UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

REGRESION NO-PARAMETRICA

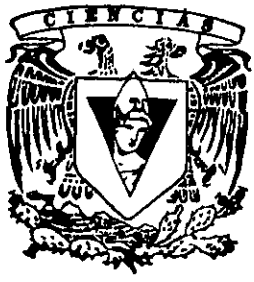
T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A

ENRIQUE BASURTO PEREZ



DIRECTOR DE TESIS:
MAT. MARGARITA E. CHAVEZ CANO

MEXICO, D.F., NOVIEMBRE 2000

VNIVERSIDAD NACIONAL
AVENMA DE
MEXICO

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis: " Regresión No-Paramétrica "

realizado por Enrique Basurto Pérez

con número de cuenta 9650356-4 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Mat. Margarita Elvira Chávez Cano	
Propietario	M. en C. Inocencio R. Madrid Ríos	
Propietario	M. en C. Jorge Iván Castro Rivadeneyra	
Suplente	M. en C. Beatriz E. Rodríguez Fernández	
Suplente	M. en C. José Antonio Flores	

Consejo Departamental de Matemáticas
M. en C. José Antonio Flores Díaz

...TEMÁTICAS

AGRADECIMIENTOS

Para el ángel delicado y fuerte que con tanto amor me procura,
a quien mi vida entera debo,
a quien mi boca se llena al decirle: "Mamá"

A mi Madre.

A quien ha llenado de luz mi camino,
a quien cargando lastres ha repartido flores,
al amigo que me dió la vida: gracias... siempre.

A mi Padre.

A quien en silencio me ha dado tanto, y sin saberlo admiro y respeto.
De quien comparto su escencia y a quien tan afín me encuentro,
soy el resultado de un proceso del cual,
me siento honrado que hayas formado parte

Arturo.

A tí que has sido más que mi hermano... mi razón de ser,
a quien Dios envió a mi mundo para ver el amor tan grande que me tiene,
a quien los momentos más bellos de mi vida debo.

Animal.

Para el angelito que con tanto cariño me cuida.
A quien con mirar su sonrisa
¡hace que valga la pena haber nacido!
Atte.: Kkn'z.

GRD.

A quienes nos han alentantado siempre a salir adelante,
a aquellos que en la distancia tengo siempre presentes.
Les ofrezco este trabajo que no valdría tanto
sin ustedes que conmigo lo valoran.

Carmen, Irma, Adelina, Claudia,
abuelito Julián, Juventino, Padrino,
Osvaldo, Arturo J., tío Samuel, Arturo S.,
abuelita Estela, Angel, Mina, Lupe, Cristy
tía Cuca, Esther, Ma. Elena, tía Chole
tío Joaquín, Josefo, Jaime, Carlos, Celes.

A la memoría de mi abuelo.

Luis Enrique Basurto Ruvalcaba.

Especialmente agradezco a aquellos que con sus aportaciones
han hecho de este, un compendio actualizado y confiable
para lectores de diversas disciplinas académicas.
Agradezco en todo lo que vale su tiempo y apoyo
para la realización de este esfuerzo.
Con toda mi admiración y respeto
este trabajo es para ustedes... ¡muchas gracias!.

Margarita Chávez, Bety Rodríguez,
Rafael Madrid, Iván Castro, José A. Flores.

Dedico también esta tesis al mejor equipo.
A aquellos que me dieron la mano y me abrieron las puertas.
A aquellos que con su ejemplo dieron dirección a mi carrera
y me sostienen su amistad aún en la distancia.
A todos ustedes dedico este esfuerzo... a GNP.

Sandra, Alma, Diana, Monsergat,
Román, Mario, Beto, Chenchan, Josué,
Oscar, Tona, Adrián, Coche Luis.

De manera especial quiero expresar mi más profundo agradecimiento,
a ustedes que han sido una gran razón para esforzarme día con día,
a ustedes que me han dado la certeza de mi inclinación profesional,
de manera muy especial... a BIMSA.

Alice, Yei,
Ricardo S., César O. de la R.

Para cada uno de los que han compartido
conmigo una parte de mi vida
cada uno de mis amigos sabe cuanto lo quiero
pues sin palabras se los he dicho tantas veces.
Ya saben... ¡para ustedes!.

Ita, Abi, Vane, Jazmín, Biela, Tatoo,
Mireya, Mónica R., Mónica L.,
Aldo, Gnomo, Jiro, Morro, Seco, Nato, Gerónimo,
Pepito, Alex, Luis, Come, Mario, Charly, Ray, Germán,
Rocío, Chucho, Mara Matilde, E. Vanessa E.,
Migue, Joaquín, Karmensa, Silvia.

ÍNDICE

PREFACIO

La teoría y los métodos de suavizamiento han tenido un intensivo e importante desarrollo a lo largo de los últimos veinte años. De manera general, podemos afirmar que dicho desarrollo se debe a dos razones: por un lado, la necesidad de contar con una herramienta de análisis de información con la característica de *flexibilidad* (entendida como: la posibilidad de construir estimadores de la variable respuesta a partir de información que influye en cada evento y la facilidad con la que es posible describir nuevas observaciones -características que en general no presentan los modelos paramétricos-); y por otro lado, el desarrollo de tecnología que ha permitido la creación de herramietas de cómputo que facilitan el cálculo de los estimadores no paramétricos (difíciles de obtener sin dichas herramientas).

Así pues, durante la década de los 80's S.XX, el avance de la tecnología dio lugar al desarrollo de técnicas matemáticas mas sofisticadas las cuales han permitido, en el caso de los modelos de regresión, explicar comportamientos mas complejos de lo que era posible utilizando los métodos tradicionales de interpolación y ajuste. Como en todos los modelos de regresión, la información recolectada será la materia prima para la obtención de los parámetros, por lo que dichas observaciones serán las que determinen: la forma que adopta el modelo y la eventual violación a los supuestos (propios de cada metodología).

Sin embargo, a pesar de su reciente desarrollo, las técnicas de suavizamiento cuentan ya con una larga tradición. En Inglaterra durante el S.XIX, el enfoque no paramétrico fue utilizado como una importante herramienta de análisis para explicar algunos fenómenos empíricos de tipo económico. Como se puede inferir, esta teoría permaneció prácticamente intacta durante poco mas de un siglo, debido a su complejidad (en el cálculo de los estimadores), respecto a aquellos modelos de tipo paramétrico que ofrecían importantes ventajas como lo son: por un lado, la simplicidad en la evaluación del modelo (i.e. la facilidad de evaluar $f(\overline{x})$ dado $\overline{x} \in R^n$ vector), y por otro, la compatibilidad con los supuestos sobre los cuales fueron construidos (normalidad, varianza constante, independiencia y homogeneidad en la distribución de las observaciones).

Este trabajo de tesis está enfocado a desarrollar aspectos estadísticos de suavizamiento kernel desde un punto de vista aplicado. Es recomendable que el lector posea conocimientos básicos en: métodos de interpolación y regresión. Sin embargo, en muchas de las ciencias y ramas de aplicación de esta teoría, como lo son: Medicina, Psicología, Matemáticas, Ingeniería, Economía, Finanzas, Administración, etc.; es muy probable que el interesado no posea los conceptos básicos mencionados. Por esto, a fin de comprender las ventajas y desventajas que presentan los métodos $No-Paramétricos$ (NP) respecto a otras técnicas de estimación y ajuste, se incluyen algunas de ellas en los capítulos II y III de este trabajo.

Es muy importante señalar además, los *alcances* de este trabajo y mencionar sus eventuales *limitaciones*. Referente a los *alcances* y como se mencionará en los objetivos, el

presente es un trabajo *descriptivo* el cual detalla las herramientas para la obtención de un modelo matemático alternativo (con base en teoría NP), esto significa que los modelos matemáticos obtenidos mediante el enfoque paramétrico no son en general, mejores ni peores que aquellos obtenidos con un enfoque no paramétrico. De hecho, es recomendable utilizar las metodologías adecuadas en función del cumplimiento de los supuestos, esto significa que si deseamos ajustar una curva de regresión y en nuestras observaciones se cumplen los supuestos de linealidad, varianza constante y normalidad en los errores; sería inútil tratar de ajustar un modelo NP, dado que la regresión lineal está especialmente construida para este tipo supuestos. Incluso en la práctica, ambos enfoques son complementarios (el paramétrico y el NP), de tal forma que los resultados de una teoría enriquezcan a los de la otra. Un ejemplo de esto puede observarse en la sección VI.2.3 y VI.5 en donde se utiliza un método backward de regresión lineal (paramétrico), para seleccionar las variables que describen la mayor variabilidad respecto del parámetro respuesta. Posteriormente se utiliza el suavizamiento kernel (no paramétrico), para obtener los estimadores que describen a cada intervalo la tendencia de las observaciones en estudio.

En cuanto a las *limitaciones* de esta teoría, es necesario mencionar que éstas pueden ser de dos tipos: las relacionadas con la teoría NP en sí y las que rebasan el alcance de este trabajo. En cuanto a las primeras y como se recordará mas adelante, en la matemática contemporánea existen una infinidad de métodos que permiten la estimación de las observaciones fuera del intervalo en donde éstas se concentran, a esto se le conoce como: *extrapolación* (o bien, *predicción*). Sin embargo, esta no es una característica de los modelos no paramétricos, la estimación de los suavizadores será posible únicamente dentro del intervalo del cual se tenga información. Otra limitación importante de esta teoría es el tipo de la variable respuesta que estima (la cual forzosamente debe ser real), y el tipo de variables independientes que requiere (donde por lo menos una deberá ser real). *Qué pasa cuando las observaciones están dadas en escala nominal u ordinal.* En general la variable dependiente y al menos una de las independientes deberán ser forzosamente escalares (reales), en caso de que algunos de los parámetros independientes esté dado en escala ordinal o nominal se recomienda segmentar las observaciones y obtener un modelo NP para cada segmentación. Esto es visto como una clara limitante de la teoría NP respecto a otras técnicas multivariadas, pero dicha deficiencia puede ser resuelta mediante el siguiente análisis. Por ejemplo, si deseamos construir un modelo NP que tenga como variable respuesta el gasto mensual en transporte y contamos con los parámetros independientes: ingreso económico y sexo (claramente ésta última dada en escala nominal), se recomienda obtener un modelo NP para hombres y otro para mujeres. Y si adicionalmente tuviéramos otra variable dependiente ordinal como por ejemplo: nivel socioeconómico dividido en: A, B y C; entonces se deben construir 6 modelos NP esto es: uno para hombres de nivel A, otro para los de nivel B y C (y los análogos, para mujeres del mismo nivel). Claramente, se deberá procurar que para cada segmentación se tenga un número de observaciones suficientemente grande para

que el análisis ha realizar tenga sentido.

Hablando de las limitaciones de este trabajo debe señalarse lo siguiente, en general, puede analizarse la validez de un modelo no paramétrico al igual que se hace con los modelos lineales a través de: la tabla de Análisis de Varianza (ANDEVA ó ANOVA), la prueba de hipótesis F (para el parámetro β_1 -de la pendiente-), el coeficiente de correlación, etc. Así mismo, pueden encontrarse herramientas análogas para analizar modelos no paramétricos, solo que dichas herramientas están fuera del alcance del presente trabajo de tesis y por lo tanto no son descritas en el mismo, si el lector está interesado en conocer a profundidad dichas herramientas de validación, se recomienda revisar la bibliografía detallada.

Quisiera expresar mi más profundo agradecimiento a quienes con sus valiosas aportaciones dieron dirección a lo expuesto en este trabajo de tesis: **Alicia de la Macorra, Rafael Madrid e Iván Castro**; y de manera muy especial a **Margarita E. Chávez Cano** quien no solamenteme me introdujo al tema y dio dirección al mismo, sino que además hiciera posible mi inclinación profesional al ámbito estadístico. También, quiero agradecer a **Lety Pang Molina** por el invaluable apoyo que siempre me ha brindado y porque a través de la información proporcionada me permitió estudiar y presentar en este trabajo de tesis el rezago social más importante de este país.

Por último, mi gratitud a la **Universidad Nacional Autónoma de México** por ser la máxima institución mexicana en ofrecer formación académica del más alto rango a nivel internacional y por promover la libertad de pensamiento y la apertura a la expresión de nuevas ideas. En particular, mi más profundo agradecimiento al profesorado de la **Facultad de Ciencias**, por su auténtica vocación académica. Agradezco en especial a: **Pedro Miramontes Vidal, Fernando Alonso Pérez Tejeda, Emilio Lluis Puebla, Rafael Campos Tenorio, Beatriz E. Rodríguez, Begoña Fernández, Ma. del Carmen Henández Ayuso y claro, Margarita E. Chávez Cano.**

A todos ellos... ¡mil gracias!.

Enrique Basurto Pérez.
México, D.F.
Octubre, 2000.

CAPÍTULO I: INTRODUCCIÓN

I.1. NATURALEZA DE LOS MODELOS MATEMÁTICOS.

A lo largo de la historia el hombre se ha preocupado por generar patrones de respuesta que den solución a sus problemas, inclusive podemos asegurar que de manera instintiva buscamos soluciones óptimas desde el momento mismo en que la problemática aparece. La optimalidad de dicha solución no depende únicamente de que el problema quede resuelto, sino de que los medios empleados vayan en relación directa a la magnitud del mismo, esto significa que no debemos utilizar demasiados recursos para la solución si es que el problema en sí no lo amerita.

La solución de problemas no es un tema exclusivo de la especie humana, los animales de todas las especies practican particulares métodos de solución e incluso al igual que los hombres, poseen la capacidad de generar soluciones cada vez mas sofisticadas ante una misma eventualidad. Sin embargo, existen dos características que en materia de conducta, distinguen al género humano de las demás especies animales: *la inteligencia y la voluntad*; todos los seres humanos nacemos potencialmente inteligentes y con capacidad de decisión. Ambas características nos permitirían generar ante un mismo problema una mejor solución que la aplicada de manera instintiva por cualquier otro animal de una especie diferente. De esta manera, poco a poco han surgido diversas aplicaciones de la voluntad e inteligencia humana a través de lo que hoy llamamos: *ciencia*. En particular, la ciencia matemática entre muchas otras cosas, busca generar patrones de respuesta óptimos a problemas generales y/o específicos. Incluso de manera natural, se elaboran: teorías, técnicas, herramientas, etc.; que tienden a mejorar los patrones de respuesta ya establecidos, por medio de metodologías más simples o más complejas que las originales. Aún mas en particular, existen ramas de la matemática que buscan:

a) Encontrar patrones que describan fenómenos que ocurren en la naturaleza.

b) Encontrar la relación entre dos o mas fenómenos entre sí.

c) Si conozco el comportamiento de un determinado fenómeno hasta este momento ¿es posible que de alguna manera logre predecir su comportamiento futuro? ¿o el anterior a mis estudios? ¿con qué confianza?.

d) Precisar en qué medida se alterará un determinado factor debido a la ocurrencia de otro relacionado ¿cómo saber si en realidad están relacionados?.

Estas y otras interrogantes se plantean día a día en forma natural por el ser humano. La ciencia matemática, ofrece efectivas respuestas a estas interrogantes mediante el estudio de: ecuaciones diferenciales, métodos de interpolación, técnicas de ajuste, modelos de predicción, técnicas de regresión, etc.

Un punto crucial en la solución matemática óptima de un problema es discernir qué método es el que debe aplicarse, para ello es imprescindible conocer los métodos disponibles, sólo de esta forma será posible ofrecer la respuesta óptima al problema en cuestión.

I.2. OBJETIVOS.

a) Proporcionar al lector técnicas de suavizamiento aplicables en diversas áreas de la ciencia como lo son: Medicina, Psicología, Matemáticas, etc.; y en diversas ramas aplicadas como: Ingeniería, Economía, Finanzas, Administración, etc.

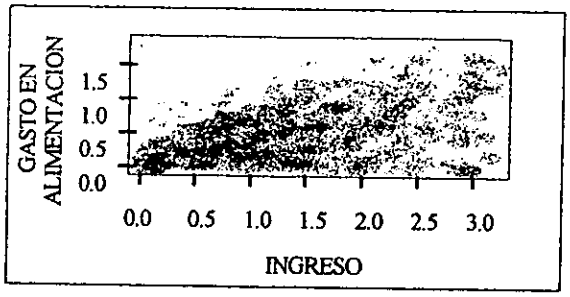
b) Que el lector distinga las ventajas que presenta la teoría no paramétrica en relación de otras técnicas de ajuste y regresión.

c) Que el lector construya modelos no paramétricos a través de suavizadores kernel y sepa interpretarlos.

El avance de la teoría No Paramétrica (NP) radica parcialmente en la *flexibilidad* del modelo construido a partir de ella, i.e. la facilidad con la que dicho modelo describe nuevas observaciones (en el capítulo IV se detallan algunas de las técnicas NP que dieron lugar a la teoría central que estudiaremos: *el suavizamiento kernel*). Por ejemplo: un modelo de Regresión Lineal es poco flexible, ya que si deseamos añadir una nueva observación, tendríamos que volver a estimar sus parámetros: ordenada al origen y pendiente (esto es, construirlo nuevamente); la situación es análoga si lo que deseamos es remover o reemplazar observaciones. Intuitivamente, en la teoría NP para: añadir, remover y/o reemplazar un nuevo dato; se verifica cuáles son los puntos adyacentes a este (vecinos), y con esta información se construye un nuevo estimador únicamente para aquellos que hayan sido afectados por la observación reemplazada. En el siguiente ejemplo se observa el efecto de construir estimadores NP por intervalos.

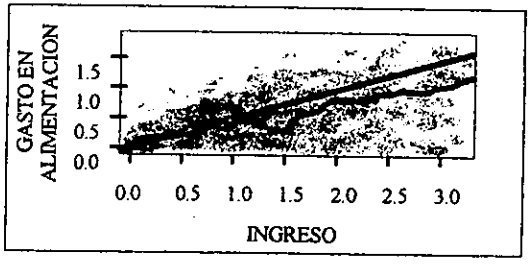
Ejemplo. Supongamos un conjunto de observaciones $\{(x_i, y_i)\}_{i=1}^{n}$ (gráfica I.2.1.), donde la pareja (x_i, y_i) representa el ingreso económico y el gasto en alimentación del $i-ésimo$

elemento de una población de interés.



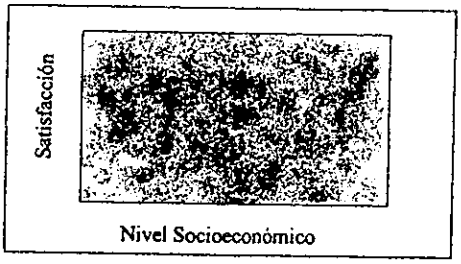
Gráfica I.2.1.

A continuación se presentan dos modelos sobre la gráfica de dispersión anterior, uno construido con base en teoría de Regresión Lineal y el otro con teoría NP; como se verá mas adelante en el capítulo V, a diferencia del modelo lineal (cuyos parámetros se obtienen tomando en cuenta todas las observaciones), los estimadores NP se obtienen tomando en cuenta únicamente los puntos cercanos que se encuentran alrededor de algún x_0. Esto significa que a cada intervalo un conjunto de puntos dentro de un intervalo (a_1,b_1) será considerado para la estimación del suavizador μ_1 correspondiente al punto $x_1 \in (a_1,b_1)$, mientras que para algún punto $x_2 \neq x_1$ el intervalo (a_2,b_2) será el que sea considerado para la estimación del suavizador correspondiente μ_2 y así sucesivamente $\forall\,(a_i,b_i) \subseteq (a,b)$, por ende un modelo NP es un conjunto de estimadores (suavizadores): $S = \{\mu(x) \mid x \in (a,b)\}$.



Gráfica I.2.2.

Ejemplo. Personas de diversos niveles socioeconómicos participan en una encuesta. Se desea observar si la satisfacción que experimenta la población hacia un cierto programa de televisión está directamentamente relacionada con el nivel socioeconómico de la población que usualmente ve dicho programa, todo esto con el objetivo de informar a sus anunciantes cuál es el perfil de ingresos de la gente que gusta de dicho programa. La información obtenida se presenta en la gráfica I.2.3.



Gráfica I.2.3.

Como se puede observar a simple vista, no es posible determinar un nivel socioeconómico específico que experimente un grado de satisfacción especial respecto al programa. Con el objetivo, de proponer una tendencia que describa a la variable respuesta: *satisfacción*, se construyen dos modelos: lineal y NP.

Gráfica I.2.4.

El modelo NP asocia un peso a cada observación y a diferencia del lineal, describe las observaciones a cada intervalo del parámetro: *nivel socioeconómico*. De acuerdo a los resultados obtenidos, se tiene evidencia para inferir que los niveles medio y medio alto son los que menos gustan del programa, en contraposición con los niveles socioeconómicos bajos y altos.

Observe que sin el modelo NP hubiera sido más complicado tener la evidencia deseada.

La flexibilidad o *versatilidad*, es sólo una de las ventajas que presentan los modelos de tipo NP; la *forma* que adoptan las curvas es otra de ellas, i.e. debido a su construcción, la morfología del modelo NP ofrecerá información adicional acerca de la concentración de las observaciones en la gráfica de dispersión además de que permite elaborar curvas de alta

complejidad a comparación de los modelos paramétricos que en muchos casos se encuentran dimensionalmente restringidos.

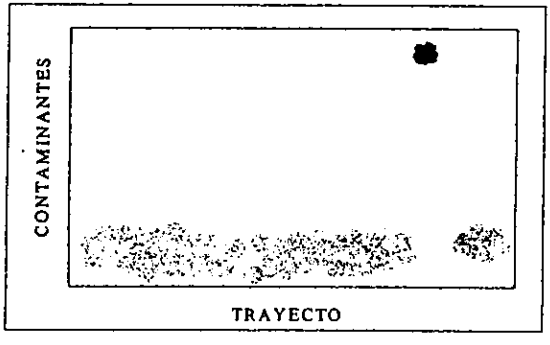
Así, las curvas NP nos permitirán inferir acerca del: *comportamiento*, *concentración*, *dispersión* y *distribución* de las observaciones en estudio (entre otras aplicaciones), con simplemente echar un vistazo a la gráfica del modelo. Además de cumplir su función de: *estimación*.

I.3. PUNTOS AISLADOS O ABERRANTES (OUTLIERS).

La curva NP modificará su estructura únicamente en las regiones en que esto sea necesario. Recordemos que bajo la presencia de outliers los modelos lineales pueden llegar a modificar su estructura, de tal forma, que se pierda por completo la explicación en sitios donde se concentra la mayor parte de la información.

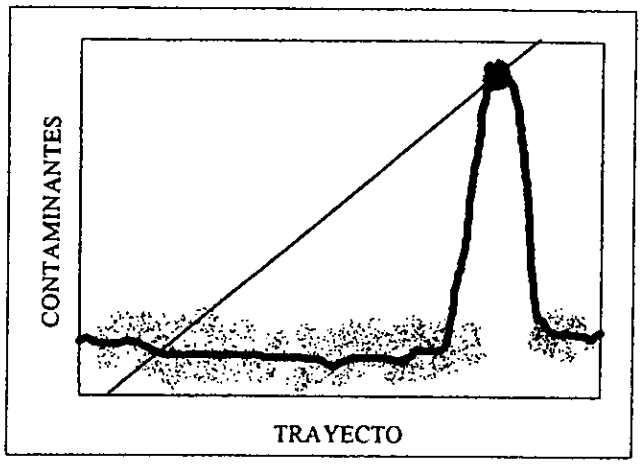
Ejemplo. El gobierno del estado de Michoacán está interesado en conocer en forma detallada el nivel de concentración de algunos residuos tóxicos en los ríos mas importantes del estado, razón por la cual ha decidido contratar a un laboratorio especializado en estudios hidrológicos. De acuerdo a la logística del estudio, los consultores contratados deciden recolectar muestras en los ríos de interés de la siguiente manera: visitarán los ríos mas representativos comenzando por el Oeste del estado y los recorrerá de tal manera que por cada río se tomen muestras a distintos niveles y alturas, de esta forma, caminado en dirección Este se obtendrán las muestras necesarias para el estudio.

El estudio comienza y se recolectan las muestras requeridas conforme a lo programado. En esos días y sin aviso a los recolectores, un barco carguero derrama cantidades industriales de combustible a la ribera de los ríos cercanos a la frontera con Oaxaca. Las muestras se toman normalmente y al revisar los resultados se obtuvo la siguiente gráfica:



Gráfica I.3.1.

Claramente, el estudio entero se vió alterado por un factor imprevisto y es necesario establecer las conclusiones del proyecto lo antes posible. Se realizan dos estudios de tendencia, uno basado en teoría lineal y otro en teoría NP:



Gráfica I.3.2.

Observe la enorme diferencia entre ambos. El modelo lineal perdió por completo su estructura y la descripción de las demás observaciones, centrándose de forma clara en la región aislada de los contaminantes. Sin embargo, el modelo NP describe perfectamente cada uno de los intervalos de interés y no solamente eso, debido a que en los demás intervalos el comportamiento de los contaminantes es estable, es posible cuantificar el impacto ecológico a consecuencia del derrame si suponemos que antes del mismo, el nivel de contaminantes de los ríos alterados era similar a los demás. Por lo tanto, se obtuvo información adicional importante del modelo NP y que de haber sido estudiado por medio de un método lineal los resultados hubiesen sido seguramente muy distintos.

Observación. Es muy importante evitar la mayor cantidad de fuentes de error que puedan eventualmente, modificar la muestra y por lo tanto el modelo de tendencia. Una ventaja importante del modelo NP es que se ajusta de manera natural a cada intervalo, tomando en cuenta únicamente la información que afecte a la estimación del suavizador en cada punto (capítulo V). Por lo tanto, en caso de no ser detectada, la información aberrante únicamente afectará dentro del intrevalo en donde ésta se encuentre y no a la muestra en su totalidad.

CAPÍTULO II: INTERPOLACIÓN

II.1. INTRODUCCIÓN.

Supongamos $f_i = f(x_i)$ valores conocidos para $(n+1)$ puntos $x_1, x_2, ..., x_{n+1}$. Queremos determinar una función P tal que:

$$P(x_i) = f_i, \text{con} : i = 1, 2, ..., n+1.$$

En caso de que tal función exista, se dice que P *interpola a* f en $x_1, x_2, ..., x_{n+1}$. Una función inter polante puede ser utilizada para estimar el valor de la función f en un cierto punto $x \in (x_1, x_2, ..., x_{n+1})$, entonces hablaremos de *interpolación*, de lo contrario de: *extrapolación*. Así mismo, podremos utilizar a P para aproximar la derivada de f en un algún punto $x \in (x_1, x_2, ..., x_{n+1})$, o bien, su integral sobre un cierto subintervalo $(x, y) \in (x_1, x_2, ..., x_{n+1})$. En estos casos es mas útil aproximar f por medio de funciones *splines* (un polinomio interpolante entre cada par de observaciones), que se construyen con la finalidad de ser fáciles de derivar y/o de integrar (estas son estudiadas en el capítulo III).

Observación. No siempre es útil aplicar métodos de interpolación si lo que conocemos son datos aproximados de la función f.

II.2. INTERPOLACIÓN POR POLINOMIOS.

¿Cuál es el grado de un polinomio interpolante?. Supongamos x_1, x_2, x_3 tres puntos cuyos correspondientes valores: f_1 , f_2 , f_3 son conocidos. En general, una línea recta no puede interpolar a cualesquiera 3 puntos, sin embargo es posible establecer un polinomio de segundo grado que interpole a los puntos $(x_1, f_1), (x_2, f_2), (x_3, f_3) \in R^2$.

De aquí en adelante, serán enunciados algunos teoremas que tienen por finalidad fortalecer conceptos fundamentales de la teoría de interpolación, demostrando solamente aquellos en los que su desarrollo origine nuevas reflexiones.

Teorema (*II*.2.1). Sean $x_1, x_2, ..., x_{n+1}$ puntos arbitrarios y distintos. Para los valores $f_1, f_2, ..., f_{n+1}$ existe un **único** polinomio P de grado $\leq n$ tal que:

$$P(x_i) = f_i \text{ , con: } i = 1, 2, ..., n+1.$$

¿Existe alguna herramienta que determine el error entre la función f y su polinomio interpolante descrito en el teorema II.2.1?. En general no es posible determinar tal diferencia, por ejemplo: si f es discontinua, el error por interpolación puede ser arbitrariamente grande. Sin embargo, es posible establecerla bajo ciertas condiciones como se describe a continuación:

Teorema (*II*.2.2). Sea f una función con $n+1$ derivadas contínuas en el intervalo $x \in (x_1, x_2, ..., x_{n+1})$. Sea P un polinomio de grado $\leq n$ que satisface:

$$P(x_i) = f(x_i) \text{ , con: } i = 1, 2, ..., n+1$$

entonces

$$f(x) - P(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}(x - x_1)(x - x_2)...(x - x_{n+1}),$$

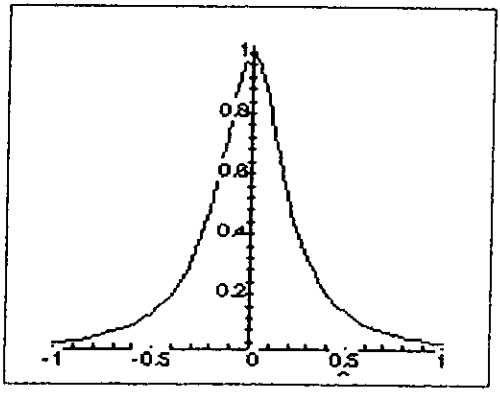
para alguna $\xi(x)$ en el intervalo $(x_1, x_2, ..., x_{n+1})$.

Observe que el error: $f(x_i) - P(x_i) = 0$, para: $i = 1, 2, ..., n+1$. En ocasiones la confiabilidad de un modelo depende del número de observaciones empleadas para el estudio en cuestión, por ello se podría pensar que el error por interpolación decrece conforme el número de puntos x_i aumenta. Sin embargo, dicha afirmación es inconsistente para el caso de los polinomios interpolantes, un contraejemplo fue el construido por Runge tomando una sucesión:

$$x_i = -1 + (i-1)\left(\frac{2}{n}\right), \text{con} : i = 1, 2, ..., n+1;$$

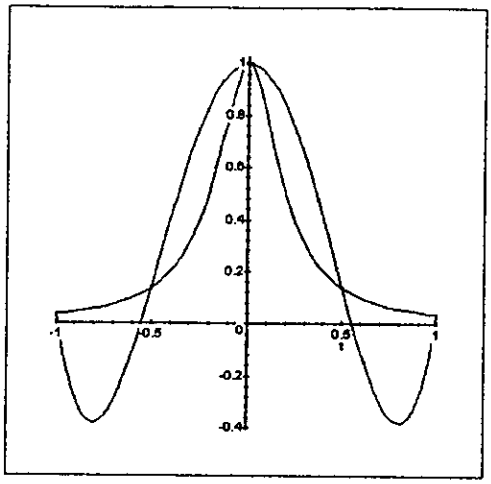
de puntos equidistantes en el intervalo $[-1, 1]$, los cuales fueron evaluados en: $f(x) = \frac{1}{1+25x^2}$.

(ver $f(x)$ de Runge en la gráfica II.2.3). A continuación, observamos los polinomios que interpolan a f en 5, 9, 13 y 21 puntos equidistantes x_i en el intervalo $[-1, 1]$. Observe que en cada caso los $(n+1)$ puntos x_i originan polinomios de grado $\leq n$.



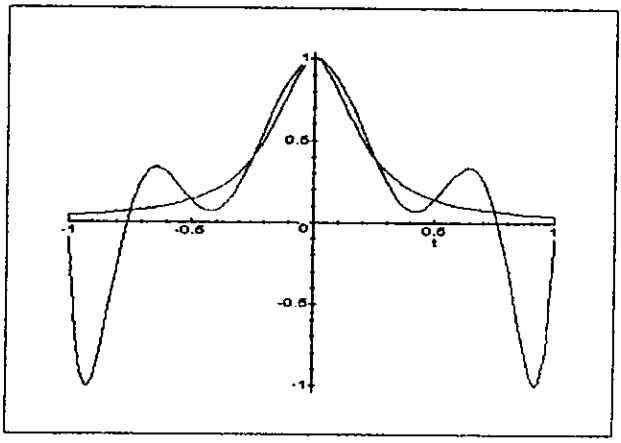
Gráfica II.2.3.

Gráfica comparativa: $f(x)$ vs $P_4(x)$, el cual interpola a $f(x)$ en 5 puntos equdistantes: -1,-0.5, 0, 0.5 y 1.



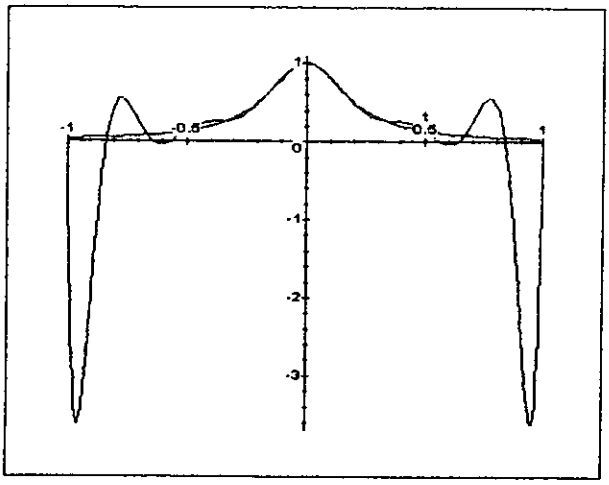
Gráfica II.2.4.

Gráfica comparativa: $f(x)$ vs $P_8(x)$:



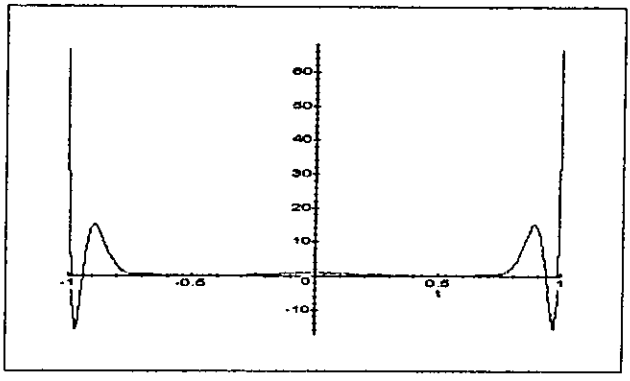
Gráfica II.2.5.

Observación. Verifique en la II.2.5 la diferencia entre f y $P_8(x)$ en comparación de la gráfica II.2.4, observe que el error aumentará a medida que se incremente el número de observaciones. Veamos con: $n+1=13$, en II.2.6: f vs $P_{12}(x)$



Gráfica II.2.6.

Por último: f vs $P_{20}(x)$,

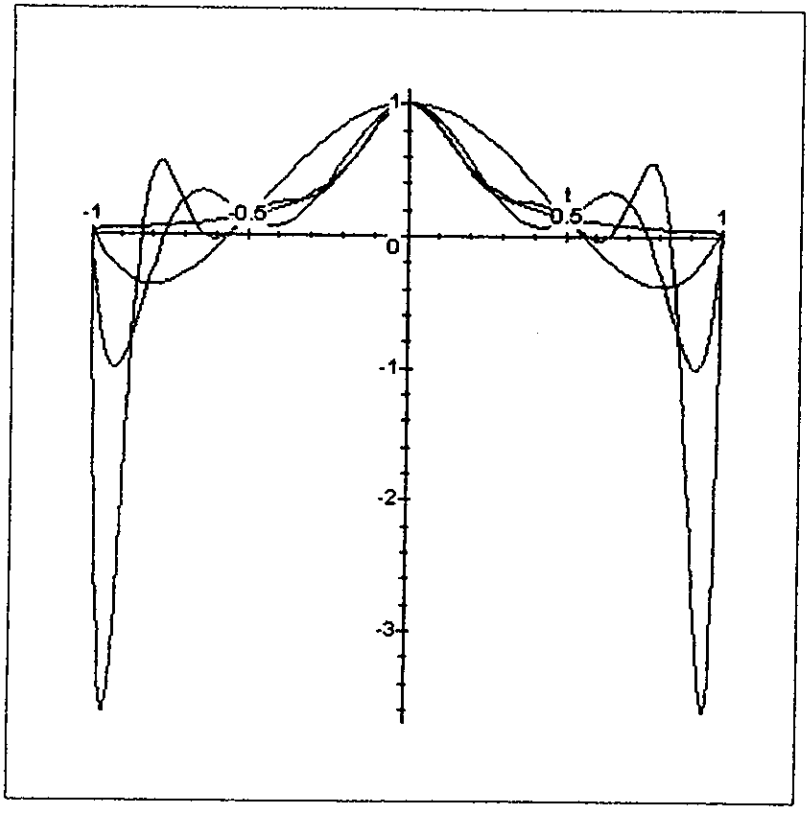


Gráfica II.2.7.

Se observa que el error es relativamente pequeño en los puntos cercanos al origen, se puede demostrar que para $f(x)$ de Runge sucede que cerca de los extremos del intervalo $[-1, 1]$:

$$lim_{n\to\infty}\,(max_{-1\leq x\leq 1}\,|f(x) - P_n(x)|) = \infty$$

La siguiente gráfica muestra en conjunto el efecto de añadir más y más datos al modelo interpolante con 5, 9 y 13 observaciones (con 21 datos se altera tanto la escala que las demás gráficas no se alcanzan a percibir).



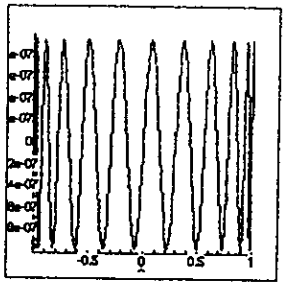
Gráfica II.2.8.

El error por interpolación puede disminuir mediante una correcta selección de observaciones. Por ejemplo la selección de Chebyshev:

$$x_k = cos\left(\frac{(2k-1)\,\pi}{2n}\right),\ \text{con } k = 1, 2, ..., n.$$

Se puede demostrar que ninguna otra selección de puntos genera un error absoluto máximo tan pequeño. La gráfica II.2.9 describe el error entre la función de Runge y un polinomio $P(x_k)$, con $k = 21$; que interpola a $f(x)$ de Runge con x_k de Chebyshev (observe

que el error máximo es del orden de: 10^{-7}).



Gráfica II.2.9.

Conclusión. Se puede demostrar en general que una selección arbitraria de puntos que interpolan a una función contínua produce un error máximo asociado que tiende a infinito. Por ello, en caso de tener un número considerablemente grande de observaciones, no es recomendable utilizar métodos de interpolación por intervalos (splines), para describir el comportamiento de una función $f(x)$ que se supone no conocida.

II.3. MÉTODOS DE INTERPOLACIÓN.

II.3.1. INTERPOLACIÓN LINEAL.

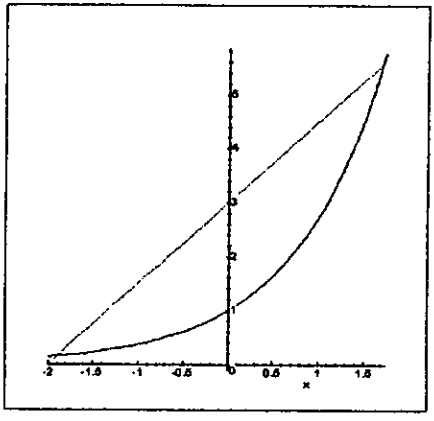
Supongamos dos puntos $(x_1, f_1), (x_2, f_2) \in R^2$ en general, es posible estimar el valor de cualquier $x \in (x_1, x_2)$ por medio de un polinomio P que interpole a la función f en $(x_1, f_1), (x_2, f_2)$. Por ejemplo, para cualquier valor de x pongamos x_0 y su imagen $f(x_0)$ será estimada por el valor $P_n(x_0) = \widehat{f(x_0)}$; sin embargo, algunas estimaciones serán mejores que otras en la medida que el polinomio se ajuste mas adecuadamente a la función f, dicho ajuste puede o no, deberse al grado n del $P_n(x)$. Cada estimación realizada va a generar un error que podremos acotar dependiendo de las características de la función, pues como ya vimos en el ejemplo de Runge, en ocasiones el error es extraordinariamente grande.

En particular para cualesquiera dos puntos: $(x_1, f_1), (x_2, f_2) \in R^2$, existe un polinomio lineal de la forma

$$P_1(x) = f_1 + \frac{x - x_1}{x_2 - x_1}(f_2 - f_1)$$

que interpola a f en f_1 y f_2. Suponga que se conocen dos puntos: $(x_1, f_1), (x_2, f_2)$; los cuales provienen de la función $f(x) = e^x$, la cual hipotéticamente desconocemos. Un

polinomio lineal que interpola a los puntos en cuestión podría ser por ejemplo: $P(x) = 1.5x + 3$. Para este ejemplo es posible que el modelo de interpolación hubiera sido mejor si tuvieramos algunas observaciones mas. Con solamente dos como es el caso, la recta interpolante no es una buena aproximación a la curva f .



Gráfica II.3.1

Breviario. Existen diversas fuentes de error que afectan el resultado de una interpolación lineal:

a) El que proviene del error por medición en el argumento.

b) El error en la evaluación del argumento en el modelo $f(x)$.

c) Error por truncamiento (aquel en el que se desprecia una parte de la expansión decimal e.g. supongamos un cierto número $x = 2.159873654201...$ pero para efectos de cálculo estaremos interesados únicamente e.g. en sus primeros tres decimales, esto es: $x' = 2.159$, por lo tanto la expansión decimal $\epsilon = 0.000873654201...$ quedará resumida en x' y el valor ϵ restante es lo que conocemos como error de truncamiento).

d) Error por redondeo, aquel en el que la información se concentra hasta un cierto decimal, si $x = 2.159873654201...$ se redondea x hasta el tercer decimal. Por lo tanto: $x' = 2.160$, donde $\eta = 0.000126345769... \simeq 0.0001$ este último valor es el que aproxima la expansión decimal η , de aquí que el valor $x = 2.159 + \eta \simeq 2.160 = x'$.

En general, el error por redondeo puede determinarse de la siguiente manera, suponga: $x = x_o.x_1x_2x_3x_4x_5x_6x_7...$ con x_o la parte entera de x. Si redondeamos a x hasta el $t-ésimo$ decimal $x' = x_o.x_1x_2x_3...x_t$, entonces:

$$x_t = \left\{ \begin{array}{c} x_t + 1 \text{ si } x_{t+1} \geq 5 \\ x_t \text{ si } x_{t+1} < 5 \end{array} \right\}$$

II.3.2. POLINOMIO INTERPOLANTE DE NEWTON

De acuerdo al Teorema II.2.1, el polinomio interpolante de una función es único para $x_1, x_2, ..., x_{n+1}$ conocidos, sin embargo existen distintas formas de obtener un mismo modelo de interpolación, esto implica que algunas metodologías son muy sencillas para la evaluación de puntos específicos del intervalo, y que existen otras que ofrecen el mismo modelo pero escrito de tal forma que su manejo y análisis es mas sencillo. El Polinomio Interpolante de Newton es una manera sencilla de encontrar una expresión explícita para el polinomio $P(x)$. Comencemos por escribir el polinomio en la forma:

$$P(x) = c_0 + c_1(x - x_1) + c_2(x - x_1)(x - x_2) + ... + c_n(x - x_1)(x - x_2)...(x - x_n)$$

donde los coeficientes $c_0, c_1, c_2, ..., c_n$ están determinados por:

$$P(x_i) = f_i, \text{con} : i = 1, 2, ..., n + 1$$

esto es, cuando:

$$\begin{aligned}
P_0(x_1) &= c_0 = f_1 \\
P_1(x_2) &= c_0 + c_1(x_2 - x_1) = f_2 \\
P_2(x_3) &= c_0 + c_1(x_3 - x_1) + c_2(x_3 - x_1)(x_3 - x_2) = f_3 \\
&\vdots \\
P_{n-1}(x_n) &= c_0 + c_1(x_n - x_1) + c_2(x_n - x_1)(x_n - x_2) + ...+ \\
& c_n(x_n - x_1)(x_n - x_2)...(x_n - x_{n-1}) = f_n
\end{aligned}$$

de aquí que:

$$\begin{aligned}
c_0 &= f_1 \\
c_1 &= \frac{f_2 - f_1}{x_2 - x_1} \\
c_2 &= \frac{f_3 - f_1 - \frac{x_3 - x_1}{x_2 - x_1}(f_2 - f_1)}{(x_3 - x_1)(x_3 - x_2)}
\end{aligned}$$

debido a que a partir de c_4 el cálculo de los coefientes se vuelve mas complicado, se suelen calcular los coeficientes c_k de forma recursiva:

$$\begin{aligned}
P_0(x) &= c_0 \\
P_k(x) &= P_{k-1}(x) + c_k(x - x_1)(x - x_2)...(x - x_k); \text{ con: } k = 1, 2, ..., n.
\end{aligned}$$

donde $P_k(x)$ es un polinomio de grado $\leq k$ que interpola a f en: $x_1, x_2, ..., x_{k+1}$. Otra manera de calcular los coeficientes recursivamente es la siguiente, tomamos dos polinomios de grado $k-1$ tales que:

$$P_{k-1}^{(1)}(x) \text{ que interpola a } f \text{ en } x_1, x_2, ..., x_k$$
$$P_{k-1}^{(2)}(x) \text{ que interpola a } f \text{ en } x_2, x_3, ..., x_{k+1}$$

así, podemos obtener un polinomio $P_k(x)$ de grado $\leq k$ de forma alterna:

$$P_k(x_i) = P_{k-1}^{(1)}(x_i) + \frac{x_i - x_1}{x_{k+1} - x_1}\left(P_{k-1}^{(2)}(x_i) - P_{k-1}^{(1)}(x_i)\right)$$

Reflexión. ¿Qué tan exacto es el polinomio interpolante de Newton con respecto a f teórica?, es posible aproximar el valor absoluto máximo del error para cuando el polinomio interpolante es estimado por el método de Newton. Por principio, observemos que cada uno los coeficientes c_k del polinomio $P(x)$ son calculados en función de los valores conocidos f_k , que a su vez están en función de los valores x_k de la siguiente manera: $f(x_k) = f_k$. De aquí que sea posible escribir a los coeficientes del polinomio de Newton:

$$c_k = f[x_1, x_2, ..., x_{k+1}]$$

Por lo tanto, podemos reescribir al polinomio:

$$P_k(x) = P_{k-1}(x) + c_k(x - x_1)(x - x_2)...(x - x_k)$$

en la forma:

$$P_k(x) = P_{k-1}(x) + f[x_1, x_2, ..., x_{k+1}](x - x_1)(x - x_2)...(x - x_k)$$

y ya que el polinomio está construído para interpolar a f en los puntos x_k con: $k = 1, 2, ..., n+1$ hagamos pues:

$$P_k(x_{k+1}) = f(x_{k+1})$$

de aquí que:

$$f(x_{k+1}) = P_{k-1}(x_{k+1}) + f[x_1, x_2, ..., x_{k+1}](x_{k+1} - x_1)(x_{k+1} - x_2)...(x_{k+1} - x_k)$$

$$\Leftrightarrow f(x_{k+1}) - P_{k-1}(x_{k+1}) = f[x_1, x_2, ..., x_{k+1}](x_{k+1} - x_1)(x_{k+1} - x_2)...(x_{k+1} - x_k)$$

Por el teorema II.2.2, si f es una función con k derivadas contínuas en el intervalo formado por los puntos $x, x_1, x_2, ..., x_k$ y P es un polinomio de grado $\leq k$ que satisface

$$P(x_i) = f(x_i) \text{ , con: } i = 1, 2, ..., k$$

entonces

$$f(x) - P(x) = \frac{f^{(k)}(\xi(x))}{k!}(x - x_1)(x - x_2)...(x - x_k)$$

De las expresiones anteriores se deriva el siguiente:

Teorema (*II*.3.2). Si f es una función con k derivadas contínuas en el intervalo formado por los puntos $x, x_1, x_2, ..., x_k$ entonces, existe ξ en este intervalo tal que:

$$f[x_1, x_2, ..., x_{k+1}] = \frac{f^{(k)}(\xi(x))}{k!}$$

Conclusión. Cuando una función es aproximada por un polinomio interpolante de Newton, es posible determinar el error entre ambas funciones como una cantidad proporcional al $k - ésimo$ coeficiente c_k correspondiente al término con exponente mas grande en el polinomio $P_k(x)$.

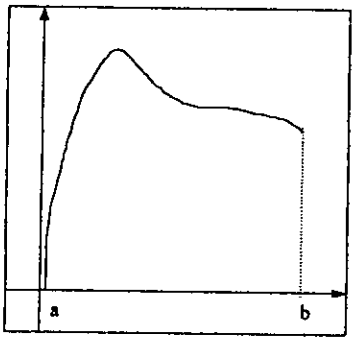
Observación. Existen otros métodos de interpolación en los que la evaluación $f(x)$ de los puntos $x \in (x_1, x_2, ..., x_{n+1})$, resulta computacionalmente mas sencilla sin embargo, el método de Newton proporciona una expresión clara del polinomio interpolante y del error que produce (es por esta razón que se expone este método y no otro).

Nota. Si el lector está interesado en profundizar sobre los conceptos paramétricos que se exponen en este documento, es conveniente que consulte la bibliografía detallada al final del texto.

CAPÍTULO III: SPLINES

III.1. INTRODUCCIÓN.

En el capítulo anterior observamos que no necesariamente podemos esperar que una función interpolante realice mejores aproximaciones conforme aumenta la cantidad de puntos conocidos. Por otro lado, podemos ver que en general un polinomio interpolante no es un buen modelo para aproximar funciones que cambian su comportamiento dependiendo de cada intervalo, tal es el caso de aquellas funciones que describen la forma de un cuerpo, un fenómeno físico, etc. Como ejemplo, consideremos la siguiente función:



Gráfica III.1.

Si trataramos de aproximar esta función por un polinomio, el ajuste sería malo para modelos de grado pequeño. Por el contrario, intentar el ajuste por medio de un polinomio de alto grado daría como resultado una aproximación oscilante. Si lo que deseamos es encontrar un modelo que se ajuste a esta función, podríamos encontrar uno construido con diferentes polinomios de grado pequeño en distintas partes del intervalo $[a, b]$.

Supongamos que el intervalo $[a, b]$ es subdividido de la siguiente manera: $a \leq x_1 < x_2 < \ldots < x_n \leq b$; la aproximación por subintervalos mas simple es aquella que únicamente utiliza líneas rectas para cada $[x_i, x_{i+1}]$. Por lo tanto, la aproximación estaría dada por una curva poligonal que por construcción, no es suave ya que su primera derivada es discontínua en los nodos $x_2, x_3, \ldots, x_{n-1}$. A menudo se utilizan polinomios de tercer grado en cada subintervalo los cuales se unen de tal forma que la función resultante sea contínua al igual que su primera y segunda derivada.

En general, la función resultante es conocida como *spline* y se denota como: $s(x)$, mas aún si esta función cumple:

$$s(x_i) = f(x_i) \text{ , para: } i = 1, 2, ..., n$$

entonces s es un *spline interpolante.*

Observación. En lo sucesivo únicamente trataremos con splines interpolantes, por lo que el adjetivo *interpolante* será omitido. Como se observa en el teorema III.3.2, de todas las funciones con segunda derivada contínua que intrepolan a una cierta función f el spline cúbico es la función que minimiza entre ambas funciones (evitando oscilaciones). A continuación se introducen los splines lineales con la finalidad de tener claros los conceptos que se exponen para la construcción de los splines cúbicos.

III.2. SPLINES LINEALES.

Sean $x_1 < x_2 < ... < x_n$ puntos conocidos a los cuales llamaremos *nodos* (*knots*). Un spline lineal es una función definida en un intervalo $[x_1, x_n]$ con las siguientes propiedades:

a) $s(x)$ es contínua en $[x_1, x_n]$.

b) $s(x)$ es una línea recta en cada subintervalo $[x_i, x_{i+1}]$, para: $i = 1, 2, ..., n-1$. Por lo tanto, la función $s(x)$ está determinada por $n-1$ líneas rectas:

$$s(x) = a_i(x - x_i) + b_i \text{ , } \forall x \in [x_i, x_{i+1}] .$$

Para encontrar $s(x)$ deberán determinarse $2(n-1)$ coeficientes, recordemos que un spline debe ser continuo en todo el intervalo $[x_1, x_n]$, para ello los polinomios lineales en los subintervalos $[x_i, x_{i+1}]$, para: $i = 1, 2, ..., n-1$; deben estar conectados entre sí i.e. el tramo $(i+1) - ésimo$ comienza donde termina el $i - ésimo$, de aquí que se generen $(n-2)$ condiciones:

$$s(x_i) = s(x_{i+1}), \text{ para : } i = 2, ..., n-2$$

Los n coeficientes restantes pueden ser e.g.

$$s(x_i) = f(x_i), \text{ para : } i = 1, 2, ..., n.$$

Lo anterior origina las siguientes condiciones:

$$a_i = \frac{f_{i+1} - f_i}{x_{i+1} - x_i} \text{ , } \quad b_i = f_i$$

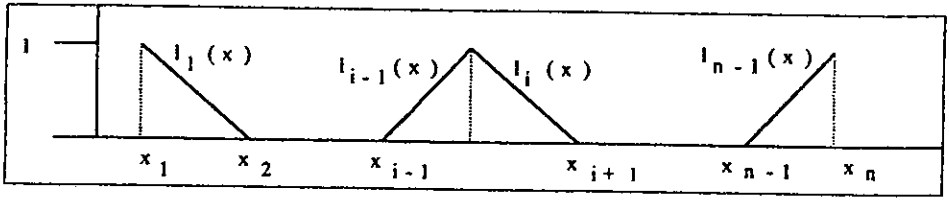
si tales condiciones se cumplen, la función $s(x)$ será contínua y compuesta por líneas rectas en cada subintervalo por lo que quedarán satisfechas las condiciones a) y b) enunciadas arriba.

Una manera alterna de representar a la función s es la siguiente:

$$s(x) = \sum_{i=1}^{n} f_i l_i(x)$$

donde $l_i(x)$ es un único spline lineal con la propiedad:

$$l_i(x_j) = \delta_{ij} = \left\{ \begin{array}{c} 0 \text{ , si } i \neq j \\ 1 \text{ , si } i = j \end{array} \right\}$$



Gráfica III.2.1.

La gráfica III.2.1. muestra las funciones base $l_i(x)$ del spline lineal $s(x)$, algebraicamente:

$$l_1(x) = \left\{ \begin{array}{c} 0 \text{ , si } x_2 \leq x \leq x_n \\ \frac{x_2 - x}{x_2 - x_1} \text{ , si } x_1 \leq x \leq x_2 \end{array} \right\}$$

$$l_i(x) = \left\{ \begin{array}{c} 0 \text{ , si } x_1 \leq x \leq x_{i-1} \\ \frac{x - x_{i-1}}{x_i - x_{i-1}} \text{ , si } x_{i-1} \leq x \leq x_i \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} \text{ , si } x_i \leq x \leq x_{i+1} \\ 0 \text{ , si } x_{i+1} \leq x \leq x_n \end{array} \right\}$$

$$l_n(x) = \left\{ \begin{array}{c} 0 \text{ , si } x_1 \leq x \leq x_{n-1} \\ \frac{x - x_{n-1}}{x_n - x_{n-1}} \text{ , si } x_{n-1} \leq x \leq x_n \end{array} \right\}$$

aquellas funciones $l_i(x)$ que forman una base para $s(x)$ son conocidas como: $B-splines\ lineales$ (para un conjunto de nodos conocido $x_1, x_2, ..., x_n$). Observe que dichas funciones son cero fuera del intervalo abierto (x_{i-1}, x_{i+1}) ver III.2.1.

Supongamos f es dos veces diferenciable, el error en la aproximación por splines lineales satisface:

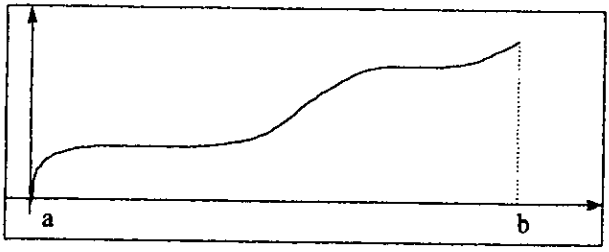
$$f(x) - s(x) = \frac{f''(\xi_i)}{2} (x - x_i)(x - x_{i+1}), \ \forall x \in [x_i, x_{i+1}].$$

Sea M una cota superior para la magnitud de $f''(x)$ en el intervalo $[x_1, x_n]$ y tomemos $[x_i, x_{i+1}] = h$ como la longitud del subintervalo mas grande, entonces:

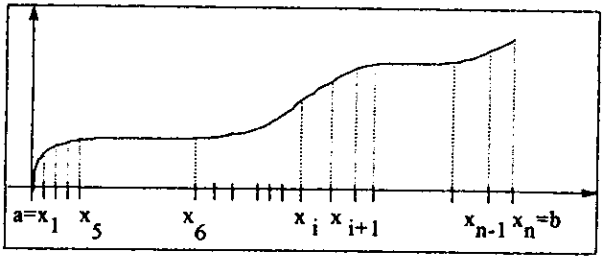
$$|f(x) - s(x)| = \frac{h^2}{8} * M.$$

Importante. La ubicación de los nodos $x_1, x_2, ..., x_n$ afecta directamente al error a través de la distancia mas larga h, de aquí que la aproximación del máximo error de $s(x)$ a f será

cada vez mejor conforme la distancia entre los nodos sea mas pequeña en los lugares en que la función f varíe mas rápidamente, este efecto se puede observar en la ubicación de los nodos en la gráfica III.2.3 a partir del comportamiento descrito en la III.2.2.



Gráfica III.2.2.



Gráfica III.2.3.

III.3. SPLINES CÚBICOS.

Sean $x_1 < x_2 < ... < x_n$ nodos conocidos, un *spline cúbico* es una función $s(x)$ definida en el intervalo $[x_1, x_n]$, con las siguientes propiedades:

a) $s(x)$, $s'(x)$ y $s''(x)$ son contínuas en el intervalo $[x_1, x_n]$.

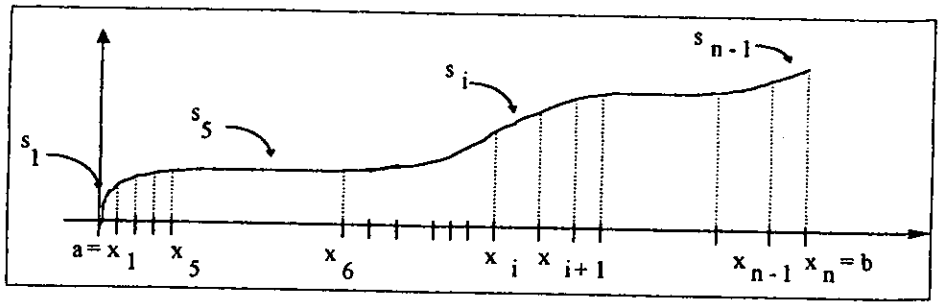
b) $s(x)$ es un polinomio cúbico en cada subintervalo $[x_i, x_{i+1}]$.

Por estas propiedades a un *spline cúbico interpolante* le pediremos lo siguiente:

$$s(x_i) = f_i, \;\; i = 1, 2, ..., n;$$

con f_i conocidos. Se definen funciones $s_i(x)$ definidas para cada subintervalo $[x_i, x_{i+1}]$:

$$s(x) = s_i(x), \;\; x_i \leq x \leq x_{i+1}.$$



Gráfica III.3.1.

La gráfica III.3.1 muestra un spline cúbico $s(x)$ construído a partir de $(n-1)$ polinomios cúbicos $s_i(x)$. De aquí que el algoritmo propuesto requiere determinar $4(n-1)$ coeficientes, 4 para cada polinomio cúbico de la forma:

$$s_i(x) = c_{0i} + c_{1i}x + c_{3i}x^2 + c_{4i}x^3, \ \forall i = 1, ..., n-1.$$

Cada polinomio debe interpolar a f en el punto inicial y final de su intervalo (primer requisito: $s(x)$ contínua), y es de aquí de donde se obtienen las $2(n-1)$ primeras condiciones:

$$s_{i-1}(x_i) = s_i(x_i) = f_i, \ \forall i = 2, ..., n.$$

Ademas se deben satisfacer los requisitos de continuidad para $s'(x)$ y $s''(x)$ en los $n-2$ nodos interiores, i.e.

$$s_{i-1}^{(k)}(x_i) = s_i^{(k)}(x_i) = f_i; \ \ k = 1, 2; \ i = 2, ..., n-1.$$

lo que da como resulatdo $2(n-2)$ condiciones. Tenemos hasta el momento: $2(n-1) + 2(n-2) = 4n-6$ de los $4n-4$ coeficientes a determinar. Los dos grados de libertad restantes se pueden obtener por medio de alguna de las siguientes opciones:

$$s''(x_1) = 0$$

$$s''(x_n) = 0$$

con estas condiciones $s(x)$ es conocido como: *spline natural*. Otra opción es agregar las condiciones:

$$s'(x_1) = f'(x_1)$$

$$s'(x_n) = f'(x_n)$$

las cuales se conocen como: *condiciones de acotamiento correcto*. O bien, se pueden agregar condiciones que involucren periodicidad como:

$$s'(x_1) = s'(x_n),$$

$$s''(x_1) = s''(x_n).$$

Cualquiera de las opciones genera un spline cúbico interpolante con $4n-4$ grados de libertad, de aquí que las condiciones del inciso a) quedan satisfechas. El inciso b) se refiere al grado de los polinomios que interpolan en los subintervalos $[x_i, x_{i+1}]$, para: $i = 1, 2, ..., n-1$; cuyos coeficientes estarán determinados por las condiciones anteriores, por ejemplo:

$$s_i(x) = f_i + s_i'(x_i)(x - x_i) + \frac{s_i''(x_i)}{2}(x - x_i)^2 + \frac{s_i^{(3)}(x_i)}{3!}(x - x_i)^3.$$

Observe que los términos del polinomio $s_i(x)$ son los cuatro primeros de la expansión de Taylor de s_i alrededor de x_i. A continuación se obtienen la primera y segunda derivada de $s_i(x)$:

$$\begin{aligned} s_i'(x) &= s_i'(x_i) + s_i''(x_i)(x - x_i) + \frac{s_i^{(3)}(x_i)}{2}(x - x_i)^2, \\ s_i''(x) &= s_i''(x_i) + s_i^{(3)}(x_i)(x - x_i) \end{aligned}$$

Denotemos h_i como la longitud del $i - ésimo$ intervalo, a saber:

$$h_i = x_{i+1} - x_i,$$

y pongamos:

$$\begin{aligned} z_i &= s_i''(x_i), \text{ para: } i = 1, 2, ..., n-1, \\ z_n &= s_{n-1}''(x_n). \end{aligned}$$

De la expresión para $s_i''(x)$ tenemos, evaluando en x_{i+1}:

$$s_i^{(3)}(x_{i+1}) = \frac{s_i''(x_{i+1}) - s_i''(x_i)}{h_i}.$$

Del requisito $s_i''(x)$ contínua en los puntos internos se tiene:

$$s_i''(x_{i+1}) = s_{i+1}''(x_{i+1}), \text{ para: } i = 1, 2, ..., n-2;$$

de aquí que:

$$s_i^{(3)}(x_i) = \frac{z_{i+1} - z_i}{h_i}, \text{ para: } i = 1, 2, ..., n-2 \quad .....(a)$$

Dicha relación es válida también para $i = n-1$, reexpresemos a $s_i(x)$ de la siguiente manera:

$$s_i(x) = f_i + s_i'(x_i)(x - x_i) + \frac{z_i}{2}(x - x_i)^2 + \frac{z_{i+1} - z_i}{6h_i}(x - x_i)^3, \text{ para: } i = 1, 2, ..., n-1 \quad .....(b)$$

y su derivada:

$$s_i'(x) = s_i'(x_i) + z_i(x - x_i) + \frac{z_{i+1} - z_i}{2h_i}(x - x_i)^2 \quad .....(c)$$

Del requisito original: $s_i(x_{i+1}) = f_{i+1}$, tenemos que:

$$f_i + s_i'(x_i)h_i + \frac{z_i}{2}h_i^2 + \frac{z_{i+1} - z_i}{6h_i}h_i^3 = f_{i+1},$$

o bien:

$$s_i'(x_i) = \frac{f_{i+1} - f_i}{h_i} - z_{i+1}\frac{h_i}{6} - z_i\frac{h_i}{3}, \text{ para: } i = 1, 2, ..., n-1 \text{ .....}(d)$$

Observe que tenemos una expresión para $s_i'(x_i)$ que hasta el momento seguimos sin poder calcular ya que es necesario encontrar una expresión para z_i , usaremos el requisito de continuidad en la primera derivada: $s_i'(x_{i+1}) = s_{i+1}'(x_{i+1})$, para: $i = 1, 2, ..., n-2$ y (c) para calcular $s_{i+1}'(x_{i+1})$, de aquí que:

$$s_i'(x_i) + z_i h_i + \frac{z_{i+1} - z_i}{2h_i}h_i^2 = s_{i+1}'(x_{i+1})$$

Sustituyendo $s_i'(x)$ y $s_{i+1}'(x_{i+1})$ en (d):

$$\frac{f_{i+1} - f_i}{h_i} - z_{i+1}\frac{h_i}{6} - z_i\frac{h_i}{3} + z_i h_i + \frac{z_{i+1} - z_i}{2}h_i = \frac{f_{i+2} - f_{i+1}}{h_{i+1}} - z_{i+2}\frac{h_{i+1}}{6} - z_{i+1}\frac{h_{i+1}}{3}$$

Simplificando:

$$h_i z_i + 2(h_i + h_{i+1})z_{i+1} + h_{i+1}z_{i+2} = 6\left(\frac{f_{i+2} - f_{i+1}}{h_{i+1}} - \frac{f_{i+1} - f_i}{h_i}\right), \text{ para : } i = 1, 2, ..., n-2.$$

Lo anterior representa un sistema de $n-2$ ecuaciones con $n-2$ incógnitas: $z_2, z_3, ..., z_{n-1}$. La matriz que describe el sistema lineal es simétrica y tridiagonal. Mas aún, de diagonal dominante lo cual implica que podemos aplicar eliminación Gaussiana sin pivoteo.

Conclusión. De esta manera, podemos calcular $s_i'(x_i)$ de (d), para: $i = 1, 2, ..., n-1$ y los demás requisitos en forma recursiva: $s_i(x)$ de (b), $s_i^{(3)}(x_i)$ de (a) y por definición $z_i = s_i''(x_i)$. En el caso que deseen añadir las condiciones de acotamiento correcto, se incluye en el sistema lineal (al principio y al final), las ecuaciones:

$$\left\{ \begin{array}{c} -\frac{h_1}{3}z_1 - \frac{h_1}{6}z_2 = \frac{f_1 - f_2}{h_1} + f'(x_1) \\ -\frac{h_{n-1}}{6}z_{n-1} - \frac{h_{n-1}}{3}z_n = \frac{f_{n-1} - f_n}{h_{n-1}} + f'(x_n) \end{array} \right\}$$

Con estas dos ecuaciones tenemos de nuevo: una matriz tridiagonal, simétrica y de diagonal dominante.

Teorema (*III*.3.2). De todas las funciones g con segunda derivada contínua en $[a, b]$, y que interpolan a f en los puntos $a = x_1 < x_2 < ... < x_n = b$, el spline cúbico natural minimiza:

$$\int_a^b (g''(x))^2 dx.$$

Lo cual significa que el spline cúbico interpolante no tendrá grandes oscilaciones, ya que la primera derivada de una función oscilante toma valores grandes tanto positivos como negativos. Por el teorema del valor medio, se observa que la segunda derivada también puede tomar valores muy grandes y por lo tanto su integral no puede ser pequeña. El teorema III.3.2 se puede probar usando el siguiente lema:

Lema (*III*.3.3). Sea s spline cúbico natural que interpola a f en los nodos:

$$a = x_1 < x_2 < ... < x_n = b$$

Para toda función g con segunda derivada contínua en $[a, b]$ que interpola a f en los nodos, la siguiente afirmación es válida:

$$\int_a^b (g''(x))^2 dx = \int_a^b (s''(x))^2 dx + \int_a^b (g''(x) - s''(x))^2 dx.$$

Demostración (*III*.3.3).

Reescribimos $\int_a^b (g''(x) - s''(x))^2 dx$ sumando y restando $\int_a^b (s''(x))^2 dx$:

$$\int_a^b (g''(x) - s''(x))^2 dx = \int_a^b (g''(x))^2 dx - \int_a^b (s''(x))^2 dx - 2\int_a^b (g''(x) - s''(x))s''(x) dx.$$

Por demostrar:

$$\int_a^b (g''(x) - s''(x))s''(x) dx = 0 \text{ .....}(e)$$

Considere (e) en el intervalo $[x_i, x_{i+1}]$. Integrando por partes:

$$I_i = \int_{x_i}^{x_{i+1}} (g''(x) - s''(x))s''(x) dx = [(g''(x) - s''(x))s''(x)]_{x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} (g'(x) - s'(x))s^{(3)}(x) dx.$$

Ya que s es un polinomio cúbico en el intervalo $[x_i, x_{i+1}]$, se deduce que $s^{(3)}(x)$ es constante. Por lo tanto:

$$\int_{x_i}^{x_{i+1}} (g'(x) - s'(x))s^{(3)}(x) dx = s^{(3)}(x_i) \int_{x_i}^{x_{i+1}} (g'(x) - s'(x)) dx = s^{(3)}(x_i) [g(x) - s(x)]_{x_i}^{x_{i+1}} \text{ .....}(f)$$

Por supuesto:

$$g(x_i) = s(x_i) = f_i, \ \forall \ i = 1, 2, ..., n,$$

por lo tanto (f) son todas cero, ya que:

$$s^{(3)}(x_i) [g(x) - s(x)]_{x_i}^{x_{i+1}} = s^{(3)}(x_i) [(g(x_{i+1}) - s(x_{i+1})) - (g(x_i) - s(x_i))] = 0.$$

De aquí que,

$$\begin{aligned}\int_a^b (g''(x) - s''(x))s''(x)dx &= \sum_{i=1}^{n-1} I_i \\ &= \sum_{i=1}^{n-1} \left[(g'(x_{i+1}) - s'(x_{i+1}))\, s''(x_{i+1}) - (g'(x_i) - s'(x_i))\, s''(x_i)\right] \\ &= (g'(x_n) - s'(x_n))\, s''(x_n) - (g'(x_1) - s'(x_1))\, s''(x_1).\end{aligned}$$

Esta última expresión es igual a cero por el supuesto de ser s un spline natural, i.e.

$$s''(x_1) = 0 \text{ y } s''(x_n) = 0.$$

por lo que el lema queda demostrado.

Nota. El teorema anterior es también válido bajo las condiciones de acotamiento correcto $s'(x_1) = f'(x_n)$ y $s'(x_n) = f'(x_n)$, si suponemos g , tal que:

$$g'(x_1) = f'(x_1) \text{ y } g'(x_n) = f'(x_n).$$

Conclusión. Un spline cúbico es la función que menor error genera al interpolar a f en los puntos conocidos x_i.

CAPÍTULO IV: ORÍGENES DE LA REGRESIÓN NO PARAMÉTRICA

IV.1. INTRODUCCIÓN.

La teoría de regresión NP es como muchas otras producto de un proceso evolutivo, en este capítulo se describen a grosso modo algunas de las teorías mas importantes que han dado origen a las modernas técnicas de suavizamiento descritas en el capítulo V.

Sin duda, la regresión lineal ha sido la base para el desarrollo NP, por ejemplo en la regresión lineal el usuario obtiene un modelo tentativo y se verifica si debe ser o no corregido esto es, verificar si es que se deben: aplicar transformaciones, discernir si es que los puntos aislados deben ser removidos o no, verificar el comportamiento del coeficiente de determinación ante los cambios realizados, verificar el impacto de dichos ajustes en los supuestos de linealidad y varianza constante, etc. A diferencia de esto, en la regresión NP el usuario comienza con un modelo que es determinado exclusivamente por los datos obteniendo estimaciones que en cierto sentido, son superiores a aquellas obtenidas a partir de un modelo paramétrico.

IV.2. RELAJANDO SUPUESTOS DE REGRESIÓN.

La regresión lineal se encuentra basada en dos supuestos básicos:

1) El modelo lineal se supone adecuado.

2) Los errores son independientes y $\varepsilon_i \sim N\left(0, Varianza\ Constante\right)$; $i = \overline{1,n}$.

¿Qué sucede si pasamos por alto el segundo supuesto?. Sin duda los procesos inferenciales establecidos para la regresión lineal no serían estrictamente aplicables, ya que estos se basan en el supuesto de errores normales y si sabemos que los errores no presentan una distribución normal, entonces sería inútil aplicar el método de mínimos cuadrados para estimar a los parámetros $\widehat{\beta_0}$ y $\widehat{\beta_1}$; correspondientes a la ordenada al origen y a la pendiente de una recta, ambos parámetros obtenidos de la derivada parcial de la función L:

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$$

donde:

$$\varepsilon = Y - \beta_0 - \beta_1 x$$

que representa la distancia vertical del valor observado Y_i a la recta ajustada $\widehat{Y}$, en cuyo caso:

$$\varepsilon_i = Y_i - \widehat{Y}_i.$$

De esta manera es como los parámetros β_0 y β_1 quedan estimados por:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

y

$$\widehat{\beta_0} = \frac{\left(\sum_{i=1}^{n} Y_i\right)}{n} - \widehat{\beta_1}\frac{\left(\sum_{i=1}^{n} x_i\right)}{n} = \overline{Y} - \widehat{\beta_1}\overline{x}.$$

Estas expresiones así como la demostración al teorema de Gauss-Markov* se encuentran frecuentemente en libros de Regresión Lineal (ver por ejemplo el tratado por Montgomery & Peck, bibliografía detallada al final de este texto).

* Los estimadores $\widehat{\beta_0}$ y $\widehat{\beta_1}$, generados por el método de mínimos cuadrados: son los mejores en el sentido de varianza mínima de todos los estimadores insesgados que sean funciones lineales de Y_i.

Ahora bien, debido a que los parámetros $\widehat{\beta_0}$ y $\widehat{\beta_1}$ fueron construidos con teoría basada en la normalidad en los errores, no es posible utilizar estos mismos parámetros en aquellos modelos cuya normalidad no puede ser asegurada. Es por esta razón que surgen nuevas herramientas estadísticas que permiten mediante una prueba de hipótesis: $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$; saber si el modelo tiene sentido o no. El Coeficiente de Correlación de Spearman, es una de las primeras herramientas matemáticas NP. Esta fue desarrollada en 1904 para verificar la relación entre dos variables y está basada en el coeficiente de correlación de Pearson:

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

El signo de este coeficiente describe cómo son las variables entre sí, por ejemplo: si $r \simeq 1 \Rightarrow$ los valores chicos de la variable X están muy relacionados con lo valores chicos de

la variable Y y viceversa: los valores grandes de X están muy relacionados con los valores grandes de Y; y si por ejemplo $r \simeq -1 \Rightarrow$ los valores chicos de la variable X están muy relacionados con lo valores grandes de la variable Y y viceversa. Spearman desarrolla una herramienta estadística basada en la idea del coeficiente de Pearson, tomando la misma información (x_i, y_i)con $i = \overline{1,n}$; asigna rangos de acuerdo a los valores de las x_i y las y_i . De aquí que a cada punto (x_i, y_i) se le asigna un valor $(R(x_i), R(y_i))$, dando como resultado el coeficiente ρ de correlación de Spearman:

$$\rho = \frac{\sum_{i=1}^{n} \left(R(x_i) - \frac{n+1}{2}\right)\left(R(y_i) - \frac{n+1}{2}\right)}{\left[\frac{n(n^2-1)}{12}\right]}$$

donde:

$$R(\overline{x}) = \sum_{i=1}^{n} \frac{R(x_i)}{n} = \sum_{i=1}^{n} \frac{i}{n} = \frac{n+1}{2}$$

y para las varianzas:

$$\sum_{i=1}^{n} \left(R(x_i) - \frac{n+1}{2}\right)^2 = \sum_{i=1}^{n} i^2 + \frac{n(n+1)^2}{4} - (n+1)\sum_{i=1}^{n} i = \frac{n(n^2-1)}{12}$$

El coeficiente ρ de Spearman es una de las primeras herramientas estadísticas que son desarrolladas para verificar la relación lineal entre variables cuando el supuesto de normalidad de los errores es insostenible.

IV.3. BOOTSTRAPPING (RECOCIDO).

Otra herramienta NP que permite saber si un modelo es adecuado o no, es la conocida como: *bootstrapping*; la cual consiste en construir un intervalo de confianza para el parámetro β_1 de un modelo lineal simple construido por mínimos cuadrados aún cuando se sospecha que el supuesto de normalidad en los errores es insostenible. Por esta razón, en realidad tampoco conocemos la distribución de $\widehat{\beta_1}$ (necesaria para obtener un intervalo de confianza).

Suponga que deseamos construir un intervalo de confianza para β_i, a falta del supuesto de normalidad. *Bootstrapping*, es una técnica de *submuestreo*, esto es, se obtienen muestras con remplazo de la población original y se calculan:

$$t_j^* = \frac{\widehat{\beta_{ij}^*} - \widehat{\beta_i}}{s_{\widehat{\beta_{ij}^*}}} \text{ , con: } j = 1, 2, ..., N$$

los percentiles de la distribución de $\widehat{\beta_i}$, donde N denota el número de submuestras bootstrap obtenidas y $\widehat{\beta_{ij}^*}$ es el estimador de $\widehat{\beta_i}$ en la $j - ésima$ muestra bootstrap ($s_{\widehat{\beta_{ij}^*}}$

es el estimador de la desviación standard del estimador bootstrap de β_i en la $j-ésima$ muestra). Por último $\widehat{\beta_i}$ denota al estimador por mínimos cuadrados de β_i de la muestra original. Un intervalo del 95% de confianza para β_i sería calculado de la siguiente manera:

$$límite\ inferior = \widehat{\beta_i} - t^*_{.975} s_{\widehat{\beta^*_{ij}}}$$

$$límite\ superior = \widehat{\beta_i} + t^*_{.025} s_{\widehat{\beta^*_{ij}}}$$

donde $t^*_{.025}$ y $t^*_{.975}$ denotan los cuantiles .025 y .975 de una distribución empírica t^*.

Es necesario tomar un número de submuestras bootstrap suficientemente grande*, para asegurar que los cuantiles que acotan al intervalo de confianza se acercan mas a los cuantiles teóricos, i.e. en cada submuestra tendremos dos datos: el límite superior y el inferior del intervalo, pero si nos fijamos en todas las submuestras en realidad habríamos generado un conjunto de límites inferiores y otro de límites superiores en cuyos rangos respectivos se encontrarán los extremos teóricos a los que se desea converger.

* se recomienda una $N = 2000$ (Hamilton 1992, p.314).

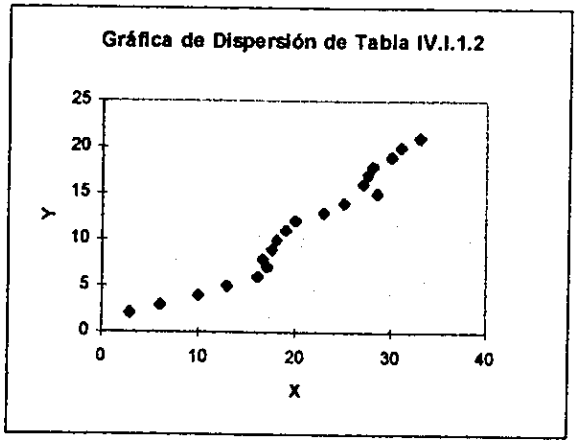
Observación. Aún sin importar cuáles serán los extremos del intervalo, el centro del mismo es el estimador por mínimos cuadrados de β_i. Esto es, entre mas anormales se distribuyan las observaciones el intervalo de confianza crecerá cada vez mas y si por el contrario, si la varianza tiende a cero entonces los límites descritos tienden a $\widehat{\beta_i}$.

IV.4. REGRESIONES MONÓTONAS.

¿Qué pasaría si adicionalmente se relaja el primer supuesto y por lo tanto, se supone que en realidad el modelo lineal no es apropiado?. Supongamos dos variables cuya relación se describe en la tabla IV.4.1.

Y	X	Y	X
3.0	2	20.0	12
6.0	3	23.0	13
10.0	4	25.0	14
13.0	5	28.5	15
16.0	6	29.0	16
17.0	7	29.0	17
17.5	8	29.5	18
17.5	9	30.0	19
18.0	10	31.0	20
19.0	11	33.0	21

Tabla IV.4.1.



Gráfica IV.4.2.

Observe la gráfica correspondiente, aparentemente existe evidencia clara de una relación lineal entre las variables X y Y. Debido a la tendencia que tienen las observaciones, sería posible aplicar eventualmente alguna(s) transformación(es) de tipo polinomial o incluso por medio de un modelo de regresión no-lineal. Sea cual fuere esta relación, es claro que las variables presentan un comportamiento monótono estrictamente creciente.

En 1979 Iman y Conover introducen el concepto de *Regresión por Rangos* como una alternativa para la formulación de modelos no-lineales. Para una regresión simple, el método consiste en convertir las variables X y Y a rangos y entonces, aplicar regresión lineal simple en los rangos mencionados. Utilizamos la información de la tabla IV.4.1 para ilustrar dicha teoría. A continuación se enuncian las fórmulas para $\widehat{\beta_0}$ y $\widehat{\beta_1}$ aplicadas a los rangos de las variables X y Y:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n(n+1)^2/4}{\sum_{i=1x_i}^{n} [R(x_i)]^2 - \frac{n(n+1)^2}{4}}$$

y

$$\widehat{\beta_0} = \frac{\left(1 - \widehat{\beta_1}\right)(n+1)}{2}$$

donde $R(x_i)$ denota el rango de x_i y $R(y_i)$ el de Y_i. Compare estas expresiones con las

dadas para la regresión lineal:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

y

$$\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\overline{x}.$$

De un regresor a otro, el cambio radica en que: $\sum_{i=1}^{n} i = n(n+1)/2$ y por lo tanto: $\overline{i} = (n+1)/2$. La ecuación de regresión $R(y)$ contra $R(x)$: $\widehat{R(y)} = 0.1105 + 0.9895\widehat{R(x)}$. La relación monótona creciente casi estricta entre X y Y se refleja en el coeficiente de determinación: $R^2 = .979$; desde luego, deseamos explicar los valores de Y y no los del $R(Y)$, para lo cual será necesario obtener $\widehat{Y}$ a partir de $\widehat{R(Y)}$. Suponga que deseamos predecir Y cuando $x_0 = 11$ (uno de los valores observados del vector X), de la tabla: $R(x_0) = 10$; de aquí que $\widehat{R(Y)} = 10.0052$. Para obtener $\widehat{Y_0}$, interpolamos en los rangos adyacentes al valor: 10.0052; es decir: 11 y 12, tomando como referencia la imagen de los rangos considerados, de esta manera un estimador $\widehat{Y}$ para el valor $x_0 = 11$ se calcularía de la siguiente manera:

$$\widehat{Y_0} = 19 - \frac{10.0052 - 10}{11 - 10}(20 - 19) = 19.0052.$$

En el ejemplo anterior no fue necesario estimar Y para cuando $x_0 = 11$ ya que este es uno de los valores observados. Si este no fuera el caso, sería necesario utilizar interpolación lineal para obtener el valor $R(x_0)$.

Comparativo. Usando regresión lineal se obtiene $\widehat{Y_0} = 19.757$ y un coeficiente de determinación: $R^2 = .951$; sin embargo, por medio de la regresión monótona por rangos, tenemos $(r_{Y,\widehat{Y_0}})^2 = .997$ y por lo tanto, se obtendrían mejores resultados por medio de un método no paramétrico aún cuando el método lineal parecía adecuado.

Observación. Debido a que .997 es un valor tan cercano a 1.0, difícilmente podríamos obtener un mejor coeficiente de determinación arreglando al modelo lineal por medio de algún tipo de transformaciones, de esta manera podríamos intuir que el método de regresión por rangos funciona como sustituto para ciertos modelos paramétricos.

CAPÍTULO V: TÉCNICAS DE SUAVIZAMIENTO

V.1. INTRODUCCIÓN.

Los modelos matemáticos de regresión están basados de manera natural en la intuición del hombre esto es, el pensamiento humano tiende a inferir a partir de información obtenida de un cierto evento en particular. Espontáneamente actuaremos en consecuencia de lo que haya ocurrido a lo largo de nuestras vidas, la información recibida en nuestro pasado, el ámbito cultural en el cual nos desarrollamos, los eventos políticos, económicos, culturales, religiosos, sociales, etc.; influirán de manera directa en qué pensamos, cómo actuamos y por consecuencia, en qué decisiones tomamos. Por ejemplo, suponga que existe una región en el mundo en donde la gente acostumbra salir a la calle con una sombrilla aún cuando no se haya presentado una sola lluvia en meses, cualquiera de nosotros juzgaría ilógico pensar de esta manera y sólo usaremos la sombrilla en el caso en que últimamente se hayan presentado lluvias constantes. La forma de reaccionar del ser humano ante cualquier eventualidad estará fuertemente influida por la información recolectada, por eventos recientes, etc.

De la misma manera, existen modelos matemáticos que buscan generar la toma de decisiones a partir de eventos ocurridos en el pasado, así mismo existen aquellos modelos cuya finalidad es inferir acerca de cómo se comportó un cierto fenómeno en un pasado anterior al momento en el cual se obtuvo la información. La estadística analiza estos y otros temas de estudio relacionados a través de los: *modelos de regresión.*

El estudio de la *teoría inferencial* se vuelve mucho más complicado entre más apegados se deseen hacer los modelos a la realidad, es decir, en la realidad no todos los eventos influyen de igual manera en la ocurrencia de determinado suceso por ende, es preciso encontrar una manera de relacionarle un peso a cada evento en la medida que este haya influido en el suceso en cuestión. Mas aún, debemos analizar qué parte de la información que tenemos, en realidad contribuye a tomar las decisiones más adecuadas. Debemos entonces, discernir qué información distrae la atención del modelo y cuál otra contribuye de manera importante en el mismo.

V.2. SUAVIZAMIENTO.

Como ya se había mencionado, es importante antes de construir un modelo matemático tomar la siguiente decisión: ¿qué método de regresión, interpolación, ajuste, etc.; es aquel que contribuye a construir un mejor modelo que describa la información recolectada y cuál

será aquel que aporte mejores inferencias acerca del fenómeno?. A menudo, es posible discernir entre un método u otro con simplemente echar un vistazo a una gráfica de dispersión, pero esto tiene varios inconvenientes, el primero es que no siempre es posible graficar la información de esta manera, el segundo es que ante una gran cantidad de puntos en una gráfica no siempre es claro observar una tendencia que permita establecer la relación entre las variables en cuestión.

Un *suavizador* (*smoother*) es una herramienta que nos permite conocer una cierta tendencia (en caso que esta exista), en la variable respuesta Y como función de uno o mas regresores.Ya hemos hablado en la introducción de este trabajo de las bondades que presentan los modelos construidos con base en la teoría NP, esto es, la construcción de curvas con las características de: flexibilidad, predicción, estimación y una correcta descripción de outliers; además de que su morfología por sí misma, suele aportar información adicional a la generada por el modelo.

En general, la teoría NP estudia dos problemas centrales:

1. Encontrar los *parámetros de suavizamiento* (*smoothing parameters*).
2. Construir las *bandas de confianza* correspondientes.

Suponga x_0 un valor de la variable independiente, la esperanza de la variable respuesta Y en ese punto:

$$\hat{Y}_0 = E\left[Y \mid x_0\right].$$

Definición. Sea $(a, b) \in X$ subconjunto de valores de la variable independiente. Decimos que S es *suavizador* si:

$$S = \left\{E\left[Y \mid x_i\right] \;\mid\; x_i \in (a, b), i = 1, ..., n\right\}.$$

En particular los *suavizadores lineales* son aquellos en los que $\hat{Y}_i$ pueden expresarse como combinación lineal de Y_i en general, cuando hablamos de regresión NP el modelo que describe las observaciones para un solo regresor se escribe:

$$Y_i = \mu(x_i) + \varepsilon_i,\ i = 1, ..., n.$$

Se puede mostrar fácilmente que un estimador $\widehat{\mu}(x_i)$, que ha sido obtenido a través de suavizamiento será sesgado. Suponga usted una colección de n observaciones de las cuales se extrae una muestra de tamaño n', la cual será usada para estimar $\mu(x_i)$.

A partir de este momento nos cuestionamos qué expresión ayuda a construir el estimador deseado utilizando la información extraída en la muestra de tamaño n', se propone

la siguiente:

$$\widehat{\mu}(x_j) = \frac{\left(Y_j + \sum_{i \neq j}^{n'} Y_i\right)}{n'}$$

la cual construye un estimador que es simplemente el promedio de las n' observaciones. Verificando el sesgo del estimador:

$$E\left(\widehat{\mu}(x_j)\right) = \frac{1}{n'}\left\{\mu(x_j) + \sum_{i \neq j}^{n'} \mu(x_i)\right\}$$

sumando un cero adecuado: $\pm\frac{n'-1}{n'}\mu(x_j)$; se obtiene:

$$E\left(\widehat{\mu}(x_j)\right) = \mu(x_j) + \frac{1}{n'}\left\{\sum_{i=1}^{n'}\left[\mu(x_i) - \mu(x_j)\right]\right\}$$

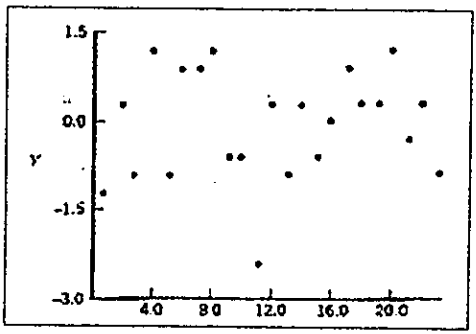
el segundo término de la última expresión determina claramente el sesgo del estimador, pero en realidad... ¿de qué manera se puede obtener un estimador adecuado $\widehat{\mu}(x_j)$? ¿de qué manera se eligen las observaciones que participan en la evaluación del estimador? ¿todas las observaciones influyen de igual manera en la estimación del ó de los parámetros?, en caso afirmativo: estaremos hablando de un modelo paramétrico tradicional; a continuación, un enunciado que tiene por objeto aclarar la idea básica del suavizamiento:

> Si μ es contínua y diferenciable, entonces las observaciones x_i alrededor de x deben contener información acerca del valor de μ en x. Por lo tanto, debe ser posible utilizar una especie de promedio local de los datos cercanos a x para construir un estimador de $\mu(x)$ (Randall Eubank, 1988).

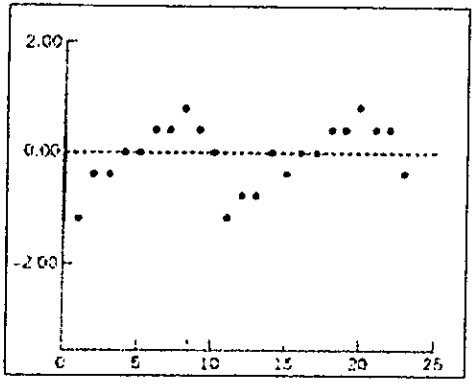
Los suavizadores han sido utilizados en estudios ajenos al de la regresión, por ejemplo: un *promedio móvil* es una técnica NP que promueve el suavizamiento de observaciones involucradas en series de tiempo e incluso en estudios de control de calidad. Sin abundar demasiado en cómo influye esta técnica dentro de las especialidades mencionadas, es importante conocer la idea de estos promedios ya que nos ayudarán a generar abstracciones análogas en la teoría del suavizamiento kernel (teoría central de este texto). Si un suavizador se utiliza para descripción o estimación de observaciones, se deben tomar en cuenta dos decisiones principales:

1) Cómo deben ser promediadas las observaciones Y_i en cada *vecindad* y,
2) Cuál debe ser el tamaño de cada vecindad.

Es importante puntualizar lo siguiente: los suavizadores deben ser utilizados con cuidado, es muy relevante discernir qué observaciones son las que intervendrán en la estimación correspondiente $\widehat{\mu}(x_j)$, ya que una selección inadecuada de vecindades podría generar estimadores en los que se observe algún tipo de tendencia a partir de información proveniente de variables que no tengan relación entre sí. Como ejemplo considere la gráfica de dispersión V.2.1,



Gráfica V.2.1.



Gráfica V.2.2.

Los puntos de la gráfica V.2.1, representan números aleatorios generados de una $N(0,1)$, a pesar de que estos poseen una naturaleza aleatoria, una selección inadecuada de vecindades daría como resultado un modelo de regresión que marca una tendencia cuando en realidad esta no existe. La gráfica V.2.2 describe un conjunto de suavizadores S (conjunto de $\widehat{\mu}(x_j)$ $\forall\, x_j \in (X_{min}, X_{max})$), esto es, promedios móviles de tamaño 3 de las observaciones de la gráfica V.2.1

V.3. SUAVIZADOR LINEAL.

El *suavizador lineal* (*running line smoother*), es en escencia muy similar al calculado por medio de promedios móviles.

Para el modelo lineal se obtienen estimadores de los parámetros β_0 y β_1 para cada ventana, los cuales corresponden a la ordenada al origen y a la pendiente del modelo de regresión lineal simple: $Y = \widehat{\beta}_0 + \widehat{\beta}_1 x$; de donde:

$$\widehat{\beta}_{1j} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n'_j}(x_i - \overline{x_j})(y_i - \overline{y_j})}{\sum_{i=1}^{n'_j}(x_i - \overline{x_j})^2},$$

$$\widehat{\beta}_{0j} = \overline{y_j} - \widehat{\beta}_{1j}\overline{x_j}$$

donde n'_j representa el tamaño de muestra de las n observaciones que caen en la $j-ésima$ ventana ($\overline{x_j}, \overline{y_j}$ los promedios correspondientes a dicha ventana), de esta manera se obtiene un conjunto de estimadores de los parámetros: $\widehat{\beta}_1 = \left\{\widehat{\beta}_{1j}, \text{ para } j = 1, ..., r\right\}$, donde r es el total de ventanas móviles (análogamente para el conjunto $\widehat{\beta}_0$). Hasta este momento hemos construido dos tipos de suavizadores: el lineal y el construido por medio de promedios móviles. Sin embargo, en la introducción de este texto se mencionó que una de las ventajas que ofrecían los métodos NP era la *flexibilidad* de la curva en cuestión ¿qué sucede si deseamos remover, añadir y/o reemplazar una observación?, en el caso de un modelo tradicional de regresión lineal los parámetros $\widehat{\beta}_0$ y $\widehat{\beta}_1$ tendrían que ser nuevamente calculados y si suponemos que el punto que entra (x^+, y^+) es distinto del que sale (x^-, y^-) entonces la curva de regresión cambiaría su estructura.

Sean $S^*_{xy}, S^*_{xx}, \overline{y}^*_k, \overline{x}^*_k$; las estadísticas actualizadas que corresponden a la vecindad de tamaño k en donde se realizó dicho cambio (las expresiones sin asteriscos representaran las estadísticas originales):

$$\overline{x}^*_k = \frac{1}{k}(k\overline{x}_k + x^+ - x^-)$$

$$\overline{y}^*_k = \frac{1}{k}(k\overline{y}_k + y^+ - y^-)$$

La expresión para S^*_{xx} se deduce de la siguiente manera, si: $S_{xx} = \sum_{i=1}^{n'}(x_i - \overline{x}'_k)^2$; con $\overline{x}'_k$ el promedio original de las observaciones en la muestra:

$$S^*_{xx} = \sum_{j=1}^{n'}\left[x_j - \overline{x}^*_k\right]^2$$

$$\Leftrightarrow S^*_{xx} = \sum_{j=1}^{n'}\left[x_j - \frac{1}{k}(k\overline{x}'_k + x^+ - x^-)\right]^2$$

$$= \sum_{j=1}^{n'} \left[(x_j - \overline{x}'_k) - \frac{1}{k}(x^+ - x^-) \right]^2$$

desarrollando el binomio de Newton y distribuyendo la suma:

$$S^*_{xx} = S_{xx} + (x^+ - \overline{x}'_k)^2 - (x^- - \overline{x}'_k)^2 - \frac{2}{k}(x^+ - x^-) \sum_{j=1}^{n'} [x_j - \overline{x}'_k] + \frac{1}{k}(x^+ - x^-)^2$$

obsérvese que los tres primeros términos del lado derecho de la igualdad corresponden a la suma: $\sum_{j=1}^{n'}(x_j - \overline{x}'_k)^2$. Simplificando la expresión anterior se obtiene:

$$S^*_{xx} = S_{xx} + (x^+ - \overline{x}'_k)^2 - (x^- - \overline{x}'_k)^2 - \frac{1}{k}(x^+ - x^-)^2$$

el último término, resultado de: $\sum_{j=1}^{n'} [x_j - \overline{x}'_k] = 0 + (x^+ - \overline{x}'_k) - (x^- - \overline{x}'_k) = (x^+ - x^-)$.

De esta manera, obtenemos la expresión para la estadística S_{xx} en caso de que se altere el conjunto de puntos de la ventana correspondiente y por tanto, haya que modificar la estimación de los parámetros lineales del suavizador correspondiente en cada vecindad, de la misma forma en el caso de promedios móviles el estimador sesgado:

$$\widehat{\mu}(x_j) = \frac{\left(Y_j + \sum_{i \neq j}^{n'} Y_i \right)}{n'}$$

se modifica de tal manera que la estimación quede en función del conjunto actualizado de puntos:

$$\widehat{\mu}^*_k(x_j) = \overline{y}^*_k = \frac{1}{k}(k\overline{y}_k + y^+ - y^-)$$

para la vecindad actualizada de tamaño k.

Análogamente para S^*_{xy}, se obtiene la estadística actualizada al nuevo conjunto de puntos a partir de:

$$S^*_{xy} = \sum_{j=1}^{n'} \left[x_j - \frac{1}{k}(k\overline{x}_k + x^+ - x^-) \right] \left[y_j - \frac{1}{k}(k\overline{y}_k + y^+ - y^-) \right]$$

distribuyendo el producto y después la suma obtenemos la expresión de la estadística actualizada para la configuración de puntos deseada,

$$S^*_{xy} = S_{xy} + (x^+ - \overline{x}_k)(y^+ - \overline{y}_k) - (x^- - \overline{x}_k)(y^- - \overline{y}_k) - \frac{1}{k}(x^+ - x^-)(y^+ - y^-)$$

con estas herramientas es posible ajustar los parámetros únicamente en las regiones en que dichas modificaciones deban realizarse, dejando intactas aquellas que no se ven afectadas directamente en la eliminación, reemplazo y/o adición del elemento en cuestión.

Es muy importante que el lector comprenda la importancia de los avances que presenta esta teoría y que imagine qué pasaría en un modelo paramétrico tradicional si por ejemplo se reemplazaran, extrajeran o añadieran observaciones. Por el momento un avance que se irá perfeccionando es el que se refiere a la: *flexibilidad*; entendida como la posibilidad de añadir, remover o reemplazar observaciones.

Hasta este momento hemos identificado dos tipos de suavizadores: el lineal y el de promedios móviles; en ambos, existe un concepto fundamental que no se ha detallado hasta el momento: el tamaño k de la vecindad., intuitivamente surgen nuevas interrogantes con respecto a este problema por ejemplo, supongamos que se ha construido una vecindad de tamaño $k = 11$ para el cálculo del estimador $\widehat{y}_{i4}$, para tales efectos tomaremos los 11 puntos que se localizan alrededor de x_{14} y a partir de esta información se construye el estimador correspondiente, ya sea por el método lineal o promediando la ordenada de los puntos $(x_9, y_9), ..., (x_{19}, y_{19})$. Si estamos de acuerdo en esta metodología ¿debemos calcular de esta manera todos los demás estimadores?, ¿es k es una constante para el cálculo de $\widehat{\mu}(x_j)$, $\forall j = 1, ..., n$?, ¿qué sucede en los extremos x_{min} y x_{max} en donde no es posible tomar vecindades simétricas?, ¿qué sucede si k es un valor par y por lo tanto tampoco es posible tomar vecindades simétricas?.

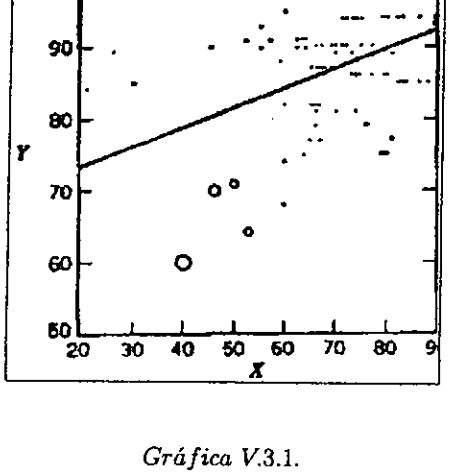
Intuitivamente, una *vecindad más cercana* (*nearest neighborhood*) es el conjunto de puntos mas cercanos a x_0 y una *vecindad simétrica más cercana* (*symmetric nearest neighborhood*), es el conjunto de puntos $\frac{k}{2}$ puntos a cada lado de x_0.

En general, las vecindades mas cercanas tendrán menor sesgo que las simétricas, sin embargo en los puntos extremos utilizar una vecindad mas cercana no siempre es una ventaja. Si suponemos que el tamaño k es constante para todos los estimadores, en los puntos extremos el modelo lineal le estaría asociando un peso muy grande a las observaciones que pueden ser muy lejanas. Por otro lado, si se elige una vecindad simétrica entonces para x_{min} y x_{max}, estarían siendo estimados por aproximadamente la mitad de los puntos de los que influirían en el mismo estimador si se manejara una vecindad mas cercana por lo tanto, es muy importante tomar la decisión de cuáles serán los puntos que estimen $\widehat{\mu}(x_j)$, $\forall j = 1, ..., n$; y que en particular dicha selección estime adecuadamente a los puntos extremos. Podríamos pensar que una vecindad mas cercana es la mas adecuada, debido a que en general reduce el sesgo en un estimador que como ya hemos visto es sesgado por naturaleza sin embargo, en los extremos no nos gustaría que el estimador se alterara demasiado debido a eventuales reemplazos de observaciones lejanas por otro lado, si el tamaño de la vecindad disminuye entonces tal vez arreglaríamos el problema en los extremos pero el suavizamiento no sería del todo correcto en las demas regiones de la dispersión es decir, no se estaría tomando en cuenta la cantidad de información suficiente para la estimación correspondiente.

Queda claro que obtener estimadores no paramétricos es un problema aún mas complejo que determinar un tamaño de vecindad adecuado y una teoría matemática congruente a la información, además de las complejidades que cada metodología trae consigo por ejemplo, un suavizador lineal es un modelo de regresión simple para cada ventana móvil por lo tanto, es necesario verificar el ajuste y revisar entre otras cosas que la regresión no esté bajo la influencia de *información aberrante* (*outliers*).

Considere la siguiente gráfica de dispersión, simbólicamente se representa el peso de cada observación en el coeficiente de correlación de acuerdo al tamaño de cada punto, las que están abiertas son aquellas que representan aportación positiva al coeficiente y por el contrario las rellenas.

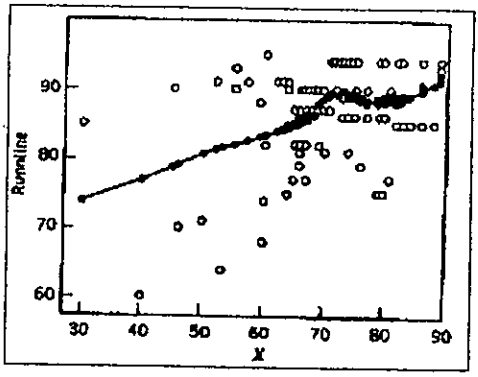
La correlación para el modelo lineal ajustado es de $R = 0.40$ sin embargo, tal significancia se debe en gran parte a los cuatro puntos ubicados en la parte inferior izquierda de la gráfica sin ellos, la correlación caería de 0.4 a 0.177 y el $p - value$ de una significancia del 0.001 a una de 0.85, de tal forma que sin estos puntos no tenemos relación alguna entre las variables X y Y .



Gráfica V.3.1.

En la construcción de un modelo NP, nos gustaría que este represente morfológicamente la relación entre las variables, que adopte una forma horizontal en caso de que no exista relación o bien, que establezca la tendencia en forma adecuada por ejemplo, construyamos una curva NP para los puntos de la dispersión anterior con vecindades de tamaño 21, 31 y 51. La siguiente gráfica es una curva NP ajustada a la misma dispersión con un suavizador

lineal con $k = 31$.



Gráfica V.3.2.

La línea no presenta una tendencia clara entre las variables excepto tal vez en valores pequeños de X y para valores mas grandes la mayoría de los puntos se localizan agrupados en líneas horizontales sin embargo, la curva presenta ciertas tendencias para dichos valores con lo cual se muestra que un suavizador lineal puede ser alterado por puntos anormales. Nuevamente, se presenta la necesidad de obtener un modelo no alterado por puntos extremos o anormales.

V.4. SUAVIZADOR LINEAL MODIFICADO.

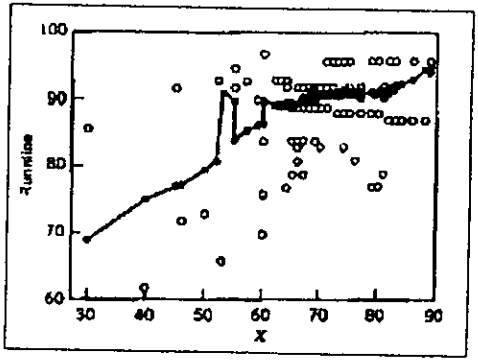
Suponga que se desea obtener un suavizador lineal que no sea alterado por outliers y/o cualquier otro tipo de información anormal, para ello debemos eliminar tal influencia en cada vecindad y este proceso debe ser cuidadoso debido a la posibilidad de *swamping* es decir, eliminar información válida érroneamente identificada como anormal. Una opción es tomar un cierto porcentaje de los puntos en cada vecindad, aquellos que parezcan mejor descritos por el modelo lineal por ejemplo, podemos utilizar el 70% de la información contenida dentro de una vecindad y después, determinar un subconjunto de puntos que minimice la función:

$$min_{\hat{\theta}} \left(\sum_{i=1}^{h} e_{(i)}^2 \right)$$

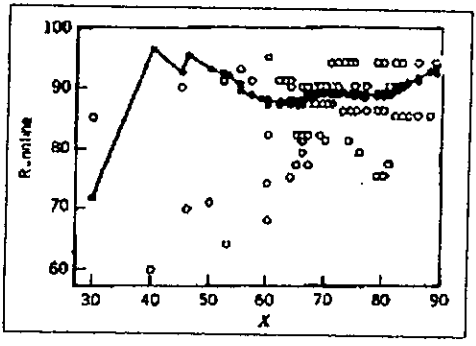
donde $e_{(1)}^2, e_{(2)}^2, ..., e_{(n)}^2$; representan los n residuales al cuadrado ordenados de menor a mayor y $h \in [0, 1]$, el porcentaje de puntos a determinar en cada vecindad de tamaño k. *Rousseeuw* y *Leroy* propusieron en 1987 la siguiente selección para h :

$$h = [n(1 - \alpha)] + 1,$$

donde $\alpha \in [0,1]$, denota la proporción de información anormal a eliminar (claramente: $\alpha + h = 1$). Si fuera posible extraer toda la información anormal sin eliminar observaciones válidas entonces obtendríamos el estimador óptimo deseado. Desafortunadamente no existe una forma práctica de obtener el mínimo de la suma de residuales al cuadrado, sino que sería necesario verificar la suma de todas las posibles combinaciones $\binom{n}{n\alpha}$ y sólo de esta forma sería posible identificar los $n\alpha$ puntos de la vecindad a remover (de acuerdo a la selección cuya suma haya sido mínima). Nuevamente se recomienda ser cuidadosos en la elección de α , los valores grandes tienen mayor probabilidad de remover información válida y los muy pequeños, mayor probabilidad de no remover información anormal.



Gráfica V.4.1.



Gráfica V.4.2.

Verifique la curvas NP de las gráficas V.4.1 y V.4.2 las cuales se ajustan a la dispersión de la gráfica V.3.1. Ambas curvas se obtuvieron por suavizamiento lineal modificado con $k = 51$. En el primer caso se removió el 30% de la información para cada vecindad ($\alpha = 0.30$), y en el segundo caso el 40%.

La diferencia entre ambas curvas es notable en los valores pequeños de X , ya que con $\alpha = 0.40$ es mas probable que los puntos anormales sean removidos en particular, aquellos ubicados en la parte inferior izquierda.

En realidad, no existe un criterio que nos ayude a discernir cuál de los modelos obtenidos es el mejor, sin embargo de acuerdo a la posición de las observaciones es posible decidir de manera visual cuál es el modelo mas razonable, en particular la curva de la gráfica V.4.1 parece la mas adecuada (salvo alrededor de $x = 55$, donde la curvatura no es lo suficientemente suave). Es interesante observar el comportamiento de ambas curvas, son escencialmente las mismas para los valores de $x \geq 65$, esto significa que los estimadores $\widehat{\mu}(x_j)$ están bien determinados para dichos valores, sin embargo para $x < 65$ no sería adecuado ajustar un modelo de regresión lineal simple antes de eliminar al menos un porcentaje de información anormal.

A manera de reflexión, considere 20 puntos que han sido ajustados por una línea recta, si todos se localizan dentro de ella estaremos de acuerdo en que no importa que tantos datos remueva ya que el modelo permanecerá inmóvil toda vez que queden al menos dos puntos distintos uno del otro. De la misma forma si existe en realidad una tendencia entre las variables, no deberá importar qué subconjunto de puntos sea el que describa al modelo, ya que este no varía demasiado con respecto al modelo descrito por los demás subconjuntos. Dicho comportamiento se verifica en el ejemplo anterior, donde las curvas estimadas para ambos casos se conservan caso contrario, en el comportamiento de las curvas mencionadas para valores de $x < 65$.

V.5. SUAVIZAMIENTO KERNEL.

Hemos descrito algunas herramientas NP que nos permiten calcularle a cada evento un estimador construido a partir de observaciones que influyen directamente en él, asumimos que no todas las observaciones deben ser contempladas para el cálculo del estimador correspondiente y fue por esta razón que buscamos una vecindad de puntos que aportara la información necesaria para una correcta estimación del evento e incluso, dentro de la misma vecindad, se removieron observaciones que alteraban dicha estimación. Es importante darse cuenta que todas las demás observaciones que queden fuera de la vecindad tendrán influencia cero en el cálculo del estimador, por lo tanto solo considera una muestra del universo de observaciones. Análogamente, dentro de la misma vecindad no todas las observaciones influyen de la misma forma y por tanto no todas deberán tener la misma importancia en el estimador calculado. El *suavizamiento kernel* es una herramienta que nos permite darle a

cada evento un peso adecuado dependiendo de su importancia e influencia en el cálculo del estimador en cuestión.

Suponga $\{(x_i, y_i)\}_{i=1}^{n}$ el conjunto de puntos recolectados para el estudio de interés y sean $\{(W_{n'i}(x))\}_{i=1}^{n'}$ los pesos asociados a cada elemento de la muestra de tamaño n' a saber, el conjunto de observaciones: $\{(x_i, y_i)\}_{i=1}^{n'}$ las cuales caen dentro de la vecindad de tamaño h. Ahora bien, dentro de esta vecindad sería posible construir una función de pesos $W_{n'i}(x)$ y que a través de un cierto parámetro c_i obliguemos a que la suma de todos los pesos dentro la vecindad:

$$\sum_{x_{min}}^{x_{max}} c_i W_{n'i}(x) = 1,$$

es claro que la intención de esta parametrización es generar una función de densidad discreta que describa la importancia de cada observación en ésa vecindad. En general, es común referirse a esta densidad como: *kernel* K ; la cual, es una función: real, contínua, acotada, simétrica (definiciones en el apéndice), y que al ser de densidad:

$$\int K(u)\, du = 1.$$

Observación. Posteriormente se analiza el caso en donde la variable independiente es en realidad un vector $x \in R^d$, con lo que el kernel K se define como una función real de variable vectorial.

Existen muchas maneras de asociar pesos a cada observación, tal vez tantas como fenómenos nos podamos imaginar, incluso en algunos casos K no será necesariamente simétrica y podrá adoptar formas variadas sin embargo, a manera de introducción trabajaremos bajo el supuesto de que el peso $W_{ji}(x_j)$ asociado a un punto (x_j, y_j), estará directamente relacionado con la distancia entre su abscisa y la respectiva al punto de interés (x_i, y_i):

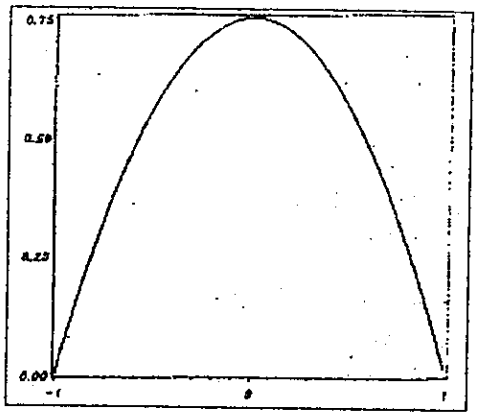
$$\frac{x_j - x_i}{h}$$

la cual es parametrizada por el tamaño de la vecindad: *ancho de banda* (h), dando como resultado una función que refleja la importancia o influencia que tiene la observación j en el cálculo del estimador del punto i :

$$K\left(\frac{x_j - x_i}{h}\right),$$

Existen funciones kernel usadas comúnmente debido a ciertas propiedades de optimalidad, una de ellas es la propuesta por Epanechnikov en 1969 (representada en la gráfica V.5.1.):

$$K(u) = \frac{3}{4}\left(1 - u^2\right) I\left(|u| \leq 1\right),$$



Gráfica V.5.1.

observe que este kernel no es diferenciable en $u = \pm 1$.

Nota : Para mayores detalles de la optimalidad de estas funciones ver: Epanechnikov (1969) y/o Bartlett (1963).

Además del K propuesto por Epanechnikov, existen muchas otras funciones que cumplen con las propiedades de un kernel, incluso es posible construir un K si se conoce la distribución de las observaciones en el intervalo de interés por esta razón, la asociación de pesos a las observaciones mencionadas es otro de los elementos que pueden ser personalizados de acuerdo a las necesidades de cada estudio.

Los sucesión de pesos $W_{n'i}(x)$, a través de suavizamiento kernel:

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

donde $K_h(u)$ es el K parametrizado por el ancho de banda. El peso $W_{n'i}(x)$ es un factor de ponderación entre cero y uno:

$$W_{n'i}(x) = \frac{K_h(x - x_i)}{\widehat{f}_h(x)}$$

donde

$$\widehat{f}_h(x) = (n')^{-1} \sum_{i=1}^{n'} K_h(x - x_i)$$

donde n' es el número de puntos dentro del ancho de banda. A la función $\widehat{f}_h(x)$ se le conoce como el *estimador de densidad kernel de Rosenblatt – Parzen*, el cual fue propuesto por dichos investigadores en 1956 y 1962, respectivamente.

Una vez asociados los pesos a cada observación, tenemos al estimador $\widehat{\mu}(x)$ que no es otra cosa mas que un promedio ponderado de las observaciones que se localizan dentro del ancho de banda de tamaño h,

$$\widehat{\mu}(x) = (n')^{-1} \sum_{i=1}^{n'} Y_i W_{n'i}(x)$$

o bien,

$$\widehat{\mu}(x) = \frac{(n')^{-1} \sum_{i=1}^{n'} Y_i K_h(x - x_i)}{(n')^{-1} \sum_{i=1}^{n'} K_h(x - x_i)}.$$

$\widehat{\mu}(x)$ es conocido como el *estimador Naradaya – Watson* debido a los investigadores que lo propusieron en 1964. Ahora bien, de este conjunto de estimadores se obtiene el: conjunto S de suavizadores kernel,

$$S = \{\widehat{\mu}(x_j) = E[Y \mid x_j],\ \forall\, j = 1, ..., n\}$$

Observación: Si dos o mas puntos comparten una misma abscisa x_i, entonces su y_i asociado será equivalente al promedio ponderado de las r ordenadas,

$$y_i = (n')^{-1} \sum_{j=1}^{r} y_{ij} W_j(x_i).$$

V.6. ASIGNACIÓN DE PESOS: EJEMPLOS.

Como se mencionó anteriormente, no existe una asignación de pesos única, veamos algunos ejemplos:

1) En ciertas aplicaciones la función de densidad marginal del parámetro independiente $f(x) = F'(x)$ es conocida, para estos casos Greblicki y Johnston (1974 y 1979, respectivamente), propusieron la siguiente asignación de pesos:

$$W_{hi}^{(1)}(x) = \frac{K_h(x - x_i)}{f(x)}.$$

A menudo nos encontramos con observaciones cuyas abscisas están separadas por distancias regulares, sin pérdida de generalidad podemos suponer en tales casos, que las abscisas de estas observaciones se extraen de una variable aleatoria uniforme en el intervalo $[0,1]$ y por lo tanto sus pesos pueden ser calculados por medio de $W_{hi}^{(1)}(x)$, donde $f = U\ [0,1]$.

2) Supongamos $\{x_i\}_{i=1}^{n}$ las abscisas correspondientes a observaciones equidistantes en $[0,1]$ cuya ubicación no presenta ningún comportamiento aleatorio. En 1972, Priestly y Chao y posteriormente Benedetti (1977), propusieron la sucesión:

$$W_{hi}^{(2)}(x) = n'\ (x_i - x_{i-1})\ K_h(x - x_i), \text{ con: } \ x_0 = 0.$$

Esta propuesta de asignación de pesos se puede verificar en términos de $W_{n'i}(x)$, en donde:

$$\widehat{f}_h(x) = [n'\ (x_i - x_{i-1})]^{-1}, \ \ \forall\, x \in [x_{i-1}, x_i].$$

3) En 1979, Gasser y Müller propusieron la versión contínua de $W_{hi}^{(2)}(x)$:

$$W_{hi}^{(3)}(x) = n \int_{S_{i-1}}^{S_i} K_h(x-u)\ du,$$

donde $x_{i-1} \leq S_{i-1} \leq x_i$ es escogido dentro del conjunto de datos ordenados del parámetro independiente X.

4) La generalización del kernel K para variables multidimensionales $x_i = (x_{i1}, x_{i2}, ..., x_{id})$:

$$K\ (u_1, ..., u_d) = \prod_{i=1}^{d} K\ (u_i),$$

para el cual la asignación de pesos $W_{ni}(x)$ se generaliza inmediatamente:

$$W_{ni}(x) = \frac{\prod_{i=1}^{d} K\ (x_j - x_{ij})}{\widehat{f}_h(x)}.$$

de igual manera el estimador kernel de Rosenblatt-Parzen utiliza el K multidimensional.

Observación. No necesariamente las sucesiones $\left\{W_{hi}^{(1)}(x)\right\}$ y $\left\{W_{hi}^{(2)}(x)\right\}$ llegan a sumar 1, sin embargo $\left\{W_{hi}^{(3)}(x)\right\}$ sí lo hace.

V.7. CASOS PARTICULARES DEL ESTIMADOR: *NARADAYA – WATSON*.

a) $\widehat{\mu}(x)$ no está definido para $\widehat{f}_h(x) = 0$, por lo que en el caso $\frac{0}{0}$ se define $\widehat{\mu}(x) = 0$.

b) A medida que disminuye el ancho de banda h, el estimador NW reproduce la información correspondiente:

$$lim_{h\to 0}\ \widehat{\mu}(x_i) = y_i$$

ya que

$$y_i = \frac{y_i K(0)}{K(0)},\ \forall\ i = 1, ..., n.$$

c) A medida que aumenta el ancho de banda h, el estimador NW asocia igual peso a todas las observaciones, quedando así, el promedio aritmético de las ordenadas dentro de la vencindad:

$$lim_{h\to\infty}\ \widehat{\mu}(x_i) = n^{-1}\sum_{i=1}^{n} y_i$$

ya que

$$n^{-1}\sum_{i=1}^{n} y_i = \frac{n^{-1}\sum\limits_{i=1}^{n} y_i K(0)}{n^{-1}\sum\limits_{i=1}^{n} K(0)},\ \forall\ i = 1, ..., n.$$

En el caso b) diremos que el estimador construido con un ancho de banda muy pequeño dará como resultado curvas *sub – suavizadas* y por el contrario, en el caso c) anchos de banda muy grandes dan como resultado curvas *sobre – suavizadas*.

Proposición. Suponiendo la naturaleza estocástica del modelo (ver apéndice), consideremos X como la variable independiente y,

a) $\int |K(u)|\, du < \infty$

b) $lim_{|u|\to\infty}\ uK(u) = 0$

c) $EY^2 < \infty$

d) $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$.

Entonces, para todo punto en el dominio de $m(x)$, $f(x)$ y $\sigma^2(x)$, con $f(x) > 0$,

$$n^{-1}\sum_{i=1}^{n} W_{hi}(x) Y_i \to^p \mu(x).$$

Esta proposición establece que el suavizador kernel converge en probabilidad a $\mu(x)$ (ver apéndice). La velocidad de esta convergencia dependerá en gran parte de la naturaleza misma del fenómeno, para tales efectos consideremos:

$$d_M(x,h) = E\left[\widehat{\mu_h(x)} - \mu(x)\right]^2$$

como el error cuadrático medio, herramienta en la que $\forall\ x$ dicha convergencia puede ser cuantificada (teorema de Gasser y Müller).

CAPÍTULO VI: APLICACIONES

VI.1. INTRODUCCIÓN.

Una vez descrita la teoría, resulta fundamental establecer un mecanismo claro que permita su aplicación utilizando herramientas accesibles tales como una calculadora científica o una hoja de cálculo. Antes de entrar al detalle del algoritmo, debemos tomar en cuenta que al momento de llevar a la práctica una teoría es necesario considerar que alrededor de cualquier fenómeno existen innumerables factores externos que pueden eventualmente alterar el modelo, por esto es indispensable que antes de adentrarnos en la construcción de un modelo matemático se establezcan elementos de control para evitar sesgos y desviaciones como consecuencia de dichos factores. Considere el siguiente ejemplo.

VI.2. CASO #1: FIDEICOMISO PARA LA LIQUIDACIÓN AL SUBSIDIO DE LA TORTILLA.

En México, el Fideicomiso para la Liquidación al Subsidio de la Tortilla (FIDELIST), es una institución sectorizada a la Secretaría de Desarrollo Social (SEDESOL), cuyo principal objetivo es otorgar un kilogramo diario de tortilla de manera gratuita a las familias mexicanas que viven en extrema pobreza. Para lograr su objetivo, el fideicomiso atiende a 1.24 millones de familias en 850 localidades en las 32 entidades federativas.

Una de las tareas fundamentales del FIDELIST es la de discernir cuáles han de ser las condiciones para que una familia forme parte del conjunto de beneficiados por el fideicomiso. Para tales fines, una de las herramientas utilizadas es la encuesta; esta permite analizar el contexto: social, económico y cultural del entrevistado y su familia, y de esta forma, discernir si es que este debe o no recibir el beneficio correspondiente.

VI.2.1. PROBLEMÁTICA

Actualmente el FIDELIST no cuenta con una herramienta matemática que permita describir la condición alimenticia de la población a partir de las variables consideradas en la

encuesta tales como: el ingreso económico, las condiciones de la vivienda y de la localidad, el nivel de estudios, el hacinamiento, etc. Idealmente la encuesta debería proporcionar elementos suficientes con los cuales se pudiera establecer cuáles son las variables más importantes a considerar. Sin embargo, la importancia de una variable no es algo que se refleje al momento de la entrevista o del vaciado de la información, en realidad el peso específico de una variable es una tarea que debe ser ponderada respecto de las demás e incluso, como ejercicio de valoración de dichas variables está el reconocer si dentro de las más importantes existen algunas muy correlacionadas entre sí, de manera que sea posible desechar algunas a fin de construir un modelo con el menor número de parámetros posible.

De lo mencionado en el párrafo anterior debemos recalcar lo siguiente: un primer estudio previo al de la construcción del modelo NP es fundamental: *la selección de las variables*; sin duda es necesario obtener información de cada uno de los parámetros que creamos que están involucrados en el fenómeno y analizar de qué manera podrán evaluarse (sobre todo cuando dichos parámetros no son fácilmente cuantificables como por ejemplo, en fenomenos de tipo psicológico en el que variables como: el stress, la angustia y la depresión, deben ser cuantificadas).

VI.2.2. OBJETIVO DEL ESTUDIO

Obtener el modelo NP asociado y mediante una muestra representativa de la población, establecer cuáles son las condiciones (variables independientes), más importantes que permiten al entrevistado recibir el apoyo alimenticio (variable respuesta).

VI.2.3. SELECCIÓN DE VARIABLES

El FIDELIST cuenta actualmente con un registro de información obtenida a través de encuestas, este registro es únicamente de la población que ha sido aprobada y que actualmente recibe un kilogramo de tortilla diario. De esta población aprobada, se eligió una muestra de tamaño 149, cada elemento fue sometido a encuesta respondiendo todas y cada una de las 114 preguntas que ahí se elaboran estas preguntas en un principio pretendieron abarcar todos y cada uno de los factores que, según los encargados de elaborarla, pensaron que estarían involucrados en la condición alimenticia de la población. Sin embargo, suponiendo que la información proporcionada por los entrevistados es útil, veraz y sin omisiones; resulta fundamental establecer si todos estos parámetros deben o no ser incluidos en el modelo, para aclarar este punto veamos el siguiente:

Ejemplo. Supongamos que se desea establecer de qué manera repercute la falta de alimentación en el aprovechamiento académico de los estudiantes de primaria del estado de Aguascalientes, para efectos de obtener la información adecuada se desarrolló un minucioso estudio de muestreo en el cual se entrevista a algunos estudiantes de primaria del estado, pero qué pasaría si se elaborara alguna pregunta como: ¿vives actualmente en el estado de Aguascalientes?; sin duda es importante que el encuestado viva en el estado ya que el estudio pretende analizar a habitantes del mismo sin embargo, considerando la población en estudio sería posible suponer que así es. Incluso si efectivamente se elaborara ésa pregunta y todos o casi todos responden afirmativamente, la pregunta no aportará información relevante ni será un factor fundamental que permita relacionar la alimentación del niño con su aprovechamiento en la escuela.

¿Qué preguntas de la encuesta son realmente decisivas?

Se realizó una encuesta nacional, los elementos de la misma, fueron integrantes de las familias que ya reciben el beneficio, de ésta manera sería posible obtener un perfil sociodemográfico de los beneficiarios actuales del fideicomiso y de ésta manera, proponer los criterios para futuros beneficiarios. Dicha encuesta contenía una variable relacionada con el gasto promedio semanal en alimentación (ALIMENTO), esta fue considerada como el parámetro dependiente en un método *backward* de regresión lineal múltiple en *Statistica*, tomando la muestra mencionada de tamaño 149 para determinar las variables con mayor influencia en la variable respuesta.

Resultados: Fueron removidas 109 variables contenidas originalmente en la encuesta y de las cuales se tenía registro en la base de datos correspondiente, únicamente 4 variables independientes describieron a la variable ALIMENTO de manera casi total con un coeficiente de determinación de 0.9954, a saber las variables mas importantes:

a) SERVICIO, indicador de los servicios de agua y electricidad; se asignaba un valor dependiendo del número de tomas de agua y número de focos en la vivienda.

b) TRANSPOR, indicador del gasto familiar en transporte público.

c) OTROS_GA, indicador de gastos familiares por motivos ajenos al transporte, la alimentación y los servicios públicos.

d) GASTO_TO, indicador del monto promedio de gastos familiares.

Nota : El *backward* de regresión lineal no es un método exclusivo para la selección de las variables mas significativas, por ejemplo existen técnicas de Análisis Multivariado como la

de Componentes Principales que proporciona de igual forma, los parámetros mas influyentes en el modelo.

VI.2.4. CONCLUSIÓN.

SERVICIO, TRANSPOR, OTROS_GA y GASTO_TO; serán las variables independientes que expliquen al parámetro respuesta: ALIMENTO.

Una vez definidas las variables involucradas, se procede a la construcción del modelo NP asociado al fenómeno, para tales fines se detallará un algoritmo general de aplicación.

Importante : En el ejemplo anterior, trabajamos bajo dos supuestos que no siempre vamos a poder asegurar:

1) Las respuestas obtenidas a través de la encuesta (o el instrumento de medición correspondiente), son fácilmente cuantificables.

2) Las respuestas obtenidas no dependieron de factores externos, por ejemplo: el número de focos, tomas de agua, número de personas que habitan en la vivienda, gasto familiar promedio en transporte, etc.; no varían en relación del estado anímico del entrevistado ni de la angustia que este haya experimentado al mometo de la misma.

VI. 3. ALGORITMO GENERAL DE APLICACIÓN.

VI.3.1. OBJETIVO.

Obtener el estimador $\widehat{\mu(x)}$ de Naradaya-Watson (NW), a partir de información recolectada:

$$\{(\overline{x_i})\}_{i=1}^{n}, \text{ con } \overline{x_i} \in R^{d+1}$$

a fin de construir el modelo NP:

$$\hat{Y} = \widehat{\mu(\overline{x})} + \varepsilon_i$$

a fin de establecer el conjunto de suavizadores S descrito en la sección V.2:

$$S = \{E[Y \mid x_i] \mid x_i \in (a,b), i = 1, ..., n\}.$$

VI.3.2. DESCRIPCIÓN.

Se desea construir un modelo de regresión NP que describa a la variable respuesta X_{d+1}, a partir de la información obtenida en las variables independientes: $X_1, X_2, ..., X_d$.

Sea K_i el kernel asociado a la $i-$*ésima* variable independiente con $i = \overline{1,d}$ y sea x_{ij} la respuesta del $i-$*ésimo* elemento de la muestra a la $j-$*ésima* variable de esta manera, se obtiene una matriz de registro de tamaño $n \times (d+1)$:

$$X = \begin{bmatrix} x_{11}, x_{12}, ..., x_{1d}, x_{1d+1} \\ x_{21}, x_{22}, ..., x_{2d}, x_{2d+1} \\ \vdots \\ x_{n1}, x_{n2}, ..., x_{nd}, x_{nd+1} \end{bmatrix}$$

PASO 0. Proponer:

$$h = [h_1, h_2, ..., h_d]$$

con: h_i , el ancho de banda para el $i-$*ésimo* parámetro independiente.

PASO 1. Se recomienda ordenar las observaciones en forma ascendente respecto a su primera variable independiente como se muestra en la matriz X (esto con el fin de localizar fácilmente los puntos en caso de que se requiera reemplazar alguna de las observaciones), con:

$$x_k = [x_{k1}, x_{k2}, ..., x_{kd}]$$

PASO 2. Hacer $k = 0$ y $S = \emptyset$.

PASO 3. Hacer $k = k + 1$.

PASO 4. Determine:

$$K_{h_i}(u) = h_i^{-1} K_i\left(\frac{u}{h_i}\right);\ \forall\ i = \overline{1,d}$$

kernel parametrizado por el ancho de banda propuesto en 0 para la variable independiente X_i, en caso de tener mas de una variable independiente, aplique lo siguiente para el kernel K multidimensional:

$$K(x_k - x_i) = \prod_{j=1}^{d} K_{h_j}(x_{kj} - x_{ij}),\ \forall\ i = 1, ..., n'$$

Nota : Para efectos de esta construcción del modelo NP, se supone que el kernel K_i describe las distribución de la $i-ésima$ variable en todo su dominio.

PASO 5. Obtenga el estimador de densidad de $Rosenblatt-Parzen$:

$$\widehat{f_{h_i}}(x_k) = \left(n'\right)^{-1} * \sum_{i=1}^{n'} K_{h_j}(x_{kj} - x_{ij}),\ \forall\, j = 1, ..., d$$

PASO 6. Obtener el peso del $i-ésimo$ elemento de la vecindad de tamaño n' en la la construcción del estimador $\widehat{\mu(x_k)}$:

$$W_{ij}(x_k) = \frac{K_{h_j}(x_{kj} - x_{ij})}{\widehat{f_{h_i}}(x_k)},\ \forall\ i = \overline{1, n'}.$$

PASO 7. Por último, obtenga el estimador de $Naradaya-Watson$ para x_k :

$$\widehat{\mu(x_k)} = (n')^{-1} * \sum_{i=1}^{n'} X_{id+1} W_{ij}(x_k),\ \forall\, j = 1, ..., d.$$

PASO 8. Verifique si el ancho de banda h propuesto es adecuado en caso contrario modifíquelo (ir a 0), de lo contrario ir al paso 9.

PASO 9. Hacer $S = S \cup \left\{\widehat{\mu(x_k)}\right\}$.

PASO 10. Si $k < n$ ir al paso 3, de lo contrario: si $k = n$, terminar el conjunto S de suavizadores kernel:

$$S = \left\{\widehat{\mu(x_k)} = E\left[Y \mid x_k\right] \mid x_k \in D_1 \times D_2 \times \cdots \times D_d\right\},$$

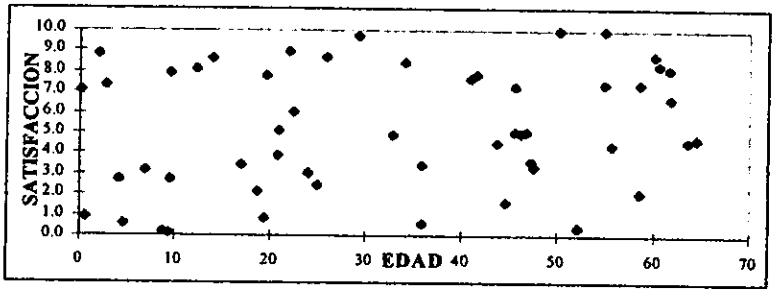
donde:

$$D_j = \left\{x_{ij} \mid x_{ij} \in \left[min_{i=\overline{1,n}}\left\{x_{ij},\ j = \overline{1,d}\right\}, max_{i=\overline{1,n}}\left\{x_{ij},\ j = \overline{1,d}\right\}\right]\right\}$$

VI.4. CASO #2: ESTUDIO DE SATISFACCIÓN.

VI.4.1. PLANTEAMIENTO.

Una productora de lacteos pretende innovar en el mercado con la línea de productos: *Super Yogo*; los cuales, recalcifican poderosamente los huesos, además de blanquear la dentadura y fortalecer la osamenta. Se desea conocer el perfil demográfico y de hábitos de consumo de aquellos individuos que experimentan una satisfacción especial a los productos lacteos y no lacteos que igualmente promuevan la recalcificación, balnqueo y fortalecimiento de los huesos tales como: pastillas, emulsiones, etc. Similares que consume actualmente. A continuación, la gráfica VI.4.1 presenta los resultados de la encuesta realizada a un grupo de consumidores de productos calcificantes de todas las edades.



Gráfica VI.4.1.

Paso a paso con el algoritmo general de aplicación descrito en la sección VI.3 se obtendrá el conjunto de suavizadores S y para ello se puede utilizar una hoja de cálculo: Excel (en cualquiera de sus versiones).

VI.4.2. DESCRIPCIÓN.

Paso 0. Debido a que la variable respuesta depende únicamente de una variable independiente, se propone únicamente un ancho de banda: $h = 20$.

Paso 1. Se ordenan las observaciones en orden ascendente respecto a la variable independiente: EDAD.

#	Edad	Satisf.	#	Edad	Satisf.	#	Edad	Satisf.	#	Edad	Satisf.	#	Edad	Satisf.	#	Edad	Satisf.
1	0	7.1	21	17	3.4	41	34	8.4	61	52	0.4	81	66	6.4	101	84	9.0
2	0	7.1	22	19	2.1	42	36	3.5	62	55	7.4	82	66	6.4	102	87	5.8
3	1	0.9	23	19	0.8	43	36	3.5	63	55	7.4	83	67	3.3	103	88	5.5
4	2	8.8	24	20	7.8	44	36	0.6	64	55	10.0	84	67	3.3	104	88	5.5
5	3	7.4	25	20	7.8	45	41	7.6	65	55	10.0	85	69	6.5	105	90	6.8
6	3	7.4	26	21	3.9	46	42	7.9	66	56	4.4	86	70	3.9	106	90	6.8
7	4	2.7	27	21	3.9	47	42	7.9	67	56	4.4	87	72	8.5	107	90	9.1
8	5	0.6	28	21	5.1	48	44	4.6	68	59	7.4	88	72	8.5	108	90	9.1
9	7	3.2	29	21	5.1	49	44	4.6	69	59	2.1	89	72	7.2	109	92	5.7
10	7	3.2	30	22	9.0	50	45	1.6	70	59	2.1	90	74	7.3	110	92	5.7
11	9	0.2	31	22	9.0	51	45	1.6	71	60	8.7	91	74	7.3	111	92	5.6
12	9	0.1	32	23	6.1	52	46	7.3	72	60	8.7	92	75	0.5	112	92	5.6
13	9	0.1	33	23	6.1	53	46	5.1	73	61	8.3	93	75	3.4	113	93	1.5
14	9	7.9	34	24	3.1	54	46	5.0	74	62	8.1	94	75	3.4	114	93	1.5
15	9	2.7	35	25	2.5	55	46	5.0	75	62	6.7	95	75	5.6	115	94	7.4
16	12	8.1	36	26	8.7	56	47	5.0	76	62	6.7	96	76	8.6	116	94	7.4
17	12	8.1	37	26	8.7	57	47	3.7	77	64	4.6	97	78	6.3	117	98	1.6
18	14	8.6	38	29	9.7	58	48	3.4	78	64	4.6	98	81	1.4	118	98	4.3
19	14	8.6	39	33	4.9	59	50	10.0	79	65	4.7	99	81	1.4	119	99	5.2
20	17	3.4	40	33	4.9	60	52	0.4	80	65	4.7	100	81	9.0	120	100	3.0

Tabla VI.4.2.

Paso 2. Sea $k = 0$ y $S = \emptyset$.

Paso 3. Sea $k = 1$.

Paso 4. Usando el kernel de Epanechnikov, se obtiene para todos los puntos dentro de la vecindad simétrica de tamaño h el $K_h(u)$ asociado. En Excel se pueden calcular todos los valores (por ejemplo), escribiendo la siguiente instrucción en la celda $E5$:.

$$= SI(ABS(E\$4-\$B5) < (\$B\$376/2), (1/\$B\$376)*(3/4*(1-((E\$4-\$B5)/\$B\$376)^2)), " ")$$

esto es, colocando todos las edades en orden ascendente en la columna B a partir de la fila 5 y ésa misma matriz transpuesta en la fila 4 a partir de la columna E; dejando fija la celda que contiene al ancho de banda (en este caso $\$B\$376$). Arrastando la fórmula desde la celda $E5$ hasta la $DT124$, se obtienen los valores: $K_{h_i}(u)$, $\forall\, i = 1, ..., n$.

Paso 5. Observe que columna a columna se han ordenado las abscisas de cada una de las observaciones obtenidas y en la columna (matriz) $E5 : E124$ están contenidos los kernel

parametrizados de los n' puntos contenidos en el ancho de banda propuesto; por esta razón el estimador RP $\widehat{f_h(x_k)}$ se calcula de la siguiente manera $\forall\ k = 1, ..., n$. En la celda $E128$ escriba lo siguiente:

$$= \frac{SUMA(E5 : E124)}{CONTAR.SI(E5 : E124, " > 0")}$$

Arrastre horizantalmente esta formula hacia la derecha hasta la celda $DT128$, de esta manera usted habrá obtenido los estimadores RP para toda observación en la muestra.

Paso 6. Colóquese en la celda $E133$ y escriba lo siguiente:

$$= SI(Y(E5 > 0, E5 < 1), E5/E\$128, " ")$$

Arrastre la fórmula a lo largo de la matriz $E133 : DT252$ de esta manera usted habrá obtenido la matriz de pesos $W_{ij}(x_k)$ para todas las observaciones en la muestra.

Paso 7.

a) Coloque en la columna C a partir de la fila 5 los resultados obtenidos en la encuesta de la variable respuesta: SATISFACCIÓN; que corresponda a la ordenada de la abscisa correspondiente de la columna B.

b) Escriba en la celda $E258$ lo siguiente:

$$= SI(E133 <> " ", E133 * C\$5, " ")$$

c) Arrastre la formula a lo largo de la matriz $E258 : DT377$.

d) Colóquese en la celda $E380$ y escriba lo siguiente:

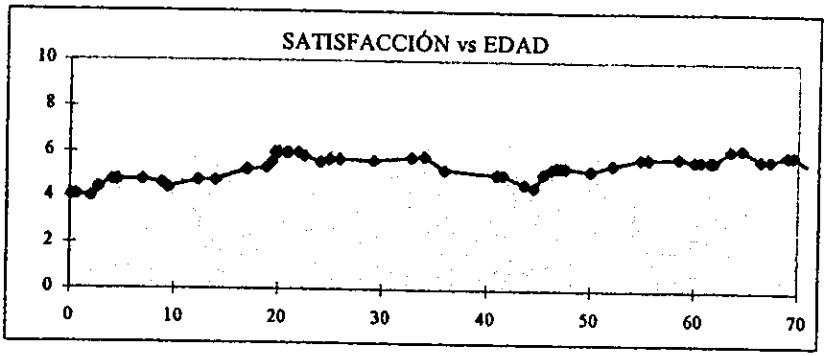
$$= \frac{SUMA(E258 : E377)}{CONTAR.SI(E5 : E124, " > 0")}$$

arrastre esta fórmula horizontalemente hacia la derecha hasta la celda $DT380$ con esto se han obtenido los estimadores NW para toda $k = 1, ..., n$; de esta manera se ha generado el conjunto de suavizadores kernel:

$$S = \{E380, D380, ..., DT380\}$$

Paso 8. Cuando se tiene un fenómeno con únicamente una variable independiente y una respuesta se recomienda graficar las abscisas contra los estimadores contenidos en el conjunto S para verificar si es que efectivamente los suavizadores estimados describen la dispersión

original, además de observar si es que existe evidencia de un sub o sobre-suavizamiento. Con $h = 20$:



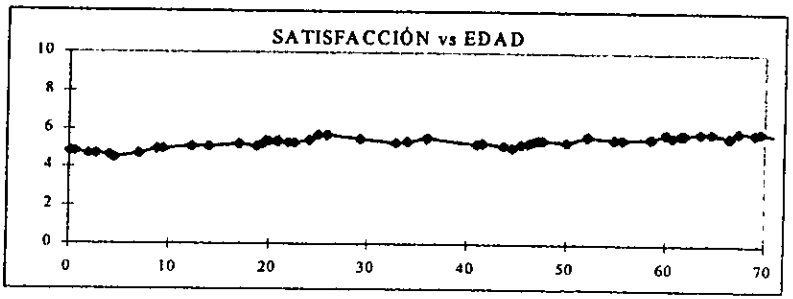
Gráfica VI.4.3.

a) Observe el comportamiento de las observaciones en la gráfica VI.4.1 respecto al comportamiento de sus suavizadores en la gráfica VI.4.3.

b) Observe además que esta gráfica es una aproximación discreta de un comportamiento que es en realidad continuo, ya que el análisis realizado no debería haberse hecho únicamente para los abscisas de las observaciones de la encuesta sino: $\forall\, x \in [0, 70]$. Sin embargo esta representación nos ayuda a ver claramente la construcción del modelo.

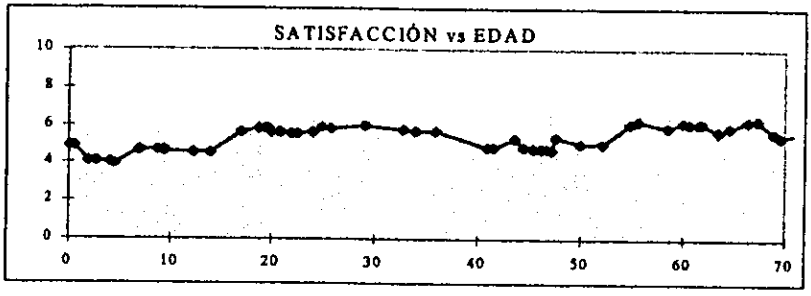
VI.4.3. COMPARATIVOS CON DISTINTOS ANCHOS DE BANDA.

Veamos ahora el comportamiento de los suavizadores obtenidos si empleamos rango de edades (anchos de banda), mayores o menores. Sea $h = 30$:



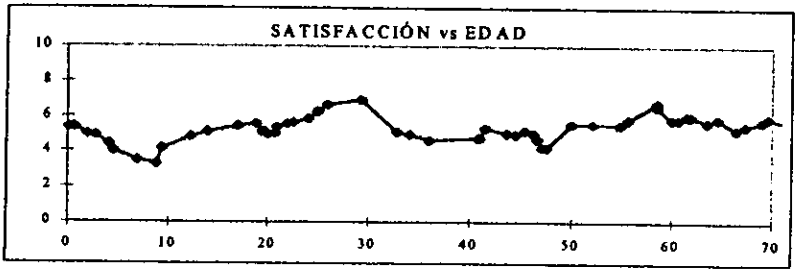
Gráfica VI.4.4.

En esta última gráfica se observa un sobre-suavizamiento respecto a la $VI.4.3$, a continuación veremos el comportamiento de los suavizadores del conjunto S con un rango de edades: $h = 15$.



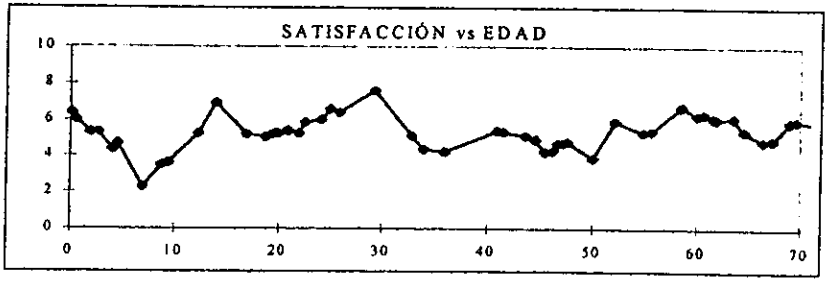
Gráfica VI.4.5.

Con un ancho de banda menor al propuesto se puede observar que este último también estaba sobre-suavizado, tratemos ahora con rango de edades: $h = 10$.



Gráfica VI.4.6.

Esta última gráfica refleja muy claramente el comportamiento de las observaciones en cada uno de los intervalos involucrados, consideremos un rango de edades menor, sea $h = 7.5$:



Gráfica VI.4.7.

Particularmente la gráfica VI.4.6 proporciona a cada rango de edades el comportamiento de la muestra y en general, si es que tomamos una muestra lo suficientemente representativa, el comportamiento de la población consumidora en general. La gráfica VI.4.7 presenta un ligero grado de sub-suavizamiento y esto sucede por haber tomado un rango de edades pequeño.

VI.4.4. CONCLUSIÓN.

En particular un ancho de banda de 10 (ó incluso hasta 7.5), genera un suavizamiento adecuado para la observaciones obtenidas, en general al analizar estas gráficas la empresa productora de lacteos puede optar por lanzar al mercado productos calcificantes para edades en que estos productos tienen mejor aceptación: personas de entre 27 y 32 años y de 55 años en adelante, aunque en realidad se observa que en general dichos productos tienen una aceptación moderada en todos los rangos de edades.

VI.5. APLICACIÓN DEL CASO #1: FIDELIST.

VI.5.1. CONSIDERACIONES DE APLICACIÓN EN MODELOS NP MULTIDIMENSIONALES.

Como vimos en la sección VI.2.4 existen 4 parámetros importantes que han permitido describir la condición alimenticia de la población encuestada casi en su totalidad, en esta ocasión usaremos la regresión NP para encontrar parámetros de decisión.

Observe que este es un caso en el cual están involucradas mas de una variable independiente, para tales efectos dichas variables deberán ser estudiadas por separado:

a) Se pueden establecer funciones kernel distintas para cada variable.

b) El ancho de banda debe ser evaluado en cada parámetro independiente.

c) Para cada variable se obtiene un conjunto $S_j, \forall j = 1, ..., d$ de suavizadores kernel los cuales permitirán evaluar a los futuros elementos de la población en cada uno de los parámetros independientemente de los restantes.

Observe que esta es otra de las ventajas de los modelos NP, al presentar el resultado final se hará un comparativo de los suavizadores obtenidos por NP, en relación a cómo fue descrita la misma información por un modelo de regresión lineal múltiple.

Observación. Es claro que en este caso el modelo no podrá apreciarse gráficamente como se hizo en la sección VI.4.3, en donde se graficaban los suavizadores obtenidos, al estar evaluando 4 variables independientes la gráfica equivalente estaría dada por suavizadores multidimensionales, en donde si para la $i-ésima$ observación la variable X_j con $j = 1,...,d$; se obtuvo un promedio ponderado:

$$s_i \in S_j \text{ , con: } i = 1,...,n$$

entonces el suavizador equivalente sería de la forma:

$$s_i = (s_{i1}, s_{i2}, s_{i3}, s_{i4}), \text{ con: } i = 1,...,n.$$

VI.5.2. OBJETIVO.

Determinar el conjunto de suavizadores kernel para cada uno de los parámetros a través del algoritmo general de aplicación descrito en la sección VI.3, esto es:

a) Se procede de manera equivalente a lo detallado en la sección VI.4.2.

b) Se utilizará el kernel de Epanechnikov para todos los parámetros (evaluando el ancho de banda para cada caso).

A fin de comprobar la validez de los resultados, a continuación se presentan los datos de la encuesta:

#	SERVICIO	TRANSPOR	OTROS_GA	GASTO_TO	ALIMENTO
1	0	0	0	400	400
2	0	0	0	800	800
3	0	0	0	0	0
4	0	20	0	20	0
5	0	0	0	800	800
6	0	0	0	300	300
7	0	0	0	500	500
8	0	0	0	0	0
9	0	80	300	1180	800
10	0	50	0	450	400
11	0	0	0	600	600
12	0	0	0	100	100
13	0	0	20	340	320
14	0	100	0	700	600
15	0	360	0	1360	1000
16	0	130	0	530	400
17	0	0	0	650	650
18	20	360	200	1580	1000
19	20	120	0	340	200
20	35	0	150	985	800
21	40	110	0	1350	1200
22	45	50	0	245	150
23	50	0	530	1060	480
24	50	100	0	950	800
25	60	0	0	860	800
26	60	0	0	210	150
27	60	80	740	1480	600
28	60	50	0	260	150
29	60	50	0	760	650
30	60	0	0	900	840

#	SERVICIO	TRANSPOR	OTROS_GA	GASTO_TO	ALIMENTO
31	60	0	0	1180	1120
32	60	20	680	1360	600
33	60	0	50	910	800
34	60	100	0	1160	1000
35	60	250	0	910	600
36	65	720	0	785	0
37	70	50	0	680	560
38	70	100	100	870	600
39	70	280	300	1450	800
40	75	0	400	1275	800
41	75	80	0	1555	1400
42	80	0	0	730	650
43	80	100	0	980	800
44	80	80	120	1080	800
45	80	100	0	780	600
46	80	0	0	200	120
47	80	200	120	1400	1000
48	80	240	50	1370	1000
49	80	0	600	1480	800
50	80	100	0	780	600
51	90	0	0	1290	1200
52	90	160	0	1450	1200
53	90	0	0	1290	1200
54	90	180	0	770	500
55	90	0	0	290	200
56	90	240	0	1230	900
57	90	20	120	830	600
58	90	0	0	1090	1000
59	90	30	0	720	600
60	90	0	60	950	800

Tabla VI.5.3.1.

#	SERVICIO	TRANSPOR	OTROS GA	GASTO TO	ALIMENTO
61	90	56	0	546	400
62	90	0	60	550	400
63	90	100	40	1230	1000
64	90	0	80	770	600
65	95	80	0	1375	1200
66	100	240	0	1540	1200
67	100	40	0	640	500
68	100	120	0	620	400
69	100	60	0	960	800
70	100	0	0	903	803
71	100	0	4800	5500	600
72	100	0	400	1100	600
73	100	280	200	2580	2000
74	100	0	0	300	200
75	100	50	0	1250	1100
76	100	50	0	950	800
77	100	50	0	950	800
78	100	250	0	650	300
79	100	0	0	500	400
80	110	80	0	590	400
81	110	150	100	960	600
82	115	45	0	1160	1000
83	115	0	80	1645	1100
84	120	240	0	1680	1320
85	120	0	0	920	800
86	120	200	0	1520	1200
87	120	280	0	2000	1600
88	120	160	0	1680	1400
89	120	0	250	2370	2000
90	120	100	80	1000	700

#	SERVICIO	TRANSPOR	OTROS GA	GASTO TO	ALIMENTO
91	120	20	50	990	800
92	120	80	240	1240	800
93	125	240	0	1165	800
94	125	120	0	1245	1000
95	125	270	0	1395	1000
96	130	220	60	1210	800
97	135	120	200	1455	1000
98	140	150	0	1290	1000
99	140	150	0	640	350
100	140	0	50	990	800
101	140	150	0	1090	800
102	143	0	0	943	800
103	150	300	0	1250	800
104	150	228	1178	2356	800
105	150	100	280	1730	1200
106	150	240	50	1840	1400
107	150	120	50	920	600
108	150	1000	50	2000	800
109	155	0	120	1075	800
110	155	400	250	1705	900
111	158	240	0	348	150
112	170	100	40	810	500
113	170	40	0	1210	1000
114	170	508	0	1678	1000
115	170	200	0	1570	1200
116	170	80	0	1450	1200
117	170	0	0	970	800
118	170	168	200	706	168
119	175	0	250	1925	1500
120	177	0	0	777	600

*Tabla VI.*5.3.1. (*continuación*)

#	SERVICIO	TRANSPOR	OTROS GA	GASTO TO	ALIMENTO
121	180	180	0	510	150
122	180	120	120	1620	1200
123	180	0	0	1380	1200
124	180	0	50	1430	1200
125	190	300	240	1830	1100
126	190	144	600	1734	800
127	200	150	0	2350	2000
128	200	80	640	2520	1600
129	200	150	0	1150	800
130	200	150	0	1150	800
131	200	550	280	2030	1000
132	200	0	0	600	400
133	210	300	0	1410	900
134	220	120	0	2340	2000
135	220	0	0	820	600
136	225	170	0	1195	800
137	230	240	80	1050	500
138	240	160	80	1480	1000
139	250	150	0	1000	600
140	260	328	0	908	320
141	270	250	220	2740	2000
142	280	0	80	1560	1200
143	310	340	13343	14993	1000
144	310	0	80	1590	1200
145	320	200	0	1520	1000
146	340	0	0	1940	1600
147	360	480	0	1040	200
148	500	100	0	1800	1200
149	650	40	0	1690	1000

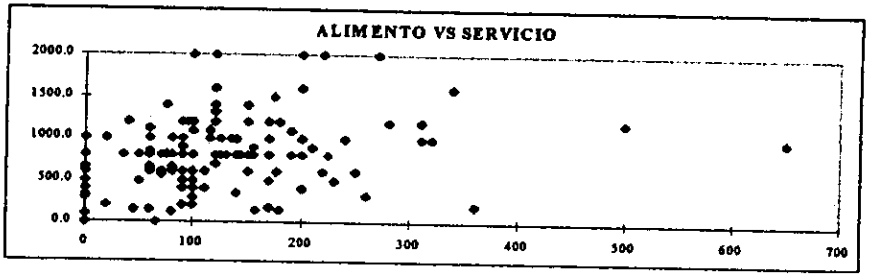
*Tabla VI.*5.3.1. (*continuación*)

Importante. Observe que a diferencia del ejemplo teórico, cuando trabajamos con datos reales se tienen inconsistencias en la información y casos de información aberrante. Por

ejemplo, en la gráfica VI.4.3.2, tenemos casos de individuos cuyo gasto en alimentación es cero y por lo tanto esta es información que debe ser revisada. También en la gráfica VI.4.3.6, existen casos de información aberrante, por ejemplo un caso en el que se gastan casi 500 pesos en transporte y menos de 100 en alimento. Es posible que este tipo de información pase de largo, pero sin duda esto afectaría la estimación de los suavizadores por lo que es muy importante analizar y validar cuidadosamente la información que nos sea proporcionada.

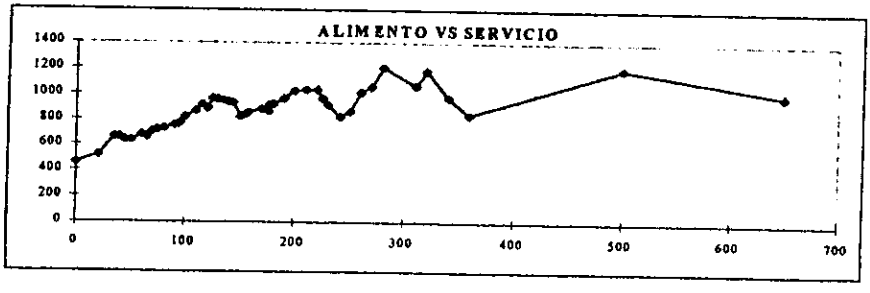
VI.5.3. DESCRIPCIÓN.

Comencemos por la primera variable independiente: SERVICIO, este parámetro contra la condición alimenticia se obtiene la siguiente gráfica:



Gráfica VI.5.3.2.

Observe en la gráfica VI.5.3.2 que existen abscisas para las cuales corresponde mas de una ordenada, en estos casos el suavizador asociado será el promedio pondeardo de las ordenadas y_i tal como se detalla en la sección V.V. A través de la metodología descrita en la sección VI.5.2 se obtiene un ancho de banda: $h = 50$.



Gráfica VI.5.3.3.

Observe el comportamiento de la gráfica VI.4.3.3 respecto de la dispersión asociada. Por otra parte, el conjunto de suavizadores kernel: $S_1, ALIMENTO\ (estimado)$.

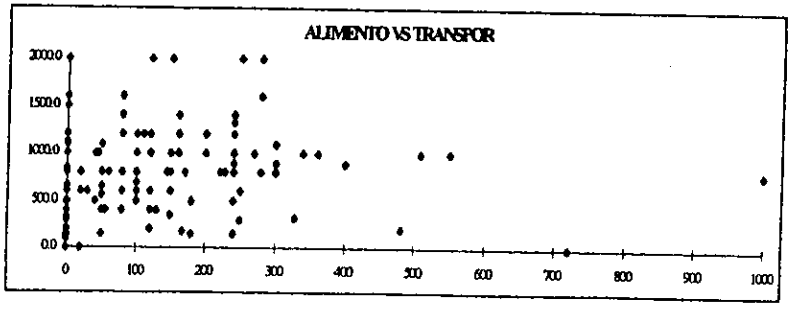
SERVICIO	ALIMENTO (estimado)	SERVICIO	ALIMENTO (estimado)
0	464.5610278	155	838.2316136
20	520.177482	158	851.3370235
35	665.9787557	170	876.0496663
40	664.8356511	175	863.2369969
45	635.3168568	177	915.606616
50	635.1458886	180	925.8665156
60	677.5680516	190	964.4466929
65	665.1031488	200	1026.28702
70	705.7757058	210	1030.222956
75	722.1649753	220	1030.583874
80	731.4549082	225	966.8393782
90	753.6430726	230	917.2519084
95	770.0325493	240	824.6929134
100	824.9016705	250	862.4347826
110	864.6118203	260	1013.913043
115	914.0041801	270	1054.042553
120	892.7687498	280	1210.285714
125	966.6319922	310	1067.567568
130	958.2987768	320	1185.106383
135	951.0078688	340	973.1343284
140	935.4018406	360	839.1304348
143	929.6596034	500	1200
150	820.2178951	650	1000

Tabla VI.5.3.4.

Estos son los parámetros estimados que definen el gasto en alimentación respecto de los servicios de drenaje, electricidad, etc. Por ejemplo, de acuerdo a la información obtenida en la encuesta, una familia que obtuvo un indicador de 155 en servicios públicos gasta aproximadamente $838.23 al mes en alimentación.

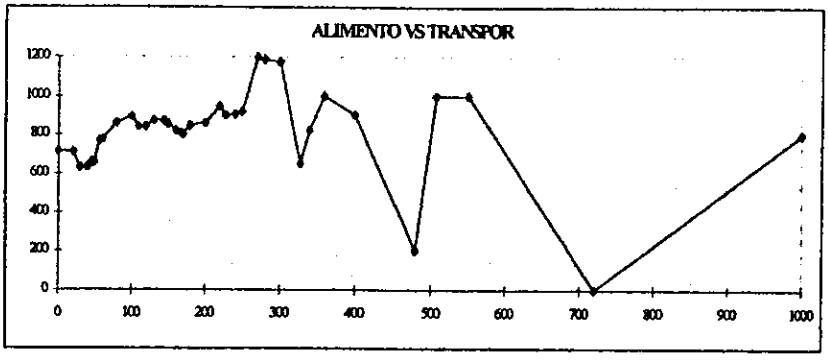
Observe que la desviación de estos resultados dependerá de si la muestra fue tomada correctamente, si los encuestados contestaron verazmente y en caso de querer ampliar este margen, del presupuesto con el que cuente el fideicomiso para el ejercicio correspondiente.

La siguiente variable a considerar es la relacionada con el gasto familiar en transporte público: TRANSPOR, a continuación, la dispersión de dicha variable contra la respuesta.



Gráfica VI.5.3.5.

Un rango de gasto en transporte adecuado: $h = 50$.



Gráfica VI.5.3.6.

Observe que aún en los casos en que se han cometido errores en el vaciado de la información el modelo únicamente se modifica en esa región y no en toda su trayectoria como ocurre normalmente con otras regresiones. Observe además que el ancho de banda no tiene que ser necesariamente el mismo para todas las variables.

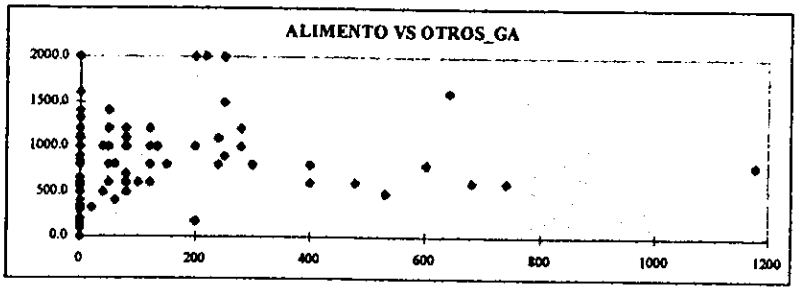
En este caso el conjunto de suavizadores: S_2,

TRANSPOR	ALIMENTO (estimado)	TRANSPOR	ALIMENTO (estimado)
0	714	180	848
20	713	200	862
30	632	220	944
40	632	228	902
45	661	240	907
50	660	250	922
56	768	270	1197
60	776	280	1190
80	862	300	1177
100	891	328	650
110	841	340	823
120	840	360	1000
130	872	400	900
144	874	480	200
150	857	508	1000
160	821	550	1000
168	808	720	0
170	805	1000	800

Tabla VI.5.3.7.

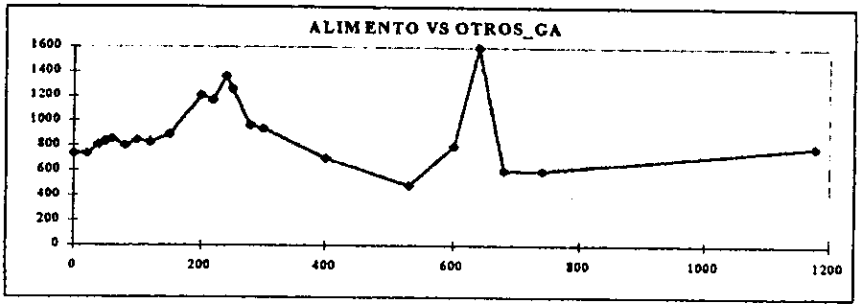
Es decir que según nuestra población encuestada una familia que obtuvo un indicador de 200 en transporte público gasta aproximadamente $862 en alimentación.

Nuevamente se realiza el análisis con el parámetro que mide los gastos mensuales de la familia por causas ajenas al transporte, la alimentación, servicios públicos y educación.



Gráfica VI.5.3.8.

El ancho de banda considerado para este caso: $h = 55$,



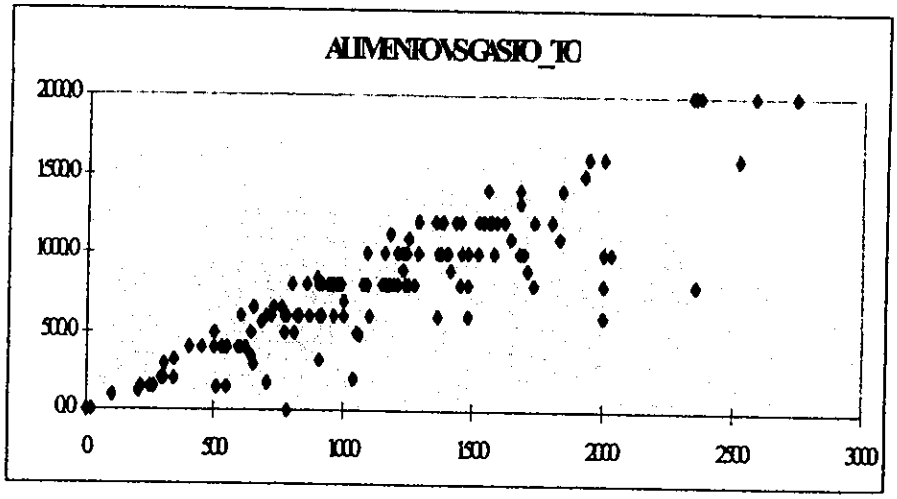
Gráfica VI.5.3.9.

El conjunto de suavizadores asociado: S_3,

OTROS_GA	ALIMENTO (estimado)	OTROS_GA	ALIMENTO (estimado)
0	738	250	1264
20	737	280	961
40	808	300	939
50	840	400	700
60	859	530	480
80	801	600	800
100	845	640	1600
120	831	680	600
150	895	740	600
200	1213	1178	800
220	1171	740	600
240	1368	1178	800

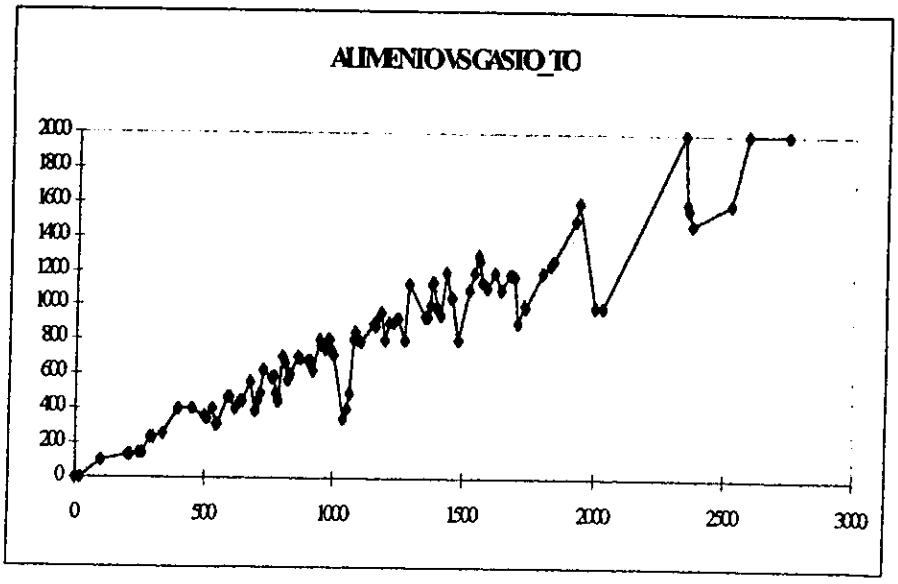
Tabla VI.5.3.10.

La interpretación es análoga que en los dos casos anteriores. Por último, la variable relacionada con el gasto total familiar: GASTO_TO,



Gráfica VI.5.3.11.

Utilizando un ancho de banda de: $h = 30$,



Gráfica VI.5.3.12.

En donde el conjunto de suavizadores: S_4,

GASTO TO	ALIMENTO (estimado)
0	0
20	0
100	100
200	134
210	136
245	150
260	150
290	232
300	235
340	260
400	400
450	400
500	358
510	342
530	400
546	316
548	316
550	316
590	464
600	469
620	400
640	449
650	451
680	560

GASTO TO	ALIMENTO (estimado)
700	388
706	442
720	490
730	626
760	586
770	590
777	488
780	483
785	445
800	708
810	670
820	568
830	600
860	706
870	694
900	676
903	674
908	677
910	676
910	676
920	631
943	800
950	774
960	769

GASTO TO	ALIMENTO (estimado)
970	752
980	800
985	800
990	753
1000	721
1040	341
1050	398
1060	489
1075	800
1080	847
1090	806
1100	792
1150	894
1160	884
1165	932
1180	960
1195	800
1210	900
1230	904
1240	932
1245	926
1250	928
1275	800
1290	1133

*Tabla VI.*5.3.13.

GASTO TO	ALIMENTO (estimado)
1350	944
1360	941
1370	1004
1375	1134
1380	1138
1395	1000
1400	969
1410	947
1430	1200
1450	1050
1455	1050
1480	800
1520	1100
1540	1200
1555	1301
1560	1268
1570	1136
1580	1128
1590	1106
1620	1200
1645	1100
1678	1187
1680	1185
1690	1177

GASTO TO	ALIMENTO (estimado)
1705	900
1730	1002
1734	998
1800	1200
1830	1241
1840	1259
1925	1500
1940	1600
2000	1000
2030	1000
2340	2000
2350	1596
2356	1562
2370	1473
2520	1600
2580	2000
2740	2000

*Tabla VI.*5.3.13. (*continuación*)

VI.6. COMPARATIVO MODELO LINEAL MULTIPLE VS MODELO NP: FIDELIST.

Con la misma información y las mismas variables independientes se determinó un coeficiente de determinación del 99.54%, esto es, las variables mencionadas influyen en ésc porcentaje en la condición alimenticia de la población encuestada. Analizando los resultados se obtuvo que estos son adecuados ya que el modelo lineal múltiple lo es:

Analysis of Variance (fidel50.sta)					
	Sums of Squares	df	Mean Squares	F	p-level
Regress.	25878966	4	6469741.5	7782.97559	0
Residual	118871.375	143	831.268372		
Total	25997838				

Tabla VI.6.

El estadístico F=7782.97 es mayor que el cuantil de una distribución F(4,143)=2.45 con una confianza del 95%, de aquí que el modelo matemático que describe a la variable ALIMENTO a través de la información registrada quedaría de la siguiente manera:

$$ALIMENTO = 2.4823 - 0.9781 * (SERVICIO) - 0.9765 * (TRANSPOR) - 0.9901 * (OTROS\_GA) + 0.9901 * (GASTO\_TO)$$

Observe que a veces las herramientas de validación como el estadístico F y el coeficiente de correlación pueden ser engañosos ya que si recordamos la dispersión de la muestra para cada parámetro contra la variable respuesta, el modelo lineal no era adecuado debido al incumplimiento de los supuestos de linealidad, varianza constante y distribución normal de los errores.

VI.7. CONCLUSIÓN.

Con los métodos de cuantificación de parámetros con los que cuenta actualmente el FIDELIST, de ahora en adelante se podrán establecer mecanismos mas ágiles para discernir si el candidato en cuestión debe o no recibir el beneficio del fideicomiso con simplemente evaluar su nivel de egresos en transporte, servicios y gastos en general; para poder predecir el costo mensual familiar de su alimentación.

CONCLUSIONES GENERALES

A lo largo de este trabajo de tesis se ha revisado una de las teorías mas recientes e importantes en materia de modelos de estimación estadística. Se detalló a profundidad sobre las ventajas que presenta en general la teoría NP en relación con otras técnicas de estimación y ajuste, tales como: la regresión lineal, los polinomios y splines interpolantes,etc. Sin embargo es necesario puntualizar que no siempre un modelo NP es mejor que aquellos construidos con las técnicas antes mencionadas ya que, en general, el término $no-paramétrico$ es aplicable en probabilidad y estadística cuando no es posible asegurar una cierta distribución de las observaciones o bien, cuando supuestos como: la normalidad, observaciones independientes e idénticamente distribuídas y varianza constante son insostenibles. Un supuesto implícito en los anteriores, nace de la naturaleza misma del fenómeno y se refiere a si el modelo lineal se supone adecuado, de aquí que cuando las observaciones a modelar no guardan una tendencia lineal es necesario aplicar métodos alternativos.

En relación al número de observaciones, recordemos que conforme se añade información un modelo paramétrico puede tener errores catastróficos como el ejemplo visto en la sección II.2 de la función propuesta por Runge esto en contraposición de la hipótesis inicial que suponía una estimación mucho mas precisa. Esta inconsistencia que presentan los modelos paramétricos es una característica a favor de los suavizadores NP pues no existe posibilidad de errores catastróficos. Observe que el ancho de banda debe contener a un número de observaciones suficiente para realizar la estimación, esto puede parecer confuso pero en realidad no existe una regla que permita discernir cuál debe ser la longitud de dicho intervalo y esto deberá ser evaluado por el interesado. Se recomienda utilizar la metodología empleada en la sección VI.4.3 en donde se propone un ancho de banda y posteriormente se procede a su ajuste.

APÉNDICE

A.1. LA NATURALEZA ESTOCÁSTICA DE LAS OBSERVACIONES

De manera natural las observaciones están relacionadas entre *sí* debido a que cada una de ellas describen respuestas de un mismo fenómeno. Dicha relación se debe a que, como en todo modelo de regresión, las observaciones provienen de eventos independientes e idénticamente distribuídos y es a partir de ellos que se busca una aproximación $\widehat{\mu}(x)$ a la curva:

$$\mu(x) = E[Y \mid X = x].$$

Mas aún, si una densidad conjunta $f(x,y)$ existe, entonces la curva puede ser calculada como:

$$\mu(x) = \frac{\int y\, f(x,y)\, dy}{f(x)},$$

con $f(x)$ la densidad marginal de X.

A.2. DEFINICIONES.

A.2.1. CONVERGENCIA EN PROBABILIDAD

Definición. Sean $X_1, X_2, X_3, \ldots$ y X variables aleatorias. Se dice que la sucesión $\{X_n\}$ *converge a* X *en probabilidad* si:

$$P(|X_n - X| \geq \varepsilon) \to 0$$

toda vez que: $n \to \infty$, $\forall\, \varepsilon > 0$. Dicha convergencia se denota de la siguiente manera:

$$X_n \to^p X \text{ , cuando: } n \to \infty.$$

A.2.2. ESPACIO MÉTRICO.

Definición. Un conjunto X cuyos elementos llamaremos *puntos*, se dice que es un *espacio métrico* si para cualesquiera dos puntos $p,\ q \in X$, $\exists$ un número real asociado $d(p,q)$, el cual se lee: *distancia de* p *a* q, tal que:

a) $d(p,q) > 0$, si $p \neq q$; $d(p,q) = 0$, si $p = q$

b) $d(p,q) = d(q,p)$
c) $d(p,q) \leq d(p,r) + d(r,q)$, $\forall\, r \in X$.

A.2.3. FUNCION CONTINUA.
Definición. Sean X y Y espacios métricos, $E \subset X$, $p \in E$,y f una función tal que:

$$f(e) = y,\ \forall\, e \in E,\ y \in Y.$$

Decimos que f es *contínua en* p si $\forall \varepsilon > 0,\ \exists\, \delta > 0$, tal que:

$$d_Y(f(x), f(p)) < \varepsilon$$

$\forall\, x \in E,\ d_X(x,p) < \delta$.

A.2.4. FUNCION ACOTADA.
Definición. Sea f una función tal que:

$$f(x) = y,\ \forall x \in E,\ y \in R^k.$$

Diremos que f es *acotada* si $\exists\, M \in R$, tal que:

$$|f(x)| \leq M,\ \ \forall\ x \in E$$

A.2.5. FUNCION REAL.
Definición. Sean X y Y espacios métricos y f una función tal que:

$$f(x) = y, \forall x \in X,\ y \in Y.$$

Diremos que f es *función real* si $X, Y \subseteq R$.

A.2.6. FUNCION SIMÉTRICA.
Definición. Decimos que una función $f : X \to Y$, es *simétrica* si:

$$f(x) = f(-x), \forall x \in X.$$

BIBLIOGRAFÍA

a) **ELDEN, Lars & WITTMEYER-KOCH, Linde.** *Numerical Analysis.* Academic Press, Inc. 1990.

b) **MONTGOMERY, Douglas C. & PECK, Elizabeth A.** *Introduction to Linear Regression Analysis.* Segunda edición. Wiley Series.

c) **HÄRDLE, Wolfgang.** *Applied Nonparametric Regression.* Cambridge University Press, 1990.

d) **RANDALL, Eubank L.** *Spline Smoothing and Nonparametric Regression.* Editorial Dekker Vol.90, 1988.

e) **RYAN, Thomas P.** *Modern Regression Methods.* Wiley Series, 1997.

f) **SIEGEL, Sidney.** *Estadística No – Paramétrica (Aplicada a las Ciencias de la Conducta).* Editorial Trillas, 1980.