

Título del trabajo: Aplicación de técnicas de calibración y modelización para la estimación del gasto turístico en España
Tutor/a: María del Mar Rueda García Correo electrónico: mrueda@ugr.es
Cotutor/a: Ramón Ferri García
Departamento responsable: Departamento de Estadística e Investigación Operativa
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): Laura Reyes Gálvez
Tipo de trabajo: 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.
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 <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: El desarrollo de nuevas técnicas para la realización de encuestas probabilísticas y no probabilísticas, unido al creciente interés en la integración de fuentes de datos de distintas procedencias, ha propiciado la aparición de una nueva clase de técnicas para mitigar los sesgos propios de su aplicación. Los habituales sesgos de cobertura y de falta de respuesta de las encuestas probabilísticas, y de autoselección en el caso de las no probabilísticas, pueden corregirse mediante métodos como la calibración, la reponderación a través de la estimación de las propensidades y las técnicas basadas en modelos predictivos. En este trabajo, se realizará un estudio y aplicación de dichas técnicas para la obtención de estadísticas de gasto turístico en España. En concreto, se emplearán datos procedentes de fuentes oficiales (encuesta EGATUR) y se estudiará el uso de otras fuentes alternativas con un gran potencial, pero cuya información es obtenida de forma no probabilística (p. ej. webs de reseñas). Los resultados obtenidos con las técnicas aplicadas, así como su precisión, se compararán con las estimaciones obtenidas por los organismos oficiales.

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 aplicar las metodologías existentes para la estimación a partir de muestras probabilísticas y no probabilísticas en un caso real de integración de datos de diversas fuentes, para así aprovechar toda la información auxiliar disponible que pueda ser utilizada por los métodos a aplicar. Los objetivos específicos son:

OE1. Uso de técnicas de calibración y estimación basada en modelos para para disminuir el error muestral en las estimaciones en subpoblaciones con tamaños muestrales pequeños, a partir de los modelos de aprendizaje más adecuados a los datos disponibles.

OE2. Uso de modelos de aprendizaje automático (Machine Learning) para predecir el gasto turístico de cada unidad de consumo en cada comunidad autónoma, y repartir así el gasto entre comunidades para realizar la estimación final de una forma más eficiente.

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. 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.
4. 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.
5. Deville, J. C., Särndal, C. E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*. 87 (418) 376–382.
6. Ferri-García, R., & Rueda, M. D. M. (2018). Efficiency of propensity score adjustment and calibration on the estimation from non-probabilistic online surveys. *SORT: statistics and operations research transactions*, 42(2), 159-182.
7. Valliant, R., Dorfman, A. H., & Royall, R. M. (2000). *Finite population sampling and inference: a prediction approach*. New York: John Wiley.
8. Wiśniowski, A., Sakshaug, J. W., Perez Ruiz, D. A., & Blom, A. G. (2020). Integrating probability and nonprobability samples for survey inference. *Journal of Survey Statistics and Methodology*, 8(1), 120-147.
9. 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*.

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.
- 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.