del valor medio es una combinación lineal de algunas funciones desconocidas. La ecuación de una observación se puede escribir como

$$Y(t) = x'\beta(t) + \sigma\epsilon(t), \quad t \in [0,1] \tag{5}$$

donde $x = \left(x_1, \dots, x_p\right)' \in \mathbb{R}^p$ es conocida, y $\beta = \left(\beta_1, \dots, \beta_p\right)' \in L_2^p$ se tiene que estimar.

Sea k el operador de covarianza de $\in$ y $\{(\xi_i, \varphi_k), k \geq 1\}$ la secuencia de pares de valores propios - funciones propias del operador de covarianza T_k. La secuencia $(\varphi_k)_{k\geq 1}$ nos permite simplificar el modelo (5) a un sistema de modelos clásicos. En particular si ϵ es Gausiana, esta simplificación preserva la independencia entre las realizaciones de proyecciones. Además nos permitirá identificar la ley de algunos estadísticos. La herramienta clave es el siguiente resultado:

Teorema 1. [Grenander U. 1981] *Sea X una variable aleatoria normal de media cero con valores en $L_2(S, \mu)$, existe una secuencia (X_k) de variables aleatorias normales independientes e idénticamente distribuidas tal que:*

$$X_k = \frac{1}{\sqrt{\lambda_k}} \langle X, \varphi_k \rangle\,, \qquad y \qquad X = \sum_{k \geq 1} \sqrt{\lambda_k}\, X_k \varphi_k\,, \quad con\ toda\ seguridad$$

Para simplificar la notación, T significará T_k y H_T significará el cierre en L_2 del subespacio cerrado por $(\varphi_k)_{k\geq 1}$, es decir,

$$\mathrm{H}_T = \left\{ f \in L_2; f = \sum_{k \geq 1} \langle f, \varphi_k \rangle \varphi_k \right\}.$$

En otras palabras H_T es el espacio de Hilbert de kernel reproducido en k. En caso de normalidad del error, asumiremos sin ninguna pérdida de generalización que la función de valor medio *m* pertenece a H_T. En realidad, cada observación *Y* es la suma de su proyección en H_T, $\mathrm{Y}_{\mathrm{H}_T}$, y $\mathrm{Y}_{(\mathrm{H}_T)^{\perp}} = Y - \mathrm{Y}_{\mathrm{H}_T}$. El error es con toda seguridad en H_T. Por lo tanto $m_{(\mathrm{H}_T)^{-}} = \mathrm{Y}_{(\mathrm{H}_T)^{\perp}}$ con toda seguridad, y de este modo sólo necesitamos hacer inferencia sobre m_{H_T}.

4. ESTIMACIÓN

Los métodos habitualmente empleados en la teoría de la estimación son los métodos de mínimos cuadrados y de la máxima verosimilitud. En la teoría de modelos lineales, bajo el supuesto de normalidad, nos llevan a los mismos estimadores que tienen algunas propiedades óptimas. Su justificación como procedimientos satisfactorios de estimación se describe en muchos textos estadísticos. Por así decirlo, el método de mínimos cuadrados está basado en minimizar la norma al cuadrado del error. En el espacio Euclídeo esto se puede conseguir calculando derivadas e igualando a cero. En el caso dimensional infinito no hay un equivalente a este procedimiento. Sin embargo, en un espacio de Hilbert, (o más en general en un espacio de Banach teniendo una base numerable), es posible aplicar el procedimiento de mínimos cuadrados. Mostraremos como aplica en L_2, y probaremos que los estimadores son óptimos en cierto sentido.

El método de máxima verosimilitud es más restrictivo que el método de mínimos cuadrados. En concreto, requiere una densidad de distribución con respecto a alguna medida. La medida de Lebesgue es la más utilizada en los modelos lineales clásicos, y es eficiente bajo el supuesto de normalidad. En el caso funcional no hay un equivalente a la medida de Lebesgue, e incluso bajo el supuesto de normalidad no existe una densidad de distribución. Presentamos algunos casos donde puede aplicarse el método de máxima verosimilitud.

Suponemos que observamos n realizaciones independientes de la función aleatoria Y de (5). En notación vectorial, el modelo puede escribirse como

$$Y(t) = X\beta(t) + \sigma\epsilon(t); \quad t \in [0,1] \,, \tag{6}$$

donde σ es un escalar positivo desconocido y $\beta = (\beta_1, \dots, \beta_p)' \in L_2^p$ se tiene que estimar, $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ es una variable aleatoria en L_2^n, y $X = (x_{ij})$ es una matriz $n \times p$ cuyos valores son números reales conocidos. Se asume que las observaciones son independientes, y tienen la misma función de covarianza integrable al cuadrado k. En otras palabras, las componentes de ϵ son de valor L_2, media cero, y variables aleatorias independientes las cuales definen el mismo operador de covarianza compacto T_k.

Para cada $t \in [0,1]$, las ecuaciones (6) son las del modelo lineal clásico y de esta manera los estimadores $\beta(t)$ están disponibles. Nuestro objetivo es encontrar un estimador de la función vectorial $\beta \in L_2^p$ que para cada $t \in [0,1]$ coincida con el estimador clásico de $\left(\beta = \left(\beta_1(t), \dots, \beta_p(t)\right)\right)$, así como la obtención de contrastes lineales sobre los parámetros del modelo $H_0{:}\,K\beta = C$.

En la sección 4.1 y 4.3 se explican los métodos de estimación empleados: el método de mínimos cuadrados y el método de máxima verosimilitud, respectivamente.

4.1. El Método de Mínimos Cuadrados

El método de mínimos cuadrados consiste en encontrar el vector $\hat{\beta}$ que minimiza la T-norma del error $e(\beta) = (Y_1 - (X\beta)_1, \ldots, Y_n - (X\beta)_n)'$ cuando es vista como una función de β. Como σ no depende de β, dividiendo $e(\beta)$ por σ si es necesario, supondremos que $\sigma = 1$.

De la definición de la T-norma en L_2^n, tenemos que

$\|e(\beta)\|_T^2 = \sum_{i=1}^n \|e_i(\beta)\|_T^2 = 1$, con $e_i(\beta) = Y_i - (X\beta)_i$, para $i = 1, \ldots, n$.

La definición del producto interno $\langle .,. \rangle_T$, y (1) muestra que

$$\|e_i(\beta)\|_T^2 = \sum_{k \geq 1} \langle e_i(\beta), \varphi_k \rangle_T^2$$

$$= \sum_{k \geq 1} [\langle TY_i, \varphi_k \rangle - \langle T(X\beta)_i, \varphi_k \rangle]$$

$$= \sum_{k \geq 1} \lambda_k \left[Y_{i(k)} - (X\beta)_{i(k)}\right].$$

Por tanto

$$\sum_{i=1}^n \|e_i(\beta)\|_T^2 = \sum_{k \geq 1} \lambda_k^2 \left(Y_{(k)} - X\beta_{(k)}\right)' \left(Y_{(k)} - X\beta_{(k)}\right). \quad (7)$$

Esto muestra que $\|e_i(\beta)\|_T^2$ es minimizado si, y sólo si, para cada $k \geq 1$, la norma Euclidea de $e_k(\beta)$ es minimizada. Para cada $k \geq 1$ fijo, la norma de $e_k(\beta)$, mirada como una función de $\beta_{(k)}$, es minimizada en

$$\hat{\beta}_{(k)} = (X'X)^{-1} X' Y_{(k)}\,.$$

Teorema 2. *El estimador de β que minimiza $\|Y - Xb\|_T$, visto como una función de b, es dado por*

$$\hat{\beta} = (X'X)^{-1}X'Y\,.$$

4.2. La Propiedad de MELI

De la teoría clásica de los modelos lineales se conoce que bajo las condiciones de Gauss-Markov, para cada $k \geq 1$, $\hat{\beta}_{(k)}$ el mejor estimador lineal insesgado (MELI) de $\beta_{(k)}$, es decir,

$$\forall a \in \mathbb{R}^p, \qquad \forall b \in \mathbb{R}^n, \qquad \mathbb{E}\left[b'Y_{(k)}\right] = a'\beta_{(k)} \Rightarrow \mathbb{V}ar\left[a'\hat{\beta}_{(k)}\right] \leq \mathbb{V}ar\left[b'Y_{(k)}\right].$$

Esto también puede ser formulado como: *Si $f \in \mathbb{R}^n$ y $g \in \mathbb{R}^p$ son dos formas lineales continuas, en sus respectivos espacios, entonces*

$$\mathbb{E}\left[\langle f, Y_{(k)}\rangle\right] = \langle g, \beta_{(k)}\rangle, \qquad \Rightarrow \mathbb{V}ar\left[\langle g, \hat{\beta}_{(k)}\rangle\right] \leq \mathbb{V}ar\left[\langle f, Y_{(k)}\rangle\right].$$

Pero $Y_{(k)}$ y $\beta_{(k)}$ son también elementos de L_2^n y L_2^p respectivamente, y formas lineales continuas en $\mathbb{R}^r$, $r = n$ o p, pueden ser vistos como elementos particulares de L_2^r. Una generalización natural de la propiedad de MELI podría ser: *Si $f \in L_2^n$ y $g \in L_2^p$ son dos formas lineales continuas, en sus respectivos espacios, entonces*

$$\mathbb{E}[\langle f, Y\rangle] = \langle g, \beta\rangle, \qquad \Rightarrow \mathbb{V}ar\left[\langle g, \hat{\beta}\rangle\right] \leq \mathbb{V}ar[\langle f, Y\rangle].$$

Sea $g \in L_2^p$, ¿para qué $f^* \in L_2^n$ es la mínima varianza observada? La respuesta se da en el siguiente teorema.

Teorema 3. *Sea $f \in L_2^n$ y $g \in L_2^p$ tal que:*

$$\mathbb{E}[\langle f, Y\rangle] = \langle g, \beta\rangle\,,$$

entonces tenemos que

$$\mathbb{V}ar\big[\langle g, \hat{\beta}\rangle\big] \leq \mathbb{V}ar[\langle f, Y\rangle]\,.$$

Demostración. Para demostrar este teorema necesitamos el siguiente lema,

Lema 1. *Para cada $f \in L_2^n$, tenemos que*

$$\mathbb{V}ar[\langle f, Y\rangle] = \sum_{k\geq 1} \mathbb{V}ar\big[\langle f_{(k)}, Y_{(k)}\rangle\big]\,.$$

[Demostración del Lema 1] Primero vemos que como $Y_{1(k)}, \dots, Y_{n(k)}$ son independientes, $\mathbb{C}ov\big[Y_{(k)}\big] = \sigma^2\lambda_k I$. De hecho, para cada $k \geq 1$, y para cada $i = 1, \dots, n$ tenemos $\mathbb{V}ar\big[Y_{i(k)}\big] = \sigma^2\lambda_k$. Para ver esto, vemos que $\mathbb{V}ar\big[Y_{i(k)}\big] = \sigma^2\mathbb{V}ar\big(\epsilon_{i(k)}\big)$, y utilizando la ecuación (3) tenemos que

$$\mathbb{V}ar\big(\epsilon_{i(k)}\big) = \mathbb{E}[\langle \epsilon_i, \varphi_k\rangle^2] = \langle T\varphi_k, \varphi_k\rangle\,.$$

Como φ_k es la función propia ortonormal de *T* asociada con λ_k, le sigue el resultado.

Empleando la definición del producto interno en L_2^n, y la in dependencia de Y_i tenemos $\mathbb{V}ar[\langle f, Y\rangle] = \sum_{i=1}^n \mathbb{V}ar[\langle f_i, Y_i\rangle]$. Pero $\mathbb{V}ar[\langle f_i, Y_i\rangle] = \sigma^2\mathbb{E}[\langle f_i, \epsilon_i\rangle^2]$, y por la ecuación (3), es igual a $\sigma^2\langle Tf_i, f_i\rangle$. Ahora utilizando la expansión de f_i en la base ortonormal (φ_k), y el hecho de que φ_k son funciones propias de *T* tenemos

$$\mathbb{V}ar[\langle f_i, Y_i \rangle] = \sigma^2 \sum_{k\geq 1} \lambda_k f_{i(k)}^2 \,,$$

donde $f_{i(k)}^2 = \langle f_i, \varphi_k \rangle$. Esto muestra que $\mathbb{V}ar[\langle f, Y \rangle] = \sigma^2 \sum_{k\geq 1} \lambda_k f'_{(k)} f_{(k)}$, y de este modo

$$\mathbb{V}ar[\langle f, Y \rangle] = \sum_{k\geq 1} \mathbb{V}ar\left[\langle f_{(k)}, Y_{(k)} \rangle\right] ,$$

el cual prueba el Lema.

Supongamos que hay un $f^* \in L_2^n$ tal que $\mathbb{E}[\langle f^*, Y \rangle] = \langle g, \beta \rangle$ y $\mathbb{V}ar[\langle f^*, Y \rangle] \leq \mathbb{V}ar[\langle h, Y \rangle]$, para cada $h \in L_2^n$ tal que $\mathbb{E}[\langle h, Y \rangle] = \langle g, \beta \rangle$.

Como la igualdad $\mathbb{E}[\langle f^*, Y \rangle] = \langle g, \beta \rangle$ debería tener para cada β, f^* tiene que satisfacer $X' f^* = g$. En otras palabras, debería satisfacer

$$\forall k \geq 1 \qquad X' f_{(k)}^* = g_{(k)} \,. \tag{8}$$

Por el Lema 1, $\mathbb{V}ar[\langle f, Y \rangle] = \sum_{k\geq 1} \mathbb{V}ar\left[\langle f_{(k)}, Y_{(k)} \rangle\right]$. De este modo, para minimizar esta varianza, es suficiente minimizar cada término de la serie. Para cada $k \geq 1$ fijado, usamos el multiplicador de Lagrange para minimizar el término k-ésimo de la serie, sujeto a las condiciones (8), y obtenemos

$$f_{(k)}^* = X(X'X)^{-1} Y_{(k)} \,.$$

Y como esto es cierto para todo $k \geq 1$, tenemos que $f^* = X(X'X)^{-1}Y$ que es lo que buscamos. □

El resultado del teorema está dado en términos producto interno habitual en L_2 para enfatizar que generaliza el resultado conocido en el caso finito dimensional. El mismo argumento empleado para la demostración se puede emplear si $\langle .,. \rangle$ es sustituido por $\langle .,. \rangle_T$. La siguiente sección está dedicada a otro método de estimación.

4.3. El Método de Máxima Verosimilitud

En esta sección asumimos que $\in$, del modelo (6) es Gausiana, y denota la $n \times 1$ matriz $X\beta$ por $m = (m_1, \dots, m_n)'$. Sean $\mathcal{L}(Y_1) = \nu$ y $\mathcal{L}(\sigma\epsilon_1) = \mu$ las distribuciones de probabilidad de Y_1 y $\sigma\epsilon_1$ respectivamente. Estas son las dos medidas de probabilidad Gausianas en $\left(L_2, B_{L_2}\right)$. Como anteriormente, T denota el operador de covarianza de ϵ_1, y $\sqrt{T}$ denota el operador definido por

$$\forall k \geq 1\,, \quad \sqrt{T}\varphi_k = \sqrt{\lambda_k}\varphi_k\,.$$

Es bien conocido [Grenander U. 1981; Kuo H. 1970] que las distribuciones de dos procesos Gausianos que tienen el mismo operador de covarianza son equivalentes u ortogonales. Se puede hallar una expresión para la derivada de Radon-Nikodym de una probabilidad con respecto a otra cuando son equivalentes. Una condición necesaria y suficiente para que exista la derivada de Radon-Nikodym de ν con respecto a μ es que m pertenezca al rango $\left(\sqrt{T}\right)$. Esta condición es en realidad equivalente a $m \in \mathrm{H}_T$ (Fernique X. 1985). Hemos visto que para estimar m no hay pérdida de generalidad en suponer que $m \in \mathrm{H}_T$. Bajo esta condición, en el próximo lema, se da la derivada de Radon-Nikodym $\partial\nu/\partial\mu$.

Lema 2. *Asumiendo que la medida de probabilidad* ν *es absolutamente continua con respecto a la medida de probabilidad* μ *entonces tenemos que*

$$\frac{\partial\nu}{\partial\mu}(x) = exp\left[-\frac{1}{2}\sum_{k\geq 1} m_k^{*2} + \sum_{k\geq 1} m_k^* x_k^*\right], \quad \forall x \in H_T ,$$

donde

$$m_j^* = \langle m, \varphi_j^* \rangle \ , \qquad x_j^* = \langle x, \varphi_j^* \rangle \ , \qquad y \ \ \varphi_j^* = \sigma\sqrt{\lambda_j}\varphi_j \ .$$

La función de verosimilitud $\ell = \partial\nu/\partial\mu$ es definida por:

$$\forall y \in L_2^n \ , \quad \ell(m, y) = exp\left[-\frac{1}{2}\sum_{j\geq 1}\sum_{i=1}^{n} m_{i(j)}^{*2} + \sum_{k\geq 1}\sum_{i=1}^{n} m_{i(j)}^* y_{i(j)}^*\right]$$

$$= \prod_{j\geq 1} exp\left\{-\frac{1}{2}\left[m_{(j)}^{*\prime} m_{(j)}^* - 2m_{(j)}^* y_{(j)}^*\right]\right\}$$

$$= \prod_{j\geq 1} exp\left\{-\frac{\sigma^2\lambda_j}{2}\left[m_{(j)}' m_{(j)} - 2m_{(j)} y_{(j)}\right]\right\} ,$$

donde $m_{(j)} = \left(m_{1(j)}, \dots, m_{n(j)}\right)'$. Sustituyendo $m(t)$ por $X\beta$ tenemos que

$$\forall y \in L_2^n \ , \quad \ell(\beta, y) = \prod_{j\geq 1} exp\left\{-\frac{\sigma^2\lambda_j}{2}\left[\beta_{(j)}' X' X \beta_{(j)} - 2\beta'_{\ (j)} X' y_{(j)}\right]\right\}.$$

Notar que, para cada $j \geq 1$,

$$\beta_{(j)}' X' X \beta_{(j)} - 2\beta'_{\ (j)} X' y_{(j)} = \left(y_{(j)} - X\beta_{(j)}\right)'\left(y_{(j)} - X\beta_{(j)}\right) - y_{(j)}' y_{(j)} \ ,$$

De manera que maximizar $\ell(\beta, y)$ con respecto de β, independientemente de σ, es equivalente a minimizar

$$\left(y_{(j)} - X\beta_{(j)}\right)'\left(y_{(j)} - X\beta_{(j)}\right).$$

De ahí tenemos el resultado siguiente

Teorema 4. Visto como una función de β, $\ell(\beta, \mathrm{Y})$ es maximizada por

$$\hat{\beta} = (X'X)^{-1}X'Y\,.$$

Comentarios de los métodos empleados para dar un estimador de β:

Comentario 1. *Los dos métodos de estimación, Mínimos Cuadrados y Máxima Verosimilitud, en el caso finito dimensional tienden al mismo estimador de* β*. Sin embargo mientras el primero siempre es aplicable el último no. En concreto,*

- *a- En el caso finito dimensional, el método de Mínimos Cuadrados no requiere normalidad. Además, no requiere ningún conocimiento de la estructura de covarianza.*
- *b- A diferencia del caso finito dimensional, en espacios dimensionales infinitos no hay equivalencia directa a la medida de Lebesque en* $\mathbb{R}^n$*. Sin embargo, si el error es Gausiano esta dificultad puede ser sorteada dejando la derivada de Radon-Nikodym (de la distribución de la observación con respecto a la distribución del error) desempeñar el papel de la función de verosimilitud. Si el error no es Gausiano no es fácil solventar esta dificultad.*

El producto interno $\langle .,. \rangle_{\mathrm{T}}$ no es artificial. Aparece de forma natural en la aproximación de la reproducción del espacio kernel de Hilbert. Además, la función de verosimilitud utilizando $\langle .,. \rangle_{\mathrm{T}}$ es maximizada en

$$\sum_{j\geq 1} \lambda Y'_{(j)} Y_{(j)} - \sum_{j\geq 1} \lambda_j \left(Y_{(j)} - X\beta_{(j)}\right)' \left(Y_{(j)} - X\beta_{(j)}\right).$$

En términos del producto interno $\langle .,. \rangle_{\mathrm{T}}$, fácilmente se puede ver que esta expresión es análoga al estadístico SCR (suma de cuadrados de la regresión) encontrado habitualmente en el análisis de la regresión o el análisis de la varianza. El estadístico SCR se define en la siguiente sección donde se desarrolla la prueba de hipótesis lineal.

5. PRUEBA DE HIPÓTESIS LINEAL

El objetivo de esta sección es hacer inferencia sobre β estableciendo una prueba para la hipótesis $H_0: K\beta = C$. Siguiendo la notación de los modelos lineales clásicos, se definirá la suma de cuadrados del error residual (SCE), la suma de cuadrados de la regresión (SCR), y se establecerán algunas de sus propiedades asumiendo que el error ϵ es Gausiano. Utilizando la definición de $\hat{\beta}$, es fácil ver que

$$Y - X\hat{\beta} = (I - X(X'X)^{-1}X')Y \ ,$$

donde I es la matriz identidad. Sea $M = \left(m_{ij}\right)_{i,j}$ que denota la matriz $(I - X(X'X)^{-1}X')$. Es idempotente, simétrica tal que $MX = O$. Para cada $k \geq 1$, $Y_{(k)}$ es un vector Gausiano $\mathrm{N}\left(\mu_{(k)}, \sigma^2 I\right)$ con media $\mu_{(k)} = X\beta_{(k)}$. Por tanto, $\sigma^{-2}Y'_{(k)}MY_{(k)}$ es una variable aleatoria $\chi^2\left(r(M)\right)$ con $r(M) = rango(M)$ grados de libertad.

5.1. La descomposición de la variabilidad: SCE y SCR

Por analogía con la teoría clásica de los modelos lineales [Searle S.R. 1971] definimos la suma de cuadrados del error (SCE) como:

$$SCE = \langle Y - \hat{Y}, Y - \hat{Y} \rangle_T \ ,$$

donde $\hat{Y} = X\hat{\beta}$. Fácilmente se puede ver que $SCE = \langle Y, MY \rangle_T$. De la definición del producto interno $\langle .,. \rangle_{\mathrm{T}}$ en L_2^n, tenemos que:

$$\langle Y, MY \rangle_T = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle Y_i, m_{ij} Y_j \rangle_T \,.$$

Empleando la expansión de los Y_i y la ortonormalidad de los φ_k, el término general del doble sumatorio se convierte en

$$\langle Y_i, m_{ij} Y_j \rangle_T = \sum_{k \geq 1} \lambda_k \langle Y_i, \varphi_k \rangle m_{ij} \langle Y_j, \varphi_k \rangle \,.$$

Reemplazando esto en la fórmula anterior, obtenemos

$$\langle Y, MY \rangle_T = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \geq 1} \lambda_k \langle Y_i, \varphi_k \rangle m_{ij} \langle Y_j, \varphi_k \rangle$$

$$= \sum_{k \geq 1} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_k \langle Y_i, \varphi_k \rangle m_{ij} \langle Y_j, \varphi_k \rangle$$

$$= \sum_{k \geq 1} \lambda_k Y'_{(k)} M Y_{(k)} \,.$$

Esto muestra que $\langle Y, MY \rangle_T = \sum_{k \geq 1} \lambda_k Y'_{(k)} M Y_{(k)}$. También definimos la suma de cuadrados total (SCT) por $SCT = \langle Y, Y \rangle_T$. Entonces la suma de cuadrados de la regresión (SCR) es la diferencia SCT – SCR, y tenemos $SCR = \langle Y, BY \rangle_T$, donde $B = X(X'X)^{-1}X'$. Es fácil ver que la matriz B es simétrica, idempotente, y ortogonal en M. Ahora podemos plantear y probar los resultados siguientes.

Teorema 5. *Asumiendo que el error ϵ es Gausiano, tenemos que*

1- *Existe una secuencia (η_k) de variables aleatorias independientes con la misma distribución de la $\chi^2_{r(M)}$, tal que la serie $\sigma^2 \sum_{k\geq 1} \lambda_k^2 \eta_k$ converge a SCE casi seguramente.*

2- *Existe una secuencia (ξ_k) de variables aleatorias reales independientes tal que cada ξ_k tiene una distribución no-central de la $\chi^{2'}\left(r(B), (2\sigma^2\lambda_k)^{-1}\mu'_{(k)}B\mu_{(k)}\right)$, y la serie $\sigma^2 \sum_{k\geq 1} \lambda_k^2 \xi_k$ converge a SCR casi seguramente. También,*

3- *La SCE y SCR son independientes.*

Demostración.

1- Hemos visto que

$$SCE = \sum_{k\geq 1} \lambda_k Y'_{(k)} M Y_{(k)} = \sigma^2 \sum_{k\geq 1} \lambda_k^2 \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right)' M \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right).$$

Como $\mu'_{(k)}M = 0$, $\eta_k = \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right)' M \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right)$ tiene una distribución de la $\chi^2_{r(M)}$ con $r(M)$ grados de libertad.

Por otro lado, $\mathbb{E}[SCE] = \mathbb{E}\langle Y, MY\rangle_T$ es finito. Por tanto la SCE es finita casi seguramente.

2- Utilizando el mismo argumento que para la SCE tenemos que

$$SCR = \sum_{k\geq 1} \lambda_k Y'_{(k)} B Y_{(k)} = \sigma^2 \sum_{k\geq 1} \lambda_k^2 \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right)' B \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right).$$

De este modo $\xi_k = \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right)' B \left(\frac{1}{\sigma\sqrt{\lambda_k}} Y_{(k)}\right)$ es una $\chi^{2\prime}$ no-central con $r(B)$ grados de libertad y un parámetro de no-centralidad $(\sigma^2\lambda_k)^{-1}\mu'_{(k)}B\mu_{(k)}$. Como $\mathbb{E}[SCR] = \mathbb{E}\langle Y, BY\rangle_T < \infty$, la SCR es finita casi seguramente.

3- Es suficiente mostrar que las secuencias $(\eta_k)_{k\geq 1}$ y $(\xi_k)_{\xi\geq 1}$ son mutuamente independientes. Como $(Y_k)_{k\geq 1}$ es una secuencia de vectores aleatorios independientes y como η_k y ξ_k son sólo funciones de Y_k, es suficiente mostrar que para cada $k \geq 1$, η_k y ξ_k son independientes.

Comentario 2. *El estadístico* $\frac{SCE}{[r(M)\sum_k \lambda_k^2]}$ *es un estimador insesgado y consistente de* σ^2*.*

5.2. Prueba de hipótesis lineal

Consideramos la hipótesis general $H_0: K\beta = C$, donde K es una matriz $m \times p$ y C es un vector de $m \times 1$ funciones dadas. Asumimos que K es de rango completo por filas, es decir, $\mathrm{rango}(K) = m$. Para probar este tipo de hipótesis necesitamos calcular un estimador de β bajo H_0. Veremos que el método de mínimos cuadrados no natural proporciona tal estimador, $\tilde{\beta}$. El estimador de mínimos cuadrados no natural se obtiene minimizando $\|e(\beta)\|_T^2$ bajo la condición $K\beta - C = 0$, que es equivalente a

$$K\beta_{(k)} - C_{(k)} = 0 \qquad \forall k \geq 1\,. \tag{9}$$

Para minimizar $\|e(\beta)\|_T^2$ sujeto a (9), es suficiente minimizar para cada $k \geq 1$, $\left(Y_{(k)} - X\beta_{(k)}\right)'\left(Y_{(k)} - X\beta_{(k)}\right)$ sujeto a la condición $K\beta_{(k)} - C_{(k)} = 0$

Usando los vectores del multiplicador de Lagrange, se puede ver fácilmente que

$$\tilde{\beta} = \hat{\beta} - (X'X)^{-1}K'[K(X'X)^{-1}K']^{-1}\left(K\hat{\beta} - C\right). \tag{10}$$

Notemos que como $rango(K) = m \leq p = rango((X'X)^{-1})$, el $rango(K(X'X)^{-1}K')$ es igual a m. Por lo tanto $K(X'X)^{-1}K'$ es de rango completo, y de este modo invertible.

Comentario 3. *Recordemos que el estimador de máxima verosimilitud no natural obtenido maximizando la función de verosimilitud sujeta a (9) es igual a $\tilde{\beta}$.*

Sea SCE_{H_0} la suma de cuadrados del error bajo la hipótesis $H_0{:}\, K\beta = C$, es decir, $SCE_{H_0} = \left\|Y - X\tilde{\beta}\right\|_T^2$.

Escribiendo $Y - X\tilde{\beta} = Y - X\hat{\beta} + X\left(\hat{\beta} - \tilde{\beta}\right)$, y como $X'\left(Y - X\hat{\beta}\right) = 0$, sigue

$$SCE_{H_0} = \left\|Y - X\tilde{\beta}\right\|_T^2 + \left\|X\left(\hat{\beta} - \tilde{\beta}\right)\right\|_T^2. \tag{11}$$

De (10) y los mismos cálculos que en análisis de la varianza clásico, obtenemos

$$X(\hat{\beta} - \tilde{\beta}) = D(Y - C^*)\ ,$$

donde

$D = X(X'X)^{-1}K'[K(X'X)^{-1}K']^{-1}[K(X'X)^{-1}K']$, y $C^* = XK'(KK')^{-1}C$.

Notamos que D es idempotente, simétrica y $DM = MD = 0$. De este modo ahora podemos establecer algunas propiedades del estadístico $Q = \|X(\hat{\beta} - \tilde{\beta})\|_T^2$. Estas propiedades pueden ser probadas siguiendo la demostración del Teorema 5.

Teorema 6. *Asumiendo que el error ϵ es Gausiano, tenemos que*

a- Existe una secuencia (ζ_k) de variables aleatorias reales independientes tal que cada ζ_k tiene una distribución no-central de la $\chi^{2'}(r(D), \delta_k)$ con $r(D)$ grados de libertad y un parámetro δ_k no centrado dado por

$$\delta_k = (2\sigma^2\lambda_k)^{-1}\left(\mathbb{E}Y_{(k)} - C^*_{(k)}\right)' D\left(\mathbb{E}Y_{(k)} - C^*_{(k)}\right)\ ,$$

Tal que la serie $\sigma^2 \sum_k \lambda_k^2 \zeta_k$ converge a $Q = \|X(\hat{\beta} - \tilde{\beta})\|_T^2$ casi seguramente.

b- Las variables aleatorias reales Q y SCE son estocásticamente independientes.

La suma de cuadrados debido a la regresión, bajo la hipótesis $H_0: K\beta = C$, está definida por $SCR_{H_0} = SCT - SCE_{H_0}$. La ecuación (11) muestra que $SCR_{H_0} - SCR = Q$. De este modo tenemos el resultado siguiente:

Teorema 7. *Para probar $H_0: K\beta = C$ versus $H_1: K\beta \neq C$, con un nivel α, existe una prueba ψ que viene dada por:*

$$\psi = \begin{cases} 1 & si \;\; S_{H_0}(Y) > C(H_0, \alpha) \\ 0 & en\; otro\; caso \end{cases},$$

La constante $C(H_0, \alpha)$ es tal que $\mathbb{P}\left\{S_{H_0}(Y) > C(H_0, \alpha), K\beta = C\right\} = \alpha$.

6. BIBLIOGRAFÍA

1. Abramovich F. and Angelini C. (2006). Testing in mixed-effects FANOVA models. J. Statist. Plann. Inference 136, 4326-4348.
2. Abramovich F., Antoniadis A., Sapatinas T. and Vidakovic B. (2002). Optimal testing in Functional Analysis of Variance models. Research Report IMAG- LMC, RR 1044-M.
3. Antoniadis A. (1984). Analysis of variance on function spaces. Math. Operationsforsch. Statist., Ser. Statist., 15, 59-71.
4. Comas C., Delicado P., Giraldo R. and Mateu J. (2009). Statistics for spatial functional data. Environmetrics.
5. Conway J.B. (1989). A Course in Functional Analysis, Second Edition. Springer-Verlag.
6. Cuesta-Albertos J. and Febrero-Bande M. (2007). A simple multiway anova for functional data. Business and Economics, 19:537–557.
7. Cuevas A., Febrero M., and Fraiman R. (2004). An anova test for functional data. Computational statistics and data analysis. 47:111-12.
8. Cuevas A., Febrero M. and Fraiman R. (2006). On the use of the bootstrap for estimating functions with functional data. Computational Statistics and Data Analysis. 51, 1063-1074.
9. Cuevas A., Febrero M. and Fraiman R. (2007). Robust estimation and classification for functional data via projection-based depth notions. Computational Statistics. 22(3), 481-496.

10. Delicado P. (2007). Functional k-sample problem when data are density functions. Computational Statistics, 22 (3) 391-410.
11. Delicado P., Giraldo R. and Mateu J. (2007). Geostatistics for functional data: An ordinary kriging approach. Environmental and Ecological Statistics.
12. Febrero M., Galeano P. and González-Manteiga W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. Environmetrics 19, 331-345.
13. Fernique X. (1985). Gaussian Random Vectors Their Reproducing Kernel and Hilbert Spaces Technical Report Series of the Laboratory for Research in Statistics and Probability, No 34. University of Otawa and Carleton University.
14. Ferraty F. and Romain Y. (2010). The Oxford Handbook of Functional Data Analysis. Oxford University Press. Oxford, England.
15. Ferraty F. and Vieu P. (2006). Non parametric functional data analysis. Theory and practice. Springer. New York.
16. Flanagan R., Ramsay J.O. and Wang X. (1995). A functional data analysis of the pinch force of human fingers. The Royal Statistical Society, 44(1):17–39.
17. Fraiman R. and Muniz G. (2001). Trimmed means for functional data. Test 10(2), 419-440.
18. Grenander U. (1981). Abstract inference. John Wiley & Sons, Inc.
19. Huang J.Z. (1998). Functional ANOVA Models for Generalized Regression. Journal of Multivariate Analysis 67, 49-71.

20. Huang J.Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. The Annals of Statistics 26, 242-272.
21. Kaufman C. and Sain S. (2010). Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions. Bayesian Analysis, 5: 123-150.
22. Kaziska D.M. (2011). Functional Analysis of Variance, Discriminant Analysis, and Clustering in a Manifold of Elastic Curves. Communications in Statistics - Theory and Methods. Vol. 40, Iss. 14, 2487-2499.
23. Kuo H. (1970). Gaussian Measures in Banach Sapces. Lecture Notes in Mathematics 463. Springer-Verlag.
24. Morris J.R. and Carroll R.J. (2006). Wavelet-based functional mixed models. J. Royal Stat. Soc. B, 68, 179-199.
25. Parzen E. (1961). An approach to time series analysis. Ann. Math. Statist., 32, 951-989.
26. Ramsay J.O. and Dalzell C.J. (1991). Some tools for functional data analysis. J Roy Statist Soc. B53:539-572.
27. Ramsay J.O. and Silverman B.W. (1997). Functional data analysis. Springer. New York: Springer-Verlag.
28. Ramsay J.O. and Silverman B.W. (2002). Applied Functional Data Analysis. New York: Springer-Verlag.

29. Roosen C. (1995). Visualization and Exploration of High-Dimensional Functions Using the Functional Anova Decomposition. Ph. D. thesis, Stanford University.
30. Sain S., Nychka D. and Mearns L. (2010). Functional ANOVA and regional climate experiments: a statistical analysis of dynamic downscaling. Environmetrics. 22 (6): 700-711.
31. Searle S.R. (1971). Linear Models. John Wiley & Sons, Inc.
32. Scheffé H. (1959). The Analysis of Variance. John Wiley & Sons, Inc.
33. Shen Q. and Faraway J. (2004). An f test for linear models with functional responses. Statistica Sinica, 14:1239–1257.
34. Spitzner D.J., Marron J.S. and Essick G.K. (2003). Mixed model functional ANOVA for studying human tactile perception. J. Amer. Stat. Assoc. 98, 263-272.
35. Stein M. (1999). Interpolation of Spatial Data. New York: Springer-Verlag.
36. Tarrío J. and Naya S. (2011). Influencia de la adición de nano y microsílice en la estabilidad térmica de una resina epoxi. Aplicaciones del ANOVA funcional. Revista Colombiana de Estadística, vol. 34, núm. 2, pp. 211-230.
37. Zoglat A. (2008). Functional Analysis of Variance. Applied Mathematical Sciences. Vol. 2, no. 23, 1115-1129.