



Propuesta de Trabajo Fin de Grado en Matemáticas (curso 2022-2023)

Responsable de tutorización: Desirée Romero Molina
Departamento: Estadística e Investigación Operativa
Correo electrónico: deromero@ugr.es

Responsable de cotutorización: Nuria Rico Castro
Departamento: Estadística e Investigación Operativa
Correo electrónico: nrico@ugr.es

(Rellenar sólo en caso de que la propuesta esté realizada a través de un estudiante)
Estudiante que propone el trabajo: Celia Floro Rodríguez

Título del trabajo: Regresión lineal múltiple. Aplicación a datos ambientales.

Tipología del trabajo (marcar una o varias de las siguientes casillas):

- Complementario de profundización
- Divulgación de las Matemáticas
- Docencia e innovación
- Herramientas informáticas
- Iniciación a la investigación

Materias del grado relacionadas con el trabajo:

Probabilidad, Inferencia Estadística, Estadística Multivariante, Estadística Computacional.

Descripción y resumen de contenidos:

El análisis de Regresión Lineal Múltiple (RLM) univariante es una técnica estadística multivariante usada en la práctica para examinar la forma en que varias variables consideradas simultáneamente (variable independientes, explicativas o regresoras) pueden usarse para comprender o predecir el comportamiento de otra variable (variable dependiente o explicada). Este modelo es la generalización para múltiples variables explicativas del modelo de Regresión Lineal Simple (RLS) que la alumna ya estudió en la asignatura de Inferencia Estadística como un caso particular de los Modelos Lineales Generales de Gauss-Markov.

El modelo de RLM pretende medir el efecto de las variables que se consideran influyentes en la investigación y representa el de las restantes variables mediante una variable aleatoria que se denomina perturbación del modelo. La dependencia entre dichas variables se establece mediante una función que se supone lineal. Al igual que en la RLS el criterio que se utilizará para determinar las relaciones entre las variables consideradas será el de mínimos cuadrados.

En este trabajo se propone estudiar el modelo RLM así como todos los problemas vinculados a la aplicación práctica de los mismos para, posteriormente, aplicar lo estudiado sobre una base de datos reales con variables de tipo medioambiental.

Se comenzará con un estudio teórico del modelo generalizando lo que el alumno estudió en el modelo RLS. A continuación se hará un estudio de las técnicas de validación del modelo mediante los contrastes de hipótesis oportunos. En este caso habrá que estudiar además el problema de multicolinealidad para el cual se estudiarán técnicas de detección y técnicas que permitan evitar dicho problema como son los métodos de introducción de variables por etapas. Finalmente se estudiará su tratamiento computacional para lo cual las competencias adquiridas en la asignatura Estadística Computacional serán fundamentales para la estudiante ya que le puede proporcionar

los conocimientos necesarios para usar el paquete R para la aplicación de los distintos modelos o incluso para poder hacer una implementación propia a la base que se le proporcionará.

Actividades a desarrollar:

La estudiante deberá comenzar recopilando bibliografía sobre el modelo de RLM para su posterior estudio, buscando todas las técnicas estadísticas vinculadas a ellos.

A continuación la estudiante profundizará sobre la base matemática que hay en estos modelos, donde tendrá que usar conocimientos de otras áreas de las matemáticas, y estudiará las técnicas estadísticas necesarias para la validación del mismo para lo cual usará lo estudiado en la asignatura de Inferencia Estadística. Además profundizará en el estudio de los Modelos Lineales Generales y en las técnicas de validación que requiere el modelo RLM.

Finalmente, la estudiante deberá aplicar este modelo, utilizando algún paquete estadístico como R, a una base de datos reales y no a meros ejemplos básicos de libros.

Si la estudiante así lo desea puede ampliar el trabajo incluyendo una implementación propia de algunas de técnicas estadísticas estudiadas en el trabajo.

Objetivos matemáticos planteados

Recopilación bibliográfica

Estudio del modelo y de las técnicas necesarias

Aplicación a datos reales

Implementación de algunas técnicas

Bibliografía para el desarrollo matemático de la propuesta:

- Anderson, T.W. (2003, 3º ed.) An introduction to Multivariate Statistical Analysis. John Wiley & Sons.
- Cuadras, C. M. (2018, 2º ed.) Nuevos métodos de Análisis Multivariante. CMC Editions.
- Peña Sánchez de Rivera, D. (2002) Análisis de Datos Multivariantes. McGraw-Hill.
- Rencher, A. C. y Schaalje, G. B. (2007) Linear Models in Statistics. John Wiley & Sons.
- Timm, N. H. (2002) Applied Multivariate Analysis, Springer.

Otras referencias (si procede):

Firma del estudiante
(solo para trabajos propuestos por estudiantes)

Firma del responsable de tutorización
(solo para trabajos propuestos por estudiantes)

Firma del responsable de cotutorización
(solo para trabajos propuestos por estudiantes)

En, Granada, a 03 de mayo de 2022