Teaching Guide				
Identifying Data				2021/22
Subject (*)	Resampling Techniques		Code	614493130
Study programme	Mestrado Universitario en Técnicas Estadísticas (Plan 2019)			
Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	Second	Optional	5
Language	SpanishGalicianEnglish			
Teaching method	Hybrid			
Prerequisites				
Department	Matemáticas			
Coordinador	Cao Abad, Ricardo	E-mail	ricardo.cao@udc.es	
Lecturers	Cao Abad, Ricardo Fernández Casal, Rubén	E-mail	ricardo.cao@udc.es ruben.fcasal@udc.es	
Web	rubenfcasal.github.io/book_remuestreo			
General description	It is intended that students acquire skills in identifying situations in which resampling methods are adequate to solve real problems with inferential tools. To do this students will have to know how the main resampling techniques can be used. This includes the bootstrap method and its applications in major areas in statistics. It is also intended that the student is able to design and implement in a computer appropriate resampling plans for a wide range of situations.			
Contingency plan	Depending on the evolution of the pandemic, teaching could be hybrid (part face-to-face and part telematics). Microsoft Teams would be used for telematic teaching. Students would be able to receive their classes from their places of residence, although they would also be able to go to the master's classrooms if their personal situation requires it, as long as the capacity of the classroom is not exceeded and so is contemplated by the university in which they have enrolled. Likewise, professors would be able to teach from their homes, university offices or from the master's classroom at UDC. Unless the number of students enrolled prevents respect for interpersonal distance, the tests that are part of the evaluation will be carried out in person. The teaching methodology and the evaluation criteria set out in this teaching guide will be used regardless of the degree of attendance under which the subject is taught. In the event that it is necessary to suspend the face-to-face teaching, this will be delivered telematically (expository and interactive) with synchronous sessions and tutorials through Teams combined with electronic material (videos and notes in electronic format). The percentages of qualification of the practical work and the written exam could be modified to adapt to the possible guidelines of the universities of A Coruña, Santiago and Vigo.			

Study programme competences / results	
Code	Study programme competences / results
A16	CE1 - Coñecer, identificar, modelar, estudar e resolver problemas complexos de estatística e investigación operativa, nun contexto científico, tecnolóxico ou profesional, xurdidos en aplicacións reais.
A18	CE3 - Adquirir coñecementos avanzados dos fundamentos teóricos subxacentes ás distintas metodoloxías da estatística e a investigación operativa, que permitan o seu desenvolvemento profesional especializado.
A19	CE4 - Adquirir as destrezas necesarias no manexo teórico-práctico da teoría de probabilidade e as variables aleatorias que permitan o seu desenvolvemento profesional no eido científico/académico, tecnolóxico ou profesional especializado e multidisciplinar.
A20	CE5 - Profundizar no coñecemento dos fundamentos teórico-prácticos especializados de modelado e estudo de distintos tipos de relacións de dependencia entre variables estatísticas.
A21	CE6 - Adquirir coñecementos teórico-prácticos avanzados de distintas técnicas matemáticas, orientadas específicamente á axuda na toma de decisións, e desenvolver a capacidade de reflexión para avaliar e decidir entre distintas perspectivas en contextos complexos.
A23	CE8 - Adquirir coñecementos teórico-prácticos avanzados das técnicas destinadas á realización de inferencias e contrastes relativos a variables e parámetros dun modelo estatístico, e saber aplicalos con autonomía suficiente nun contexto científico, tecnolóxico ou profesional.
A24	CE9 - Coñecer e saber aplicar con autonomía en contextos científicos, tecnolóxicos ou profesionais, técnicas de aprendizaxe automático e técnicas de análise de datos de alta dimensión (big data).

A25	CE10 - Adquirir coñecementos avanzados sobre metodoloxías para a obtención e o tratamento de datos derivados de distintas fuentes, como enquisas, internet, ou entornos ?na nube".
B1	CB6 - Posuír e comprender coñecementos que acheguen unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación
B2	CB7 - Que os estudantes saiban aplicar os coñecementos adquiridos e a súa capacidade de resolución de problemas en ámbitos novos ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
B3	CB8 - Que os estudantes sexan capaces de integrar coñecementos e enfrontarse á complexidade de formular xuízos a partir dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos
B4	CB9 - Que os estudantes saiban comunicar as súas conclusións e os coñecementos e razóns últimas que as sustentan a públicos especializados e non especializados dun modo claro e sen ambigüidades
B5	CB10 - Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun modo que haberá de ser en gran medida autodirixido ou autónomo.
B17	CG1 - Coñecer, comprender e saber aplicar os principios, metodoloxías e novas tecnoloxías na estatística e a investigación operativa en contextos científico/académicos, tecnolóxicos ou profesionais especializados e multidisciplinares, así como adquirir as destrezas e competencias descritas nos objectivos generales do título.
B18	CG2 - Desenvolver autonomía para identificar, modelar e resolver problemas complexos da estatística e da investigación operativa en contextos científico/académicos, tecnolóxicos ou profesionais especializados e multidisciplinares.
B19	CG3 - Desenvolver a capacidade para realizar estudos e tarefas de investigación e transmitir os resultados a públicos especializados, académicos e xeneralistas.
B20	CG4 - Integrar coñecementos avanzados e enfrontarse á toma de decisións a partir de información científica e técnica.
B21	CG5 - Desenvolver a capacidade de aplicación de algoritmos e técnicas de resolución de problemas complexos no eido da estatística e a investigación operativa, manexando o software especializado axeitado.
C11	CT1 - Desenvolver firmes capacidades de razoamento, análise crítica e autocrítica, así como de argumentación e de síntese, contextos especializados e multidisciplinais.
C12	CT2 - Desenvolver destrezas avanzadas no manexo de Tecnoloxías da Información e a Comunicación (TIC), tanto para a obtención de información como para a difusión do coñecemento, nun ámbito científico/académico, tecnolóxico ou profesional especializado e multidisciplinar.
C13	CT3 - Ser capaz de resolver problemas complexos en novos escenarios mediante a aplicación integrada dos coñecementos.
C14	CT4 - Desenvolver unha sólida capacidade de organización e planificación do estudo, asumindo a responsabilidade do seu propio desenvovemento profesional, para a realización de traballos en equipo e de xeito autónomo.
C15	CT5 - Desenvolver capacidades para o aprendizaxe e a integración no traballo en equipos multidisciplinais, nos ámbitos científico/académico, tecnolóxico e profesional.

Learning outcomes			
Learning outcomes	Study programme competences / results		
Coñecer os fundamentos teóricos das técnicas de remuestreo.	AC16 AC18 AC19 AC20 AC21 AC23 AC24 AC25	BJ1 BJ2 BJ3 BJ4 BJ5 BJ17 BJ18 BJ19 BJ20 BJ21	CJ11 CJ12 CJ13 CJ14 CJ15

Saber aplicar de xeito autónomo os principios do bootstrap aos principais problemas de inferencia estatística.	AC16 AC18 AC19 AC20 AC21 AC23 AC24 AC25	BJ1 BJ2 BJ3 BJ4 BJ5 BJ17 BJ18 BJ20 BJ21	CJ11 CJ12 CJ13 CJ14 CJ15
Ser capaz de deseñar e validar algoritmos bootstrap para a resolución de problemas de inferencia non paramétrica sobre as funcións de densidade e regresión.	AC16 AC18 AC19 AC20 AC21 AC23 AC24 AC25	BJ1 BJ2 BJ3 BJ4 BJ5 BJ17 BJ18 BJ19 BJ20 BJ21	CJ11 CJ12 CJ13 CJ14 CJ15

Contents	
Topic	**Sub-topic**
1. Motivation of the Bootstrap principle.	Uniform bootstrap. Bootstrap distribution calculation: exact distribution and Monte Carlo approximation. Examples. Tools available in R. Parallel computing.
2. Application to the estimation of the precision and bias of an estimator.	Application of the Bootstrap to estimate the precision and the bias of an estimator. Examples. The Jackknife method. Motivation of the Jackknife method. Jackknife estimation of the precision and bias of an estimator. Bootstrap / Jackknife relationship in these estimation problems. Examples. Simulation studies.
3. Variations of the uniform Bootstrap.	Parametric Bootstrap, symmetrized Bootstrap, smoothed Bootstrap, weighted Bootstrap and biased Bootstrap. Discussion and examples. Validity of the Bootstrap approach. Examples.
4. Applications of Bootstrap to construct confidence intervals.	Percentile method, percentile-t method, symmetrized percentile-t method . Examples. Simulation studies.
7. Bootstrap applications in hypothesis testing.	P-value approximation by resampling. Parametric bootstrap tests. Permutations tests. Semi-parametric bootstrap tests.
5. Bootstrap and nonparametric density estimation.	Bootstrap approximation for the distribution of the Parzen-Rosenblatt estimator. The Bootstrap in the selection of the smoothing parameter.
6. Bootstrap for regression function estimation.	The Bootstrap in Regression and Correlation. Bootstrap and nonparametric estimation of the regression function. Bootstrap approximation of the distribution of the Nadaraya-Watson estimator. Different resampling methods and results.
8. Bootstrap for censored data.	Introduction to censored data. Bootstrap resampling plans in the presence of censorship. Relations among them. Implementation in R.
9. Bootstrap with dependent data.	Introduction to the usual conditions of dependency and dependent data models. Parametric models of dependence. General dependence situations: Moving Block Bootstrap, Stationary Bootstrap and Subsampling method. Implementation in R. The bootstrap in Spatial Statistics.

Planning				
Methodologies / tests	**Competencies / Results**	**Teaching hours (in-person & virtual)**	**Student?s personal work hours**	**Total hours**

Oral presentation	A7 A13 B3 B4 B5 B8 C9 C11 C15	21	31.5	52.5
ICT practicals	A19 A21 A24 A25 B1 B2 B17 B19 B20 C12 C14 C15	14	28	42
Multiple-choice questions	A9 A11 A12 A14 A15 A16 A20 A23 B3 B9 B10 B15 B16 B18 B21 C6 C13	1	11.5	12.5
Problem solving	A18 B5 C11 C14 C15	4	8	12
Personalized attention		6	0	6
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	**Description**
Oral presentation	Presentation with computer by teleconference
ICT practicals	Resampling algorithm implementation
Multiple-choice questions	Multiple-choice test on concepts.
Problem solving	Design of resampling plans. Bias and variance calculation for the bootstrap analogues.

Personalized attention	
Methodologies	**Description**
ICT practicals Problem solving	Attendance and participation in lectures. Written multiple choice test. Participation in workshops and seminars. Practicals to be performed by the student.

Assessment			
Methodologies	**Competencies / Results**	**Description**	**Qualification**
ICT practicals	A19 A21 A24 A25 B1 B2 B17 B19 B20 C12 C14 C15	Using the software R to implement the bootstrap method in some setup.	40
Problem solving	A18 B5 C11 C14 C15	Original work on the bootstrap on some relevant setup.	10
Multiple-choice questions	A9 A11 A12 A14 A15 A16 A20 A23 B3 B9 B10 B15 B16 B18 B21 C6 C13	Comprehension Test.	40
Oral presentation	A7 A13 B3 B4 B5 B8 C9 C11 C15	Presentation of the original work on the bootstrap on some relevant setup.	10

Assessment comments

The assessment will be carried out using R practicals, a student work (in groups), as well as a written concept test. The concept test score will be 40% of the total qualification, the R practicals will correspond to 40% of the global score, while the remaining 20% will correspond to the student work (in groups), that has to be presented orally. To pass the subject is necessary to obtain a score of at least 5 out of 10 overall. On July opportunity, students could avoid those tests with scores of at least 4 out of 10 in January tests. Only students that didn't take any test will be qualified as NON ATTENDANT in the first opportunity (January-February). In July opportunity only students that didn't take the final exam will be qualified as NON ATTENDANT.

Sources of information	
Basic	Bibliografía básicaCao, R. y Fernández-Casal, R. (2020). Técnicas de Remuestreo. Libro online: https://rubenfcasal.github.io/book_remuestreo.Davison, A.C. and Hinkley, D.V. (1999). 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. (1996). The Jackknife and Bootstrap. Springer Verlag.
Complementary	Bibliografía complementariaAkritas, 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. (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.Cao, R. and Prada-Sánchez, J.M. (1993). Bootstrapping the mean of a symmetric population. Statistics & Probability Letters 17, 43-48.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. (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.Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1, 54-77.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. (1988a). Theoretical comparison of bootstrap confidence intervals. Ann. Statist. 16, 927-953.Hall, P. (1988b). Rate of convergence in bootstrap approximations. Ann. Probab. 16, 4, 1665-1684.Hall. P. (1992). The Bootstrap and Edgeworth Expansion. Springer Verlag.Hall, P. and Martin, M.A. (1988). On bootstrap resampling and iteration. Biometrika 75, 661-671.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. (1994a). The stationary bootstrap. J. Amer. Statist. Assoc. 89, 1303-1313.Politis, D.N. and Romano, J.R. (1994b). 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.

Recommendations

Subjects that it is recommended to have taken before
Estatística Matemática/614468102
Modelos de Probabilidade/614468103
Estatística Aplicada/614468104
Modelos de Regresión/614468105
Análise Exploratoria de Datos (data mining)/614468106
Estatística non Paramétrica/614468109
Simulación Estatística/614468113
Subjects that are recommended to be taken simultaneously
Series de Tempo/614427111
Fiabilidade e Modelos Biométricos/614427116
Subjects that continue the syllabus
Contrastes de Especificación/614468123
Datos Funcionais/614468124
Proxecto Fin de Carreira ou Traballo Tutelado/614468128
Other comments

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.