



1. DATOS BÁSICOS DEL TFG:

Título: Métodos de predicción para mitigar la falta de respuesta y la auto-selección en encuestas

Descripción general (resumen y metodología):

Dentro de las fuentes de error habituales en las encuestas, dos de las más habituales son la falta de respuesta, que se produce cuando una persona seleccionada para participar en una encuesta decide no responder a ciertas preguntas (falta de respuesta parcial) o directamente no responde a ninguna (falta de respuesta total), y la auto-selección de la muestra, que ocurre cuando la encuesta no sigue un diseño muestral en el que se seleccione a cada persona con una cierta probabilidad, sino que son las propias personas las que se auto-seleccionan para participar. Estos errores pueden derivar en un sesgo importante en los estimadores habituales. Diversas técnicas han sido desarrolladas en la literatura para mitigar este tipo de errores, entre las que destaca el método de imputación. Dicha imputación puede aplicarse a aquellas personas que deciden no responder o también aplicarse a otras fuentes de datos diferentes (p. ej. otras encuestas probabilísticas o incluso censos poblacionales) basándose en modelos entrenados a partir de la muestra disponible, que predicen el valor más probable de la variable objetivo en aquellos individuos donde ésta no ha sido observada. En este Trabajo Fin de Grado se investigarán los métodos de imputación más habituales en encuestas probabilísticas y no probabilísticas, y se estudiará su aplicación en un caso real.

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

Objetivos planteados:

El objetivo principal de este trabajo es estudiar las metodologías existentes relativas a la imputación en encuestas (probabilísticas y no probabilísticas), aplicarlas en un caso real y, en caso de que fuera posible, desarrollar alguna metodología novedosa en este campo. Los objetivos específicos son:

OE1. Uso de técnicas de imputación para la falta de respuesta (parcial y total) en encuestas probabilísticas: modelos semi-paramétricos y no paramétricos, Predictive Mean Matching, algoritmos predictivos de aprendizaje automático (Machine Learning) e imputación múltiple.

OE2. Uso de técnicas de imputación en el contexto de encuestas no probabilísticas con auto-selección: Statistical Matching o Mass Imputation, estimadores doblemente robustos y estimadores modelo-basados, modelo-asistidos y modelo-calibrados aplicados a censos poblacionales de variables auxiliares.

Bibliografía básica:

- Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International statistical review*, 78(1), 40-64.
- Buelens, B., Burger, J., & van den Brakel, J. A. (2018). Comparing inference methods for non-probability samples. *International Statistical Review*, 86(2), 322-343.
- Chen, S., & Haziza, D. (2019). Recent developments in dealing with item non-response in surveys: A critical review. *International Statistical Review*, 87, 192-218.
- Chen, S., Yang, S., & Kim, J. K. (2022). Nonparametric mass imputation for data integration. *Journal of survey statistics and methodology*, 10(1), 1-24.
- Ferri-García, R., Castro-Martín, L., & del Mar Rueda, M. (2021). Evaluating Machine Learning methods for estimation in online surveys with superpopulation modeling. *Mathematics and Computers in Simulation*, 186, 19-28.

- Haziza, D., & Beaumont, J. F. (2007). On the construction of imputation classes in surveys. *International Statistical Review*, 75(1), 25-43.
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley & Sons.
- Särndal, C. E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.
- Yang, S., & Kim, J. K. (2018). Integration of survey data and big observational data for finite population inference using mass imputation. *arXiv preprint arXiv:1807.02817*.

Recomendaciones y orientaciones para el estudiante:

Se recomienda que el estudiante haya cursado las materias Muestreo Estadístico, Diseño de Encuestas, Modelos Lineales y Minería de Datos del Grado en Estadística.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: RAMÓN FERRI GARCÍA

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: rferri@ugr.es

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

Nombre y apellidos: MARÍA DEL MAR RUEDA GARCÍA

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: mrueda@ugr.es

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: MARIA ELENA GARCIA MONTAÑEZ

Correo electrónico: elenagarcia19@correo.ugr.es