

PROGRAMA DE CURSO

Código	Nombre			
MA5501	Estadística Computacional			
Nombre en Inglés				
Computational Statistics				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
5.0	10	3.0	1.5	5.5
Requisitos			Carácter del Curso	
MA3401/MA3401/MA34A			Electivo (Licenciatura/Magíster Ing. Mat.)	
Resultados de Aprendizaje				
<p>El alumno aprende métodos computacionales para el análisis estadístico de datos y su aplicación. Se estudia modelos lineales generalizados, junto con métodos de clasificación, estimación de distribuciones, procesamiento y filtraje de datos, análisis multivariado y técnicas de simulación. Durante el curso, el alumno obtiene capacidad en el lenguaje estadístico de código abierto R.</p>				

Metodología Docente	Evaluación General
<p>Clases de cátedra expositivas. Clases auxiliares dedicadas a resolver los problemas.</p>	<p>2 controles parciales y un examen final.</p>

Resumen de Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Introducción a R y Procesamiento de Datos	4.0
2	Modelos Lineales Generalizados	4.0
3	Corrección para Test Múltiple	1.5
4	Clasificación y Clustering	4.0
5	Perspectivas	1.5
TOTAL		15.0

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas	
1	Introducción a R y procesamiento de datos	4.0	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1.1 Tipos de variables y su manipulación. Vectores, marices, arreglos, Listas y data frames. 1.2 Importación, exportación, cargando y guardando conjuntos de datos. 1.3 Aritmética y cálculos básicos en R. 1.4 Gráficos. 1.5 Funciones, estructura de programas en R, control de flujo. 1.6 Estadísticas básicas in R		El estudiante aprende los elementos de programación en R.	2, 3, 4, 12

Número	Nombre de la Unidad	Duración en Semanas	
2	Modelos lineales generalizados	4.0	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
2.1 Revisión de modelos lineales clásicos como fundación de modelos más generalizados. 2.2 Definiciones y teoría básica de modelos lineales generalizados. 2.3 Ejemplos de modelos lineales generalizados, ej. Binomial, Poisson, exponencial, Gamma. 2.4 Diagnósticos y tests para determinar la bondad de ajuste.		El estudiante aprende de modelos lineales generalizados y como aplicarlos a datos usando R.	3, 10, 11

Número	Nombre de la Unidad	Duración en Semanas	
3	Corrección para test múltiple	1.5	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
3.1 Revisión de tests de hipótesis. 3.2 Corrección de Bonferroni, Método de Sidák, método de Holm-Bonferroni, false discovery rate (FDR).		El estudiante comprende por qué hay que compensar para múltiple tests y métodos para hacerlo.	13

Número	Nombre de la Unidad	Duración en Semanas	
4	5.1 Clasificación	4.0	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
4.1 CART 4.2 k-means clustering 4.3 Support Vector Machines 4.4 k-Nearest Neighbors 4.5 Perceptrones 4.6 Clustering Jerárquico		El estudiante aprende métodos de clasificación y como usar R para aplicarlos.	6, 7, 8, 9

Número	Nombre de la Unidad	Duración en Semanas	
5	5.2 Perspectivas	1.5	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
5.3 Distribuciones de cola larga 5.4 Correlación v/s causalidad 5.5 Introducción a los sistemas caóticos 5.6 Tamaño de la muestra y nivel de significación		El propósito de esta unidad es reflexionar acerca de los límites de validez del análisis estadístico y cómo verificar que las hipótesis requeridas se cumplen.	1

Bibliografía

1. Peter Dalgaard. *Introductory Statistics with R*. Springer, 2nd edition, 2008. ISBN 978-0-387-79053-4.
2. Deepayan Sarkar. *Lattice: Multivariate Data Visualization with R*. Springer, New York, 2008. ISBN 978-0-387-75968-5.
3. William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S. Fourth Edition*. Springer, New York, 2002. ISBN 0-387-95457-0.
4. John Maindonald and John Braun. *Data Analysis and Graphics Using R*. Cambridge University Press, Cambridge, 2nd edition, 2007. ISBN 978-0-521-86116-8.
5. Paul Teetor. *R Cookbook*, first edition. O'Reilly, 2011. ISBN 978-0-596-80915-7.
6. Hastie, Tibshirani and Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2008.
7. Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification*, 2nd Edition. Wiley, 2001. ISBN: 978-0-471-05669-0
8. Breiman L, Friedman J, Ohlsen R, Stone C: *Classification and regression trees*. hWadsworth Advanced Books and Software, 1984.
9. Jorge Luis Borges, *Funes el memorioso*, en *Ficciones*, 1944.
10. Brian S. Everitt and Torsten Hothorn. *A Handbook of Statistical Analyses Using R*. Chapman and Hall, Boca Raton, Florida, 2006.
11. Annette J. Dobson. *An Introduction to Generalized Linear Models*. Chapman and Hall, London, 1990.
12. W. N. Venables and D. M. Smith. *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics*. <http://cran.r-project.org/doc/manuals/R-intro.html>.
13. Warren Ewens and Gregory Grant. *Statistical Methods in Bioinformatics*. Springer, New York, 2nd edition, 2005.

Vigencia desde:	2011
Elaborado por:	Andrés Aravena y Andrew Hart
Revisado por:	Jefe Docente - Iván Rapaport