



Línea de Trabajo fin de Máster 2025-2026

Máster Universitario en Estadística Aplicada.	
Título	Simulación de datos sintéticos complejos para trabajar en encuestas
Tipo	INVESTIGACIÓN ☐ ORIENTACIÓN PRÁCTICA ☒
Número de alumnos admitidos	1
Profesor(es)/ email	María del Mar Rueda García (mrueda@ugr.es) Jorge Luis Rueda Sánchez (jorgerueda@ugr.es)
Descripción	La producción de conjuntos de datos sintéticos se ha propuesto como una solución para la investigación con datos reales mientras que se protege el anonimato de los datos. Esto se realiza generando archivos de uso público a partir de datos protegidos y como herramienta para crear "conjuntos de datos aumentados" que sirvan como entrada para modelos de micro simulación. Los datos sintéticos se han convertido en un instrumento importante para las evaluaciones ex ante del impacto de las políticas. El rendimiento y la aceptabilidad de esta herramienta dependen en gran medida de la calidad de las poblaciones sintéticas, es decir, de la similitud estadística entre la población sintética y la real de interés. Se han desarrollado múltiples herramientas para generar datos sintéticos.
Objetivos particulares	Estudiar los métodos actuales de creación de poblaciones sintéticas Implementación de código de los citados métodos propuestos en el entorno de programación R, así como elaboración de funciones. Realización de una aplicación práctica.
Prerrequisitos y recomendaciones	Es necesario que el alumno disponga de una formación en inferencia en poblaciones finitas. Por ello el alumno debería de cursar la materia del máster "Encuestas por muestreo. Aplicaciones económicas, sociales y medioambientales", y sería conveniente que cursara también la materia "Aspectos Computacionales en la Estimación de Errores en Encuestas por Muestreo".
Plan de trabajo	En este TFM el alumno hará una descripción de los diversos enfoques para creación de datos sintéticos y también del software disponible. Se hará una aplicación práctica con datos reales del INE
Competencias generales y específicas	CB: 6, 7, 8, 9, 10 CG: 1, 2, 3, 6, 9 CE: 10, 13, 15, 16, 17, 18, 19, 20, 22, 23, 26, 28, 29
Bibliografía	Alfons A, Kraft S (2013). <i>simPopulation: Simulation of Synthetic Populations for Surveys Based on Sample Data</i> . R package version 0.4.1, URL https://CRAN.R-project.org/package=simPopulation . Templ, M., Meindl, B., Kowarik, A., & Dupriez, O. (2017). Simulation of Synthetic Complex Data: The R Package simPop. <i>Journal of Statistical Software</i> , 79(10), 1–38. https://doi.org/10.18637/jss.v079.i10