REGRESIÓN APLICADA USANDO R

Edgar Acuña Fernandez

Departamento de Ciencias Matemáticas

Universidad de Puerto Rico

Recinto Universitario de Mayaguez

Enero 15, 2015

PREFACIO

La razón principal de escribir este libro es la carencia de un texto completo de regresión que cubra las diversas técnicas de regresión, especialmente aquellas que han tomado auge en la última década. Un par de buenos libros de regresión son el "Classical and Modern Regression with applications" de Myers (2000) y el "Applied linear Regression" de Weisberg (2005), pero ambos cubren muy poco material u omiten temas importantes en regresión tales como: selección de variables, regresión logística, regresión robusta y la muy importante área de regresión no paramétrica. Existen por otro lado buenos textos cubriendo solamente Regresión Robusta como el "Robust Regression and Outlier Detection" de Rousseeuw y Leroy (2005) y otros que tratan exclusivamente Regresión noparamétrica como el "Applied Nonparametric Regression" de Haerdle (1994). El objetivo de este texto es cubrir la parte más transcendental de los libros antes mencionados.

En el transcurso de los quince años que he venido desarrollando el texto he usado varios programa estadísticos tales como: MINITAB, SAS, MATLAB, S-PLUS y últimamente R. La meta final es desarrollar todo el texto usando el programa gratuito R, el cual está disponible en www.r-project.org. Aún quedan en el texto algunas salidas de MINITAB. Las salidas de SAS, MATLAB y S-Plus han sido prácticamente eliminadas.

Aunque el texto es en regresión aplicada también se ha tratado de probar varias identidades y propiedades de estimadores que aparecen en regresión. Sin embargo no es nuestra intención llenar el texto con demostraciones teóricas. Dos buenos textos donde se ve el lado teórico de Regresión son "Linear Regression Analysis" de Seber (2003) y "Linear Statistical Inference and its Applications" de Rao(2008).

El texto está organizado en 9 capítulos. El primer capítulo se enfoca en regresión lineal simple y el segundo en regresión lineal múltiple. En el tercer capítulo se discute los diversos métodos de diagnosticar si las suposiciones del modelo de regresión se están cumpliendo o no. En el capítulo 4 se estudian diferentes transformaciones que se pueden hacer de las variables predictoras y de la variable de respuesta con la finalidad de mejorar el modelo de regresión para que haga un mejor ajuste de los datos. En el capítulo 5 se discute modelos de regresión considerando la presencia de variables categóricas. Aquí se incluye el estudio de la regresión logística. El capítulo 6 está dedicada al importante problema de selección de variables en regresión y en el problema 7 se discute la forma de detectar y resolver el problema de multicolinealidad entre las variables predictoras. Los capítulos 8 y 9 están dedicados a regresión robusta y regresión noparamétrica respectivamente.

Los conjuntos de datos que aparecen en este texto pueden ser obtenidos en el siguiente sitio de la internet en http://academic.uprm.edu/eacuna/class6205.html.

Finalmente, deseo agradecer la ayuda de mi pasados asistentes de investigación por colaborar conmigo en la depuración de errores presentes en el texto, así como en la edición de algunos capítulos y en la preparación de las transparencias del texto.

Por favor para reportar cualquier sugerencia o error mandarme un e-mail a edgar.acuna@upr.edu.

Mayagüez, Febrero 14, 2018

CONTENIDO