Servy. Elsa
Garcia, María del Carmen
Paccapelo, Valeria
Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística

REGRESIÓN NO PARAMÉTRICA: UNA APLICACIÓN

1.- INTRODUCCIÓN

La teoría clásica de la regresión se basa, en gran parte, en el supuesto que las observaciones son independientes y se encuentran idéntica y normalmente distribuidas. Si bien existen muchos fenómenos del mundo real que pueden modelarse de esta manera, para el tratamiento de ciertos problemas, la normalidad de los datos es insostenible. En el intento de eliminar esa restricción se diseñaron métodos que hacen un número mínimo de supuestos sobre los modelos que describen las observaciones.

La teoría de los métodos no paramétricos trata, esencialmente, el desarrollo de procedimientos de inferencia estadística, que no realizan una suposición explícita con respecto a la forma funcional de la distribución de probabilidad de las observaciones de la muestra. Si bien en la Estadística no paramétrica también aparecen modelos y parámetros, ellos están definidos de una manera más general que en su contrapartida paramétrica.

La regresión no paramétrica es una colección de técnicas para el ajuste de funciones de regresión cuando existe poco conocimiento a priori acerca de su forma. Proporciona funciones suavizadas de la relación y el procedimiento se denomina suavizado.

Los fundamentos de los métodos de suavizado son antiguos pero sólo lograron el estado actual de desarrollo gracias a los avances de la computación y los estudios por simulación han permitido evaluar sus comportamientos.

La técnica más simple de suavizado, los promedios móviles, fue la primera en usarse, sin embargo han surgido nuevas técnicas como la estimación mediante núcleos ("kernel") o la regresión local ponderada. Estos estimadores de regresión no paramétrica son herramientas poderosas para el análisis de datos, tanto como una técnica de estimación para resumir una relación compleja que no puede ser aprehendida por un modelo paramétrico, como para suplementar (o complementar) un análisis de regresión paramétrico.

En este este trabajo se presenta una aplicación de estos métodos para el ajuste de modelos de regresión para explicar el ingreso del jefe del hogar, a partir de información suministrada por la Encuesta Permanente de Hogares, relevada por el INDEC, para el aglomerado Rosario en la segunda onda de 2002. El mismo se realiza en el marco del proyecto "Métodos no paramétricos y semiparamétricos para el análisis de regresión con datos univariados y multivariados".

2. Regresión no paramétrica

En los análisis paramétricos se comienza haciendo supuestos rígidos sobre la estructura básica de los datos, luego se estiman de la forma más eficiente posible los parámetros que definen la estructura y por último se comprueba si los supuestos iniciales se cumplen.

La regresión no paramétrica, en cambio, desarrolla un "modelo libre" para predecir la respuesta sobre el rango de valores de los datos. Básicamente está constituida por métodos que proporcionan una estimación suavizada de la relación para un conjunto de valores (denominado ventana) de la variable explicativa. Estos valores son ponderados de modo que, por ejemplo, los vecinos más cercanos tengan mayor peso que los más alejados dentro de una ventana de datos. Se pueden utilizar diversas funciones de ponderación, que son los pesos en que se basan los estimadores. La combinación de la función de ponderación y el ancho de la ventana inciden sobre la bondad de la estimación resultante.

La mayor parte de las publicaciones sobre regresión no paramétrica consideran el caso de un solo regresor a pesar de que, a simple vista no pareciera de gran utilidad, ya que las aplicaciones más interesantes involucran varias variables explicativas. Sin embargo, la regresión no paramétrica simple es importante por dos motivos:

- En etapas preliminares del análisis de datos o en pruebas de diagnóstico se utilizan gráficos de dispersión en los cuales puede ser muy útil ajustar una "curva suavizada". Por ejemplo, para explorar la forma de la función respuesta, para confirmar una función respuesta en particular que haya sido ajustada a los datos, para obtener estimaciones de la respuesta media sin especificar la forma de la función respuesta, para estudiar el cumplimientos de supuestos, etc.
- Forma la base a partir de la cual se extienden los conceptos para regresión no paramétrica múltiple.

El análisis de regresión considera que una variable respuesta Y es función de un conjunto de variables explicativas X_1, X_2,…,X_p. En general, se asume que la relación entre las variables es de tipo lineal y $g(.)$ es una función que depende de parámetros (β_j)

$$y_i = g(\mathbf{x}_i') + \ _i$$

```
ERROR: rangecheck
OFFENDING COMMAND: .buildcmap

STACK:

-dictionary-
/WinCharSetFFFF-V2TT3491565Ct
/CMap
-dictionary-
/WinCharSetFFFF-V2TT3491565Ct
```