



<b>Título del trabajo:</b> Técnicas de machine learning para el tratamiento de sesgos en encuestas		
<b>Tutor/a:</b> María del Mar Rueda García		
<b>Cotutor/a:</b> Ramón Ferri García		
<b>Departamento responsable:</b> Estadística e Investigación Operativa		
<b>Perfil y número de estudiantes al que va dirigido (máximo 2):</b> 1 <i>(Rellenar sólo en caso de que la propuesta esté realizada a través de estudiante)</i>		
<b>Estudiante que propone el trabajo (Nombre, Apellidos, DNI):</b>		
<b>Tipo de trabajo</b> <i>(consultar (*))</i> 1		
<b>Competencias</b> <i>(estas son las mínimas; consultar (**) si se considera añadir otras)</i> <i>Competencias generales:</i> G02, G03, G04, G05, G06, G07, G08, G09. <i>Competencias específicas:</i> E01, E02, E03, E04, E05, E08, E09, E10.		
<b>Resultados de aprendizaje</b> <i>(estos son los mínimos; añadir otros si se considera)</i> <ul style="list-style-type: none"><li>• Adquirir competencias globales ligadas al desarrollo y aplicación de los conocimientos del Grado.</li><li>• Adquirir competencias ligadas a la búsqueda y organización de información y documentación relevante sobre el tema objeto de estudio.</li><li>• 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).</li><li>• Saber presentar, de forma escrita y oral, la memoria, los resultados y las conclusiones del trabajo realizado.</li></ul>		
<b>Antecedentes y resumen del tema propuesto:</b> <p>Las encuestas no probabilísticas realizadas a través de teléfonos móviles o internet tienen una cobertura insuficiente de importantes subgrupos de población dando lugar a estimaciones sesgadas. Además, se añade el problema de la no representatividad de la muestra.</p> <p>En este trabajo se estudiará el uso de un conjunto de ajustes post hoc como forma de reducir el sesgo en las estimaciones de muestras no probabilísticas; estos ajustes utilizan datos auxiliares en un esfuerzo por tratar con la selección y otros sesgos. El ajuste mediante Propensity Score Adjustment (PSA) desarrollado en Lee (2006) es el método más usado en la práctica. El alumno estudiará el uso de técnicas <i>de machine-learning</i> para ajustar probabilidades de participar y así corregir sesgos.</p> <p>Las técnicas estudiadas se pondrán en práctica con la elaboración de una encuesta por internet y el análisis de los resultados, en algún tema de interés para la comunidad.</p>		
<b>Breve descripción de las actividades presenciales y no presenciales a realizar:</b>		
Actividades presenciales (15-30%)	Planteamiento, orientación y supervisión	40 horas
	Exposición del trabajo	20 horas
	Otras:	
Actividades no presenciales (70-85%)	Preparación del trabajo	120 horas
	Elaboración de la memoria	120 horas
	Otras:	
Total (12 ECTS)		300 horas
<b>Objetivos que se pretenden alcanzar:</b> <p>El alumno debe ser capaz de encontrar y utilizar bibliografía científica relevante dentro del área de estudio para conocer las técnicas indicadas, y encontrar y familiarizarse con el software exis-</p>		



tente y ser capaz de utilizarlo para propósitos concretos. El alumno debe ser capaz de poner en práctica los métodos estudiados con su aplicación a una encuesta real.

**Bibliografía básica para la puesta en marcha del trabajo:**

1. Flach, Peter (2012) Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press. ISBN 978-1-107-42222-3.
2. Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of official statistics*, 22(2), 329.
3. Lee, S., & Vaillant, R. (2009): Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research*, 37,319-343
4. Schonlau, M., Van Soest, A., Kapteyn, A., & Couper, M. (2009). Selection bias in web surveys and the use of propensity scores. *Sociological Methods & Research*, 37(3), 291-318.
5. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ..., Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.

**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 2018/2019**

**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.