



ENTORNOS DE COMPUTACIÓN ESTADÍSTICA

Máster en Estadística Aplicada. Curso 2021-22

DESCRIPCIÓN

El desarrollo de la Estadística ha ido siempre ligado al de la Informática, la cual proporciona herramientas que si bien pueden ser demandadas por otras áreas de la ciencia, es indudable que en la Estadística ofrecen un instrumento de trabajo indispensable para la correcta aplicación de las técnicas estadísticas. La Estadística es una ciencia aliada de la investigación científica, Neyman (1955) proclamó a la Estadística como *“la sirvienta de todas las ciencias”*. Mediante la Estadística se evalúan cuantitativamente hipótesis de investigación, se desarrollan modelos predictivos, se estiman parámetros y se analizan experimentos, entre otras aplicaciones. Muchos conocimientos implican procesar y combinar datos de diversas maneras, tanto numérica como simbólicamente y la Informática nos permite realizar estos cálculos y gestionar los datos convenientemente, estudiando cómo se puede representar con provecho la información o conocimientos y prestando especial atención a los límites de lo que puede ser convenientemente manipulado desde el punto de vista computacional.

El rápido y sostenido incremento en la potencia de cálculo de la computación desde la segunda mitad del siglo XX ha tenido un sustancial impacto en la práctica de la Estadística. La complejidad de los computadores unida a los algoritmos numéricos apropiados, han causado un renacer del interés en nuevos métodos y modelos estadísticos complicados que antes eran descartados debido a la intensidad de cálculo que requerían. La revolución en los computadores tiene implicaciones en el futuro de la Estadística, con un nuevo énfasis en estadísticas experimentales. Actualmente existe un gran número de programas estadísticos disponibles para los investigadores, algunos de ellos genéricos y otros más específicos para tareas concretas. Así mismo existen gran cantidad de programas libres (*free software*) como el entorno de análisis y programación estadística R, de difusión gratuita y código abierto. Dado que diferentes programas implementan distintos algoritmos para llevar a cabo los mismos tipos de análisis, los usuarios se benefician de una comparación entre los programas más usados.

La asignatura **Entornos de Computación Estadística** en el **Máster en Estadística Aplicada** se estructura según el programa de contenidos que se recoge más adelante. En términos generales se persigue una formación del estudiante en el manejo de herramientas informáticas adecuadas para el tratamiento de datos procedentes de

cualquier disciplina. Además, se pretende capacitar al estudiante para evaluar los distintos programas actualmente disponibles en el mercado, realizando una comparativa de las características y facilidades que cada uno ofrece para la resolución de problemas concretos.

REQUISITOS Y RECOMENDACIONES

No existen requisitos para la matrícula en esta asignatura. No obstante se trabaja tomando como base métodos estadísticos desde niveles básicos de la Estadística Descriptiva, Cálculo de Probabilidades e Inferencia Estadística, hasta métodos más complejos del Análisis Multivariante. En este sentido se recomienda poseer unos conocimientos previos de Estadística de nivel medio, por ejemplo haber cursado alguna titulación como Diplomado en Estadística, Licenciado en Ciencias y Técnicas Estadísticas y Grado en Estadística o Grado en Matemáticas. En otro caso es necesario haber cursado alguna asignatura básica de Estadística existente en la mayor parte de las titulaciones de ciencias experimentales. El aprendizaje del estudiante en las herramientas estadísticas computacionales se adaptará al nivel de conocimientos estadísticos que este posea. En cuanto a las habilidades computacionales, igualmente el desarrollo de la asignatura se adapta a los conocimientos previos del estudiante y a sus intereses particulares. De este modo se presenta una asignatura que persigue un aprendizaje significativo del estudiante en la materia teniendo en cuenta su nivel e intereses.

OBJETIVOS PARTICULARES

- Conocer la evolución histórica de la Estadística Computacional y prever su desarrollo.
- Conocer la implicación de la Estadística Computacional en el desarrollo de la Estadística y viceversa.
- Conocer la metodología de análisis estadístico computacional.
- Conocer un programa de computación estadística de tipo 1 (SPSS, SAS).
- Introducción a los lenguajes de programación orientada a objetos (POO).
- Modificación de hojas de datos residentes en un archivo secuencial.
- Aplicaciones estadísticas con Visual Basic
- Modificación de hojas de datos residentes en una base de datos
- Conocer el entorno de programación y análisis estadísticos R.
- Desarrollar análisis de datos con R así como programación un nivel medio.
- Resolver problemas clásicos de la Estadística con varios programas (SPSS, SAS, R etc.).
- Diseñar un curso básico de Estadística práctica asistida por R y/u otro software.
- Realizar informes estadísticos en R utilizando los libros Sweave o Knitr.

La asignatura se estructura en seis bloques temáticos:

- **Bloque 1.** Computación Estadística y Estadística Computacional. Evolución histórica.

- **Bloque 2.** Lenguajes de programación: Visual Basic.
- **Bloque 3.** Metodología del Análisis Estadístico Computacional con SPSS.
- **Bloque 4.** Metodología del Análisis Estadístico Computacional con los entornos de análisis y programación estadística R y S+.
- **Bloque 5.** Casos prácticos: análisis estadístico de datos reales. Comparación de software.
- **Bloque 6.** Resolución práctica de problemas clásicos de la Estadística con R.
 - Contrastes de hipótesis e intervalos de confianza paramétricos y no paramétricos.
 - Análisis de regresión.
 - Métodos estadísticos multivariantes.

PROGRAMA DETALLADO

Bloque 1. Computación Estadística y Estadística Computacional. Evolución histórica.

Bloque 2. Lenguajes de programación: Visual Basic.

Tema 1. Introducción a Visual Basic.

Tema 2. Elementos del lenguaje.

Tema 3. Estructuras de datos: matrices.

Tema 4. Procedimientos y funciones.

Tema 5. Entrada y salida de datos: acceso secuencial, aleatorio y binario.

Tema 6. Acceso a bases de datos.

Bloque 3. Metodología del Análisis Estadístico Computacional con SPSS.

Tema 1. Introducción a SPSS.

Tema 2. Descripción de SPSS.

Tema 3. Lenguaje de órdenes de SPSS.

Tema 4. Programación con Visual Basic para aplicaciones en SPSS.

Tema 5. Algunos análisis estadísticos con SPSS.

Bloque 4. El entorno de análisis y programación estadística R.

Tema 1. R un entorno de análisis y programación estadísticos.

Tema 2. Algunas clases de objetos comunes.

Tema 3. Funciones.

Tema 4. Uso de archivos externos.

Tema 5. Gráficos.

Tema 6. Fórmulas y modelos.

Tema 7. Análisis estadísticos elementales.

Bloque 5. Casos prácticos: análisis estadístico de datos reales. Comparación de software.

Bloque 6. Resolución práctica de problemas clásicos de la estadística con R.

Tema 1. Contrastes de hipótesis e intervalos de confianza paramétricos y no paramétricos.

Tema 2. Análisis de la varianza.

Tema 3. Análisis de Regresión.

Tema 4. Métodos multivariantes: análisis factorial, análisis clúster, etc.

METODOLOGÍA

La asignatura se desarrolla de modo virtual. El trabajo del estudiante en esta asignatura viene definido por el interés que este tenga en la misma, por lo que se dirige desde tres posibles itinerarios de aprendizaje:

- Estadística computacional básica
- Estadística computacional para la docencia y
- Estadística computacional para la investigación

El estudiante deberá elegir uno de estos itinerarios para cursar la asignatura. La asignatura se concibe desde la idea de *“qué tengo que hacer y para ello qué necesito estudiar y aprender”*. Así, dependiendo de la elección, el estudiante profundizará en algunos de los bloques temáticos del programa general de contenidos. La elección del itinerario se hará al comienzo del curso, a través de la actividad 0, si bien el estudiante podrá cambiar justificadamente su elección a medida que desarrolla su trabajo.

A continuación se describe la orientación y metodología de cada itinerario.

Itinerario I (Estadística computacional básica)

Este itinerario está dirigido a estudiantes que no hayan cursado asignaturas de Estadística Computacional previamente. Desde la concepción de la asignatura antes presentada, el estudiante tiene que ser capaz de realizar unas actividades relativas a una parte de los contenidos. Se propone un programa de contenidos básicos centrado fundamentalmente en dos de los bloques temáticos antes descritos. En primer lugar, y

de forma obligatoria para todos los estudiantes de este itinerario, el bloque relativo al aprendizaje del lenguaje de programación y un programa estadística R (bloque 4). Y en segundo lugar, el estudiante deberá elegir entre uno de los dos bloques siguientes: SPSS (bloque 3) o Visual Basic (bloque 2).

Para cada bloque de contenidos, se ofrecen al estudiante a través de la plataforma materiales para el trabajo personal. Los materiales están diseñados para el estudio en el ordenador y contienen actividades de autoaprendizaje que el estudiante irá realizando con el *software* concreto. Además se proponen un total de dos actividades para la evaluación final de la asignatura (más detalles en el apartado evaluación). Estas actividades se entregarán en los plazos fijados y a través de los enlaces habilitados para ello en la plataforma virtual.

Itinerario II (Estadística computacional para la docencia)

Dirigido a estudiantes con intereses profesionales en la enseñanza de la Estadística (a cualquier nivel: secundaria, universidad etc.). Se pretende que el estudiante diseñe un material docente computacional que permita la realización de prácticas en ordenador para el alumnado de un curso de estadística básico (o posiblemente de una temática más específica). En este sentido se requiere que el estudiante tenga la formación teórica en la temática estadística concreta y se pretende capacitarlo para una adecuada elaboración de materiales prácticos.

Para la consecución de los objetivos planteados, el estudiante que elija este itinerario deberá profundizar en los contenidos recogidos en los bloques temáticos 4, 5 y posiblemente 6. Se han diseñado dos actividades de aprendizaje que permitirán a su vez la evaluación del estudiante en este itinerario. En la primera actividad deberá demostrar sus destrezas en el manejo del entorno de computación estadística R, para lo cual deberá realizar ejercicios de nivel básico/medio con este programa. Esta actividad será común a la primera actividad del itinerario I. En la segunda actividad se solicitará al estudiante que elabore un material de prácticas docente original en R. Para ello hará uso de libros específicos para la docencia en estadística como el libro TeachingDemos u otros apropiados. El material se elaborará preferentemente en LaTeX utilizando los libros Sweave o Knitr de R. Estos libros proporcionan un marco flexible para mezclar texto y código R, con el objetivo de generar documentos de forma automática. El estudiante dispondrá en la plataforma de documentos de ayuda, ejemplos y enlaces para facilitar el uso de estas herramientas. Más detalles sobre las actividades se pueden encontrar en PRADO.

Itinerario III (Estadística computacional para la investigación)

Está dirigido a estudiantes que hayan cursado previamente asignaturas de Estadística Computacional y quieran profundizar en la metodología computacional. Desde este itinerario se pretende facilitar una iniciación a la investigación en algún tema avanzado de Estadística. Se requieren habilidades computacionales y conocimientos estadísticos y matemáticos a un nivel medio/alto. El estudiante que elija este itinerario habrá de poseer conocimientos de nivel medio de R y algún otro programa estadístico como SPSS, SAS u otros. Los materiales que se pondrán a disposición del estudiante en este caso

suponen un trabajo de investigación desde su inicio, lo que supone una búsqueda bibliográfica y documentación a través de Internet.

Se han diseñado dos actividades de aprendizaje que permitirán la evaluación en este itinerario. En la primera actividad el estudiante deberá demostrar sus destrezas en el manejo de R, para lo cual resolverá una serie de ejercicios de nivel básico/medio, que serán comunes para los tres itinerarios. La segunda actividad consistirá en: (1) elegir un tema concreto de la estadística; (2) estudiar el marco teórico y explorar los libros específicos que hay disponibles en R para la práctica; y (3) elaborar un documento donde se describa el trabajo realizado, proporcionando ejemplos con datos reales o simulados. Idealmente el estudiante elegiría un tema en el que esté iniciando tareas de investigación, o por el que tenga un interés particular, y en ningún caso podrá ser un tema de Estadística básica como los desarrollados en los materiales de este curso. El documento a entregar se elaborará obligatoriamente en LaTeX utilizando los libros Sweave o Knitr de R. Más detalles sobre las actividades se pueden encontrar en PRADO.

ATENCIÓN TUTORIAL

El curso se desarrolla de modo virtual y requiere un trabajo autónomo del estudiante, guiado por los materiales ofrecidos en PRADO y de acuerdo al calendario del curso.

Para facilitar el seguimiento de los estudiantes las profesoras del curso ofrecerán **una hora semanal de atención colectiva a través de Google Meet**. El enlace de conexión se publicará en PRADO y para acceder los estudiantes deberán utilizar de forma obligatoria sus credenciales de la UGR.

Para la **primera parte de curso** (del 11 de octubre al 12 de noviembre) esta hora semanal será atendida por la profesora María Dolores Martínez Miranda. Para la **segunda parte del curso** (del 15 de noviembre al 21 de diciembre) lo hará la profesora Yolanda Román Montoya. El día y hora en que tendrán lugar estas horas de atención se comunicará a través de PRADO con suficiente antelación.

La asistencia a dichas horas es totalmente voluntaria, se desarrollarán de acuerdo a las dificultades y cuestiones planteadas por los estudiantes que asistan. En este sentido se conciben como tutorías colectivas y no será procedente a su grabación. Adicionalmente, estas horas podría complementarse si fuera necesario con tutorías individuales que el estudiante podrá solicitar y acordar con las profesoras por email a lo largo del curso.

CRITERIOS DE EVALUACIÓN

Convocatoria de junio

La evaluación de la asignatura dependerá del itinerario seguido para su curso (elección que se realizará en la actividad 0 desde la plataforma virtual), si bien desde cualquiera de ellos el estudiante podrá obtener la máxima puntuación (10 puntos).

De forma común a todos los itinerarios el estudiante será evaluado atendiendo a los siguientes criterios:

- **Resolución de las actividades de propuestas** (hasta **9 puntos**). Desde cada itinerario se proponen al estudiante dos actividades obligatorias a entregar mediante la plataforma en las fechas fijadas (ver calendario).
- **Autoevaluación** (hasta **1 punto**). El estudiante deberá elaborar un informe final del trabajo que ha realizado en la asignatura, así como una autoevaluación de su trabajo personal y de la asignatura. Este informe se entregará a través de la plataforma después de entregar la segunda actividad.

Aquellos estudiantes que no hayan podido entregar las actividades en los plazos establecidos podrán hacerlo durante el periodo de recuperación.

Convocatoria de septiembre

Los estudiantes que no hayan aprobado la asignatura en la convocatoria de junio, dispondrán de una convocatoria en septiembre. El modo de evaluación en dicha convocatoria será únicamente el correspondiente al itinerario básico. Para aquellos estudiantes que soliciten su evaluación en esta convocatoria, los profesores elaborarán una actividad consistente en la resolución de diversos ejercicios sobre los contenidos básicos del curso.

CALENDARIO

La asignatura Entornos de Computación Estadística está incluida en los cursos de la fase 1 del máster, que comienzan el 11 de octubre de 2021 y finalizan el 21 de diciembre de 2021. Durante el curso los estudiantes deberán resolver y entregar dos actividades de evaluación obligatorias, la **fecha límite para la entrega de dichas actividades para todos los itinerarios** será el **15 de noviembre de 2021**, para la primera, y el **21 de diciembre de 2021** para la segunda. La fecha límite para entregar el informe final será también el **21 de diciembre de 2021**.

A partir del 7 de enero se inicia un **periodo de recuperación** para aquellos que se hayan demorado en la entrega de las actividades y que **finaliza el 14 de enero de 2022**. Los estudiantes que no hayan entregado las actividades, en las fechas fijadas durante el curso, deberán hacerlo durante el periodo de recuperación, a través del enlace que se habilitará en la plataforma para tal fin. Después de esta fecha no se admitirá la entrega de ninguna actividad para su evaluación en la convocatoria de junio.

Los estudiantes que no hayan aprobado la asignatura en la convocatoria de junio, dispondrán de una **convocatoria en septiembre**. Para la evaluación en dicha convocatoria, se enviarán por correo electrónico a los estudiantes implicados con las

actividades a realizar. Para ello **será imprescindible que el estudiante que se acoja a la convocatoria de septiembre contacte previamente con el profesor responsable del curso a través del correo electrónico solicitándolo**. Las actividades resueltas deberán entregarse en la **primera semana de septiembre**, para lo que se darán más detalles en su momento.

BIBLIOGRAFÍA

En formato libro se sugieren las siguientes fuentes bibliográficas:

1. Arriaza Gómez, J., Fernández Palacín, F., López Sánchez, M.A., Muñoz Márquez, M., Pérez Plaza, S. y Sánchez Nava, S. (2008). Estadística Básica con R y R-Commander. Disponible libremente on-line (<http://knuth.uca.es/ebrcmdr>).
2. Ceballos, F.J. (1996). Enciclopedia de Visual Basic. Ed. Rama
3. Chambers – Extending R (2016), CRC Press
4. Crawley, M.J. (2012). The R book, 2nd ed. John Wiley & Sons Inc.
5. de Vries, Meys (2016) . R For Dummies, 2nd ed. John Wiley & Sons Inc.
6. Faraway, J.J. (2002). Practical Regression and Anova using R. Ed. el autor.
7. Fischetti (2015). Data Analysis with R. Packt Publishing
8. Gentle, J.E. (2002). Elements of Computational Statistics. Springer-Verlag, New York, Inc.
9. Gillespie, Lovelace (2016). Efficient R Programming. O'Reilly
10. Harriger, A.R., Lisack, S.K., Gotwals, J.K., y Lutes, K.D. (2004). Introduction to computer programming with Visual Basic 6. A Problem-Solving Approach. Series in Programming and Development. E&T.
11. Johnson, Morgan (2016) - Survey Scales. A Guide to Development, Analysis, and Reporting. The Guilford Press
12. Lizasoan, L. y Joaristi, L. (1999). SPSS para Windows. Ed. Rama
13. O'Regan (2012). A Brief History of Computing, 2nd ed. Springer
14. Pérez, C. (2004). Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS. Ed. Pearson.Prentice Hall.
15. Pérez, C. (2001). Técnicas estadísticas con SPSS. Ed. Prentice.
16. Ren – Learning R Programming (2016), Packt Publishing
17. Robbins (2013). Effective graphs with Microsoft R Open. Microsoft
18. Spector, P. (2008). Data Manipulation with R. Springer Science+Business Media, LLC.
19. Soh (2016). Understanding Test and Exam Results Statistically. Springer
20. Tattar, Ramaiah, Manjunath (2016). A Course in Statistics with R. Wiley
21. Van der Linden ed. (2016). Handbook of Item Response Theory, Volume Two. Statistical Tools. CRC Press
22. Visual Basic. Manual del usuario.
23. SPSS para Windows. Manual del usuario
24. Wickham, H. (2016). ggplot2. Elegant Graphics for Data Analysis, 2nd ed. Springer.
25. Xie, Y. (2015). Dynamic Documents with R and knitr. Chapman and Hall. CRC Press.
26. Zhang (2016). R for Programmers. Mastering the Tools. CRC Press

Además se recomiendan las siguientes publicaciones on-line y direcciones de Internet.

R

Página principal, descarga y documentación: <http://www.r-project.org/>

SPSS

Página principal y descarga: <http://www.spss.com>

SAS

Página principal y descarga: <http://www.sas.com/index.html>

Visual Basic

Página principal y descarga:

Visual Basic 2008 <http://www.microsoft.com/express/>

Documento de ayuda en español:

<http://mat21.etsii.upm.es/ayudainf/aprendainf/VisualBasic6/vbasic60.pdf>

MatLab

Página principal y descarga: <http://www.mathworks.com/>,

<http://www.mathworks.com/cmspro/req15288.html>

Documento de ayuda en español:

<http://mat21.etsii.upm.es/ayudainf/aprendainf/Matlab70/matlab70primero.pdf>

Comparativa de programas:

<http://www.ecologiaaustral.com.ar/files/ea190c83d4.pdf>

http://en.wikipedia.org/wiki/Comparison_of_statistical_packages

http://en.wikibooks.org/wiki/Statistics/Numerical_Methods/Numerical_Comparison_of_Statistical_Software

Elaboración de informes estadísticos en LaTeX y R:

http://cran.r-project.org/doc/contrib/Rivera-Tutorial_Sweave.pdf

http://www.ub.edu/stat/docencia/EADB/Manual_Sweave.pdf

<http://www.maths.tcd.ie/~dwilkins/LaTeXPrimer/>

<http://beckmw.wordpress.com/2014/02/24/brief-introduction-on-sweave-and-knitr-for-reproducible-research/>