

PROGRAMA OFICIAL DE POSTGRADO EN ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

DATOS IDENTIFICATIVOS DE LA MATERIA

Código de la materia: 614493130

Nombre de la materia: TÉCNICAS DE REMUESTREO

Número de créditos ECTS: 5

Curso académico: 2020/2021

Profesorado:

Ricardo Cao Abad (coordinador, 3 créditos)

Rubén Fernández Casal (2 créditos)

OBJETIVOS DE LA MATERIA

Se pretende que el/la alumno/a adquiera destreza en la identificación de situaciones en las que los métodos de remuestreo son herramientas inferenciales adecuadas para resolver problemas reales. Para ello se tratará de que el/la alumno/a conozca el funcionamiento de las principales técnicas de remuestreo, entre las que se destaca el método bootstrap, así como sus aplicaciones en los principales ámbitos de la estadística. Asimismo se persigue que el/la alumno/a sea capaz de diseñar e implementar en ordenador planes de remuestreo adecuados para un amplio abanico de situaciones.

CONTENIDOS DE LA MATERIA

Tema 1. Motivación del principio Bootstrap.

Motivación del principio Bootstrap. El Bootstrap uniforme. Cálculo de la distribución Bootstrap: distribución exacta y distribución aproximada por Monte Carlo. Ejemplos. Herramientas disponibles en R. Computación en paralelo.

Tema 2. Aplicación a la estimación de la precisión y el sesgo de un estimador.

Aplicación del Bootstrap a la estimación de la precisión y el sesgo de un estimador. Ejemplos. El método Jackknife. Motivación del método Jackknife. Estimación Jackknife de la precisión y el sesgo de un estimador. Relación Bootstrap/Jackknife en dicha estimación. Ejemplos. Estudios de simulación.

Tema 3. Modificaciones del Bootstrap uniforme.

Bootstrap paramétrico, simetrizado, suavizado, ponderado y sesgado. Discusión y ejemplos. Validez de la aproximación Bootstrap. Ejemplos.

Tema 4. Aplicación del Bootstrap a la construcción de intervalos de confianza.

Métodos percentil, percentil-t, percentil-t simetrizado. Ejemplos. Estudios de simulación.

Tema 5. El Bootstrap en la estimación de la densidad.

Bootstrap y estimación no paramétrica de la densidad. Aproximación Bootstrap de la distribución del estimador de Parzen-Rosenblatt. El Bootstrap en la selección del parámetro de suavizado.

Tema 6. El Bootstrap en la estimación de la función de regresión.

El Bootstrap en Regresión y Correlación. Bootstrap y estimación no paramétrica de la función de regresión. Aproximación Bootstrap de la distribución del estimador de Nadaraya-Watson. Distintos métodos de remuestreo y resultados para ellos.

Tema 7. Aplicaciones del Bootstrap en contrastes de hipótesis.

Aproximación del p-valor mediante remuestreo. Contrastes bootstrap paramétricos. Contrastes de permutaciones. Contrastes bootstrap semiparamétricos.

Tema 8. El Bootstrap con datos censurados.

Introducción a los datos censurados. Remuestreos Bootstrap en presencia de censura. Relaciones entre ellos. Implementación en R.

Tema 9. El Bootstrap con datos dependientes.

Introducción a las condiciones de dependencia y modelos habituales de datos dependientes. Modelos paramétricos de dependencia. Situaciones de dependencia general: el Bootstrap por bloques, el Bootstrap estacionario y el método del submuestreo. Implementación en R. El bootstrap en Estadística Espacial.

BIBLIOGRAFÍA BÁSICA Y COMPLEMENTARIA

Básica

Cao, R. y Fernández-Casal, R. (2020). Técnicas de Remuestreo. Libro online: https://rubenfcasal.github.io/book_remuestreo.

- Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.
- Efron, B. (1979). Bootstrap Methods: Another look at the Jackknife. *Ann. Statist.*, 7, 1-26.
- Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer Verlag.

Complementaria

- Akritas, M. G. (1986). Bootstrapping the Kaplan--Meier estimator. *J. Amer. Statist. Assoc.* 81, 1032-1038.
- Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 12, 470-482.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3, 123-148.
- Cao, R. (1990). Órdenes de convergencia para las aproximaciones normal y bootstrap en la estimación no paramétrica de la función de densidad. *Trabajos de Estadística*, vol. 5, 2, 23-32.
- Cao, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Ann. Statist.* 19, 2226-2231.
- Cao, R. and Prada-Sánchez, J.M. (1993). Bootstrapping the mean of a symmetric population. *Statistics & Probability Letters* 17, 43-48.
- Cao, R. (1993). Bootstrapping the mean integrated squared error. *Jr. Mult. Anal.* 45, 137-160.
- Cao, R. (1999). An overview of bootstrap methods for estimating and predicting in time series. *Test*, 8, 95-116.
- Cao, R. and González-Manteiga, W. (1993). Bootstrap methods in regression smoothing. *J. Nonparam. Statist.* 2, 379-388.
- Chernick, M. R. (2008). *Bootstrap methods: A guide for practitioners and researchers*. John Wiley & Sons.
- Chernick, M. R., and LaBudde, R. A. (2014). *An introduction to bootstrap methods with applications to R*. John Wiley & Sons.
- Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* 76, 312-319.
- Efron, B. (1982). The Jackknife, the Bootstrap and other Resampling Plans. CBMS-NSF. Regional Conference series in applied mathematics.
- Efron, B. (1983). Estimating the error rate of a prediction rule: improvements on cross-validation. *J. Amer. Stat. Assoc.* 78, 316-331.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 54-77.

- Efron, B. (1987). Better Bootstrap confidence intervals (with discussion), *J. Amer. Stat. Assoc.* 82, 171-200.
- Efron, B. (1990). More Efficient Bootstrap Computations. *J. Amer. Stat. Assoc.* 85, 79-89.
- Freedman, D.A. (1981). Bootstrapping regression models. *Ann. Statist.* 9, 6, 1218-1228.
- García-Jurado, I. González-Manteiga, W., Prada-Sánchez, J.M., Febrero-Bande, M. and Cao, R. (1995). Predicting using Box-Jenkins, nonparametric and bootstrap techniques. *Technometrics* 37, 303-310.
- Hall, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* 14, 1431-1452.
- Hall, P. (1988-a). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* 16, 927-953.
- Hall, P. (1988-b). Rate of convergence in bootstrap approximations. *Ann. Probab.* 16, 4, 1665-1684.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Verlag.
- Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* 19, 778-796.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* 17, 1217-1241.
- Mammen, E. (1992). *When does Bootstrap Work?*. Springer Verlag.
- Navidi, W. (1989). Edgeworth expansions for bootstrapping regression models. *Ann. Statist.* 17, 4, 1472-1478.
- Politis, D.N. and Romano, J.R. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* 89, 1303-1313.
- Politis, D.N. and Romano, J.R. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statist. Sin.* 4, 461-476.
- Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling*. Springer Verlag.
- Reid, N. (1981). Estimating the median survival time. *Biometrika* 68, 601-608.
- Stine, R.A. (1987). Estimating properties of autoregressive forecasts. *J. Amer. Statist. Assoc.* 82, 1072-1078.
- Thombs, L.A. and Schucany, W.R. (1990). Bootstrap prediction intervals for autoregression. *J. Amer. Statist. Assoc.* 85, 486-492.
- Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* 14, 1261-1350.

COMPETENCIAS GENERALES Y ESPECÍFICAS

Competencias generales:

- G1 - Capacidad para iniciar la investigación y para participar en proyectos de investigación que pueden culminar en la elaboración de una tesis doctoral.
- G2 - Capacidad de aplicación de algoritmos de resolución de los problemas y manejo del software adecuado.
- G3 - Capacidad de trabajo en equipo y de forma autónoma
- G4 - Capacidad de formular problemas en términos estadísticos, y de resolverlos utilizando las técnicas adecuadas.
- G6 - Capacidad de identificar y resolver problemas
- G10 - Capacidad de integrarse en un equipo multidisciplinar para el análisis experimental
- G11 - Adquirir destreza para el desarrollo de software
- G12 - Capacidad de análisis estadístico crítico de las muestras, los planteamientos y resultados
- G13 - Redacción de informes estadísticos con precisión, orden y claridad
- G14 - Representar un problema real mediante un modelizado estadístico adecuado.
- G15 - Diseñar un plan de observación o recogida de datos que permita abordar el problema de interés

Competencias específicas:

- E2 - La adquisición de los conocimientos de estadística e investigación operativa necesarios para la incorporación en equipos multidisciplinarios pertenecientes a diferentes sectores profesionales.
- E3 - Capacidad para comprender, plantear, formular y resolver aquellos problemas susceptibles de ser abordados a través de modelos de la estadística y de la investigación operativa.
- E4 - Conocer las aplicaciones de los modelos de la estadística y la investigación operativa.
- E5 - Conocer algoritmos de resolución de los problemas y manejar el software adecuado.
- E11 - Modelar la dependencia entre una variable respuesta (dependiente) y varias variables explicativas (independientes).
- E12 - Realizar inferencias respecto a los parámetros que aparecen en el modelo.
- E19 - Tratamiento de datos y análisis estadístico de los resultados obtenidos.
- E22 - Capacidad de identificar y resolver problemas que requieran el uso de técnicas del análisis de series de tiempo.
- E27 - Obtener los conocimientos precisos para un análisis crítico y riguroso de los resultados.
- E28 - Complementar el aprendizaje de los aspectos metodológicos con apoyo de software.
- E78 - Fomentar la sensibilidad hacia los principios del pensamiento científico, favoreciendo las actitudes asociadas al desarrollo de los métodos matemáticos, como: el cuestionamiento de las ideas intuitivas, el análisis crítico de las afirmaciones, la capacidad de análisis y síntesis o la toma de decisiones racionales

E82 - El estudiante será capaz de comprender la importancia de la Inferencia Estadística como herramienta de obtención de información sobre la población en estudio, a partir del conjunto de datos observados de una muestra representativa de ésta. Para ello deberá reconocer la diferencia entre estadística paramétrica y no paramétrica.

E84 - Ser capaz de manejar diverso software (en particular R) e interpretar los resultados que proporcionan éstos en los correspondientes estudios prácticos.

E86 - Soltura en el manejo de la teoría de la probabilidad y las variables aleatorias.

METODOLOGÍA DOCENTE: ACTIVIDADES DE APRENDIZAJE Y SU VALORACIÓN EN CRÉDITOS ECTS

Tres quintas partes de la docencia presencial se impartirán mediante exposiciones orales del profesor mientras que el resto corresponderá a prácticas, propuestas por el profesor (se hará uso del paquete estadístico R; por tanto, es necesario que los alumnos dispongan en el aula de un ordenador). El total de ambas actividades tendrá una valoración de 2 créditos ECTS. Los 3 créditos ECTS restantes corresponderán a estudio personal (2 créditos) y realización de trabajos en grupo que el alumno tendrá que elaborar a lo largo del curso (1 créditos).

CRITERIOS Y MÉTODOS DE EVALUACIÓN

La evaluación se realizará por medio de prácticas en R, un trabajo en grupo del/a alumno/a, así como una prueba escrita de conceptos. La calificación de la prueba de conceptos representará el 40% de la calificación global, las pruebas de prácticas en R corresponderá al 40% mientras que el 20% restante corresponderá al trabajo en grupo, que ha de ser presentado en público por los alumnos.

En la siguiente tabla se adjunta una lista de las metodologías que se utilizarán para las diferentes competencias.

EVALUACIÓN			
Metodologías	Competencias	Descripción	Calificación
Prácticas a través de TIC	E3 E5 E19 E28 E84 G2 G11	Prácticas con R	40
Prueba de respuesta múltiple	E2 E4 E11 E12 E22 E27 E78 E82 E86 G4 G6 G14	Prueba de comprensión de los conceptos impartidos.	40
Trabajo en grupo	E3 E11 E19 E22 E27 E78 G1 G2 G3 G4 G6 G11 G12 G13 G14 G15	Trabajo original sobre el bootstrap en algún contexto de interés que se presentará en público	20

TIEMPO DE ESTUDIO Y DE TRABAJO PERSONAL QUE DEBE DEDICAR UN ESTUDIANTE PARA SUPERAR LA MATERIA

Docencia presencial: 35 h (24 h de lección magistral, 11 h de prácticas con R)

Tutorías (presenciales o bien online): 12 h

Realización de pruebas de evaluación: 3 h.

Estudio y trabajo personal: 75 h

RECOMENDACIONES PARA EL ESTUDIO DE LA MATERIA

Es necesario que el alumno posea conocimientos (teóricos y aplicados) básicos de Estadística. Para superar con éxito la materia es aconsejable la asistencia a las clases, siendo fundamental el seguimiento diario del trabajo realizado en el aula. Disponer de conocimientos generales del entorno estadístico R facilitará el aprovechamiento de las clases y la realización de las prácticas.

RECURSOS PARA EL APRENDIZAJE

Bibliografía, apuntes y ordenador. Uso del repositorio de material docente del máster, de los apuntes online (https://rubenfcasal.github.io/book_remuestreo; disponible en el repositorio Github: https://github.com/rubenfcasal/book_remuestreo) y del software estadístico R.

OBSERVACIONES

Un prerequisite necesario es haber cursado al menos una de las materias *Estadística Aplicada* o *Estadística Matemática* del presente máster. Es recomendable (aunque no necesario) que el alumno haya cursado la materia optativa de *Simulación Estadística*.

El desarrollo de los contenidos de la materia se realizará teniendo en cuenta que las competencias a adquirir por el alumnado deben cumplir con el nivel MECES3. Los contenidos que se incluyen en esta materia, orientada a las técnicas de remuestreo, se estudiarán tanto de manera teórica como aplicada las principales aplicaciones del bootstrap a la inferencia estadística, tanto en problemas paramétricos, como no paramétricos.

Para los casos de realización fraudulenta de ejercicios o pruebas, será de aplicación lo recogido en las respectivas normativas de las universidades participantes en el Máster en Técnicas Estadísticas.

Esta guía y los criterios y metodologías en ella descritos están sujetos a las modificaciones que se deriven de normativas y directrices de las universidades participantes en el Máster en Técnicas Estadísticas.

COVID19.

Debido a la incidencia de la pandemia en España la docencia se plantea como híbrida (en parte presencial y en parte telemática). Para la docencia telemática se empleará Microsoft Teams. Los alumnos podrán recibir sus clases desde sus lugares de residencia, aunque también podrán acudir a las aulas del máster si su situación personal lo requiere, mientras no se supere la capacidad del aula y así lo contemple la universidad en la que se haya matriculado. Igualmente, los profesores, podrán impartir clase desde sus domicilios, despachos de la universidad o desde el aula del máster en la UDC.

Salvo que el número de alumnos matriculados impida respetar la distancia interpersonal, las pruebas que formen parte de la evaluación serán llevadas a cabo de forma presencial.

La metodología docente y los criterios de evaluación expuestos en esta guía docente se utilizarán independientemente del grado de presencialidad bajo el que se imparta la asignatura. En caso de que sea necesario suspender la docencia presencial, esta se impartirá de forma telemática (expositiva e interactiva) con sesiones síncronas y tutorías a través de Teams combinadas con material electrónico (videos y apuntes en formato electrónico). Los porcentajes de calificación de los trabajos prácticos y del examen escrito podrían verse modificados para adaptarse a las eventuales directrices de las universidades de A Coruña, Santiago y Vigo.