UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

Propuesta TFG. Curso 2026/2027

GRADO: Grado en Matemáticas

CÓDIGO DEL TFG: 270-074-2026/2027

1. DATOS BÁSICOS DEL TFG:

Título: Multivariable regression versus the G-Formula

Descripción general (resumen y metodología):

Utilizar modelos de regresión para ajustar por los factores de confusión requiere asumir que no hay modificación del efecto y debemos elegir una medida que sea en términos de probabilidad reversible en escala, como la diferencia de riesgo, en lugar de la razón del producto cruzado.
Una alternativa al uso del ajuste de regresión multivariable es la Fórmula G (una generalización de la estandarización con respecto a la distribución de factores de confusión). En 1986, un artículo fundamental demostró que bajo supuestos (intercambiabilidad condicional, positividad, consistencia y no interferencia), se podía obtener una estimación consistente de la estimación marginal de la razón del producto cruzado utilizando la fórmula G. El cálculo de G, basado en la estimación de los componentes de la fórmula G, permite un efecto del tratamiento que puede variar según los niveles de los factores de confusión (permite modelizar la heterogeneidad del tratamiento).

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

Evaluar mediante la experimentación de Monte Carlo comparada, las características y propiedades estadísticas del estimador no paramétrico G-Formula, para la estimación de un efecto de un tratamiento ajustado por confusores mediante una medida de orden multiplicativo, frente a la modelización del mismo efecto con métodos clásicos de regresión multivariable (regresión logística). Ver: https://github.com/migariane/HETMOR-Causal-Inference

Bibliografía básica:

1. Spiegelman D, Zhou X. Evaluating Public Health Interventions: 8. Causal Inference for Time-Invariant Interventions. Am J Public Health. 2018:e1-e4.
2. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Statistical Science. 1999;14(1):29-46.
3. Sjolander A, Dahlqwist E, Zetterqvist J. A Note on the Noncollapsibility of Rate Differences and Rate Ratios. Epidemiology. 2016;27(3):356-359.
4. Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The parametric g-formula for time-to-event data: intuition and a worked example. Epidemiology. 2014;25(6):889-897.
5. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. International journal of epidemiology. 1986;15(3):413--419
6. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol. 2011;173(7):731-738.
7. Sofrygin O, van der Laan MJ, Neugebauer R (2015). simcausal: Simulating Longitudinal Data with Causal Inference Applications. R package version 0.5.
8. Luque Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. Targeted maximum likelihood estimation for a binary treatment: A tutorial. Stat Med. 2018;37:2530-2546.

Recomendaciones y orientaciones para el estudiante:

Experiencia con programación en R y afinidad por la bioestadística

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: MIGUEL ÁNGEL LUQUE FERNÁNDEZ

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: mluquefe@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: LUCAS FEHLAU ARBULU

Correo electrónico: lucasfa@correo.ugr.es