

**Proyecto/Guía docente de la asignatura**

Asignatura	COMPUTACIÓN ESTADÍSTICA		
Materia	Probabilidad y Estadística		
Módulo			
Titulación	GRADO EN ESTADÍSTICA (549) INdat: Título combinado Grado en INFORMÁTICA + Grado en ESTADÍSTICA		
Plan	549	Código	47089
Periodo de impartición	C2	Tipo/Carácter	Optativa
Nivel/Ciclo	Grado	Curso	2º
Créditos ECTS	6 ECTS		
Lengua en que se imparte	Español		
Profesor/es responsable/s	María Teresa González Arteaga		
Datos de contacto (E-mail, teléfono...)	Despacho 229 en la Facultad de Ciencias. teresag@eio.uva.es		
Horario de tutorías	Consultar la web de la UVa		
Departamento	Estadística e Investigación Operativa		



1. Situación / Sentido de la Asignatura

1.1 Contextualización

Como objetivo general pretendemos que el alumnado conozca los paquetes estadísticos R y SAS como entornos de trabajo y los maneje con la soltura suficiente como para ser capaz de plantearse utilizar por sí mismo nuevas funciones y nuevas librerías cuando sea necesario en el futuro.

La elección del R no es casual, sino que responde a sus características. Primeramente, debemos destacar que R es un entorno y un lenguaje para la computación estadística que proporciona una gran variedad de técnicas estadísticas y gráficas (test de hipótesis, modelos lineales y no lineales, series temporales, clasificación, etc.). En segundo lugar, de gran interés resulta que este software es de distribución libre al cual se puede acceder fácilmente en la web (<http://www.cran.r-project.org/>).

Por otro lado, este entorno permite que los futuros profesionales accedan fácilmente a muchas librerías para enfrentar otros métodos estadísticos tanto ya estudiados durante el Grado como los que aún no conocen. Estas librerías les serán de fácil uso una vez conocido el manejo esencial del R que planteamos en esta asignatura.

El término “entorno” hace referencia a un sistema completamente diseñado y coherente, antes que como una agregación incremental de herramientas muy específicas e inflexibles, como ocurre frecuentemente con otros programas de análisis de datos. Además, se utiliza RStudio como IDE.

La elección de SAS tampoco es casual. Este entorno es uno de los más utilizados por muchas empresas e instituciones estando entre los más importantes en el mercado. Por estas mismas razones haremos una breve incursión en el paquete SPSS.

La asignatura está totalmente orientada a las aplicaciones y por ello todas las horas lectivas se impartirán en el laboratorio y casi todo el trabajo que el alumnado tendrá que realizar será de índole práctico. Además, utilizaremos la herramienta de e-learning MOODLE que la Universidad de Valladolid ofrece como *campusvirtual.uva.es* para que el alumnado tenga mayor facilidad para la disponibilidad de los distintos recursos que la profesora proporciona, para realizar sus tareas, así como, para relacionarse con el resto del grupo.

1.2 Relación con otras materias

En esta asignatura se introduce software estadístico que será utilizado en la mayoría de las asignaturas posteriores a la misma en el plan de estudios de Grado en Estadística y en el programa de estudios conjunto del Grado en Estadística y del Grado en Informática (INdat).

1.3 Prerrequisitos

Se recomienda tener conocimientos básicos de probabilidad, estadística descriptiva, álgebra y cálculo.

2. Competencias

Según la descripción de la memoria del Grado en Estadística: G1, G2, G3, E1, E3, E6, E7, I1, I2, I3, I4, I5, I6, I7, P2, P3, P4, S1, S2, S3, S4, S5.

2.1 Generales

- G1. Capacidad para la gestión de la información.
- G2. Capacidad para la abstracción y el razonamiento crítico.
- G3. Capacidad para la puesta al día y el auto-aprendizaje.

2.2 Específicas

- E1. Recogida y tratamiento de datos.
- E3. Ajuste de modelos estadísticos y de investigación operativa.
- E6. Presentación y comunicación de resultados.
- E7. Dominio de conceptos y herramientas informáticas a nivel de usuario avanzado.

3. Objetivos

Resultados de aprendizaje recogidos en la memoria del grado:

- Capacidad para manejar conjuntos de datos en el entorno de SAS y en R.
- Saber hacer análisis descriptivos, así como ajustar modelos de regresión lineal simple con SAS y R.
- Saber utilizar las herramientas básicas de simulación de SAS y R.
- Capacidad para realizar gráficos “elaborados” con R.
- Capacidad para programar, a nivel básico, en SAS y en R.



4. Contenidos y/o bloques temáticos

Bloque 1: “Computación Estadística”

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

La contextualización puede leerse en el apartado correspondiente a la asignatura completa que aparece más arriba.

b. Objetivos de aprendizaje

- Manejar con soltura el programa R como entorno de programación para análisis de datos y la realización de gráficos sencillos y elaborados.
- Realizar el análisis estadístico básico de datos con R y desarrollar la capacidad de abordar otros más complejos de forma autónoma por el alumnado.
- Conocer e implementar algoritmos para generar muestras de distintas distribuciones aleatorias (continuas, discretas, multivariantes).
- Utilizar la simulación de distribuciones con fines específicos dentro de la Estadística como por ejemplo para aplicar procedimientos de inferencia estadística.
- Introducir el entorno de SAS para manejar conjuntos de datos y realizar análisis descriptivos.
- Introducir el software SPSS para manejar conjuntos de datos y realizar análisis descriptivos.

c. Contenidos

Temas a desarrollar en el programa de la asignatura:

1. Introducción al sistema SAS: lectura de datos y manejo de archivos. Paso data. Procedimientos básicos de SAS. Programación con SAS. Introducción a macros con R.
2. Introducción a SPSS.
3. Introducción al entorno R: manejo de objetos en R. Manejo de archivos. Funciones básicas. Programación con R. Gráficos tradicionales con R.
4. Uso avanzado de R: Librería Lattice. Librería ggplot2. Algunas aplicaciones estadísticas con R. Representación de datos en mapas.

d. Métodos docentes

La asignatura se desarrollará mediante la realización de diversas actividades, clases en el laboratorio, tanto teóricas como prácticas, entrega de ejercicios, trabajos y cuestionarios, exposición oral de ejercicios resueltos, tutorías individualizadas y examen final.

En esta asignatura utilizaremos la herramienta de e-learning MOODLE que está implementada en el campus virtual de la UVa. Esta herramienta permitirá una relación más fluida en la realización de actividades y en la disposición de los distintos recursos para el alumnado.

Todas las actividades tienen como objetivo principal el de potenciar el aprendizaje de los alumnos, facilitando la adquisición de cuantos conocimientos y competencias precisas. Las diferentes actividades estarán sujetas a un proceso de evaluación continua, y algunas permitirán dar la certificación necesaria del aprendizaje. Véase el apartado dedicado a la evaluación del aprendizaje.

En esta asignatura se utilizan programas sobre el sistema operativo Windows. Estos programas son R, SAS básico y SPSS.

Las diferentes actividades que se realizarán a lo largo del curso en el ámbito de esta asignatura se detallan a continuación.

Clases:

La teoría básica necesaria será expuesta en clase por la profesora de la asignatura y se ilustrará continuamente su aplicación mediante ejemplos, lo cual llevará a que no podamos diferenciar claramente entre clases de teoría y clases prácticas. Todas las horas lectivas se desarrollarán en el laboratorio.

La participación activa del alumnado será necesaria en todos los casos.

Pruebas y ejercicios:

En las fechas señaladas en el calendario de actividades se solicitará a cada alumno/a la realización de ejercicios y/o la respuesta a un cuestionario. Estas dos pruebas tendrán lugar dentro del horario de clases previsto. Estas pruebas se realizarán en equipos con sistema operativo Windows con el software utilizado durante el curso en las aulas de la Facultad.

Se llevará a cabo un seguimiento de la asignatura a través de la participación habitual en las clases y prácticas de la asignatura. Se tendrá en consideración el trabajo realizado a lo largo de todo el curso, la actitud y el interés manifestado por aprender. A lo largo del curso se pedirá al alumnado que exponga oralmente algunos ejercicios y/o que responda a algún cuestionario. También se podrá solicitar la grabación de la exposición de tareas.

**Examen Final:**

El alumnado tendrá que dar respuesta a varios ejercicios que se resolverán en el laboratorio utilizando el paquete R y/o SAS básico y/o SPSS con cierto material de apoyo que se indicará con antelación. El examen se realizará en equipos con sistema operativo Windows disponibles en las aulas de la Facultad.

Tutorías:

Las tutorías podrán ser atendidas en el horario que se disponga para ello al inicio del curso o a cualquier otra hora, previa cita con la profesora.

En todas las actividades realizadas se llevará un control de asistencia.

e. Plan de trabajo

Programa detallado de la asignatura:**1. Introducción al sistema SAS: lectura de datos y manejo de archivos. Paso data. Procedimientos básicos de SAS. Programación con SAS.**

- Nociones generales sobre el entorno SAS.
- Paso data:
 - o Creación de ficheros de datos sas. Lectura de datos en formatos no SAS. Almacenamiento de datos en ficheros externos. Operaciones con ficheros sas.
 - o Sentencias en el paso data: crear nuevas variables, if, do, output, ...
- Generalidades del paso proc.
- Procedimientos básicos en SAS: Proc PRINT, FREQ, MEANS, SUMMARY, UNIVARIATE, TABULATE, GPLOT, ...
- Introducción a la realización de mapas temáticos con SAS: proc MAPS
- Introducción al uso de macros.

Bibliografía:

Pérez, C. "El Sistema estadístico SAS". 2001 de. Prentice Hall

"Introducción al SAS". Joan Valls, Llorenç Badiella.

Manuales de SAS.

2. Introducción al SPSS.

- Conceptos básicos: editor de datos, visor de resultados. Gestión de archivos, manejo de datos, realización de tablas y gráficos básicos.

Bibliografía:

Guía del usuario de IBM SPSS Statistics 24 Core System. IBM (disponible on line)

3. Introducción al paquete R: Objetos en R. Manejo de archivos. Funciones básicas. Programación con R. Gráficos tradicionales con R.

- Instalación. Primeros pasos.
- Objetos (vectores, matrices, arrays, marcos de datos, listas)
- Operaciones básicas con objetos. Funciones básicas.
- Manejo de archivos. Importar y exportar datos. Ficheros de código.
- Gráficos tradicionales con R.
- Programación básica con R.
- Realización de informes con R con rmarkdown.

Web: <http://www.cran.r-project.org/>

Bibliografía:

Dalgaard, P (2001) "Introductory statistics with R"

Gil Bellosta, C.J.. 2018. "R para profesionales de los datos: una introducción" Paradis E . "R para principiantes"

R Core Team. 2001. "Introducción a R"

4. Uso avanzado de R: Librería Lattice. Librería ggplot2. Algunas aplicaciones estadísticas con. Representación de datos en mapas.

- Análisis descriptivo. Regresión lineal simple. Funciones *lm* y *loess*.
- Manejo de las distribuciones de probabilidad clásicas con R.
- Función *density*.
- Concepto de número pseudoaleatorio. Generadores de $U(0, 1)$. Algunos métodos de generación de muestras.
- Método de Montecarlo para simulación de la distribución de estadísticos.
- Librería *lattice* y librería *ggplot2*.
- Introducción a la realización de mapas temáticos.

Bibliografía:

- G. Golemund and H. Wickham. 2017. "R for Data Science"
- Maindonald, J., Braun, J. 2007. "Data analysis and graphics using R. An example-based approach"
- Martinez W., Martínez A. 2002. "Computational statistics handbook with MATLAB".

f. Evaluación

Se detalla a continuación el procedimiento para asignar la calificación final.

- El 40% de la nota se obtiene con la **evaluación continua** a través de las actividades realizadas a lo largo del cuatrimestre.
 - Se realizarán dos pruebas con peso en la calificación final del 10% para la prueba primera y 15% para la segunda en las fechas señaladas en el calendario general de actividades que se comunica a principios del cuatrimestre. *Estas pruebas se realizan en equipos con sistema operativo Windows y sólo se permite utilizar el software indicado en el aula.*
 - Se hará un seguimiento de la participación habitual en las clases y prácticas de la asignatura. Se puntuará globalmente con una calificación que tendrá un peso del 15% de la calificación final. Para esta calificación se tendrá en cuenta el trabajo realizado a lo largo de todo el curso, la actitud y el interés manifestado por aprender. Esto puede incluir la exposición oral de algunos ejercicios o la entrega de alguna grabación de la misma.

Llamémos A a la puntuación obtenida en la evaluación continua.

- El 60% de la nota se obtiene con un **examen final**. Sea B la puntuación de este examen en una escala de 0 a 10.
- Para realizar la media ponderada de la puntuación de la evaluación continua y la puntuación del examen final ambas calificaciones deben ser mayores de 3.
- Si se cumple la condición anterior **Nota final = 0.40 * A + 0.60 * B**
- Para aprobar B debe ser mayor de 3 y la Nota final debe superar el valor de 5.

En la convocatoria extraordinaria la calificación será la máxima entre la nota obtenida en el examen correspondiente y la calificación obtenida con el procedimiento descrito para la convocatoria ordinaria.

g. Bibliografía básica

- Dalgaard, P (2001) "Introductory statistics with R". Springer
- Gil Bellosta, C.J. 2018. "R para profesionales de los datos: una introducción" (en la web https://www.datanalytics.com/libro_r/)
- *Guía del usuario de IBM SPSS Statistics 24 Core System. IBM (disponible on line)*



- Paradis E . 2002. “R para principiantes”. (Texto disponible en la web de R)
- Pérez, C. 2001. “El Sistema estadístico SAS”. Prentice Hall
- R Core Team. 2001. “Introducción a R”. (Texto disponible en la web de R)
- R Core Team. 2019. “An n Introduction to R”. (Texto disponible en la web de R)
- Valls, J., Badiella LI. “Introducción al SAS”. Servei d'Estadística UAB

h. Bibliografía complementaria

- Gentle, J.E., 1998. "Random Number Generation and Monte Carlo Methods". Springer-Verlag.
- G. Golemund and H. Wickham. 2017. “R for Data Science”. O'Reilly Media (también en la web <https://r4ds.had.co.nz/>)
- Maindonald, J., Braun,J. 2007. “Data analysis and graphics using R. An example-based approach”Cambridge University press.
- Martínez W., Martínez A. 2002. “Computational statistics handbook with MATLAB”. Chapman & Hall.
- Ríos Insua, D., Ríos Insua, S., Martín, J, Jiménez, A. 2008. “Simulación. Métodos y aplicaciones”. Ed. Rama

i. Recursos necesarios

Software estadístico especificado en los temas del programa.

j. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
6	C2

5. Métodos docentes y principios metodológicos

Ver sección 4 apartado d (Métodos docentes) de esta guía.

6. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases en aula	-	Estudio autónomo / Trabajo personal en el laboratorio	82
Laboratorio/clase con ordenador	45		
Seminarios		Elaboración de trabajos	4
Tutorías personalizadas		Preparación de presentaciones Orales o seminarios	4
Presentación de trabajos	11		
Realización de exámenes	4		
Total presencial	60	Total no presencial	90

7. Sistema y características de la evaluación

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Participación en clase y seguimiento de la asignatura, exposiciones orales	15%	
Prueba 1	10%	
Prueba 2	15%	
Examen final	60%	

CRITERIOS DE CALIFICACIÓN

- **Convocatoria ordinaria:**
Para aprobar, tanto la calificación de la evaluación continua (A) como del examen final (B) deben ser mayores de 3 y la *Nota final* = $0.40 * A + 0.60 * B$ debe ser mayor de 5. La calificación será la *Nota final* si se cumplen los requisitos descritos y en otro caso la calificación será el mínimo (A, B)
- **Convocatoria extraordinaria:**
La calificación es el máximo entre la nota obtenida en el examen correspondiente y la calificación obtenida con el procedimiento descrito para la convocatoria ordinaria.

8. Consideraciones finales

Las clases y actividades se desarrollarán en el laboratorio.

Se recomienda tener capacidad de leer inglés técnico.