



Línea de Trabajo fin de Máster

(Fecha última actualización: 01/10/2022)

Máster en Estadística. CURSO ACADÉMICO 2020-2021	
Título	Integración de datos de encuestas probabilísticas y no probabilísticas con Kernel Weighting.
Tipo	INVESTIGACIÓN <input checked="" type="checkbox"/> ORIENTACIÓN PRÁCTICA <input type="checkbox"/>
Profesor(es)/ email	María del Mar Rueda García mrueda@ugr.es Beatriz Cobo Rodríguez beacr@ugr.es
Descripción	En los últimos años, se han proporcionado diferentes métodos para combinar información de múltiples fuentes de datos. Nos centramos en el caso de muestras probabilísticas y no probabilísticas que comparten el mismo cuestionario, combinando ambas para maximizar la eficiencia de las estimaciones con la ayuda de métodos de aprendizaje automático. En este trabajo, se revisarán algunos métodos de integración de datos, como el propensity score adjustment y kernel weighting, se estudiarán las características de cada enfoque, incluidos sus beneficios y limitaciones, y se realizarán algunos resultados empíricos.
Objetivos particulares	El alumno aprenderá cómo realizar estimaciones mediante la combinación de encuestas obtenidas mediante diversos procedimientos de muestreo, estudiando cómo construir los pesos, cómo mitigar el sesgo y disminuir la varianza. También aprenderá técnicas de estimación de la varianza, técnicas de machine learning y profundizará en el manejo de R.
Prerrequisitos y recomendaciones	Es necesario que el alumno disponga de una formación avanzada 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	<ul style="list-style-type: none"> • Revisión bibliográfica de la literatura sobre kernel weighting. • Revisión bibliográfica de la literatura sobre la técnica de integración de datos de encuestas • Propuesta de nuevos métodos de integración. Estudio de sus propiedades • Estudios de simulación para la comprobación de los desarrollos explicados en los apartados anteriores mediante el software estadístico R
Competencias generales y específicas	OE1 Ser capaz de diseñar el proceso de adquisición de datos. OE7. Estar capacitado para realizar el análisis de los modelos y de los resultados obtenidos. OE8. Ser capaz de encontrar y utilizar bibliografía científica relevante dentro del campo de investigación de la Estadística. OE9. Ser capaz de realizar trabajos de investigación y extraer conclusiones
Bibliografía	Beaumont, J. F. and Rao, J.N.K. (2021). Pitfalls of making inferences from non-probability samples: Can data integration through probability samples provide remedies. The Survey Statiscian, International Association of Survey Statiscians, 83, 11-22 Castro-Martín, L., Rueda, M.d.M. and Ferri-García, R. (2020). Inference from non-probability surveys with statistical matching and propensity score adjustment using modern prediction techniques. Mathematics 8,87 Kern, C., Li, Y. and Wang, L. (2020) Boosted Kernel Weighting—Using statistical learning to improve inference from nonprobability samples. Journal of Survey Statistics and Methodology, 9(5),1088–1113 Kim, J. K. and Tam, S.M. (2021). Data Integration by Combining Big Data and Survey Sample Data for Finite Population Inference.International Statistical Review, 89, (2), 382–401.