



Título del trabajo: Técnicas modernas de análisis de datos aplicadas a la estimación en encuestas por muestreo.
Tutor/a: María del Mar Rueda Correo electrónico: mrueda@ugr.es
Cotutor/a: Ramón Ferri García
Departamento responsable: Estadística e I.O
Perfil y número de estudiantes al que va dirigido (máximo 2): 1 <i>(Rellenar sólo en caso de que la propuesta esté realizada a través de estudiante)</i>
Estudiante que propone el trabajo (Nombre, Apellidos, DNI): Jorge Luis Rueda Sánchez,
Tipo de trabajo: Trabajos de inicio a la investigación.
Competencias <i>Competencias generales:</i> G02, G03, G04, G05, G06, G07, G08, G09. <i>Competencias específicas:</i> E01, E02, E03, E04, E05, E08, E09, E10.
Resultados de aprendizaje <i>(estos son los mínimos; añadir otros si se considera)</i> <ul style="list-style-type: none">• Adquirir competencias globales ligadas al desarrollo y aplicación de los conocimientos del Grado.• Adquirir competencias ligadas a la búsqueda y organización de información y documentación relevante sobre el tema objeto de estudio.• Aplicar el “pensamiento estadístico” y tener capacidad para enfrentarse a las distintas etapas de un estudio estadístico (desde el planteamiento del problema hasta la exposición de resultados).• Saber presentar, de forma escrita y oral, la memoria, los resultados y las conclusiones del trabajo realizado.
Antecedentes y resumen del tema propuesto: <p>Las encuestas, desde su desarrollo teórico en la primera mitad del siglo XX, suponen uno de los métodos principales de recogida de información en la investigación empírica. La teoría del muestreo estadístico permite la obtención de estimaciones precisas a partir de una muestra obtenida de una población. Sin embargo, en las últimas décadas los métodos habituales de administración de encuestas (telefónicas o mediante entrevista personal) han sufrido un declive en las tasas de respuesta y un aumento de sus costes que ha propiciado la aparición de otras metodologías como las encuestas online.</p> <p>Estas nuevas metodologías permiten una recogida de información más rápida, con un coste considerablemente más bajo y una mayor cantidad de posibilidades para la elaboración del cuestionario. Sin embargo, las características de estas encuestas conllevan la aparición de sesgos de selección, debido a que la población objeto de estudio puede no estar correctamente representada en el marco muestral (error de cobertura) o la muestra puede haberse seleccionado de forma incidental, mediante técnicas como el muestreo “bola de nieve” o la autoselección. Los sesgos producidos por estas circunstancias pueden afectar seriamente la fiabilidad de las estimaciones.</p> <p>En este TFG se desarrollarán metodologías para mitigar los sesgos de selección asociados a las nuevas técnicas de encuesta. Para ello, se emplearán técnicas de Machine Learning para modelizar la probabilidad de respuesta a una encuesta con mecanismo de autoselección, así</p>



como la falta de respuesta en encuestas con diseño probabilístico. También se utilizarán para modelizar las variables objeto de estudio, de cara a su predicción en una muestra de referencia o en la población para la estimación basada en modelos o modelo asistida. Se investigará también el uso de técnicas de selección de variables para una mejor especificación de los modelos que consiga una mayor mitigación del sesgo de selección. La eficacia de estas técnicas se estudiará mediante su aplicación en conjuntos de datos de encuesta, tanto simulados como reales.

Breve descripción de las actividades presenciales y no presenciales a realizar:

Actividades presenciales (15-30%)	Planteamiento, orientación y supervisión	50 horas
	Exposición del trabajo	10 horas
	Otras:	
Actividades no presenciales (70-85%)	Preparación del trabajo	140 horas
	Elaboración de la memoria	100 horas
	Otras:	
Total (12 ECTS)		300 horas

Objetivos que se pretenden alcanzar:

El objetivo principal de este trabajo es desarrollar nuevas metodologías para disminuir de manera eficiente el efecto de los diferentes sesgos en las estimaciones a partir de encuestas probabilísticas y no probabilísticas mediante el uso de información auxiliar de diferentes tipos. Los objetivos específicos son:

OE1. Uso de técnicas de Machine Learning para ajustar probabilidades de participar y de falta de respuesta, así como la estimación paramétrica basada en el modelo y modelo asistida.

OE2. Uso de algoritmos de selección de variables para la correcta configuración de las covariables en los modelos de estimación de probabilidades de participar y de falta de respuesta, así como en la estimación paramétrica basada en el modelo y modelo asistida.

Bibliografía básica para la puesta en marcha del trabajo:

1. Bethlehem, J. (2010) Selection bias in Web Surveys. *International Statistical Review*, 78(2), 161-188.
2. Breidt, F. J., Opsomer, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science*, 32(2), 190-205.
3. Buelens, B., Burger, J., van den Brakel, J. A. (2018). Comparing inference methods for non-probability samples. *International Statistical Review*, 86(2), 322-343.
4. Buskirk, T. D., Kirchner, A., Eck, A., & Signorino, C. S. (2018). An introduction to machine learning methods for survey researchers. *Survey Practice*, 11(1), 1-10.
5. Castro Martín, L., Ferri García, R., Rueda, M. (2019). NonProbEst: Estimation



in Nonprobability Sampling. R package version 0.2.0. <https://CRAN.R-project.org/package=NonProbEst>

6. Castro-Martín, L., Rueda, M. D. M., & Ferri-García, R. (2020). Inference from Non-Probability Surveys with Statistical Matching and Propensity Score Adjustment Using Modern Prediction Techniques. *Mathematics*, 8(6), 879.
7. Deville, J. C., Särndal, C. E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*. 87 (418) 376–382.
8. Elliott, M. R., Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 32(2), 249-264.
9. Valliant, R. (2020). Comparing alternatives for estimation from nonprobability samples. *Journal of Survey Statistics and Methodology*, 8(2), 231-263.

Tipo de trabajo (*):

1. Estudio de profundización en algún tema concreto de Estadística, o como proyecto de aplicación de la misma a estudios o problemas de otros ámbitos científicos o sociales.
2. Realización completa de todas las fases de un proyecto estadístico, bien con auxilio de prácticas en empresas o con prácticas propuestas y dirigidas por el tutor.
3. Estudio de casos, teóricos o prácticos, relacionados con la Estadística.
4. Elaboración de un informe o un proyecto de naturaleza profesional.
5. Elaboración de un plan de empresa.
6. Simulación de encargos profesionales.
7. Trabajos bibliográficos sobre el estado actual de una temática relacionada con la Estadística.
8. Creación y/o empleo de herramientas informáticas para su uso en Estadística.
9. Trabajos de inicio a la investigación.
10. Trabajos cuya finalidad sea la divulgación de la Estadística en diversos contextos.
11. Trabajos sobre Historia de la Estadística.
12. Trabajos relacionados con la docencia de la Estadística.

Competencias (**)

Competencias generales:

G01. Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.



- G02.** Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.
- G03.** Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- G04.** Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.
- G05.** Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- G06.** Saber utilizar herramientas de búsqueda de recursos bibliográficos.
- G07.** Poder comunicarse en otra lengua de relevancia en el ámbito científico.
- G08.** Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.
- G09.** Fomentar y garantizar el respeto a los Derechos Humanos, a los principios de accesibilidad universal, igualdad, y no discriminación; y los valores democráticos, de la cultura de la paz y de igualdad de género.

Competencias específicas:

- E01.** Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.
- E02.** Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.
- E03.** Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.
- E04.** Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.
- E05.** Comprender la importancia de la Investigación Operativa como metodología de optimización, toma de decisiones y diseño de modelos particulares para la resolución de problemas en situaciones específicas.
- E06.** Comprender y utilizar básicamente el lenguaje matemático.
- E07.** Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.
- E08.** Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.
- E09.** Conocer los conceptos básicos y habilidades propias de un ámbito científico o social en el que la Estadística o la Investigación operativa sean una herramienta fundamental.



UNIVERSIDAD
DE GRANADA

**GRADO EN ESTADÍSTICA
PROPUESTA DE TEMA PARA TRABAJOS FIN DE GRADO
CURSO ACADÉMICO 2021/2022**

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.