

Estimación y contrastes de hipótesis en modelos de regresión desde la perspectiva del análisis funcional

por

Luis Alberto Ramil Novo y Wenceslao González Manteiga

En este trabajo se analizan los funcionales que en un modelo de regresión dan lugar a la estimación de la función de regresión, m , la varianza del término de error, y la discrepancia de m con respecto a un modelo de regresión paramétrico. El estudio se centra en diversos aspectos comunes de las técnicas de estimación y contraste de hipótesis en los contextos paramétrico y no paramétrico, como son la definición de los estimadores, las ecuaciones normales, el análisis de la varianza, los F tests o los tests de razón de verosimilitudes.

Palabras clave: modelos de regresión; estimadores mínimo cuadráticos; estimadores no paramétricos; contrastes de bondad de ajuste.

1 . INTRODUCCIÓN

Muchos problemas físicos, químicos, biológicos, etc. se describen haciendo uso de relaciones teóricas entre variables, $y = m(x)$, que no se pueden conocer de manera exacta. En la práctica, la aproximación de la función m se hace en base a un conjunto de observaciones $\{(X_i, Y_i)\}_{i=1}^n$, que con frecuencia conllevan un error implícito, ε , de carácter aleatorio,

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \dots, n; \quad (1)$$

En este trabajo consideraremos que las observaciones (X_i, Y_i) son bidimensionales, y errores

$$\varepsilon_i \text{ independientes, de media nula, y varianza constante } \sigma^2. \quad (2)$$

Al modelo matemático (1) que describe la relación entre la variable (respuesta) Y y la variable (explicativa) X se le conoce como modelo de regresión, y a la función m que describe la relación entre ambas variables, se le denomina función de regresión. En el análisis estadístico, tanto la función de regresión, como la varianza del error, σ^2 , son objeto de interés.

En este trabajo se analizan métodos de estimación de m y σ^2 paramétricos y no paramétricos, y contrastes de hipótesis sobre la forma funcional de m basados en el análisis de la varianza. Veamos las ideas de estos procedimientos

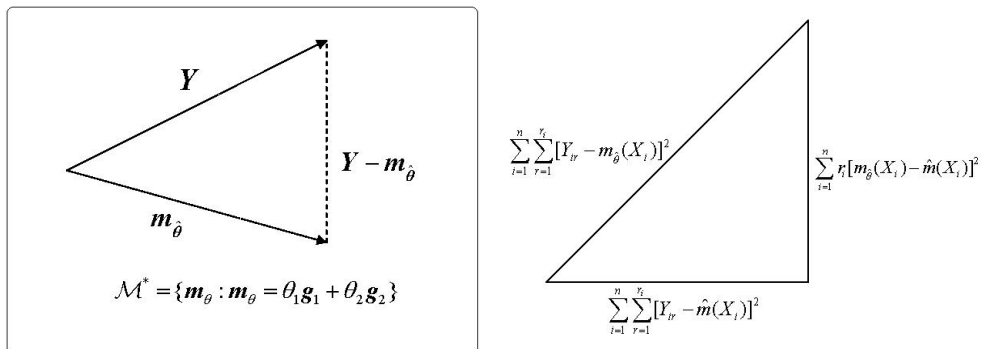


Figura 1: Izquierda: interpretación del estimador mínimo cuadrático. Derecha: análisis de la varianza asociado al F clásico de falta de ajuste.

con un ejemplo relevante, el modelo de regresión lineal general, definido por la igualdad (1), la hipótesis (2) y la relación,

$$m \in \mathcal{M}_0 = \{m_{\theta} : m_{\theta}(x) = \theta_1 g_1(x) + \cdots + \theta_p g_p(x), \theta_j \in \mathbb{R}, j = 1, \dots, p\}, \quad (3)$$

en donde $g_1, \dots, g_p$ son funciones conocidas, y $\theta_1, \dots, \theta_p$ son parámetros desconocidos.

El método de estimación de los parámetros θ_j más utilizado, es el método de los mínimos cuadrados (véase, por ejemplo, Seber (1977)), consistente en minimizar con respecto a $\theta = (\theta_1, \dots, \theta_p)^t$ la *suma de cuadrados*,

$$S(\theta) = \sum_{i=1}^n [Y_i - (\theta_1 g_1(X_i) + \cdots + \theta_p g_p(X_i))]^2.$$

Se requiere que $\mathbf{g}_1 = (g_1(X_1), \dots, g_1(X_n))^t, \dots, \mathbf{g}_p = (g_p(X_1), \dots, g_p(X_n))^t$ sean linealmente independientes y que $n \geq p$ para garantizar la existencia y unicidad de la solución, $\hat{\theta}_1, \dots, \hat{\theta}_p$. El vector $\mathbf{m}_{\hat{\theta}} = (m_{\hat{\theta}}(X_1), \dots, m_{\hat{\theta}}(X_n))^t$ (con $m_{\hat{\theta}}(X_i) = \hat{\theta}_1 g_1(X_i) + \cdots + \hat{\theta}_p g_p(X_i)$), es entonces la única proyección ortogonal de $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ sobre $\mathcal{M}^*$, el espacio vectorial generado por los vectores $\mathbf{g}_1, \dots, \mathbf{g}_p$ (véase la Figura 1 - izquierda, correspondiente al caso $p = 2$, y la Sección 3).

Si los errores ε_i siguen una distribución normal (además de satisfacer (2)), dos métodos clásicos (véase Seber (1977)) para contrastar la hipótesis *nula* de validez del modelo lineal definido en (3), $H_0 : m \in \mathcal{M}_0$, frente a la hipótesis *alternativa*, $H_1 : m \notin \mathcal{M}_0$, son los siguientes.

El primero, válido cualquiera que sea m , requiere que para cada X_i estén disponibles varias observaciones independientes (réplicas) de la variable Y ,

$Y_{ir} = m(X_i) + \varepsilon_{ir}$, $i = 1, \dots, n$; $r = 1, \dots, r_i$. El contraste de hipótesis se lleva a cabo mediante el F test de falta de ajuste, basado en el estadístico,

$$F = \frac{N - n}{n - p} \frac{\sum_{i=1}^n \sum_{r=1}^{r_i} [Y_{ir} - m_{\hat{\theta}}(X_i)]^2 - \sum_{i=1}^n \sum_{r=1}^{r_i} [Y_{ir} - \hat{m}(X_i)]^2}{\sum_{i=1}^n \sum_{r=1}^{r_i} [Y_{ir} - \hat{m}(X_i)]^2}, \quad (4)$$

en donde $\hat{m}(X_i) = \bar{Y}_i = \sum_{r=1}^{r_i} Y_{ir}/r_i$ representa una estimación de m bajo H_1 , y $N = \sum_{i=1}^n r_i$, es el tamaño de la muestra. Bajo H_0 , F sigue una distribución F con $n - p$ y $N - n$ grados de libertad, por lo que se rechaza H_0 si $p = \Pr(F_{n-p, N-n} \geq F_{obs})$ es un valor pequeño. Aquí, y en lo sucesivo, F_{obs} denota el valor observado del correspondiente estadístico F .

El segundo método requiere poder asumir que, para cierto $q > p$,

$$m \in \mathcal{M} = \{m_{\theta} : m_{\theta} = \theta_1 g_1(x) + \dots + \theta_q g_q(x), \theta_j \in \mathbb{R}, j = 1, \dots, q\}, \quad (5)$$

en donde $g_1, \dots, g_q$ son funciones conocidas, y $\theta_1, \dots, \theta_q$ son parámetros desconocidos. La hipótesis H_0 postula entonces que m pertenece a un subespacio de $\mathcal{M}$ de menor dimensión, $\mathcal{M}_0$. Entre los tests clásicos de reducción de la dimensionalidad, figuran el test de razón de verosimilitudes y el F test que pasamos a describir.

Consideremos por simplicidad que sólo está disponible una observación de la variable Y por cada valor de la variable X . Si la función de regresión es una función g , la función de densidad de los errores es $f(\varepsilon_i; g, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-\varepsilon_i^2/[2\sigma^2]) = (2\pi\sigma^2)^{-1/2} \exp(-[Y_i - g(X_i)]^2/[2\sigma^2])$, y la función de verosimilitud se define como,

$$L(g, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i; g, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n [Y_i - g(X_i)]^2/[2\sigma^2]\right). \quad (6)$$

El test de razón de verosimilitudes (que se analiza en la Sección 8.1) se basa en el estadístico,

$$\Lambda = \frac{\sup_{g \in \mathcal{M}_0, \sigma^2} L(g, \sigma^2)}{\sup_{g \in \mathcal{M}, \sigma^2} L(g, \sigma^2)} = \left(\frac{\sum_{i=1}^n [Y_i - m_{\hat{\theta}}(X_i)]^2}{\sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2} \right)^{-n/2} = \left(\frac{q-p}{n-q} F + 1 \right)^{-n/2},$$

en donde ahora, $\hat{m}(x) = \hat{\theta}_1 g_1(x) + \dots + \hat{\theta}_q g_q(x)$ y,

$$F = \frac{n - q}{q - p} \frac{\sum_{i=1}^n [y_i - m_{\hat{\theta}}(x_i)]^2 - \sum_{i=1}^n [y_i - \hat{m}(x_i)]^2}{\sum_{i=1}^n [y_i - \hat{m}(x_i)]^2}, \quad (7)$$

es el estadístico de contraste del F test equivalente (que se examina en la Sección 6.2).

Los resultados de los F tests suelen presentarse en las denominadas tablas de análisis de la varianza (o tablas ANOVA) que recogen la relación entre

las sumas de cuadrados utilizadas en la definición del estadístico F . Así por ejemplo, la tabla ANOVA correspondiente al F test (4) es la siguiente (véase la Figura 1 - derecha; en la tabla, $\hat{\sigma}_{FA}^2$ representa una medida de la discrepancia de m con respecto al modelo lineal, y $\hat{\sigma}_{EP}^2$ es la estimación de σ^2).

Tabla 1. *Análisis de la varianza asociado al F test de falta de ajuste*

Variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado medio	Estadístico F
Falta de ajuste	$SS_{FA} = \sum_{i=1}^n r_i [\hat{m}(X_i) - m_{\hat{\theta}}(X_i)]^2$	$n - p$	$\hat{\sigma}_{FA}^2 = \frac{SS_{FA}}{n - p}$	$F = \frac{\hat{\sigma}_{FA}^2}{\hat{\sigma}_{EP}^2}$
Error puro	$SS_{EP} = \sum_{i=1}^n \sum_{r=1}^{r_i} [Y_{ir} - \hat{m}(X_i)]^2$	$N - n$	$\hat{\sigma}_{EP}^2 = \frac{SS_{EP}}{N - n}$	
Residual (bajo H_0)	$SS_R = \sum_{i=1}^n \sum_{r=1}^{r_i} [Y_{ir} - m_{\hat{\theta}}(X_i)]^2$	$N - p$		

En lo que sigue presentamos una caracterización de la función de regresión y de la varianza de los errores (Sección 2) que servirá para introducir, de forma unificada, los estimadores mínimo cuadráticos (Sección 3), y los estimadores no paramétricos *spline suaves* - *smoothing spline* en la terminología anglosajona (Sección 4). La caracterización de ambos estimadores como las funciones que minimizan determinados funcionales, nos permitirá presentar de forma novedosa, y unificando el enfoque paramétrico y no paramétrico (Secciones 5-9), el análisis de la varianza de la regresión, los F tests y el test de razón de verosimilitudes. En la Sección 5 presentamos los fundamentos en los que se basan el análisis de la varianza y los F tests, basados en estimadores mínimo cuadráticos, que analizamos en la Sección 6, y los basados en estimadores *spline* suaves, que se describen en la Sección 7. En la Sección 8 se analizan los tests de razón de verosimilitudes equivalentes a F tests, considerando estimadores mínimo cuadráticos y estimadores *spline* suaves. Por último, en la Sección 9 se analizan estimadores no paramétricos alternativos al estimador *spline* suave para llevar a cabo el análisis de regresión, el análisis de la varianza o los F tests.

1.1 . EJEMPLO

Los métodos de estimación y los contrastes de hipótesis que se presentan en las secciones sucesivas, se ilustrarán con el análisis de los datos representados en la Figura 2.

El gráfico de la izquierda representa los ratios peso/altura de los niños entre cero y seis años de edad de una muestra de familias (Eubank (1988), Pág. 5), $\{(x_i, y_i) : i = 1, \dots, 72\}$, en donde $x_i = i - 0,5$ es la edad en meses, e $y_i = m(x_i) + \varepsilon_i$ es el ratio asociado a x_i . En los gráficos se aprecia que la evolución del ratio peso/altura es, en términos medios, una función lineal de la edad en meses a partir del primer año ($x \geq 12,5$, gráfico de la derecha), pero no desde el nacimiento.

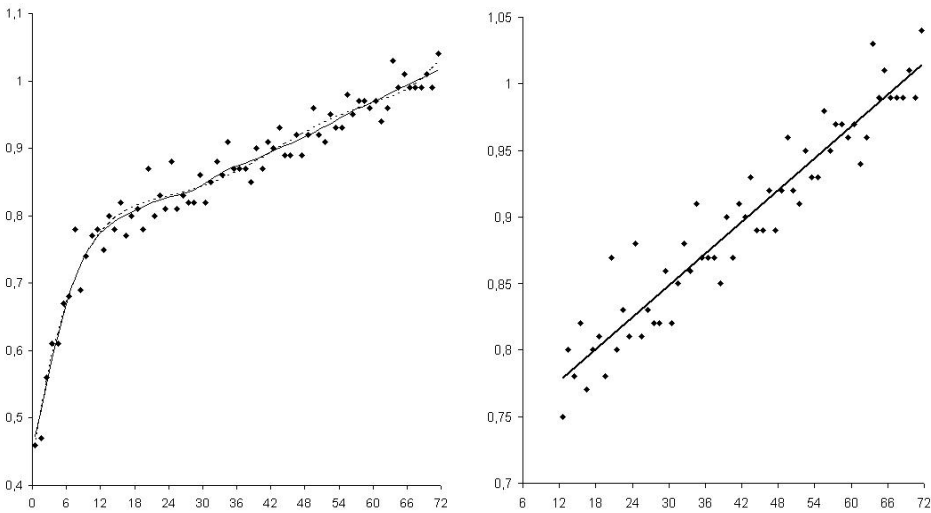


Figura 2: Ratio peso/altura (en libras por pulgada) frente a la edad en meses. Izquierda: muestra completa, de cero a seis años. Incluye la estimación mínimo cuadrática de un polinomio de grado cinco (trazo discontinuo) y el estimador spline suave cúbico de parámetro $\lambda = 0,033^4$ (trazo continuo). Derecha: submuestra correspondiente a una edad superior o igual a un año. Incluye la recta de regresión mínimo cuadrática.

El análisis estadístico, además de las estimaciones paramétrica y no paramétrica de m correspondientes al conjunto de datos global que se muestran en la gráfica de la izquierda, abarca el contraste de la hipótesis de linealidad, $H_0 : m(x) = \theta_0 + \theta_1 x$, a partir del primer año. En la Sección 6.3 se considera el análisis paramétrico clásico, y en la Sección 8.3 el análisis no paramétrico.

2 . LA FUNCIÓN DE REGRESIÓN Y LA VARIANZA DEL ERROR

Consideremos un vector aleatorio (X, Y) tal que $E[Y^2] < \infty$ e Y se puede expresar como $Y = m(X) + \varepsilon$, en donde el error $\varepsilon = Y - m(X)$ es una variable aleatoria de media nula independiente de X .

2.1 . CARACTERIZACIÓN DE LA FUNCIÓN DE REGRESIÓN

Consideremos el problema de predecir Y en función de X . Aplicando las propiedades de la esperanza condicionada, la función de regresión se expresa como la esperanza de Y condicionada a $\{X = x\}$, $m(x) = E[Y/X = x]$. Esta función es también la que proporciona la mejor predicción de Y (véase Shao

(2003, Cáp. 1)), en el sentido de que m minimiza el error cuadrático medio,

$$V(g) = E[(Y - g(X))^2],$$

sobre el espacio de funciones,

$$\mathcal{M} = \{g : g \text{ es una función de Borel con } E[g^2(X)] < \infty\} \equiv \mathcal{B}. \quad (8)$$

Además m es la única solución del problema de minimización anterior, por lo que,

$$m = \arg \min_{g \in \mathcal{M}} V(g).$$

Las definiciones de los estimadores mínimo cuadráticos y *spline* suaves de la función de regresión se fundamentan como veremos en la caracterización anterior.

De acuerdo con esta definición de m , fijada $g \in \mathcal{M}$, la función,

$$N(\alpha) = E[(m(X) + \alpha g - Y)^2] = V(m + \alpha g) \quad \alpha \in \mathbb{R},$$

tiene un mínimo en $\alpha = 0$. La condición necesaria de mínimo, $N'(0) = 0$, conduce a las igualdades,

$$E[(Y - m(X))g(X)] = 0 \quad \text{para toda } g \in \mathcal{M}. \quad (9)$$

Las denominadas ecuaciones normales de la regresión, que son la base para la estimación de la función de regresión, se obtienen, como veremos, de forma análoga a las igualdades anteriores.

2.2 . LA VARIANZA DEL ERROR

La caracterización de m como la función que proporciona el valor mínimo de $V(g)$,

$$E[(Y - m(X))^2] = \min_{g \in \mathcal{M}} E[(Y - g(X))^2], \quad (10)$$

muestra que σ^2 es el valor mínimo del funcional $V(g)$,

$$\sigma^2 = V(m) = \min_{g \in \mathcal{M}} V(g).$$

Esta expresión servirá como referencia para definir los estimadores mínimo cuadráticos y *spline* suaves de σ^2 . De la igualdad (9) se deduce que σ^2 se puede expresar también como,

$$\sigma^2 = \text{cov}(\varepsilon, Y) = E[\varepsilon \cdot Y]. \quad (11)$$

Varios de los estimadores de σ^2 que veremos admiten asimismo una motivación similar.

3 . ESTIMADORES MÍNIMO CUADRÁTICOS

Consideremos una muestra aleatoria $(X_1, Y_1), \dots, (X_n, Y_n)$ de un vector aleatorio (X, Y) que satisface la ecuación (1). Denotamos por $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ y $\mathbf{g} = (g(X_1), \dots, g(X_n))^t$ para cualquier función g ; y por simplicidad, sin que ello suponga pérdida de generalidad, supondremos que el soporte de X es el intervalo $[0, 1]$.

3.1 . ESTIMACIÓN DE LA FUNCIÓN DE REGRESIÓN

En la Estadística clásica o paramétrica, se considera que la función de regresión pertenece a un espacio de funciones paramétrico, $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$. Aquí consideraremos un modelo de regresión lineal general, definido por la igualdad (1), la hipótesis (2) y la relación (5). Supondremos además que $\mathbf{g}_1, \dots, \mathbf{g}_q$ son vectores linealmente independientes.

El estimador mínimo cuadrático de m se obtiene minimizando sobre $\mathcal{M}$ la versión muestral de $V(g)$,

$$\hat{V}(g) = \frac{1}{n} \sum_{i=1}^n [Y_i - g(X_i)]^2 = \frac{1}{n} \|\mathbf{Y} - \mathbf{g}\|^2 \quad \text{para } g \in \mathcal{M},$$

en donde $\|\cdot\|$ es la norma de $\mathbb{R}^n$ inducida por el producto interior $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle = \sum_{i=1}^n z_{1i} z_{2i}$, $\|\mathbf{z}\|^2 = \sum_{i=1}^n z_i^2$. El estimador mínimo cuadrático $\hat{m}$, es la única función en la que se alcanza el mínimo, por lo que,

$$\hat{m} = \arg \min_{g \in \mathcal{M}} \hat{V}(g) = \arg \min_{g \in \mathcal{M}} \frac{1}{n} \|\mathbf{Y} - \mathbf{g}\|^2, \quad \text{y} \quad \|\mathbf{Y} - \hat{\mathbf{m}}\|^2 = \min_{g \in \mathcal{M}} \|\mathbf{Y} - \mathbf{g}\|^2.$$

En consecuencia, el vector de estimaciones $\hat{\mathbf{m}}$ es la proyección de $\mathbf{Y}$ sobre el espacio vectorial generado por $\mathbf{g}_1, \dots, \mathbf{g}_q$,

$$\mathcal{M}^* = \{\mathbf{m}_\theta : \mathbf{m}_\theta = \theta_1 \mathbf{g}_1 + \dots + \theta_q \mathbf{g}_q, \theta_j \in \mathbb{R}, j = 1, \dots, q\}.$$

Por ser $\mathcal{M}$ de dimensión finita, una función $g = \sum_{j=1}^q \theta_j g_j$, queda caracterizada por el vector de parámetros $\theta = (\theta_1, \dots, \theta_q)^t$ por lo que $V(g)$ y $\hat{V}(g)$ son en realidad funciones de θ ,

$$V(\theta) = E[(Y - \sum_{j=1}^q \theta_j g_j(X))^2] \quad \text{y} \quad \hat{V}(\theta) = \frac{1}{n} \sum_{i=1}^n [Y_i - \sum_{j=1}^q \theta_j g_j(X_i)]^2.$$

3.2 . ECUACIONES NORMALES

Las ecuaciones normales que caracterizan al estimador se deducen de la siguiente manera. Fijada $g \in \mathcal{M}$ y la función,

$$\hat{N}(\alpha) = \hat{V}(\hat{m} + \alpha g) = \hat{E}[(\hat{m}(X) + \alpha g - Y)^2] \quad \alpha \in \mathbb{R},$$

la condición necesaria de mínimo, $N'(0) = 0$, equivale a las ecuaciones normales de la regresión lineal (análogas muestrales a las ecuaciones (9)),

$$\langle \mathbf{Y} - \hat{\mathbf{m}}, \mathbf{g} \rangle = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)] g(X_i) = 0 \quad \text{para toda } g \in \mathcal{M}.$$

De estas ecuaciones se deduce que $\hat{\mathbf{m}} = \mathbf{P}\mathbf{Y}$ en donde, $\mathbf{P} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ es la llamada matriz de proyección, que es una matriz simétrica e idempotente, y $\mathbf{X} = (\mathbf{g}_1, \dots, \mathbf{g}_q)$.

3.3 . ESTIMACIÓN DE LA VARIANZA

El estimador mínimo cuadrático de la varianza del error se obtiene al evaluar $\hat{V}(g)$ en $\hat{m}$,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2 = \min_{g \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n [Y_i - g(X_i)]^2. \quad (12)$$

El estimador resultante admite pues las siguientes interpretaciones,

$$\hat{\sigma}^2 = \hat{V}(\hat{m}) = \min_{g \in \mathcal{M}} \hat{V}(g) = \min_{g \in \mathcal{M}} \frac{1}{n} \|\mathbf{Y} - \mathbf{g}\|^2 = \frac{1}{n} \|\mathbf{Y} - \hat{\mathbf{m}}\|^2,$$

en donde $\|\mathbf{Y} - \hat{\mathbf{m}}\|^2 = \|\mathbf{Y} - \mathbf{P}\mathbf{Y}\|^2 = \mathbf{Y}^t(\mathbf{I} - \mathbf{P})^t(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}^t(\mathbf{I} - \mathbf{P})\mathbf{Y}$. Además, la expresión $\hat{\sigma}^2 = n^{-1}\mathbf{Y}^t(\mathbf{I} - \mathbf{P})\mathbf{Y} = n^{-1} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]Y_i$, permite interpretar $\hat{\sigma}^2$ como una versión muestral de la ecuación (11).

4 . ESTIMADORES SPLINE SUAVES

El estudio de métodos de aproximación de funciones, cuyas aplicaciones son numerosas en diversas ciencias (Química, Física,...), tiene ya una extensa y rica historia tanto en el campo de la Estadística como en el de la Matemática Aplicada. En términos generales se pueden distinguir dos tipos de problemas de aproximación de funciones (véase Schumaker (1993)).

El primero lo constituyen los problemas descritos en la introducción (Sección 1). Estos problemas, conocidos como problemas de ajuste de datos, son frecuentes en la mayoría de las ciencias experimentales. En muchos casos requieren una solución de naturaleza estadística.

Los problemas del segundo tipo, usualmente, no tienen componente estadística y son problemas que se resuelven por métodos puramente analíticos (ecuaciones diferenciales, integrales, etc.). En muchos de estos problemas está resuelto el estudio de existencia y unicidad de la solución, pero sólo se pueden resolver de forma explícita algunos casos sencillos, por lo que en la

práctica se recurre a una solución aproximada. El estudio de las clases de funciones razonables para llevar a cabo la aproximación es el objetivo de la Teoría de la Aproximación, mientras que el Análisis Numérico se ocupa del diseño y análisis de algoritmos de aproximación eficientes basados en las clases de funciones apropiadas.

Una clase de funciones que han resultado de gran utilidad tanto en problemas de ajuste de datos como en la obtención de soluciones numéricas son las funciones *spline*. Nos ocupamos aquí de las denominadas funciones *spline* suaves, cuyo uso en inferencia estadística no paramétrica ha dado lugar a numerosos artículos científicos y varias monografías (véase, por ejemplo, Eubank (1988)).

4.1 . ESTIMACIÓN DE LA FUNCIÓN DE REGRESIÓN

Consideremos el espacio de funciones de Sobolev de orden $(2, p)$, que denotamos por $\mathcal{W}_2^p[0, 1]$ y definimos como

$$\left\{ g : g^{(j)} \text{ es absolutamente continua para } 0 \leq j \leq p-1, \int_0^1 [g^{(p)}(x)]^2 dx < \infty \right\}.$$

Si para una función $g \in \mathcal{W}_2^p[0, 1]$, consideramos $J(g) = \int_0^1 [g^{(p)}(x)]^2 dx$ para medir el grado de suavidad de la función, y $\hat{V}(g)$ como medida del grado de ajuste de la función a los datos, el funcional $E(g) = (1 - \alpha)\hat{V}(g) + \alpha J(g)$, con $0 < \alpha < 1$ una constante, pondera ambos aspectos.

Se denomina estimador *spline* suave de orden $(2, p)$ a la función $\hat{m}$ que se obtiene minimizando sobre $\mathcal{M} = \mathcal{W}_2^p[0, 1]$ el funcional ($\lambda > 0$ una constante fija),

$$\hat{V}(g, \lambda) = \frac{1}{n} \sum_{i=1}^n [Y_i - g(X_i)]^2 + \lambda \int_0^1 [g^{(p)}(x)]^2 dx.$$

O sea,

$$\hat{m} = \arg \min_{g \in \mathcal{M}} \hat{V}(g, \lambda).$$

Obsérvese que $\hat{V}(g, \lambda)$ es $\hat{V}(g)$ más el término de penalización $\lambda J(g)$, que garantiza la existencia y unicidad de la solución si $n \geq p$. Además $\hat{V}(g, \lambda)$ equivale a $E(g)$ haciendo $\lambda = \alpha/(1 - \alpha)$. Por tanto, la constante λ , conocida como parámetro de suavización, mide la relación entre el peso que se da a la suavidad de la curva y el peso que se da a la bondad del ajuste, para obtener $\hat{m}$. Cuanto mayor sea λ , más suave será el estimador $\hat{m}$ resultante.

El estimador *spline* suave admite una interpretación geométrica similar a la del estimador mínimo cuadrático, en el espacio producto $\mathcal{H} = \mathbb{R}^n \times L_2[0, 1]$. En $\mathcal{H}$, los datos quedan caracterizados por el par formado por el vector $\mathbf{Y}$ y la función cero, $\mathbf{Y}^* = (\mathbf{Y}, 0)$, y las funciones quedan caracterizadas por la

ecuación diferencial que define el par $(\mathbf{g}, g^{(p)})$ (véanse Cox (1983, 1984)). El espacio $\mathcal{H}$ es un espacio de Hilbert con el producto interior $\langle (\mathbf{z}_1, g_1), (\mathbf{z}_2, g_2) \rangle_{\mathcal{H}} = \sum_{i=1}^n z_{1i} z_{2i} + n\lambda \int_0^1 g_1(x) g_2(x) dx$, que induce la norma,

$$\|(\mathbf{z}, g)\|_{\mathcal{H}}^2 = \sum_{i=1}^n z_i^2 + n\lambda \int_0^1 g^2(x) dx, \quad (\mathbf{z}, g) = ((z_1, \dots, z_n)^t, g) \in \mathcal{H}.$$

En consecuencia,

$$\hat{V}(g, \lambda) = \hat{V}(g) + \lambda \int_0^1 [g^{(p)}(x)]^2 dx = \frac{1}{n} \|(\mathbf{Y} - \mathbf{g}, g^{(p)})\|_{\mathcal{H}}^2 = \frac{1}{n} \|\mathbf{Y}^* - g^*\|_{\mathcal{H}}^2,$$

mide la distancia entre $\mathbf{Y}^*$ y g^* ($d^2(\mathbf{Y}^*, g^*) = \|\mathbf{Y}^* - g^*\|_{\mathcal{H}}^2 = n\hat{V}(g, \lambda)$). El estimador *spline* suave visto en el espacio $\mathcal{M}^* = \mathcal{W}_2^{p*}[0, 1] = \{g^* = (\mathbf{g}, g^{(p)}) : g \in \mathcal{W}_2^p[0, 1]\}$, que es un espacio de Hilbert (véase Cox (1984)), es

$$\hat{m}^* = \arg \min_{g^* \in \mathcal{M}^*} \frac{1}{n} \|\mathbf{Y}^* - g^*\|_{\mathcal{H}}^2,$$

ya que, por la propia definición de $\hat{m}$,

$$\begin{aligned} \frac{1}{n} \|(\mathbf{Y} - \hat{\mathbf{m}}, \hat{m}^{(p)})\|_{\mathcal{H}}^2 &= \min_{g \in \mathcal{M}} \frac{1}{n} \|(\mathbf{Y} - \mathbf{g}, g^{(p)})\|_{\mathcal{H}}^2 \\ \Leftrightarrow \frac{1}{n} \|\mathbf{Y}^* - \hat{m}^*\|_{\mathcal{H}}^2 &= \min_{g^* \in \mathcal{M}^*} \frac{1}{n} \|\mathbf{Y}^* - g^*\|_{\mathcal{H}}^2. \end{aligned}$$

En consecuencia, $\hat{m}^*$ es la proyección de $\mathbf{Y}^*$ sobre $\mathcal{M}^* = \mathcal{W}_2^{p*}[0, 1]$ (véase la Figura 3 - izquierda).

El estimador *spline* suave pertenece al espacio $\mathcal{NS}^{2p}(X_1, \dots, X_n)$ de *splines* naturales s tales que s es una función polinómica de grado $2p-1$ en cada subintervalo $[X_i, X_{i+1})$, las derivadas de s hasta el orden $2p-2$ existen y son continuas, s tiene derivada de orden $2p-1$ que es una función a trozos con saltos en los X_i y s es un polinomio de grado p fuera de $[X_{(1)}, X_{(n)}]$. Demmler y Reinsch (1975) proporcionaron una base $\{\varphi_{kn}\}_{k=1}^n$ del espacio $\mathcal{NS}^{2p}(X_1, \dots, X_n)$ que satisface $n^{-1} \sum_{i=1}^n \varphi_{jn}(X_i) \varphi_{kn}(X_i) = \delta_{jk}$ y $\int_0^1 \varphi_{jn}^{(p)}(x) \varphi_{kn}^{(p)}(x) dx = \gamma_{kn} \delta_{jk}$ con $0 \leq \gamma_{1n} \leq \dots \leq \gamma_{pn} \leq \gamma_{p+1,n} \leq \dots \leq \gamma_{nn}$. En términos de la base $\{\varphi_{kn}\}_{k=1}^n$, el estimador *spline* viene dado por,

$$\hat{m}(x) = \sum_{k=1}^n \frac{\hat{\alpha}_{kn}}{1 + \lambda \gamma_{kn}} \varphi_{kn}(x)$$

(en donde $\hat{\alpha}_{kn} = n^{-1} \sum_{i=1}^n \varphi_{kn}(X_i)Y_i$). Así $\hat{m}(x) = n^{-1} \sum_{j=1}^n W_\lambda(x, X_j)Y_j$ es un estimador lineal, con una función de pesos,

$$W_\lambda(x, X_j) = \sum_{k=1}^n \frac{\varphi_{kn}(x)\varphi_{kn}(X_j)}{1 + \lambda\gamma_{kn}},$$

que se corresponde con el estimador *spline* suave de orden $(2, p)$ para el conjunto de datos $\{X_i, n\delta_{ij}\}_{i=1}^n$ (véase Silverman (1984)).

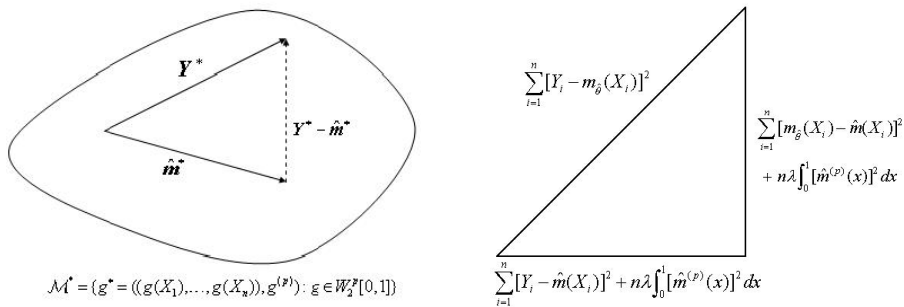


Figura 3: Izquierda: interpretación del estimador spline suave. Derecha: análisis de la varianza asociado al F test natural (basado en el estimador spline suave).

El vector de estimaciones resultante es $\hat{\mathbf{m}} = \mathbf{H}\mathbf{Y}$ en donde, $\mathbf{H} = \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^t = [n^{-1}W_\lambda(X_i, X_j)]$, siendo $\mathbf{\Phi} = n^{-1/2}(\varphi_{1n}, \dots, \varphi_{nn})$ y $\mathbf{\Gamma} = \text{diag}((1 + \lambda\gamma_{kn})^{-1})$.

4.2 . ECUACIONES NORMALES

En lo que respecta a las ecuaciones normales, podemos definir una versión continua del estimador *spline* suave que admite un sistema de ecuaciones “normales” que caracterizan al estimador (véanse Cox (1983, 1984)). Se denomina estimador *spline* continuo de orden $(2, p)$ a la única función $s(x)$ que minimiza sobre $\mathcal{M} = \mathcal{W}_2^p[0, 1]$ el funcional,

$$\begin{aligned} V(g, \lambda) &= E[(Y - g(X))^2] + \lambda \int_0^1 [g^{(p)}(x)]^2 dx \\ &= E[(m(X) - g(X))^2] + \lambda \int_0^1 [g^{(p)}(x)]^2 dx + \sigma^2 \end{aligned}$$

(véase la igualdad (15), página 345). Así pues,

$$s = \arg \min_{g \in \mathcal{M}} V(g, \lambda).$$

Fijada $g \in \mathcal{M}$, y la función,

$$N(\alpha) = V(s + \alpha g) \quad \alpha \in \mathbb{R},$$

la condición necesaria de mínimo, proporciona las siguientes ecuaciones normales para el estimador *spline* continuo,

$$E[(Y - s(X))g(X)] = \int_0^1 [m(x) - s(x)]g(x)dF(x) = \lambda \int_0^1 s^{(p)}(x)g^{(p)}(x)dx,$$

para toda $g \in \mathcal{M} = \mathcal{W}_2^p[0, 1]$. Si X tiene densidad $f(x)$, la igualdad

$$\int_0^1 [m(x) - s(x)]s(x)dF(x) = \lambda \int_0^1 [s^{(p)}(x)]^2 dx,$$

conduce a la siguiente ecuación diferencial que caracteriza también al estimador *spline* continuo,

$$(-1)^p \lambda \frac{d^{2p}s(x)}{dx^{2p}} + s(x)f(x) = m(x)f(x),$$

con $s^{(j)}(0) = s^{(j)}(1) = 0$, $j = p, \dots, 2p - 1$.

Para el estimador *spline* suave, $\hat{m}$, es posible obtener unas ecuaciones similares (Eubank (1988, Cáp. 5)). Fijada $g \in \mathcal{M}$, y la función,

$$\hat{N}(\alpha) = \hat{V}(\hat{m} + \alpha g, \lambda) \quad \alpha \in \mathbb{R},$$

la condición necesaria de mínimo, $N'(0) = 0$, equivale a las siguientes igualdades que caracterizan al estimador,

$$\langle \mathbf{Y}^* - \hat{m}^*, g^* \rangle_{\mathcal{H}} = \sum_{i=1}^n [Y_i - \hat{m}(X_i)]g(X_i) - n\lambda \int_0^1 \hat{m}^{(p)}(x)g^{(p)}(x)dx = 0, \quad (13)$$

para toda $g \in \mathcal{M} = \mathcal{W}_2^p[0, 1]$. En efecto (véase Eubank (1988, Cáp. 5)), expresando $\hat{m}$ en función de la base de Demmler y Reinsch (1975), $\hat{m}(x) = \sum_{j=1}^n \theta_j \varphi_{jn}(x)$, se obtiene que,

$$\int_0^1 \hat{m}^{(p)}(x)g^{(p)}(x)dx = (-1)^p(2p-1)! \sum_{i=1}^n g(X_i) \sum_{j=1}^n \theta_j \delta_{ij},$$

para toda $g \in \mathcal{M} = \mathcal{W}_2^p[0, 1]$. De lo anterior se deduce que la condición necesaria de mínimo, $N'(0) = 0$, equivale a las siguientes ecuaciones,

$$Y_i - \sum_{j=1}^n \theta_j \varphi_{jn}(X_i) = n\lambda(-1)^p(2p-1)! \sum_{j=1}^n \theta_j \delta_{ij}, \quad i = 1, \dots, n,$$

cuya solución es única en los θ_j .

4.3 . ESTIMACIÓN DE LA VARIANZA

Como generalización natural del estimador de la varianza mínimo cuadrático, consideraremos el estimador que resulta de evaluar el funcional $\hat{V}(g, \lambda)$ en $\hat{m}$ (por analogía con (10) y (12)),

$$\begin{aligned}\hat{\sigma}_{\mathcal{H}}^2(\lambda) &= \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2 + \lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx \\ &= \min_{g \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n [Y_i - g(X_i)]^2 + \lambda \int_0^1 [g^{(p)}(x)]^2 dx.\end{aligned}$$

Este estimador admite las siguientes interpretaciones,

$$\hat{\sigma}_{\mathcal{H}}^2(\lambda) = \hat{V}(\hat{m}, \lambda) = \min_{g \in \mathcal{M}} \hat{V}(g, \lambda) = \min_{g^* \in \mathcal{M}^*} \frac{1}{n} \|\mathbf{Y}^* - g^*\|_{\mathcal{H}}^2 = \frac{1}{n} \|\mathbf{Y}^* - \hat{m}^*\|_{\mathcal{H}}^2,$$

en donde $\|\mathbf{Y}^* - \hat{m}^*\|_{\mathcal{H}}^2 = \mathbf{Y}^t(\mathbf{I} - \mathbf{H})\mathbf{Y}$. En efecto, por las ecuaciones normales (13),

$$\mathbf{Y}^t(\mathbf{I} - \mathbf{H})^t \mathbf{H} \mathbf{Y} = \sum_{i=1}^n [Y_i - \hat{m}(X_i)] \hat{m}(X_i) = n\lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx, \quad (14)$$

por lo que,

$$\begin{aligned}\|\mathbf{Y}^* - \hat{m}^*\|_{\mathcal{H}}^2 &= \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2 + \sum_{i=1}^n [Y_i - \hat{m}(X_i)] \hat{m}(X_i) \\ &= \mathbf{Y}^t(\mathbf{I} - \mathbf{H})^t (\mathbf{I} - \mathbf{H}) \mathbf{Y} + \mathbf{Y}^t(\mathbf{I} - \mathbf{H})^t \mathbf{H} \mathbf{Y} = \mathbf{Y}^t(\mathbf{I} - \mathbf{H}) \mathbf{Y}.\end{aligned}$$

De esta igualdad se deduce $\hat{\sigma}_{\mathcal{H}}^2(\lambda) = n^{-1} \mathbf{Y}^t(\mathbf{I} - \mathbf{H}) \mathbf{Y} = n^{-1} \sum_{i=1}^n [Y_i - \hat{m}(X_i)] Y_i$, que permite interpretar $\hat{\sigma}_{\mathcal{H}}^2(\lambda)$ como una versión muestral de la ecuación (11).

5 . FUNDAMENTOS DEL ANÁLISIS DE LA VARIANZA Y DEL F TEST

Es bien conocido (véase Shao (2003), Pág. 1) que para cualquier función $g \in \mathcal{M} = \mathcal{B}$,

$$E[(Y - g(X))^2] = E[(m(X) - g(X))^2] + E[(Y - m(X))^2]. \quad (15)$$

Consideremos un espacio de funciones paramétrico de interés, $\mathcal{M}_0 = \{m_\theta : \theta \in \Theta\}$. Para medir la distancia de las funciones de $g \in \mathcal{M}_0$ a m haremos uso del funcional,

$$D(g) = E[(m(X) - g(X))^2] = \int [m(x) - g(x)]^2 dF(x).$$

Si $\hat{m}$ es un estimador de m , $D(\hat{m})$ se conoce como error cuadrático integrado, y se denota por $ISE(\hat{m})$.

5.1 . EL ANÁLISIS DE LA VARIANZA TEÓRICO

Consideremos la función m_0 (supondremos que existe y es única) que minimiza $D(g)$ sobre $\mathcal{M}_0$. La función m_0 también minimiza $V(g)$ sobre $\mathcal{M}_0$ ya que $V(g) = D(g) + V(m)$, y $V(m) = \sigma^2$ es constante. Así pues,

$$m_0 = \arg \min_{g \in \mathcal{M}_0} V(g) = \arg \min_{g \in \mathcal{M}_0} D(g).$$

Consideremos el contraste de hipótesis,

$$H_0 : m \in \mathcal{M}_0 \quad \text{frente a} \quad H_1 : m \notin \mathcal{M}_0. \quad (16)$$

Los F tests se basan en la siguiente igualdad, a la que denominaremos análisis de la varianza teórico (o abreviadamente, ANOVA teórico),

$$E[(Y - m_0(X))^2] = E[(m(X) - m_0(X))^2] + E[(Y - m(X))^2].$$

En lo sucesivo denotaremos $\sigma_0^2 = E[(Y - m_0(X))^2]$ y $d^2 = E[(m(X) - m_0(X))^2]$ ($= ISE(m_0)$). Las diferentes interpretaciones del ANOVA teórico se recogen en la siguiente tabla.

Tabla 2. *Análisis de la varianza teórico*

¿Qué mide?	Parámetro	Valor esperado	Valor del funcional	Mínimo del funcional
Distancia a $\mathcal{M}_0$	d^2	$E[(m(X) - m_0(X))^2]$	$D(m_0)$	$\min_{g \in \mathcal{M}_0} D(g)$
Varianza bajo H_1	σ^2	$E[(Y - m(X))^2]$	$V(m)$	$\min_{g \in \mathcal{M}} V(g)$
Varianza bajo H_0	σ_0^2	$E[(Y - m_0(X))^2]$	$V(m_0)$	$\min_{g \in \mathcal{M}_0} V(g)$

El ANOVA teórico se puede expresar también en términos de covarianzas si la función,

$$r(x) = m(x) - m_0(x) = E[U/X = x], \quad \text{en donde } U = Y - m_0(X),$$

es ortogonal a las funciones constantes, o sea, si $E[r(X)] = \int r(x)dF(x) = 0$. Considérense para ello los funcionales,

$$\mathcal{D}(g) = E[(Y - g(X))(m(X) - g(X))] \quad \text{y} \quad \mathcal{V}(g) = E[(Y - m_0(X))(Y - g(X))],$$

definidos sobre el espacio de funciones (8). Se tiene que $\mathcal{V}(m_0) = V(m_0) = \sigma_0^2$, y teniendo en cuenta la igualdad (9), se observa que $\mathcal{D}(g)$ coincide con $D(g)$, y por tanto $\mathcal{D}(m_0) = d^2$, y $\mathcal{V}(m) = E[(Y - m_0(X))(Y - m(X))] = E[(Y - m(X))Y] = E[(Y - m(X))^2] = V(m) = \sigma^2$. En consecuencia, $\sigma_0^2 = d^2 + \sigma^2$ equivale a la igualdad $\mathcal{V}(m_0) = \mathcal{D}(m_0) + \mathcal{V}(m)$.

Si asumimos que $E[r(X)] = 0$, o equivalentemente que $E[U] = 0$ (ya que $U = r(X) + \varepsilon$), los términos del ANOVA teórico se corresponden con las siguientes covarianzas: $d^2 = \sigma_{U,r(X)}$, $\sigma^2 = \sigma_{U,\varepsilon}$ y $\sigma_0^2 = \sigma_{U,U}$. Esta interpretación se recoge en la siguiente tabla.

Tabla 3. *Análisis de la varianza teórico en términos de covarianzas*

Parámetro	Covarianza	Valor del funcional
d^2	$\sigma_{U,r(X)} = \text{cov}(Y - m_0(X), m(X) - m_0(X))$	$\mathcal{D}(m_0)$
σ^2	$\sigma_{U,\varepsilon} = \text{cov}(Y - m_0(X), Y - m(X))$	$\mathcal{V}(m)$
σ_0^2	$\sigma_{U,U} = \text{cov}(Y - m_0(X), Y - m_0(X))$	$\mathcal{V}(m_0)$

Cualquiera de las dos expresiones del análisis de la varianza teórico (Tablas 2 y 3) admite una versión muestral. Mostramos por su interés, la correspondiente a la Tabla 2, basada en los funcionales,

$$ASE(g) = \frac{1}{n} \sum_{i=1}^n [m(X_i) - g(X_i)]^2 \text{ y } \Delta(g) = \frac{1}{n} \sum_{i=1}^n [Y_i - m(X_i)][m(X_i) - g(X_i)].$$

Si $\hat{m}$ es un estimador de m , $ASE(\hat{m})$ se conoce como error cuadrático medio. La igualdad general, $\hat{V}(g) = ASE(g) + 2\Delta(g) + \hat{V}(m)$, aplicada a m_0 es,

$$\begin{aligned} \hat{V}(m_0) &= \frac{1}{n} \sum_{i=1}^n [Y_i - m(X_i)]^2 + \frac{1}{n} \sum_{i=1}^n [m(X_i) - m_0(X_i)]^2 \\ &\quad + \frac{2}{n} \sum_{i=1}^n [Y_i - m(X_i)][m(X_i) - m_0(X_i)]. \end{aligned} \tag{17}$$

Denotando por $\sigma_n^2 = \hat{V}(m)$, $\sigma_{0n}^2 = \hat{V}(m_0)$, $d_n^2 = ASE(m_0)$ y $\delta_n = 2\Delta(m_0)$, la igualdad anterior, a la que denominaremos ANOVA teórico muestral, se expresa como $\sigma_{0n}^2 = d_n^2 + \delta_n + \sigma_n^2$.

5.2 . FUNDAMENTOS DEL F TEST

El análisis de la varianza teórico de la Tabla 2 sugiere redefinir el contraste de hipótesis (16) en términos de,

$$C = \frac{d^2}{\sigma^2} = \frac{\sigma_0^2 - \sigma^2}{\sigma^2},$$

ya que bajo H_0 , y sólo bajo H_0 , $C = 0$, y por tanto el contraste se puede expresar de forma equivalente como,

$$H_0 : C = 0 \quad \text{frente a} \quad H_1 : C > 0.$$

En términos de los funcionales D y V , la primera expresión de C se corresponde con,

$$C = \frac{D(m_0)}{V(m)} = \frac{\min_{g \in \mathcal{M}_0} D(g)}{\min_{g \in \mathcal{M}} V(g)} = \frac{E[(m(X) - m_0(X))^2]}{E[(Y - m(X))^2]}, \quad (18)$$

y la segunda con,

$$\begin{aligned} C &= \frac{V(m_0) - V(m)}{V(m)} = \frac{\min_{g \in \mathcal{M}_0} V(g) - \min_{g \in \mathcal{M}} V(g)}{\min_{g \in \mathcal{M}} V(g)} \\ &= \frac{E[(Y - m_0(X))^2] - E[(Y - m(X))^2]}{E[(Y - m(X))^2]}. \end{aligned} \quad (19)$$

Las versiones muestrales correspondientes a estas expresiones de C son,

$$\hat{C} = \frac{d_n^2}{\sigma_n^2} = \frac{n^{-1} \sum_{i=1}^n [m(X_i) - m_0(X_i)]^2}{n^{-1} \sum_{i=1}^n [Y_i - m(X_i)]^2}$$

y

$$\tilde{C} = \frac{\sigma_{0n}^2 - \sigma_n^2}{\sigma_n^2} = \frac{n^{-1} \sum_{i=1}^n [Y_i - m_0(X_i)]^2 - n^{-1} \sum_{i=1}^n [Y_i - m(X_i)]^2}{n^{-1} \sum_{i=1}^n [Y_i - m(X_i)]^2}.$$

Si $E[\varepsilon^4] < +\infty$, la esperanza de $\hat{C}$ admite la siguiente aproximación (Johnson, Kotz y Kemp (1992, Cap. 1)),

$$E[\hat{C}] = E\left[\frac{d_n^2}{\sigma_n^2}\right] \div \frac{E[d_n^2]}{E[\sigma_n^2]} \left[1 + \frac{\text{var}(\sigma_n^2)}{(E[\sigma_n^2])^2} - \frac{\text{cov}(d_n^2, \sigma_n^2)}{E[d_n^2] \cdot E[\sigma_n^2]}\right] = \frac{d^2}{\sigma^2} [1 + O(\frac{1}{n})] \rightarrow \frac{d^2}{\sigma^2}$$

en donde $a_n = O(n^{-1})$ significa que $a_n/n^{-1} \rightarrow c$, siendo c una constante. Un resultado similar es válido para $\tilde{C}$.

La idea en la que se basan los F tests es en utilizar versiones muestrales de C , bien sea del tipo $\hat{C} = \hat{d}^2/\hat{\sigma}^2$, bien del tipo, $\hat{C} = (\hat{\sigma}_0^2 - \hat{\sigma}^2)/\hat{\sigma}^2$, basadas en estimadores $\hat{m}_0$ y $\hat{m}$ en vez de m_0 y m , para definir estadísticos de contraste del tipo,

$$F = c \cdot \frac{\hat{d}^2}{\hat{\sigma}^2} \quad \text{o} \quad F = k \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2},$$

en donde $c = c_1/c_2$ y $k = k_1/k_2$ son cocientes de constantes tales que c_1 y k_1 dependen de $\hat{\sigma}^2$, mientras que c_2 y k_2 dependen de $\hat{d}^2$ y $\hat{\sigma}_0^2 - \hat{\sigma}^2$ respectivamente.

Argumentos análogos a los utilizados con las expresiones de C basadas en la Tabla 2, pueden emplearse a las expresiones basadas en la Tabla 3, o a las que resultan de combinar ambos enfoques, tales como

$$C = \frac{\mathcal{D}(m_0)}{\mathcal{V}(m)} \quad \text{y} \quad C = \frac{\mathcal{D}(m_0)}{\mathcal{V}(m)}. \quad (20)$$

Cada una de estas expresiones (18)-(20) dará lugar, como veremos, a una generalización del F test clásico.

Tabla 4. *Fundamentos del F test*

Hipótesis nula	Hipótesis alternativa	Comparación relativa de funciones de regresión	Comparación relativa de varianzas	Comparación de covarianzas
$H_0 : C = 0$	$H_1 : C > 0$	$C = \frac{d^2}{\sigma^2}$	$C = \frac{\sigma_0^2 - \sigma^2}{\sigma^2}$	$C = \frac{\sigma_{u,r}(X)}{\sigma_{u,\varepsilon}}$

5.3 . EL ANÁLISIS DE LA VARIANZA MUESTRAL

Las versiones muestrales de los funcionales $D(g)$ y $\Delta(g)$ basadas en un estimador $\hat{m}$ cualquiera son,

$$\hat{D}(g) = \frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - g(X_i)]^2 \quad \text{y} \quad \hat{\Delta}(g) = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)][\hat{m}(X_i) - g(X_i)].$$

Se tiene así la igualdad general, $\hat{V}(g) = \hat{D}(g) + 2\hat{\Delta}(g) + \hat{V}(\hat{m})$. Esta igualdad aplicada a $\hat{m}_0$, el estimador mínimo cuadrático de m_0 , es,

$$\begin{aligned} \hat{V}(\hat{m}_0) &= \frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)]^2 \\ &+ \frac{2}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)][\hat{m}(X_i) - \hat{m}_0(X_i)] + \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2. \end{aligned}$$

La igualdad anterior (análoga a la (17)) se puede considerar como el análisis de la varianza asociado a $\hat{C} = [\hat{V}(\hat{m}_0) - \hat{V}(\hat{m})]/\hat{V}(\hat{m})$ y $\hat{C} = \hat{D}(\hat{m}_0)/\hat{V}(\hat{m})$, las versiones muestrales de C correspondientes a las expresiones (18) y (19). Su interpretación se presenta en la tabla siguiente. En lo sucesivo denotaremos por $\hat{d}^2 = \hat{D}(\hat{m}_0)$, $\hat{\sigma}^2 = \hat{V}(\hat{m})$, $\hat{\sigma}_0^2 = \hat{V}(\hat{m}_0)$ y $\hat{\delta} = \hat{\Delta}(\hat{m}_0)$.

Tabla 5. *Interpretación del análisis de la varianza muestral con estimadores cualesquiera. Comparación de ajustes y comparación de varianzas*

Estimador	Media muestral	Valor del funcional	Expresión vectorial
$\hat{d}^2$	$\frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)]^2$	$\hat{D}(\hat{m}_0)$	$\frac{1}{n} \hat{\mathbf{m}} - \hat{\mathbf{m}}_0 ^2$
$\hat{\delta}$	$\frac{2}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)][Y_i - \hat{m}(X_i)]$	$2\hat{\Delta}(\hat{m}_0)$	$\frac{2}{n} (\hat{\mathbf{m}} - \hat{\mathbf{m}}_0)^t (\mathbf{Y} - \hat{\mathbf{m}})$
$\hat{\sigma}^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2$	$\hat{V}(\hat{m})$	$\frac{1}{n} \mathbf{Y} - \hat{\mathbf{m}} ^2$
$\hat{\sigma}_0^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)]^2$	$\hat{V}(\hat{m}_0)$	$\frac{1}{n} \mathbf{Y} - \hat{\mathbf{m}}_0 ^2$

Las versiones muestrales de (20), se obtienen a partir de los funcionales,

$$\tilde{\mathcal{D}}(g) = \frac{1}{n} \sum_{i=1}^n [y_i - g(x_i)][\hat{m}(x_i) - g(x_i)] \quad y \quad \tilde{\mathcal{V}}(g) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}_0(x_i)][y_i - g(x_i)].$$

El estimador de d^2 a considerar es,

$$\tilde{d}^2 = \tilde{\mathcal{D}}(\hat{m}_0) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}_0(X_i)][\hat{m}(x_i) - \hat{m}_0(X_i)] = \hat{D}(\hat{m}_0) + \hat{\Delta}(\hat{m}_0). \quad (21)$$

Es decir, $\tilde{d}^2$ es la distancia entre ajustes $\hat{d}^2 = \hat{D}(\hat{m}_0)$ más el término “de penalización” $\tilde{\delta} = \hat{\Delta}(\hat{m}_0)$. La igualdad general $\hat{V}(g) = \tilde{\mathcal{D}}(g) + \hat{\Delta}(g) + \hat{V}(\hat{m})$ aplicada a $\hat{m}_0$, $\hat{V}(\hat{m}_0) = \tilde{\mathcal{D}}(\hat{m}_0) + \hat{\Delta}(\hat{m}_0) + \hat{V}(\hat{m})$, proporciona el análisis de la varianza muestral asociado a $\hat{C} = \tilde{\mathcal{D}}(\hat{m}_0)/\hat{V}(\hat{m}) = \tilde{d}^2/\hat{\sigma}^2$, que se presenta en la siguiente tabla.

Tabla 6. Interpretación del análisis de la varianza con cualquier estimador. Comparación de ajustes “penalizada” y comparación de varianzas

Esti- mador	Media muestral	Valor del funcional	Expresión vectorial
$\tilde{d}^2$	$\frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}_0(x_i)][\hat{m}(x_i) - \hat{m}_0(x_i)]$	$\tilde{\mathcal{D}}(\hat{m}_0)$	$\frac{1}{n}(\mathbf{Y} - \hat{\mathbf{m}}_0)^t(\hat{\mathbf{m}} - \hat{\mathbf{m}}_0)$
$\tilde{\delta}$	$\frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)][Y_i - \hat{m}(X_i)]$	$\hat{\Delta}(\hat{m}_0)$	$\frac{1}{n}(\hat{\mathbf{m}} - \hat{\mathbf{m}}_0)^t(\mathbf{Y} - \hat{\mathbf{m}})$
$\hat{\sigma}^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2$	$\hat{V}(\hat{m})$	$\frac{1}{n} \ \mathbf{Y} - \hat{\mathbf{m}}\ ^2$
$\hat{\sigma}_0^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)]^2$	$\hat{V}(\hat{m}_0)$	$\frac{1}{n} \ \mathbf{Y} - \hat{\mathbf{m}}_0\ ^2$

El estimador de σ^2 “penalizado” es,

$$\tilde{\sigma}^2 = \tilde{\mathcal{V}}(\hat{m}) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}_0(x_i)][y_i - \hat{m}(x_i)] = \hat{V}(\hat{m}) + \hat{\Delta}(\hat{m}_0).$$

Si además tenemos en cuenta que $\tilde{\mathcal{V}}(\hat{m}_0) = \hat{V}(\hat{m}_0) = \hat{\sigma}_0^2$, se obtiene la igualdad, $\tilde{\mathcal{V}}(\hat{m}_0) = \tilde{\mathcal{D}}(\hat{m}_0) + \tilde{\mathcal{V}}(\hat{m})$, que representa el análisis de la varianza asociado a $\hat{C} = \tilde{\mathcal{D}}(\hat{m}_0)/\tilde{\mathcal{V}}(\hat{m}) = \tilde{d}^2/\tilde{\sigma}^2$. Se presenta en la tabla siguiente, análoga a la Tabla 3.

Tabla 7. Interpretación del análisis de la varianza basado en covarianzas. Comparación de ajustes “penalizada” y estimación de la varianza “penalizada”

Esti- mador	Media muestral	Valor del funcional	Expresión vectorial
$\tilde{d}^2$	$\frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}_0(x_i)][\hat{m}(x_i) - \hat{m}_0(x_i)]$	$\tilde{\mathcal{D}}(\hat{m}_0)$	$\frac{1}{n}(\mathbf{Y} - \hat{\mathbf{m}}_0)^t(\hat{\mathbf{m}} - \hat{\mathbf{m}}_0)$
$\tilde{\sigma}^2$	$\frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}_0(x_i)][y_i - \hat{m}(x_i)]$	$\tilde{\mathcal{V}}(\hat{m})$	$\frac{1}{n}(\mathbf{Y} - \hat{\mathbf{m}}_0)^t(\mathbf{Y} - \hat{\mathbf{m}})$
$\hat{\sigma}_0^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)]^2$	$\tilde{\mathcal{V}}(\hat{m}_0)$	$\frac{1}{n} \mathbf{Y} - \hat{\mathbf{m}}_0 ^2$

En las Secciones 6-9 examinamos las características particulares de las descomposiciones de la varianza de las Tablas 5-7 basadas en estimadores paramétricos y no paramétricos de m . A cada una de las tablas analizadas le asociaremos una segunda tabla (a la que denominaremos “resultados del análisis de la varianza y del F test”), orientada a la presentación de los resultados de un análisis de datos.

5.4 . MATRICES DE ESTIMACIÓN

Las tres descomposiciones de la varianza propuestas (Tablas 5-7) se basan en la forma en que se utiliza $\tilde{\delta} = \hat{\Delta}(\hat{m}_0)$ ($= \tilde{d}^2 - \hat{d}^2$) para definir el estimador del cociente C (véase la Tabla 4). Cabe la posibilidad de no utilizar $\tilde{\delta}$ (con $\tilde{C} = \hat{D}(\hat{m}_0)/\hat{V}(\hat{m}) = \hat{d}^2/\hat{\sigma}^2$), de utilizar $\tilde{\delta}$ sólo para estimar d^2 (con $\tilde{C} = [\hat{V}(\hat{m}_0) - \hat{V}(\hat{m})]/\hat{V}(\hat{m}) = (\hat{\sigma}_0^2 - \hat{\sigma}^2)/\hat{\sigma}^2$ y con $\hat{C} = \tilde{\mathcal{D}}(\hat{m}_0)/V(\hat{m}) = \tilde{d}^2/\hat{\sigma}^2$), o de utilizar $\tilde{\delta}$ para estimar tanto d^2 como σ^2 (con $\hat{C} = \tilde{\mathcal{D}}(\hat{m}_0)/\tilde{\mathcal{V}}(\hat{m}) = \tilde{d}^2/\tilde{\sigma}^2$).

Las relaciones entre las diferentes versiones muestrales de C se pueden deducir del análisis de la varianza muestral inicial de la Tabla 5. Dicha descomposición se corresponde, salvo el factor n^{-1} , con la siguiente igualdad, en donde $\mathbf{H}$ representa la matriz que proporciona el vector de ajustes del estimador $\hat{m}$, $\hat{\mathbf{m}} = \mathbf{H}\mathbf{Y}$, y $\mathbf{P}_0$ la matriz que proporciona el vector de ajustes mínimo cuadráticos, $\hat{\mathbf{m}}_0 = \mathbf{P}_0\mathbf{Y}$,

$$\begin{aligned} \mathbf{Y}^t(\mathbf{I} - \mathbf{P}_0)\mathbf{Y} &= \mathbf{Y}^t(\mathbf{H} - \mathbf{P}_0)^t(\mathbf{H} - \mathbf{P}_0)\mathbf{Y} \\ &+ 2\mathbf{Y}^t(\mathbf{H} - \mathbf{P}_0)^t(\mathbf{I} - \mathbf{H})\mathbf{Y} + \mathbf{Y}^t(\mathbf{I} - \mathbf{H})^t(\mathbf{I} - \mathbf{H})\mathbf{Y}. \end{aligned}$$

Las matrices que proporcionan los diferentes estimadores de d^2 , σ^2 y σ_0^2 se resumen en la siguiente tabla ($\hat{\theta} = n^{-1}\mathbf{Y}^t\mathbf{M}\mathbf{Y}$ si θ es uno de estos parámetros, y $\mathbf{M}$ su matriz de estimación asociada). Se incluyen también los correspondientes a $\tilde{\delta}$ y $\hat{\delta} = 2\tilde{\delta}$. En el estudio de estos estadísticos basados en diferentes

estimadores $\hat{m}$, la diferencia más notoria es, como veremos, la interpretación del término $\tilde{\delta} = \hat{\Delta}(\hat{m}_0)$.

Tabla 8. *Matrices de estimación*

Estimador	Matriz de estimación	Estimador	Matriz de estimación
$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V}_0 \mathbf{Y}$	$\mathbf{V}_0 = \mathbf{I} - \mathbf{P}_0$	$\hat{d}^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{D} \mathbf{Y}$	$\mathbf{D} = (\mathbf{H} - \mathbf{P}_0)^t (\mathbf{H} - \mathbf{P}_0)$
$\hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V} \mathbf{Y}$	$\mathbf{V} = (\mathbf{I} - \mathbf{H})^t (\mathbf{I} - \mathbf{H})$	$\hat{d}^2 = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{D}} \mathbf{Y}$	$\tilde{\mathbf{D}} = (\mathbf{I} - \mathbf{P}_0)(\mathbf{H} - \mathbf{P}_0)$
$\hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{V}} \mathbf{Y}$	$\tilde{\mathbf{V}} = (\mathbf{I} - \mathbf{P}_0)(\mathbf{I} - \mathbf{H})$	$\tilde{\delta} = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{B}} \mathbf{Y}$	$\tilde{\mathbf{B}} = (\mathbf{I} - \mathbf{H})^t (\mathbf{H} - \mathbf{P}_0)$
$\hat{\sigma}_0^2 - \hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}^t \nabla \mathbf{Y}$	$\nabla = \mathbf{V}_0 - \mathbf{V}$	$\hat{\delta} = \frac{1}{n} \mathbf{Y}^t \mathbf{B} \mathbf{Y}$	$\mathbf{B} = 2(\mathbf{I} - \mathbf{H})^t (\mathbf{H} - \mathbf{P}_0)$

En general, si la matriz $\mathbf{H}$ es simétrica, $\hat{\Delta}(\hat{m}_0)$ se puede interpretar como el término atribuible al sesgo del estimador $\hat{m}$. En efecto, el vector que proporciona el sesgo en los puntos del diseño es $\mathbf{b} = \mathbf{m} - E[\hat{\mathbf{m}}] = (\mathbf{I} - \mathbf{H})\mathbf{m}$. Así pues, una comparación del sesgo de $\hat{m}$ bajo H_1 , frente al sesgo de $\hat{m}$ bajo H_0 sería, $\hat{\mathbf{b}} - \hat{\mathbf{b}}_0 = (\mathbf{I} - \mathbf{H})\hat{\mathbf{m}} - (\mathbf{I} - \mathbf{H})\hat{\mathbf{m}}_0 = (\mathbf{I} - \mathbf{H})(\mathbf{H} - \mathbf{P}_0)\mathbf{Y}$. En general, $\hat{\Delta}(\hat{m}_0) = \hat{\Delta}_1 - \hat{\Delta}_0$, en donde,

$$\hat{\Delta}_1 = \frac{1}{n} \mathbf{Y}^t (\mathbf{I} - \mathbf{H})^t \mathbf{H} \mathbf{Y} \quad \text{y} \quad \hat{\Delta}_0 = \frac{1}{n} \mathbf{Y}^t \mathbf{P}_0 (\mathbf{I} - \mathbf{H}) \mathbf{Y}.$$

Si $\mathbf{H}$ es simétrica, $\hat{\Delta}(\hat{m}_0) = n^{-1} \mathbf{Y}^t (\hat{\mathbf{b}} - \hat{\mathbf{b}}_0)$, $\hat{\Delta}_1 = n^{-1} \mathbf{Y}^t \hat{\mathbf{b}}$ y $\hat{\Delta}_0 = n^{-1} \mathbf{Y}^t \hat{\mathbf{b}}_0$, por lo que $\hat{\Delta}_1$ puede atribuirse al sesgo de $\hat{m}$ bajo H_1 , y $\hat{\Delta}_0$ al sesgo de $\hat{m}$ bajo H_0 .

Los términos $\hat{\Delta}_1$ y $\hat{\Delta}_0$ en determinados casos se anulan. En general, $\hat{\Delta}_1$ se anula si $\mathbf{H}^t \mathbf{H} = \mathbf{H}$, y $\hat{\Delta}_0$ se anula si $\mathbf{P}_0 \mathbf{H} = \mathbf{P}_0$. Con modelos de regresión $\mathcal{M}_0$ lineales como el (3), $\mathbf{m}_\theta = \mathbf{X}\theta$, con lo que el estimador $\hat{m}$ es insesgado bajo H_0 ($\Leftrightarrow \mathbf{b}_0 = \mathbf{m}_\theta - E[\hat{\mathbf{m}}] = (\mathbf{I} - \mathbf{H})\mathbf{X}\theta = 0$ para todo θ) si y sólo si $\mathbf{H}\mathbf{X} = \mathbf{X}$, o equivalentemente, si $\mathbf{H}\mathbf{P}_0 = \mathbf{P}_0$. Por tanto, si $\mathbf{H}$ es simétrica, y $\hat{m}$ es insesgado bajo H_0 , $\hat{\Delta}(\hat{m}_0) = \hat{\Delta}_1$.

Pasamos ahora a estudiar el análisis de la varianza y los estadísticos de contraste de los F tests basados en los estimadores introducidos en las Secciones 3 y 4. Para describir la distribución de los estadísticos F , supondremos que $(X_1, Y_1), \dots, (X_n, Y_n)$ son una muestra de un modelo de regresión $(Y_i = m(X_i) + \varepsilon_i)$ de diseño fijo, es decir, consideraremos que los valores X_i de la variable explicativa son valores prefijados, por lo que la aleatoriedad de los Y_i proviene sólo de los errores ε_i . Asumiremos, no obstante, que los valores X_i están generados por una función de distribución absolutamente continua F , en el sentido de que $F(X_i) = (i - 1/2)/n$, $i = 1, \dots, n$. En un diseño de

este tipo, la función $F_n(x) = n^{-1} \sum_{i=1}^n I_{\{X_i \leq x\}}$ converge uniformemente a la función F , por lo que en este sentido, podemos considerar que F es la función de distribución de los X_i .

6 . EL ANÁLISIS DE LA VARIANZA Y EL F TEST PARAMÉTRICOS

Consideremos el contraste de la hipótesis de validez del modelo lineal (3), $H_0 : m \in \mathcal{M}_0$, frente a la hipótesis alternativa, $H_1 : m \notin \mathcal{M}_0$, suponiendo que $m \in \mathcal{M}$, el espacio de funciones lineales más amplio definido en (5).

Sean $\hat{m}_0 = m_{\hat{\theta}_0}$ y $\hat{m} = m_{\hat{\theta}}$ en donde $\hat{\theta}_0$ y $\hat{\theta}$ son los estimadores mínimo cuadráticos de θ correspondientes a los espacios $\mathcal{M}_0$ y $\mathcal{M}$ respectivamente. Denotaremos por $\hat{\mathbf{m}}_0 = \mathbf{P}_0 \mathbf{Y}$, y $\hat{\mathbf{m}} = \mathbf{P} \mathbf{Y}$, los correspondientes vectores de estimaciones.

El estimador $\hat{m}_0$, que por definición es la función que minimiza $\hat{V}(g)$ sobre $\mathcal{M}_0$, es también la función que minimiza $\hat{D}(g)$ sobre $\mathcal{M}_0$ ya que $\mathbf{P}_0 \mathbf{P} = \mathbf{P}_0$, y por tanto $\mathbf{P}_0 \hat{\mathbf{m}} = \mathbf{P}_0 \mathbf{P} \mathbf{Y} = \mathbf{P}_0 \mathbf{Y} = \hat{\mathbf{m}}_0$. Así pues,

$$\hat{m}_0 = \arg \min_{g \in \mathcal{M}_0} \hat{V}(g) = \arg \min_{g \in \mathcal{M}_0} \hat{D}(g).$$

6.1 . EL ANÁLISIS DE LA VARIANZA CLÁSICO

Por ser $\hat{\mathbf{m}}$ la proyección de $\mathbf{Y}$ sobre $\mathcal{M}^* = \{\mathbf{m}_\theta : \mathbf{m}_\theta = \sum_{j=1}^q \theta_j \mathbf{g}_j\}$,

$$\frac{1}{n} ||\mathbf{Y} - \hat{\mathbf{m}}_0||^2 = \frac{1}{n} ||\hat{\mathbf{m}} - \hat{\mathbf{m}}_0||^2 + \frac{1}{n} ||\mathbf{Y} - \hat{\mathbf{m}}||^2,$$

o sea, $\hat{V}(\hat{m}_0) = \hat{D}(\hat{m}_0) + \hat{V}(\hat{m})$. Las diferentes interpretaciones de la descomposición de la varianza anterior se resumen en la tabla siguiente.

Tabla 9. Interpretación del análisis de la varianza paramétrico

Estimador	Media muestral	Valor del funcional	Mínimo del funcional	Distancia
$\hat{d}^2$	$\frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)]^2$	$\hat{D}(\hat{m}_0)$	$\min_{g \in \mathcal{M}_0} \hat{D}(g)$	$\frac{1}{n} \hat{\mathbf{m}} - \hat{\mathbf{m}}_0 ^2$
$\hat{\sigma}^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2$	$\hat{V}(\hat{m})$	$\min_{g \in \mathcal{M}} \hat{V}(g)$	$\frac{1}{n} \mathbf{Y} - \hat{\mathbf{m}} ^2$
$\hat{\sigma}_0^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)]^2$	$\hat{V}(\hat{m}_0)$	$\min_{g \in \mathcal{M}_0} \hat{V}(g)$	$\frac{1}{n} \mathbf{Y} - \hat{\mathbf{m}}_0 ^2$

Obsérvese que $\hat{\Delta}(\hat{m}_0) = 0$, y por consiguiente las tres versiones del ANOVA muestral (Tablas 5-7) conducen a los mismos estimadores de d^2 y σ^2 .

6.2 . EL F TEST CLÁSICO

El F test clásico o paramétrico (ecuación (7)) se basa en el ANOVA muestral de la Tabla 9. El estadístico de contraste es,

$$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\hat{d}^2}{\hat{\sigma}^2} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2},$$

en donde $\mathbf{D} = \mathbf{P} - \mathbf{P}_0 = \mathbf{V}_0 - \mathbf{V}$, $\mathbf{V} = \mathbf{I} - \mathbf{P}$ y $\mathbf{V}_0 = \mathbf{I} - \mathbf{P}_0$ son las matrices que proporcionan los estimadores de d^2 , σ^2 y σ_0^2 . En términos de los funcionales de la distancia y la varianza, la primera expresión se corresponde con,

$$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\hat{D}(\hat{m}_0)}{\hat{V}(\hat{m})} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\min_{g \in \mathcal{M}_0} \hat{D}(g)}{\min_{g \in \mathcal{M}} \hat{V}(g)},$$

y la segunda con,

$$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{V}(\hat{m}_0) - \hat{V}(\hat{m})}{\hat{V}(\hat{m})} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\min_{g \in \mathcal{M}_0} \hat{V}(g) - \min_{g \in \mathcal{M}} \hat{V}(g)}{\min_{g \in \mathcal{M}} \hat{V}(g)}. \quad (22)$$

Como es bien conocido (Seber (1977)), si los errores ε_i siguen una distribución normal (y verifican la condición (2)), bajo H_0 , el estadístico F sigue una distribución F_{ν_1, ν_2} con $\nu_1 = \text{tr}(\mathbf{D}) = q - p$ y $\nu_2 = \text{tr}(\mathbf{V}) = n - q$ grados de libertad.

La presentación de los resultados del análisis de la varianza y del F test puede hacerse mediante las tablas siguientes (similares a las propuestas por Motulsky y Christopoulos (2002)).

Tabla 10. Resultados del análisis de la varianza y del F test paramétricos

ANOVA muestral		Estadístico de contraste	
Trazas		Comparación relativa de ajustes	Comparación relativa de varianzas
$\text{tr}(\mathbf{D}) = q - p$	$\hat{d}^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{D} \mathbf{Y}$	$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\hat{d}^2}{\hat{\sigma}^2} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2}$	
$\text{tr}(\mathbf{V}) = n - q$	$\hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V} \mathbf{Y}$		
$\text{tr}(\mathbf{V}_0) = n - p$	$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V}_0 \mathbf{Y}$	Nivel de significación $p = \Pr(F > F_{obs}/H_0)$	

Obsérvese que $p = \Pr(F_{q-p, n-q} > F_{obs})$.

6.3 . ANÁLISIS ESTADÍSTICO PARAMÉTRICO DEL EJEMPLO

La representación gráfica correspondiente a los datos $\{(x_i, y_i)\}_{i=1}^{72}$ del ejemplo (Figura 2-izquierda), permite ver que el ajuste mínimo cuadrático

de un polinomio de grado cinco, $m_{\hat{\theta}}(x)$, proporciona un buen ajuste de los datos, por lo que podemos asumir que $m \in \mathcal{M} = \{m_{\theta} : m_{\theta}(x) = \theta_0 + \theta_1x + \theta_2x^2 + \theta_3x^3 + \theta_4x^4 + \theta_5x^5\}$. Además, un análisis de los residuos $e_i = y_i - m_{\hat{\theta}}(x_i)$, $i = 1, \dots, 72$, muestra que la hipótesis de errores independientes con distribución normal de media nula y varianza constante es asumible.

La función $m_{\hat{\theta}}$ se puede obtener a partir de la variable $Z = X/72$, definida en el intervalo $[0, 1]$ (ya que $X \in [0, 72]$). El ajuste mínimo cuadrático correspondiente a los datos $\{(z_i, y_i) : z_i = x_i/72 \ i = 1, \dots, 72\}$, es $\hat{m}(z) = 0,43213416 + 3,97785273z - 16,07455397z^2 + 31,6283053z^3 - 28,7795343z^4 + 9,84891005z^5$. La función $m_{\hat{\theta}}$ es $m_{\hat{\theta}}(x) = \hat{m}(x/72)$.

Consideremos ahora el conjunto de datos $\{(x_i, y_i)\}_{i=13}^{72}$ para contrastar la hipótesis nula de linealidad en el intervalo $[12, 72]$, $H_0 : m \in \mathcal{M}_0 = \{m_{\theta} : m_{\theta}(x) = \theta_0 + \theta_1x\}$, frente a la hipótesis alternativa de que m no es una recta, $H_1 : m \notin \mathcal{M}_0$ (asumiendo que $m \in \mathcal{M}$). El F test clásico, proporciona los resultados siguientes.

Resultados del análisis de la varianza y del F test clásicos (Tabla 10)

Trazas		ANOVA muestral		Estadístico de contraste	
				Comparación de ajustes y de varianzas	
$\text{tr}(\mathbf{D}) = 4$	$\hat{d}^2 = 0,00000661$			$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\hat{d}^2}{\hat{\sigma}^2} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2} =$ $= 0.19137342$	
$\text{tr}(\mathbf{V}) = 54$	$\hat{\sigma}^2 = 0,00046658$				
$\text{tr}(\mathbf{V}_0) = 58$	$\hat{\sigma}_0^2 = 0,00047319$			Nivel de significación	
				$p = \Pr(F > F_{obs}/H_0) = 0,94189272$	

Se concluye pues que no existen evidencias para rechazar la hipótesis de linealidad en el rango de edad de uno a seis años ($p = 0,94189272$), lo que está en consonancia con lo que se aprecia en la Figura 2 - derecha.

Si considerásemos, sin embargo, el conjunto total de datos, $\{(x_i, y_i)\}_{i=1}^{72}$, el resultado del F test correspondiente sería claramente significativo, ya que entonces $F_{obs} = 62,45946091$, por lo que se rechaza la hipótesis de linealidad desde el nacimiento, en concordancia con lo que se aprecia en la gráfica de la izquierda de la Figura 2.

7 . EL ANÁLISIS DE LA VARIANZA Y EL F TEST SPLINE

Supongamos que podemos asumir que $m \in \mathcal{M} = \mathcal{W}_2^p[0, 1]$. En esta sección nos centraremos en el contraste de hipótesis de un modelo de regresión polinómico,

$H_0 : m \in \mathcal{M}_0 = \{m_{\theta} : m_{\theta}(x) = \theta_0 + \theta_1x + \dots + \theta_{p-1}x^{p-1}\}$ frente a $H_1 : m \notin \mathcal{M}_0$.
(23)

Como estimadores de m consideraremos, $\hat{m}_0 = m_{\hat{\theta}}$, en donde $\hat{\theta}$ es el estimador mínimo cuadrático de θ , y $\hat{m}$ el estimador *spline* suave de orden $(2, p)$ y parámetro de suavización λ .

La distancia de las funciones de $g \in \mathcal{M}_0$ a $\hat{m}$ se medirá mediante la distancia entre g^* y $\hat{m}^*$ en el espacio $\mathcal{M}^* = \mathcal{W}_2^{p*}[0, 1] = \{g^* = (\mathbf{g}, g^{(p)}) : g \in \mathcal{W}_2^p[0, 1]\}$, es decir, mediante la distancia inducida por la norma de $\mathcal{H}$. Introducimos para ello el funcional ($g \in \mathcal{M}_0$),

$$\hat{D}(g, \lambda) = \frac{1}{n} \|\hat{m}^* - g^*\|_{\mathcal{H}}^2 = \frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - g(X_i)]^2 + \lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx.$$

El estimador $\hat{m}_0$ que por definición es la función que minimiza $\hat{V}(g)$ sobre $\mathcal{M}_0$, es también la función que minimiza $\hat{V}(g, \lambda)$ sobre $\mathcal{M}_0$ ya que $\int_0^1 [g^{(p)}(x)]^2 dx = 0$ si $g \in \mathcal{M}_0$ (y por tanto, $\hat{V}(g, \lambda) = \hat{V}(g)$ en $\mathcal{M}_0$). La función que minimiza $\hat{D}(g, \lambda)$ sobre $\mathcal{M}_0$ es también $\hat{m}_0$ ya que $\mathbf{H}\mathbf{X} = \mathbf{X}$ (véase Eubank (1984)), y por tanto, $\mathbf{H}\mathbf{P}_0 = \mathbf{P}_0$ en donde $\mathbf{P}_0$ denota la matriz que proporciona el vector de ajustes polinómico, $\hat{\mathbf{m}}_0 = \mathbf{P}_0 \mathbf{Y}$. En consecuencia, $\mathbf{P}_0 \hat{\mathbf{m}} = \mathbf{P}_0 \mathbf{H}\mathbf{Y} = \mathbf{P}_0 \mathbf{Y} = \hat{\mathbf{m}}_0$. Así pues,

$$\hat{m}_0 = \arg \min_{g \in \mathcal{M}_0} \hat{V}(g, \lambda) = \arg \min_{g \in \mathcal{M}_0} \hat{D}(g, \lambda).$$

7.1 . EL ANÁLISIS DE LA VARIANZA SPLINE

El estimador *spline* suave visto en $\mathcal{H} = \mathbb{R}^n \times L_2[0, 1]$, $\hat{m}^*$, es la proyección de $\mathbf{Y}^*$ sobre $\mathcal{M}^* = \mathcal{W}_2^{p*}[0, 1]$. Así pues, considerando $\mathbf{Y}$ y $\hat{m}_0$ en $\mathcal{H}$, obtenemos la descomposición de la varianza que generaliza a la clásica,

$$\frac{1}{n} \|\mathbf{Y}^* - \hat{m}_0^*\|_{\mathcal{H}}^2 = \frac{1}{n} \|\hat{m}^* - \hat{m}_0^*\|_{\mathcal{H}}^2 + \frac{1}{n} \|\mathbf{Y}^* - \hat{m}^*\|_{\mathcal{H}}^2,$$

o sea, $\hat{V}(\hat{m}_0, \lambda) = \hat{D}(\hat{m}_0, \lambda) + \hat{V}(\hat{m}, \lambda)$. En lo sucesivo denotaremos por

$$\hat{d}_{\mathcal{H}}^2(\lambda) = \hat{D}(\hat{m}_0, \lambda) = \frac{1}{n} \|\hat{m}^* - \hat{m}_0^*\|_{\mathcal{H}}^2 = \frac{1}{n} \sum_{i=1}^n [\hat{m}(x_i) - \hat{m}_0(x_i)]^2 + \lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx,$$

el término asociado a la distancia entre $\hat{m}^*$ y $\hat{m}_0^*$. (Los términos asociados a la varianza bajo H_0 y bajo H_1 , ya vistos, son $\hat{\sigma}_0^2 = n^{-1} \|\mathbf{Y}^* - \hat{m}_0^*\|_{\mathcal{H}}^2$ y $\hat{\sigma}_{\mathcal{H}}^2(\lambda) = n^{-1} \|\mathbf{Y}^* - \hat{m}^*\|_{\mathcal{H}}^2$).

Las interpretaciones de la descomposición de la varianza anterior se recogen en la tabla siguiente (véase también la Figura 3 - derecha).

Tabla 11. *Interpretación del análisis de la varianza spline*

Estimador	Media muestral	Valor del funcional	Mínimo del funcional	Distancia
$\hat{d}_{\mathcal{H}}^2(\lambda)$	$\frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)]^2 + \lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx$	$\hat{D}(\hat{m}_0, \lambda)$	$\min_{g \in \mathcal{M}_0} \hat{D}(g, \lambda)$	$\frac{1}{n} \ \hat{m}^* - \hat{m}_0^*\ _{\mathcal{H}}^2$
$\hat{\sigma}_{\mathcal{H}}^2(\lambda)$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2 + \lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx$	$\hat{V}(\hat{m}, \lambda)$	$\min_{g \in \mathcal{M}} \hat{V}(g, \lambda)$	$\frac{1}{n} \ \mathbf{Y}^* - \hat{m}^*\ _{\mathcal{H}}^2$
$\hat{\sigma}_0^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)]^2$	$\hat{V}(\hat{m}_0, \lambda)$ $= \hat{V}(\hat{m}_0)$	$\min_{g \in \mathcal{M}_0} \hat{V}(g, \lambda)$ $= \min_{g \in \mathcal{M}_0} \hat{V}(g)$	$\frac{1}{n} \ \mathbf{Y}^* - \hat{m}_0^*\ _{\mathcal{H}}^2$ $= \frac{1}{n} \ \mathbf{Y} - \hat{\mathbf{m}}_0\ ^2$

Obsérvese que, por las ecuaciones normales (13) y por ser $\hat{m}_0^{(p)}(x) \equiv 0$, se tiene $\sum_{i=1}^n [Y_i - \hat{m}(X_i)] \hat{m}_0(X_i) = 0$, por lo que (véase (14)),

$$\hat{\Delta}(\hat{m}_0) = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)] \hat{m}(X_i) = \lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx \tag{24}$$

En consecuencia, $\hat{d}_{\mathcal{H}}^2(\lambda) = \hat{D}(\hat{m}_0) + \hat{\Delta}(\hat{m}_0) = \tilde{d}^2$ y $\hat{\sigma}_{\mathcal{H}}^2(\lambda) = \hat{V}(\hat{m}) + \hat{\Delta}(\hat{m}_0) = \tilde{\sigma}^2$. Por tanto, el análisis de la varianza de la tabla anterior se corresponde con el de la Tabla 7.

7.2 . EL F TEST NATURAL

El F test natural (Ramil Novo y González Manteiga (2000)) se basa en la descomposición de la varianza de la Tabla 11. El estadístico de contraste de este test es,

$$F = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\tilde{\mathbf{D}})} \cdot \frac{\hat{d}_{\mathcal{H}}^2(\lambda)}{\hat{\sigma}_{\mathcal{H}}^2(\lambda)} = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\mathbf{V}_0 - \tilde{\mathbf{V}})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}_{\mathcal{H}}^2(\lambda)}{\hat{\sigma}_{\mathcal{H}}^2(\lambda)},$$

en donde $\tilde{\mathbf{D}} = \mathbf{H} - \mathbf{P}_0 = \mathbf{V}_0 - \tilde{\mathbf{V}}$, $\tilde{\mathbf{V}} = \mathbf{I} - \mathbf{H}$ y $\mathbf{V}_0 = \mathbf{I} - \mathbf{P}_0$ son las matrices que proporcionan $\hat{d}_{\mathcal{H}}^2(\lambda)$, $\hat{\sigma}_{\mathcal{H}}^2(\lambda)$ y $\hat{\sigma}_0^2$. En términos de los funcionales de la distancia y la varianza, la primera expresión se corresponde con,

$$F = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\tilde{\mathbf{D}})} \cdot \frac{\hat{D}(\hat{m}_0, \lambda)}{\hat{V}(\hat{m}, \lambda)} = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\tilde{\mathbf{D}})} \cdot \frac{\min_{g \in \mathcal{M}_0} \hat{D}(g, \lambda)}{\min_{g \in \mathcal{M}} \hat{V}(g, \lambda)},$$

y la segunda con,

$$F = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\mathbf{V}_0 - \tilde{\mathbf{V}})} \cdot \frac{\hat{V}(\hat{m}_0, \lambda) - \hat{V}(\hat{m}, \lambda)}{\hat{V}(\hat{m}, \lambda)}$$

$$= \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\mathbf{V}_0 - \tilde{\mathbf{V}})} \cdot \frac{\min_{g \in \mathcal{M}_0} \hat{V}(g, \lambda) - \min_{g \in \mathcal{M}} \hat{V}(g, \lambda)}{\min_{g \in \mathcal{M}} \hat{V}(g, \lambda)}.$$

La presentación de los resultados del análisis de la varianza *spline* suave y del F test natural puede hacerse mediante unas tablas análogas a las propuestas en el caso paramétrico.

Tabla 12. *Resultados del análisis de la varianza y del F test spline*

Trazas	ANOVA muestral	Estadístico de contraste Comparación relativa de ajustes o de varianzas
$\text{tr}(\tilde{\mathbf{D}}) = \text{tr}(\mathbf{H}) - p$	$\hat{d}_{\mathcal{H}}^2(\lambda) = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{D}} \mathbf{Y}$	$F = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\tilde{\mathbf{D}})} \frac{\hat{d}_{\mathcal{H}}^2(\lambda)}{\hat{\sigma}_{\mathcal{H}}^2(\lambda)}$
$\text{tr}(\tilde{\mathbf{V}}) = n - \text{tr}(\mathbf{H})$	$\hat{\sigma}_{\mathcal{H}}^2(\lambda) = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{V}} \mathbf{Y}$	$= \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\mathbf{V}_0 - \tilde{\mathbf{V}})} \frac{\hat{\sigma}_0^2 - \hat{\sigma}_{\mathcal{H}}^2(\lambda)}{\hat{\sigma}_{\mathcal{H}}^2(\lambda)}$
$\text{tr}(\mathbf{V}_0) = n - p$	$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V}_0 \mathbf{Y}$	Nivel de significación $p = \Pr(F > F_{obs}/H_0)$

Bajo H_0 , y ciertas condiciones adicionales (véase Ramil Novo y González Manteiga (2000)), la distribución del estadístico $S = [\text{tr}(\tilde{\mathbf{D}})/(2\text{tr}(\tilde{\mathbf{D}}^2))][F - 1]$ converge a una distribución $N(0, 1)$. En consecuencia, con tamaños de muestra grandes, $p \approx \Pr(N(0, 1) > S_{obs})$, en donde, de ahora en adelante, S_{obs} denota el valor observado en la muestra de datos del correspondiente estadístico S .

8 . EL TEST DE RAZÓN DE VEROSIMILITUDES

Analizamos aquí los F tests equivalentes al test de razón de verosimilitudes, basados en estimadores mínimo cuadráticos y *spline* suaves.

Si los errores siguen una distribución normal, la función de verosimilitud (ecuación (6)) alcanza su máximo en la misma función que la función soporte,

$$l(g, \sigma^2) = \ln(L(g, \sigma^2)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{g}\|^2.$$

8.1 . EL TEST DE RAZÓN DE VEROSIMILITUDES PARAMÉTRICO

Asumamos que $m \in \mathcal{M}$, el espacio de funciones lineales (5). Consideremos el contraste de hipótesis del modelo lineal $\mathcal{M}_0$ definido en (3), $H_0 : m \in \mathcal{M}_0$, frente a $H_1 : m \notin \mathcal{M}_0$.

Si $\hat{\theta}_0$ y $\hat{\theta}$ son los estimadores mínimo cuadráticos de θ correspondientes a $\mathcal{M}_0$ y $\mathcal{M}$, el máximo de l en $\mathcal{M}_0$ se alcanza en $\hat{m}_0 = m_{\hat{\theta}_0}$, y el máximo en $\mathcal{M}$ se alcanza en $\hat{m} = m_{\hat{\theta}}$. En efecto, para cualquier $g \in \mathcal{M}_0$, $\hat{V}(\hat{m}_0) \leq \hat{V}(g)$ (o equivalentemente $\|\mathbf{Y} - \hat{\mathbf{m}}_0\|^2 \leq \|\mathbf{Y} - \mathbf{g}\|^2$), y por tanto, $L(g, \sigma^2) \leq L(\hat{m}_0, \sigma^2)$. Un razonamiento análogo es válido para $\hat{m}$: para cualquier $g \in \mathcal{M}$, $\hat{V}(\hat{m}) \leq \hat{V}(g)$, y por tanto, $L(g, \sigma^2) \leq L(\hat{m}, \sigma^2)$. Como función de σ^2 , el máximo (que es la solución de la ecuación $\partial l / \partial \sigma^2 = 0$) se alcanza para $\hat{\sigma}_0^2 = \hat{V}(\hat{m}_0)$ ($= \|\mathbf{Y} - \hat{\mathbf{m}}_0\|^2 / n$) y $\hat{\sigma}^2 = \hat{V}(\hat{m})$ ($= \|\mathbf{Y} - \hat{\mathbf{m}}\|^2 / n$) en $\mathcal{M}_0$ y $\mathcal{M}$ respectivamente. Como resultado, el estadístico de razón de verosimilitudes es,

$$\begin{aligned} \Lambda &= \frac{\sup_{g \in \mathcal{M}_0, \sigma^2} L(g, \sigma^2)}{\sup_{g \in \mathcal{M}, \sigma^2} L(g, \sigma^2)} = \left(\frac{\hat{V}(\hat{m}_0)}{\hat{V}(\hat{m})} \right)^{-n/2} \\ &= \left(\frac{\|\mathbf{Y} - \hat{\mathbf{m}}_0\|^2}{\|\mathbf{Y} - \hat{\mathbf{m}}\|^2} \right)^{-n/2} = \left(\frac{\text{tr}(\mathbf{P} - \mathbf{P}_0)}{\text{tr}(\mathbf{I} - \mathbf{P})} F + 1 \right)^{-n/2}, \end{aligned}$$

en donde,

$$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\hat{D}(\hat{m}_0)}{\hat{V}(\hat{m})} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{V}(\hat{m}_0) - \hat{V}(\hat{m})}{\hat{V}(\hat{m})},$$

es el estadístico de contraste del test F (ecuaciones (7) y (22)) y por tanto los tests basados en Λ y F son equivalentes.

8.2 . EL TEST DE RAZÓN DE VEROSIMILITUDES SPLINE

En esta sección volvemos a considerar el contraste de un modelo de regresión polinómico (23). Bajo H_0 , el estimador de m será $\hat{m}_0 = m_{\hat{\theta}}$, en donde $\hat{\theta}$ es el estimador mínimo cuadrático de θ .

Bajo H_1 , el estimador de m será la función $\hat{m} = \hat{m}_\rho$ que minimiza sobre $\mathcal{W}_2^p[0, 1]$ el funcional $\hat{V}(g) = n^{-1} \sum_{i=1}^n [Y_i - g(X_i)]^2$ sujeta a las restricciones $J(g) = \int_0^1 [g^{(p)}(x)]^2 dx \leq \rho$ para toda $g \in \mathcal{W}_2^p[0, 1]$. Así pues,

$$\hat{m}_\rho = \arg \min_{g \in \mathcal{M}} \hat{V}(g) \quad \text{en donde } \mathcal{M} = \left\{ g : g \in \mathcal{W}_2^p[0, 1], \int_0^1 [g^{(p)}(x)]^2 dx \leq \rho \right\}. \quad (25)$$

La restricción $J(g) \leq \rho$, garantiza que la solución $\hat{m}_\rho$ es única si $n \geq p$, y además ρ determina el grado de suavidad de la curva. Cuanto más pequeño sea ρ , más suave será el estimador resultante. El estimador $\hat{m}_\rho$ es además equivalente a un estimador *spline* suave (véase Eubank (1988, Cáp. 5)). Más en concreto, existe una constante ρ_0 , tal que para cualquier $0 \leq \rho \leq \rho_0$, la solución $\hat{m}_\rho$ del problema anterior es el estimador *spline* suave $\hat{m}_\lambda$ de orden $(2, p)$ correspondiente al parámetro de suavización $\lambda = \lambda(\rho)$ determinado por $J(\hat{m}_\lambda) = \rho$. Recíprocamente, dado $\lambda > 0$, $\hat{m}_\lambda$ es la solución del problema de minimización anterior correspondiente a $\rho = J(\hat{m}_\lambda)$.

Dado que para cualquier $g \in \mathcal{M}$, $\hat{V}(\hat{m}) \leq \hat{V}(g)$ (o equivalentemente $\|\mathbf{Y} - \hat{\mathbf{m}}\|^2 \leq \|\mathbf{Y} - \mathbf{g}\|^2$), se tiene que $L(g, \sigma^2) \leq L(\hat{m}, \sigma^2)$. Como función de σ^2 , el máximo se alcanza para $\hat{\sigma}^2(\lambda) = \hat{V}(\hat{m}) = \|\mathbf{Y} - \hat{\mathbf{m}}\|^2 / n$. Como resultado, el estadístico de razón de verosimilitudes es,

$$\Lambda = \frac{\sup_{\mathcal{M}_0, \sigma^2} L(g, \sigma^2)}{\sup_{\mathcal{M}, \sigma^2} L(g, \sigma^2)}$$

$$= \left(\frac{\hat{V}(\hat{m}_0)}{\hat{V}(\hat{m})} \right)^{-n/2} = \left(\frac{\|\mathbf{Y} - \hat{\mathbf{m}}_0\|^2}{\|\mathbf{Y} - \hat{\mathbf{m}}\|^2} \right)^{-n/2} = \left(\frac{\text{tr}(\mathbf{V}_0 - \mathbf{V})}{\text{tr}(\mathbf{V})} F + 1 \right)^{-n/2},$$

en donde,

$$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \frac{\hat{V}(\hat{m}_0) - \hat{V}(\hat{m})}{\hat{V}(\hat{m})} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2(\lambda)}{\hat{\sigma}^2(\lambda)}$$

$$= \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \frac{\|\mathbf{Y} - \hat{\mathbf{m}}_0\|^2 - \|\mathbf{Y} - \hat{\mathbf{m}}\|^2}{\|\mathbf{Y} - \hat{\mathbf{m}}\|^2},$$

siendo $\mathbf{V} = (\mathbf{I} - \mathbf{H})^t(\mathbf{I} - \mathbf{H}) = (\mathbf{I} - \mathbf{H})^2$ y $\mathbf{V}_0 = \mathbf{I} - \mathbf{P}_0$ las matrices que proporcionan los estimadores $\hat{\sigma}^2(\lambda)$ (basado en $\hat{\mathbf{m}} = \mathbf{H}\mathbf{Y}$ con $\hat{m} = \hat{m}_{\lambda(\rho)}$) y $\hat{\sigma}_0^2$ respectivamente. Por tanto el test de razón de verosimilitudes es equivalente al test basado en el estadístico F anterior (estudiado por Ramil Novo y González Manteiga (2000)), pero no al F test natural. Ambos, el numerador y el denominador, son distintos a los del F test natural ya que $\mathbf{V}_0 - \mathbf{V} = (\mathbf{I} - \mathbf{P}_0) - (\mathbf{I} - \mathbf{H})^2 \neq \mathbf{H} - \mathbf{P}_0 = \tilde{\mathbf{D}}$ y $\mathbf{V} = (\mathbf{I} - \mathbf{H})^2 \neq \mathbf{I} - \mathbf{H} = \tilde{\mathbf{V}}$.

8.2.1 . EL ANÁLISIS DE LA VARIANZA SPLINE (TEST DE RAZÓN DE VEROSIMILITUDES)

La descomposición de la varianza de la Tabla 13 (que se corresponde con la Tabla 5), es útil para la interpretación del F test de razón de verosimilitudes *spline* suave. En la tabla, $\lambda = \lambda(\rho)$, $\hat{m} = \hat{m}_\rho (= \hat{m}_{\lambda(\rho)})$, $\hat{\delta}(\lambda) = 2\hat{\Delta}(\hat{m}_0)$, y $\mathcal{M}$ es el espacio definido en (25).

Tabla 13. Interpretación del análisis de la varianza spline (test de razón de verosimilitudes)

Estimador	Media muestral	Valor del funcional	Mínimo del funcional	Expresión vectorial
$\hat{d}^2(\lambda)$	$\frac{1}{n} \sum_{i=1}^n [\hat{m}(X_i) - \hat{m}_0(X_i)]^2$	$\hat{D}(\hat{m}_0)$	$\min_{g \in \mathcal{M}_0} \hat{D}(g)$	$\frac{1}{n} \ \hat{\mathbf{m}} - \hat{\mathbf{m}}_0\ ^2$
$\hat{\delta}(\lambda)$	$\frac{2}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)] \hat{m}(X_i)$	$2\hat{\Delta}(\hat{m}_0)$	$\min_{g \in \mathcal{M}_0} 2\hat{\Delta}(g)$	$\frac{2}{n} \hat{\mathbf{m}}^t (\mathbf{Y} - \hat{\mathbf{m}})$
$\hat{\sigma}^2(\lambda)$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2$	$\hat{V}(\hat{m})$	$\min_{g \in \mathcal{M}} \hat{V}(g)$	$\frac{1}{n} \ \mathbf{Y} - \hat{\mathbf{m}}\ ^2$
$\hat{\sigma}_0^2$	$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)]^2$	$\hat{V}(\hat{m}_0)$	$\min_{g \in \mathcal{M}_0} \hat{V}(g)$	$\frac{1}{n} \ \mathbf{Y} - \hat{\mathbf{m}}_0\ ^2$

Obsérvese que $\hat{\Delta}(g)$ es constante en $\mathcal{M}_0$. La presentación del F test puede hacerse como sigue a continuación.

Tabla 14. Resultados del análisis de la varianza y del F test de razón de verosimilitudes spline

Trazas	ANOVA muestral	Estadístico de contraste
$\text{tr}(\mathbf{D}) = \text{tr}(\mathbf{H}^2) - p$	$\hat{d}^2(\lambda) = \frac{1}{n} \mathbf{Y}^t \mathbf{D} \mathbf{Y}$	Comparación relativa de varianzas
$\text{tr}(\mathbf{B}) = 2\text{tr}(\mathbf{H}) - 2\text{tr}(\mathbf{H}^2)$	$\hat{\delta}(\lambda) = \frac{1}{n} \mathbf{Y}^t \mathbf{B} \mathbf{Y}$	$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2(\lambda)}{\hat{\sigma}^2(\lambda)}$
$\text{tr}(\mathbf{V}) = n - 2\text{tr}(\mathbf{H}) + \text{tr}(\mathbf{H}^2)$	$\hat{\sigma}^2(\lambda) = \frac{1}{n} \mathbf{Y}^t \mathbf{V} \mathbf{Y}$	Nivel de significación
$\text{tr}(\mathbf{V}_0) = n - p$	$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V}_0 \mathbf{Y}$	$p = \Pr(F > F_{\text{obs}}/H_0)$

Bajo H_0 , y ciertas condiciones adicionales (véase Ramil Novo y González Manteiga (2000)), la distribución del estadístico $S = [\text{tr}(\mathbf{D})/(2\text{tr}(\mathbf{D}^2))][F - 1]$ converge a una distribución $N(0, 1)$. En consecuencia, para tamaños muestrales grandes, $p \approx \Pr(N(0, 1) > S_{\text{obs}})$.

8.3 . ANÁLISIS ESTADÍSTICO NO PARAMÉTRICO DEL EJEMPLO

Consideremos de nuevo el conjunto de datos $\{(x_i, y_i)\}_{i=1}^{72}$ de la Figura 2-izquierda. La función de regresión se puede estimar sin asumir una forma paramétrica, lo que es menos restrictivo. Es necesario, no obstante, asumir ciertas condiciones de suavidad acerca de m . Consideramos aquí el estimador *spline*, y los valores $z_i = x_i/n, i = 1, \dots, 72$, pertenecientes al intervalo $[0, 1]$. Si

$\hat{m}_\lambda(z)$ es el ajuste correspondiente a $\{(z_i, y_i)\}_{i=1}^{72}$, el ajuste correspondiente a los datos originales es $\hat{m}_\lambda(x/72)$. Seleccionando un estimador $\hat{m}_\lambda$ *spline* suave cúbico (o *spline* suave de orden $(2, p)$ con $p = 2$), es necesario asumir que $m \in \mathcal{W}_2^2[0, 1] = \{g : g, g' \text{ son absolutamente continuas y } \int_0^1 [g''(x)]^2 dx < \infty\}$ (véase la Sección 4.1). La definición alternativa de la Sección 8.2, requiere poder considerar que $m \in \mathcal{M} = \{g : g \in \mathcal{W}_2^2[0, 1], \int_0^1 [g''(x)]^2 dx \leq \rho\}$ en el caso del *spline* cúbico ($p = 2$). En el primer caso, la estimación óptima se obtiene considerando el parámetro λ_o que minimiza la denominada función de validación cruzada generalizada (véase Eubank (1988, Cáp. 5)),

$$GCV(\lambda) = \hat{V}(\hat{m}_\lambda) / [1 - \frac{1}{n} \text{tr}(\mathbf{H}(\lambda))]^2 = \frac{n^{-1} \sum_{i=1}^n [Y_i - \hat{m}_\lambda(X_i)]^2}{[n^{-1} \text{tr}(\mathbf{I} - \mathbf{H}(\lambda))]^2}.$$

Evaluando $GCV(\lambda)$ en la sucesión de valores $\lambda(i) = [h(i)]^4$, $i = 0, \dots, 3900$, en donde $h(i) = 0,01 + 0,0001 \times i$, se obtuvo como valor aproximado de λ_o , $\lambda = 0,033^4$. El estimador correspondiente a este parámetro de suavización es el representado en la Figura 2 - izquierda. Dicho estimador coincide con el estimador $\hat{m}_\rho$ basado en el parámetro $\rho = J(\hat{m}_\lambda) = \int_0^1 [\hat{m}_\lambda''(x)]^2 dx$. Por tanto, $\rho = \tilde{\delta}/\lambda = 60,97473862$ (véase (24)).

Para contrastar la hipótesis nula de linealidad de uno a seis años, podemos considerar los valores $\{x_i\}_{i=13}^{72}$ recentrados y reescalados en el intervalo $[0, 1]$, $z_i = (x_i - 12)/60$, $i = 13, \dots, 72$. La hipótesis nula es entonces, $H_0 : m(z) = \theta_0 + \theta_1 z$, y la hipótesis alternativa, $H_1 : m(z)$ no es una recta.

Para aplicar el F test natural, consideramos el estimador *spline* correspondiente a los datos $\{(x_i, y_i)\}_{i=13}^{72}$ con el parámetro de suavización $\lambda = 0,033^4$ (el obtenido por el criterio GCV para el conjunto global de datos). En el F test de razón de verosimilitudes se consideró el mismo estimador *spline* suave cúbico, correspondiente ahora a $\rho = 9,72666347$.

Resultados del análisis de la varianza y del F test natural
(Tabla 12, *spline* suave cúbico)

Trazas		ANOVA muestral	Estadístico de contraste
$\text{tr}(\tilde{\mathbf{D}}) = 8,18515954$		$\hat{d}_{\mathcal{H}}^2(\lambda) = 0,00002235$	$\tilde{F} = \frac{\text{tr}(\tilde{\mathbf{V}})}{\text{tr}(\tilde{\mathbf{D}})} \cdot \frac{\hat{d}_{\mathcal{H}}^2(\lambda)}{\hat{\sigma}_{\mathcal{H}}^2(\lambda)} = 0,30172341$
$\text{tr}(\tilde{\mathbf{V}}) = 49,81484046$		$\hat{\sigma}_{\mathcal{H}}^2(\lambda) = 0,00045084$	
$\text{tr}(\mathbf{V}_0) = 58$		$\hat{\sigma}_0^2 = 0,00047319$	Nivel de significación
			$p = \Pr(\tilde{F} > \tilde{F}_{obs}/H_0) \approx 0,99223815$

Resultados del análisis de la varianza y del F test tipo razón de verosimilitudes

(Tabla 14, *spline suave cúbico*)

Trazas	ANOVA muestral	Estadístico de contraste
		Comparación relativa de varianzas
$\text{tr}(\mathbf{D}) = 6,13381337$	$\hat{d}^2(\lambda) = 0,00001082$	$F = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2(\lambda)}{\hat{\sigma}^2(\lambda)} =$
$\text{tr}(\mathbf{B}) = 4,10269234$	$\hat{\delta}(\lambda) = 0,00002307$	
$\text{tr}(\mathbf{V}) = 47,76349429$	$\hat{\sigma}^2(\lambda) = 0,00043930$	$= 0,35991668$
$\text{tr}(\mathbf{V}_0) = 58$	$\hat{\sigma}_0^2 = 0,00047319$	Nivel de significación
		$p = \Pr(F > F_{obs}/H_0) \approx$
		$\approx 0,98580437$

Ambos tests dan un resultado que no es significativo, como podemos ver en las tablas anteriores, por lo que no hay razones para rechazar la hipótesis de linealidad. En estos tests, se observa además una mayor evidencia que con el F test paramétrico de que un polinomio de grado uno es suficiente para obtener un buen ajuste.

Si incluyésemos en el análisis todos los datos, $\{(x_i, y_i)\}_{i=1}^{72}$, el resultado sería significativo en ambos F tests no paramétricos con parámetro $\lambda = 0,033^4$. Los correspondientes valores de los estadísticos de contraste son $\tilde{F}_{obs} = 25,18443044$ y $F_{obs} = 22,78932854$. En consecuencia, se rechaza la hipótesis de linealidad en el rango de edad de cero a seis años.

9 . EXTENSIONES BASADAS EN OTROS ESTIMADORES NO PARAMÉTRICOS

Una parte importante de las analogías entre el proceso de estimación y de contraste de hipótesis mínimo cuadráticos y *spline* debería de poder extenderse a otros estimadores no paramétricos. Mostramos aquí algunas posibilidades.

En la Sección 5.3 se consideró el estimador $\tilde{d}^2$ (ecuación (21)), motivado por la interpretación de $d^2 = E[r^2(X)]$ como una covarianza, asumiendo que $E[r(X)] = 0$ ($r(X) = m(X) - m_0(X)$) y que X es una variable aleatoria. Bajo estas condiciones, la expresión de $d^2 = E[r^2(X)]$ admite también las interpretaciones,

$$E[r^2(X)] = E[Ur(X)] = \text{cov}(U, r(X)) = \max_{g \in \mathcal{M}} E[Ug(X)] = \max_{g \in \mathcal{M}} \text{cov}(U, g(X)), \tag{26}$$

en donde $\mathcal{M} = \{g : g \text{ es una función de Borel tal que } \|g\|^2 = \int g^2(x) dF(x) \leq d^2\}$. Así pues,

$$C = \frac{d^2}{\sigma^2} = \frac{\max_{g \in \mathcal{M}} E[Ug(X)]}{\sigma^2}.$$

Es más, si en la definición de $\mathcal{M}$ consideramos $\|g\|^2 \leq c$ con c una constante fija positiva arbitraria, el máximo se alcanzaría con una función proporcional a $r(x)$ (Bradley y McClelland (1996a)).

La expresión (26) sugiere la posibilidad de considerar el estimador,

$$\hat{d}_U^2 = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)] \hat{r}(X_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n W_{hj}(X_i) [Y_i - \hat{m}_0(X_i)] [Y_j - \hat{m}_0(X_j)], \quad (27)$$

en donde $\hat{U} = Y - \hat{m}_0(X)$ ($\hat{m}_0$ el estimador mínimo cuadrático) y donde $\hat{r}(x) = \sum_{j=1}^n W_{hj}(x) [Y_j - \hat{m}_0(X_j)]$ es un estimador no paramétrico de $u(x) = E[\hat{U}/X = x]$. Por último, la matriz de estimación de $\hat{d}_U^2 = n^{-1} \mathbf{Y}^t \mathbf{\Omega} \mathbf{Y}$ es $\mathbf{\Omega} = (\mathbf{I} - \mathbf{P}_0) \mathbf{H} (\mathbf{I} - \mathbf{P}_0)$.

9.1 . ESTIMADOR POLINÓMICO LOCAL Y ESTIMADOR TIPO NÚCLEO

Una función real se denomina analítica en un punto x si en un entorno del punto se puede desarrollar en serie de potencias (serie de Taylor). Así pues, si m es analítica, y z pertenece al entorno de x en el que el desarrollo es válido,

$$m(z) = \sum_{k=0}^{\infty} \frac{m^{(k)}(x)}{k!} (z - x)^k.$$

Si $m \in \mathcal{M} = \mathcal{C}^{q+1}[0, 1] = \{g : g, g', \dots, g^{(q+1)} \text{ son funciones continuas}\}$, m se puede aproximar por el polinomio de Taylor de grado q , $m(z) \simeq \theta_0(x) + \theta_1(x)(z - x) + \dots + \theta_q(x)(z - x)^q$, en donde $\theta_j(x) = m^{(j)}(x)/j!$, $j = 0, \dots, q$. Se denomina estimador polinómico local de grado q (orden $q + 1$) de $\theta(x) = (\theta_0(x), \dots, \theta_q(x))^t$, al vector $\hat{\theta}(x) = (\hat{\theta}_0(x), \dots, \hat{\theta}_q(x))^t$ que minimiza el funcional,

$$\hat{V}(\theta(x)) = \frac{1}{n} \sum_{i=1}^n [Y_i - \sum_{j=0}^q \theta_j(x)(X_i - x)^j]^2 K_h(X_i - x),$$

en donde los pesos se definen como,

$$K_h(X_i - x) = \frac{1}{h} K\left(\frac{X_i - x}{h}\right),$$

siendo K una función núcleo, que determina la forma de los pesos, y h el denominado parámetro *ventana*, que sirve para controlar la suavidad del estimador resultante. En general una función núcleo es una función continua, simétrica con respecto al cero, y tal que $\int K(x)dx = 1$. En la práctica, sin embargo, el núcleo K es generalmente una función de densidad con soporte un intervalo finito. Así por ejemplo, si el soporte de K es $[-1, 1]$, el soporte de $K_h(X_i - x)$ es $[X_i - h, X_i + h]$ y $\int_{X_i - h}^{X_i + h} K_h(X_i - x)dx = 1$, por lo que el parámetro ventana

se puede interpretar como el parámetro de escala que ajusta la forma y el tamaño de la función núcleo al entorno de X_i , $[X_i - h, X_i + h]$.

El estimador polinómico local de grado q de la función de regresión es $\hat{m}(x) = \hat{\theta}_0(x)$. Un caso particular de interés es el correspondiente a $q = 0$. El estimador resultante, al que se conoce como estimador tipo núcleo de Nadaraya-Watson, es

$$\hat{m}(x) = \frac{1}{n\hat{f}(x)} \sum_{i=1}^n K_h(X_i - x)Y_i, \text{ en donde } \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_h(X_i - x), \quad (28)$$

y $\hat{m}(x) = 0$ si $\hat{f}(x) = 0$. En la definición anterior, el estimador tipo núcleo de la función de densidad de X , $\hat{f}(x)$, juega el papel de ajustar los pesos que proporciona el núcleo K a la frecuencia de aparición de valores de X en torno a x .

El vector de ajustes del estimador polinómico local de grado q ,

$$\hat{\mathbf{m}} = (\hat{\theta}_0(X_1), \dots, \hat{\theta}_0(X_n))^t = \mathbf{H}\mathbf{Y},$$

se puede obtener a partir de las matrices $\mathbf{P}(x)$ que proporcionan los ajustes locales,

$$\mathbf{P}(x) = \mathbf{X}(x)[\mathbf{X}^t(x)\mathbf{\Pi}(x)\mathbf{X}(x)]^{-1}\mathbf{X}^t(x)\mathbf{\Pi}(x),$$

en donde,

$$\mathbf{X}(x) = \begin{pmatrix} 1 & X_1 - x & \dots & (X_1 - x)^q \\ \vdots & \vdots & & \vdots \\ 1 & X_n - x & \dots & (X_n - x)^q \end{pmatrix} \quad \text{y} \quad \mathbf{\Pi}(x) = \text{diag}\left[\frac{1}{h}K\left(\frac{X_i - x}{h}\right)\right].$$

En efecto, $\hat{m}(X_i) = \hat{\theta}_0(X_i) = e_{i,n}^t \mathbf{P}(X_i) \mathbf{Y}$ en donde $e_{i,n} = (0, \dots, 1, 0, \dots, 0)^t$ (el uno en la posición i -ésima). Por consiguiente, la entrada (i, j) de la matriz $\mathbf{H} = [W_h(X_i, X_j)]$ se puede obtener ajustando localmente un polinomio de grado q al conjunto de datos $\{X_i, \delta_{ij}\}_{i=1}^n$, y evaluando a continuación dicho ajuste en el punto X_i , o sea, $W_h(X_i, X_j) = e_{i,n}^t P(X_i) e_{j,n} =$ elemento i -ésimo de la j -ésima columna de $P(X_i)$.

9.2 . ANÁLISIS DE LA VARIANZA BASADO EN EL ESTIMADOR POLINÓMICO LOCAL

Consideremos el contraste de hipótesis (23) de un modelo de regresión polinómico. Examinamos aquí las descomposiciones de la varianza basadas en el estimador polinómico local análogas a las estudiadas con el estimador *spline* suave.

Denotemos por $\mathbf{H}$ la matriz de pesos del estimador polinómico local de grado $q \geq p$. El análisis de la varianza de la Tabla 5, $\hat{V}(\hat{m}_0) = \hat{D}(\hat{m}_0) +$

$2\hat{\Delta}(\hat{m}_0) + \hat{V}(\hat{m})$, basado en la matriz $\mathbf{H}$ requiere el estimador de σ^2 basado en el estimador polinómico local, $\hat{\sigma}^2 = \hat{V}(\hat{m})$, estudiado por Ruppert; Wand; Holst y Hössjer (1997). Como estimador de d^2 se utiliza la comparación directa de ajustes, $\hat{d}^2 = \hat{D}(\hat{m}_0)$, basada en el estimador polinómico local, estadístico que fue estudiado por Alcalá; Cristóbal y González Manteiga (1999).

Consideremos ahora el análisis de la varianza de la Tabla 6, $\hat{V}(\hat{m}_0) = \tilde{D}(\hat{m}_0) + \hat{\Delta}(\hat{m}_0) + \hat{V}(\hat{m})$. El estimador $\hat{d}_V^2$ basado en el estimador polinómico local de grado $q \geq p$ aplicado a los residuos $\hat{U}_i = Y_i - \hat{m}_0(X_i)$ del modelo paramétrico fue estudiado por Liu, Stengos y Li (2000). Este estadístico está basado en la matriz $\mathbf{\Omega} = (\mathbf{I} - \mathbf{P}_0)\mathbf{H}(\mathbf{I} - \mathbf{P}_0)$. Dado que se verifica la propiedad $\mathbf{H}\mathbf{P}_0 = \mathbf{P}_0$, la matriz $\mathbf{\Omega}$ coincide con $\tilde{\mathbf{D}} = (\mathbf{I} - \mathbf{P}_0)(\mathbf{H} - \mathbf{P}_0)$. Por consiguiente, con el estimador polinómico local considerado, $\hat{d}_V^2$ coincide con el estadístico $\tilde{d}^2 = \tilde{D}(\hat{m}_0)$.

A diferencia de lo que ocurría con el estimador *spline* suave considerado en las Secciones 7 y 8.2, la matriz $\mathbf{H}$ del estimador polinómico local considerado no es simétrica, por lo que $\tilde{\mathbf{D}} = (\mathbf{I} - \mathbf{P}_0)(\mathbf{H} - \mathbf{P}_0)$ no coincide con $\mathbf{H} - \mathbf{P}_0$. El estimador ponderado de la varianza, $\tilde{\sigma}^2$, tampoco se corresponde con la versión muestral de (11), $n^{-1} \sum_{i=1}^n [Y_i - \hat{m}(X_i)]Y_i$.

Considérense las expresiones de las matrices de estimación $\mathbf{V}$ y $\tilde{\mathbf{D}}$ (definidas en la Tabla 8),

$$\mathbf{V} = \mathbf{I} - \mathbf{H} - \mathbf{H}^t(\mathbf{I} - \mathbf{H}) \text{ y } \tilde{\mathbf{D}} = \mathbf{H} - \mathbf{P}_0 + \mathbf{P}_0(\mathbf{I} - \mathbf{H}),$$

válidas para cualquier estimador $\hat{m}$. La expresión de $\mathbf{V}$ muestra que,

$$\begin{aligned} \frac{1}{n} \mathbf{Y}^t(\mathbf{I} - \mathbf{H})\mathbf{Y} &= \frac{1}{n} \mathbf{Y}^t(\mathbf{I} - \mathbf{H})^t(\mathbf{I} - \mathbf{H})\mathbf{Y} + \frac{1}{n} \mathbf{Y}^t\mathbf{H}^t(\mathbf{I} - \mathbf{H})\mathbf{Y} \\ &= \tilde{\sigma}^2 + \hat{\Delta}_1 = \tilde{\sigma}^2 + \hat{\Delta}(\hat{m}_0) + \hat{\Delta}_0 = \tilde{\sigma}^2 + \hat{\Delta}_0. \end{aligned}$$

De la expresión de $\tilde{\mathbf{D}}$ se deduce que,

$$\frac{1}{n} \mathbf{Y}^t(\mathbf{H} - \mathbf{P}_0)\mathbf{Y} = \frac{1}{n} \mathbf{Y}^t\tilde{\mathbf{D}}\mathbf{Y} - \frac{1}{n} \mathbf{Y}^t\mathbf{P}_0(\mathbf{I} - \mathbf{H})\mathbf{Y} = \tilde{d}^2 - \hat{\Delta}_0 \quad (= \hat{d}^2 + \hat{\Delta}(\hat{m}_0) - \hat{\Delta}_0)$$

Considerando la matriz $\mathbf{P}_0$ correspondiente al modelo de regresión polinómico $\mathcal{M}_0$ definido en (23), y la matriz $\mathbf{H}$ del estimador polinómico local de grado $q \geq p$, el término $\hat{\Delta}_0$ no se anula, a diferencia de lo que ocurría con el estimador *spline* suave considerado en la Tabla 11.

9.2.1 . F TESTS BASADOS EN EL ESTIMADOR POLINÓMICO LOCAL

El F test basado en la comparación relativa de las varianzas definido en términos del estimador polinómico local fue propuesto por Cleveland y

Devlin (1988). Fan, Zhang y Zhang (2001) estudiaron la correspondiente generalización del test de razón de verosimilitudes, equivalente a dicho F test. La presentación de los resultados puede hacerse en una tabla como la siguiente (análoga a la Tabla 14).

Tabla 15. *Resultados del análisis de la varianza y del F test basado en el estimador polinómico local. Comparación de ajustes y comparación de varianzas*

Trazas	ANOVA muestral	Estadístico de contraste
$\text{tr}(\mathbf{D}) = \text{tr}(\mathbf{H}^t \mathbf{H}) - p$	$\hat{d}^2(h) = \frac{1}{n} \mathbf{Y}^t \mathbf{D} \mathbf{Y}$	Comparación de ajustes $F_D = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\hat{d}^2(h)}{\hat{\sigma}^2(h)}$
$\text{tr}(\mathbf{B}) = 2\text{tr}(\mathbf{H}) - 2\text{tr}(\mathbf{H}^t \mathbf{H})$	$\hat{\delta}(h) = \frac{1}{n} \mathbf{Y}^t \mathbf{B} \mathbf{Y}$	Comparación de varianzas $F_V = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{V}_0 - \mathbf{V})} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2(h)}{\hat{\sigma}^2(h)}$
$\text{tr}(\mathbf{V}) = n - 2\text{tr}(\mathbf{H}) + \text{tr}(\mathbf{H}^t \mathbf{H})$	$\hat{\sigma}^2(h) = \frac{1}{n} \mathbf{Y}^t \mathbf{V} \mathbf{Y}$	
$\text{tr}(\mathbf{V}_0) = n - p$	$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V}_0 \mathbf{Y}$	Niveles de significación $p_D = \Pr(F_D > F_{obs}/H_0)$ $p_V = \Pr(F_V > F_{obs}/H_0)$

El F test basado en la comparación de ajustes admite dos versiones, en función de qué estimador $\hat{d}^2$ o $\tilde{d}^2$ se utilice. La Tabla 15 muestra los resultados del ANOVA y el F test asociados a $\hat{d}^2$. El análisis de la varianza asociado al F test basado en $\tilde{d}^2$ se presenta en la Tabla 16.

Tabla 16. *Resultados del análisis de la varianza basado en el estimador polinómico local. Comparación de ajustes “penalizada” y comparación de varianzas.*

Trazas	ANOVA muestral	Estadístico de contraste
$\text{tr}(\tilde{\mathbf{D}}) = \text{tr}(\mathbf{H}) - p$	$\tilde{d}^2(h) = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{D}} \mathbf{Y}$	Comparación de ajustes “penalizada” $\tilde{F} = \frac{\text{tr}(\mathbf{V})}{\text{tr}(\mathbf{D})} \cdot \frac{\tilde{d}^2(h)}{\hat{\sigma}^2(h)}$
$\text{tr}(\tilde{\mathbf{B}}) = \text{tr}(\mathbf{H}) - \text{tr}(\mathbf{H}^t \mathbf{H})$	$\tilde{\delta}(h) = \frac{1}{n} \mathbf{Y}^t \tilde{\mathbf{B}} \mathbf{Y}$	Nivel de significación $p = \Pr(\tilde{F} > \tilde{F}_{obs}/H_0)$
$\text{tr}(\mathbf{V}) = n - 2\text{tr}(\mathbf{H}) + \text{tr}(\mathbf{H}^t \mathbf{H})$	$\hat{\sigma}^2(h) = \frac{1}{n} \mathbf{Y}^t \mathbf{V} \mathbf{Y}$	
$\text{tr}(\mathbf{V}_0) = n - p$	$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{Y}^t \mathbf{V}_0 \mathbf{Y}$	

Bajo H_0 , y ciertas condiciones adicionales, las distribuciones de $S = [\text{tr}(\mathbf{D})/(2\text{tr}(\mathbf{D}^2))][F_D - 1]$, $S = [\text{tr}(\mathbf{V}_0 - \mathbf{V})/(2\text{tr}((\mathbf{V}_0 - \mathbf{V})^2))][F_V - 1]$ y $S = [\text{tr}(\tilde{\mathbf{D}})/(2\text{tr}(\tilde{\mathbf{D}}^2))][\tilde{F} - 1]$, convergen a una distribución $N(0, 1)$ (véanse las referencias ya citadas). En consecuencia, para tamaños de muestra grandes, $p \approx \Pr(N(0, 1) > S_{obs})$.

9.3 . ESTIMADORES TIPO NÚCLEO

Consideremos la matriz de pesos $\mathbf{H}$ del estimador tipo núcleo de Nadaraya-Watson, $\hat{m}$ (ecuación (28)), y el contraste de la hipótesis nula del modelo lineal general (3), $H_0 : m \in \mathcal{M}_0$, frente a $H_1 : m \notin \mathcal{M}_0$.

El estimador $\hat{\sigma}^2 = \hat{V}(\hat{m})$ basado en el estimador tipo núcleo de Nadaraya-Watson, fue estudiado por Hall y Marron (1990). El F test basado en la comparación relativa de varianzas, $\hat{C} = (\hat{\sigma}_0^2 - \hat{\sigma}^2)/\hat{\sigma}^2$, definido en términos de dicho estimador fue estudiado por Dette (1999). El análisis de la varianza asociado es el de la Tabla 5.

Consideremos ahora el análisis de la varianza de la Tabla 6, $\hat{V}(\hat{m}_0) = \tilde{D}(\hat{m}_0) + \hat{\Delta}(\hat{m}_0) + \hat{V}(\hat{m})$. Zheng (1996) estudió un test basado en un estimador similar al $\hat{d}_V^2(h)$ de la ecuación (27) utilizando el estimador tipo núcleo de Nadaraya-Watson. El estimador $\hat{d}_V^2(h)$ no proporciona sin embargo una descomposición de la varianza basada en el estimador tipo núcleo del tipo $\hat{\sigma}_0^2 = \tilde{d}^2(h) + \tilde{\delta}(h) + \hat{\sigma}^2(h)$, puesto que la matriz de pesos del estimador de Nadaraya-Watson no verifica que $\mathbf{H}\mathbf{P}_0 = \mathbf{P}_0$, lo que implica que $\hat{d}_V^2$ y $\tilde{d}^2$ (ecuaciones (27) y (21)), no coinciden. Un estadístico de contraste similar a $\tilde{d}^2$ es el siguiente, estudiado por Bradley y McClelland (1996b),

$$\tilde{d}_{BM}^2 = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_0(X_i)][\hat{m}_i(X_i) - \hat{m}_0(X_i)]\hat{f}_i(X_i),$$

en donde $\hat{m}(x)$ es un estimador tipo núcleo de Nadaraya-Watson de la función de regresión, y $\hat{f}(x)$ es un estimador tipo núcleo de la función de densidad de X . La versión empleada de estos estimadores no permite, sin embargo, una expresión del tipo $n^{-1}\mathbf{Y}^t(\mathbf{I} - \mathbf{P}_0)\mathbf{W}(\mathbf{H} - \mathbf{P}_0)\mathbf{Y}$ con $\mathbf{W} = \text{diag}[\hat{f}_i(X_i)]$, ya que cada estimación $\hat{m}_i(X_i)$ requiere determinar un parámetro ventana.

9.4 . ESTIMADORES TIPO SERIE

La generalización más simple de los espacios de funciones lineales, paramétricos, al contexto no paramétrico, la proporcionan los espacios de funciones infinito dimensionales que constituyen un espacio de Hilbert. En un

espacio de Hilbert infinito dimensional las funciones admiten una representación del tipo $g = \sum_{k=1}^{\infty} \alpha_k \varphi_k$ en donde $\{\varphi_k\}_{k=1}^{\infty}$ es un sistema ortonormal completo de funciones. Un espacio de Hilbert importante es,

$$\mathcal{L}_2[0, 1] = \left\{ g : \|g\|^2 = \int g^2(x) dF(x) < \infty \right\}.$$

Dado un sistema ortogonal completo de $\mathcal{L}_2[0, 1]$, $\{\varphi_k\}_{k=1}^{\infty}$, denotando por $\alpha_k = \int_0^1 m(x) \varphi_k^*(x) dx$, $k = 1, 2, \dots$, los denominados coeficientes de Fourier,

$$m(x) \sim \sum_{k=1}^{\infty} \alpha_k \varphi_k(x),$$

es el desarrollo en serie de Fourier de m . Para cada N (si $n \geq N$) se obtiene un estimador por series de Fourier $\hat{m}_N = \sum_{k=1}^N \hat{\alpha}_k \varphi_k(x)$, en donde $\hat{\alpha}_k = n^{-1} \sum_{i=1}^n \varphi_k(X_i) Y_i$, que se puede caracterizar como,

$$\hat{m}_N = \arg \min_{g \in \mathcal{M}_N} \frac{1}{n} \sum_{i=1}^n [Y_i - g(X_i)]^2,$$

pues se obtiene minimizando $\hat{V}(g)$ sobre $\mathcal{M}_N = \{g : g = \sum_{k=1}^N \alpha_k \varphi_k, \alpha_k \in \mathbb{R}\}$.

Un caso importante es el estimador por series de Fourier clásico, asociado a $\varphi_k(x) = \exp(2\pi i k x)$, $k = 0, \pm 1, \dots$. Este estimador es un estimador lineal, $\hat{m}_N(x) = \sum_{j=1}^n W_N(x, X_j) Y_j$, con función de pesos $W_N(x, X_j) = n^{-1} K_N(x - X_j)$, en donde $K_N(x) = \sin(\pi(2N+1)x) / \sin(\pi x)$ si $x \notin \mathbb{Z}$, y $K_N(x) = 2N+1$ si $x \in \mathbb{Z}$. El vector de estimaciones es $\hat{\mathbf{m}} = \mathbf{H}\mathbf{Y}$, en donde,

$$\mathbf{H} = \Phi(\Phi^* \Phi)^{-1} \Phi^*$$

es una matriz real, simétrica e idempotente, y $\Phi = (\varphi_{-N}, \dots, \varphi_N)$. Dado que,

$$\hat{\mathbf{m}}_N = \arg \min_{g \in \mathcal{M}_N} \hat{V}(g) = \arg \min_{g \in \mathcal{M}_N} \|\mathbf{Y} - \mathbf{g}\|^2 = \arg \min_{g \in \mathcal{M}_N} \left\| \mathbf{Y} - \sum_{k=-N}^N \alpha_k \varphi_k \right\|^2,$$

($\mathcal{M}_N = \{g : g = \sum_{k=-N}^N \alpha_k \varphi_k, \alpha_k \in \mathbb{R}\}$) $\hat{\mathbf{m}}_N$ es la proyección de $\mathbf{Y}$ sobre el espacio vectorial generado por $\varphi_{-N}, \dots, \varphi_N$.

El estimador de la varianza análogo a los considerados con otros estimadores es,

$$\hat{V}(\hat{m}) = \min_{g \in \mathcal{M}_N} \hat{V}(g) = \min_{g \in \mathcal{M}_N} \frac{1}{n} \|\mathbf{Y} - \mathbf{g}\|^2 = \frac{1}{n} \|\mathbf{Y} - \hat{\mathbf{m}}_N\|^2,$$

en donde, $\|\mathbf{Y} - \hat{\mathbf{m}}_N\|^2 = \|\mathbf{Y} - \mathbf{H}(\mathbf{N})\mathbf{Y}\|^2 = \mathbf{Y}^t (\mathbf{I} - \mathbf{H}(\mathbf{N}))^t (\mathbf{I} - \mathbf{H}(\mathbf{N})) \mathbf{Y} = \mathbf{Y}^t (\mathbf{I} - \mathbf{H}(\mathbf{N})) \mathbf{Y}$.

AGRADECIMIENTOS

Los autores quieren expresar su agradecimiento al *referee* de este trabajo. Sus comentarios contribuyeron notablemente a mejorar la presentación del mismo. Asimismo quieren mostrar su gratitud a los editores por el interés mostrado por el trabajo.

Este trabajo ha sido parcialmente financiado por el *Ministerio Español de Ciencia y Tecnología* por medio del Proyecto MCYT-BFM2002-03213, por el *Ministerio de Educación y Ciencia* mediante el proyecto MTM2005-00820. y por la Xunta de Galicia, a través del proyecto PGIDIT03PXIC20702PN.

REFERENCIAS

- [1] J.T. ALCALÁ, J.A. CRISTÓBAL Y W. GONZÁLEZ MANTEIGA, Goodness-of-fit test for linear models based on local polynomials. *Statist. Probab. Lett.* **42** (1999) 39–46.
- [2] R. BRADLEY, R. MCCLELLAND, A kernel test for neglected nonlinearity. *Studies in Nonlinear Dynamics and Econometrics* **1** (1996a) 2, 119–130.
- [3] R. BRADLEY Y R. MCCLELLAND, A general nonparametric test for misspecification. *Technical Report*, US Bureau of Labor Statistics, 1996b, Washington, DC.
<http://www.bls.gov/ore/abstract/ec/ec960030.htm>
- [4] W.S. CLEVELAND Y S.J. DEVLIN, Locally weighted regression: an approach to regression analysis by local fitting, *J. Amer. Statist. Assoc.* **83** (1988) 403, 596–610.
- [5] D.D. COX, Asymptotics for M -type smoothing splines, *Ann. Statist.* **11** (1983) 2, 530–551.
- [6] D.D. COX, Multivariate smoothing spline functions, *SIAM J. Numer. Anal.* **21** (1984) 4, 789–813.
- [7] A. DEMMLER Y C. REINSCH, Oscillation matrices with spline functions, *Numer. Math.* **24** (1975) 375–382.
- [8] H. DETTE, A consistent test for the functional form of a regression based on a difference of variance estimators, *Ann. Statist.* **27** (1999) 1012–1050.
- [9] R. L. EUBANK, The hat matrix for smoothing splines, *Statist. Probab. Lett.* **2** (1984) 9–14.
- [10] R. L. EUBANK, *Spline Smoothing and Nonparametric Regression*. New York, 1988, Marcel Dekker.
- [11] J. FAN, C. ZHANG, Y J. ZHANG, Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** (2001) No. 1, 153–193.
- [12] P. HALL Y J.S. MARRON, On variance estimation in nonparametric regression. *Biometrika* **77** (1990) 2, 415–419.

- [13] N. JOHNSON, S. KOTZ Y A.W. KEMP, *Univariate discrete distributions*. New York, 1992, John Wiley & Sons.
- [14] Z. LIU, T. STENGOS Y Q. LI, Nonparametric model check based on local polynomial fitting. *Statist. Probab. Lett.* **48** (2000) 327–334.
- [15] H. MOTULSKY Y A. CRISTOPOULOS, *Comparing models and curves (Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting)*. GraphPad Software, 2002.
- [16] L.A. RAMIL NOVO Y W. GONZÁLEZ MANTEIGA, F tests and regression analysis of variance based on smoothing spline estimators. *Statistica Sinica* **10** (2000) 3, 819–837.
- [17] D. RUPPERT, M.P. WAND, U. HOLST Y O. HÖSSJER, Local polynomial variance-function estimation. *Technometrics* **39** (1997) 3, 262–273.
- [18] L.L. SCHUMAKER, *Spline Functions: Basic Theory*. Malabar, Florida, 1993, Krieger Publishing Company.
- [19] G.A.F. SEBER, *Linear Regression Analysis*. New York, 1977, John Wiley & Sons.
- [20] J. SHAO, *Mathematical Statistics*, 2ª edición, New York, 2003, Springer.
- [21] B.W. SILVERMAN, Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12** (1984) 898–916.
- [22] J.X. ZHENG, A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** (1996) 263–289.

Luis Alberto Ramil Novo
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias
27002 Lugo
Correo electrónico: lramil@lugo.usc.es

Wenceslao González Manteiga
Departamento de Estadística e Investigación Operativa
Facultad de Matemáticas
Campus Sur
15706 Santiago de Compostela (A Coruña)
Correo electrónico: wenceslao@usc.es