CURSO
ESTADÍSTICA
APLICADA CON R

Organizado por:

Instituto IMDEA Alimentación

Facultad de Ciencias de la

Universidad Autónoma de Madrid

MADRID, de Septiembre a Diciembre de 2015

CURSO DE FORMACIÓN CONTINUA: ESTADÍSTICA APLICADA CON R

PRESENTACIÓN

La **Unidad de Bioestadística** del **Instituto IMDEA Alimentación** y los **departamentos de Matemáticas y Ecología** de la **Facultad de Ciencias** de la **Universidad Autónoma de Madrid** organizan conjuntamente un **programa de formación continua en Estadística Aplicada con el software R**, compuesto por diferentes módulos que pueden ser cursados en su totalidad o separadamente, dependiendo de los diferentes intereses y conocimientos de los alumnos.

En este programa se han reunido una gran variedad de técnicas estadísticas que han sido seleccionadas por su importancia en el **análisis de datos moderno**. Comprende un amplio abanico de técnicas, desde las más sencillas, como las herramientas descriptivas de análisis de datos o inferencia básica, hasta **métodos más complejos y especializados**, como los modelos de regresión (lineales, generalizados, aditivos y mixtos), el análisis multivariante, el análisis de supervivencia y las técnicas supervisadas para la construcción de modelos predictivos usadas habitualmente en minería de datos (data mining). Se han **organizado en módulos** que puedan resultar de interés a determinados colectivos profesionales, por eso algunos temas aparecen repetidos parcialmente en varios módulos, aunque son tratados desde distintos puntos de vista prácticos en cada uno de ellos, de tal forma que el alumno puede asistir a los 7 módulos que comprende el programa completo o bien seleccionar aquellos que mejor se adapten a sus intereses.

R se está convirtiendo en el **software estadístico de referencia** en la mayoría de los centros de investigación y universidades, y su uso se está extendiendo también dentro de la empresa privada. **R es software libre, gratuito** y está incorporando más rápidamente que otros programas las **técnicas avanzadas de análisis de datos** que se están desarrollando en los últimos años. Otro aspecto importante que convierte a R en una **herramienta muy potente** es que incorpora un **lenguaje de programación** sencillo y muy flexible, que permite tener un control total sobre el análisis que se está desarrollando.

El módulo 1 incluye los conocimientos básicos del lenguaje R y algunas de las técnicas básicas de estadística, y debería ser cursado por todos los alumnos que no tengan conocimientos de dicho lenguaje. El resto de módulos pueden ser cursados una vez que se ha asistido al de introducción, o también por alumnos que tengan ya conocimientos de R y de estadística básica.

DIRIGIDO A

Este programa de formación continua va dirigido a **profesionales** de distintos ámbitos, **investigadores, profesores y alumnos** que quieran conocer el software R y su aplicación práctica con diferentes técnicas estadísticas

Para acceder al programa se debe ser graduado, licenciado o estudiante de master

ESTRUCTURA

Módulos	Fechas	Núm. Horas	Precio	Créditos
1. Introducción a R	24 y 25 Septiembre 2015	10	110€	1
2. Métodos de Regresión con R	15 y 16 Octubre 2015	10	110€	1
3. Métodos de Regresión Avanzados para la Investigación en Ciencias Naturales con R	28, 29 y 30 Octubre 2015	20	220€	2
4. Estadística Aplicada a la Investigación Biomédica con R	11, 12 y 13 Noviembre 2015	21	280€	2
5. Modelos Mixtos / Jerárquicos / Multinivel con R	18, 19 y 20 Noviembre 2015	20	220€	2
6. Estadística Multivariante con R	26 y 27 Noviembre 2015	10	110€	1
7. Técnicas Estadísticas de Data Mining con R	14, 15, 16 y 17 Diciembre 2015	28	420€	3

CRÉDITOS

El número máximo de créditos convalidables será de 3

DIRECTORES

- José Ramón Berrendero Díaz (Profesor Titular en el Dpto. de Matemáticas de la UAM)
- Jesús Herranz Valera (Bioestadístico en el IMDEA Alimentación)

PERSONAL DOCENTE

- Amparo Baíllo Moreno (Profesora Titular en el Dpto. de Matemáticas de la UAM)
- José Ramón Berrendero Díaz (Profesor Titular en el Dpto. de Matemáticas de la UAM)
- Jesús Herranz Valera (Bioestadístico en el IMDEA Alimentación)
- Carlos Pérez Carmona (Investigador postdoctoral Marie Curie, University of South Bohemia)
- Javier Seoane Pinilla (Profesor Titular en el Dpto. de Ecología de la UAM)

INSCRIPCIÓN Y DATOS PRÁCTICOS

Las **inscripciones** en los diferentes módulos se realizarán en la página web de la **Fundación de la Universidad Autónoma de Madrid**:

http://matriculas.fuam.es/matriculauam/Convocatorias.action

Los **módulos 1, 2, 3, 5, y 6** se impartirán en la **Facultad de Ciencias de la UAM.** Las clases se desarrollarán en aulas de informática provistas de los equipos y programas necesarios, aunque se **recomienda** que los participantes traigan su propio **ordenador portátil**

Los **módulos 4 y 7** se impartirán en el Instituto IMDEA Alimentación, y en ellos es **imprescindible** que los participantes traigan su propio **ordenador portátil**

BECAS

Entre los alumnos matriculados en cada módulo se concederán dos becas que cubrirán el total del módulo correspondiente. Los criterios de selección de becarios son:

1. Estar desempleado (40%)
2. Situación socio-económica (30%)
3. Expediente académico (30%)

Las solicitudes de las becas se harán por correo electrónico dirigido al personal docente del módulo. En el correo se proporcionará información sobre los puntos anteriores, indicando la situación laboral del alumno: personal fijo, contratado eventual o becario. En caso necesario, se podrá solicitar una declaración jurada.

La asignación de las becas se realizará después de finalizada la matrícula de cada módulo, y el importe de la matrícula será devuelto posteriormente a los alumnos que hayan sido becados.

MÓDULO 1: INTRODUCCIÓN A R

PRESENTACIÓN Y OBJETIVOS

El objetivo de este módulo es proporcionar una introducción a R para alumnos que no han tenido contacto previo con el programa. Además, se aprovechará esta introducción para ilustrar algunas nociones básicas de descripción de datos e inferencia estadística.

Se tratarán las propiedades de los principales tipos de objetos en R, algunas herramientas de representación gráfica y se estudiarán algunos ejemplos de programación de funciones sencillas. Finalmente, se introducirán algunas herramientas que permiten usar R conjuntamente con editores de texto como Word o Latex, lo que facilita la reproducción y presentación de los resultados obtenidos.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Introducción a R**
Docentes	Amparo Baíllo (Dpto. Matemáticas, UAM) amparo.baillo@uam.es y José Ramón Berrendero (Dpto. Matemáticas, UAM), joser.berrendero@uam.es
Fechas del Curso	24 y 25 de septiembre de 2015. Horario: 14:30-19:45
Inscripciones	Hasta el 23/09/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Universidad Autónoma de Madrid Ctra. Madrid-Colmenar Viejo (M-607), km 15 Facultad de Ciencias (C/ Francisco Tomás y Valiente, 7) (por confirmar) Madrid http://www.uam.es
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	110€
Créditos	1 ECTS
Requisitos previos	No se requieren conocimientos previos de R. Para el máximo aprovechamiento del módulo, es conveniente tener conocimientos elementales de estadística (en caso de duda contacten con los profesores del módulo)
Observaciones	Los conceptos teóricos, datos y código para el seguimiento del módulo se harán accesibles en documentos en red. Las clases se desarrollarán en aulas de informática provistas de los equipos y programas necesarios, aunque se **recomienda** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 1: INTRODUCCIÓN A R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 6 sesiones, con una duración variable de entre 1 y 2 horas, dependiendo de cada tema.

En cada sesión se explican los principales conceptos, tanto estadísticos como relativos al funcionamiento del programa, y se ilustran de manera práctica con ejemplos realizados con R. Finalmente, se proponen ejercicios a los alumnos para que puedan practicar los conocimientos adquiridos.

1. Características básicas del programa
2. Principales tipos de objetos
3. Ejemplos sencillos de simulación e inferencia estadística
4. Gráficos
5. Introducción a la programación en R
6. Elaboración de informes dinámicos

MÓDULO 2: MÉTODOS DE REGRESIÓN CON R

PRESENTACIÓN Y OBJETIVOS

El objetivo de este módulo es que los alumnos adquieran los conocimientos necesarios para aplicar con R diversos métodos de regresión que permiten explicar una variable respuesta de interés en función de un conjunto de variables explicativas. La primera parte del módulo se dedica a los modelos lineales de regresión, incluyendo el diagnóstico de sus hipótesis. La segunda parte se dedica a extensiones de gran importancia en la práctica: los modelos lineales generalizados y los métodos no paramétricos.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Métodos de Regresión con R**
Docentes	Amparo Baíllo (Dpto. Matemáticas, UAM) amparo.baillo@uam.es y José Ramón Berrendero (Dpto. Matemáticas, UAM), joser.berrendero@uam.es
Fechas del Curso	14 y 15 de octubre de 2015. Horario: 14:30-19:45
Inscripciones	Hasta el 13/10/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Universidad Autónoma de Madrid Ctra. Madrid-Colmenar Viejo (M-607), km 15 Facultad de Ciencias (C/ Francisco Tomás y Valiente, 7) (por confirmar) Madrid http://www.uam.es
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	110€
Créditos	1 ECTS
Requisitos previos	Para el máximo aprovechamiento del módulo se requieren conocimientos básicos de R y de inferencia estadística (en caso de duda contacten con los profesores del módulo)
Observaciones	Los conceptos teóricos, datos y código para el seguimiento del módulo se harán accesibles en documentos en red. Las clases se desarrollarán en aulas de informática provistas de los equipos y programas necesarios, aunque se **recomienda** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 2: MÉTODOS DE REGRESIÓN CON R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 5 sesiones de 2 horas.

En cada sesión se explican los principales conceptos, tanto estadísticos como relativos al funcionamiento del programa, y se ilustran de manera práctica con ejemplos realizados con R. Finalmente, se proponen ejercicios a los alumnos para que puedan practicar los conocimientos adquiridos.

1. Modelos de regresión
2. Modelos de diseño de experimentos
3. Diagnóstico del modelo y tratamiento de datos atípicos
4. Modelos lineales generalizados
5. Métodos no paramétricos en regresión

MÓDULO 3: MÉTODOS DE REGRESIÓN AVANZADOS PARA LA INVESTIGACIÓN EN CIENCIAS NATURALES CON R

PRESENTACIÓN Y OBJETIVOS

El objetivo principal de este módulo es proporcionar a los alumnos los conocimientos de las técnicas estadísticas que extienden el análisis de regresión a las situaciones encontradas más frecuentemente en los campos de estudio de las ciencias naturales.

El módulo parte de la exposición de los Modelos Lineales Generalizados (**GLM**), incidiendo en las regresiones de Poisson, logística, binomial y binomial negativa, para continuar con sus extensiones a los casos en los que aparece un gran número de ceros, frecuentes en los análisis de conteos de organismos (**modelos inflados por ceros**). Finalmente, se abordarán las extensiones no lineales (Modelos Aditivos Generalizados, **GAM**). Los modelos se explican desde un punto de vista práctico y el módulo incluye la consideración de estrategias para la construcción de modelos, la interpretación de interacciones y la inferencia multimodelo.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Métodos de Regresión Avanzados para la Investigación en Ciencias Naturales con R**
Docentes	Javier Seoane (Dpto Ecología, UAM) javier.seoane@uam.es y Carlos Pérez (UAM-Universidad South Bohemia), carlos.perez@uam.es
Fechas del Curso	28, 29 y 30 de octubre de 2015. Horario: 9:00 – 18:00
Inscripciones	Hasta el 27/10/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Universidad Autónoma de Madrid Ctra. Madrid-Colmenar Viejo (M-607), km 15 Facultad de Ciencias (C/ Francisco Tomás y Valiente, 7) (por confirmar) Madrid http://www.uam.es
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	220€
Créditos	2 ECTS
Requisitos previos	Para el máximo aprovechamiento del módulo, se requiere conocer al menos los rudimentos del lenguaje R y la regresión lineal (en caso de duda contacten con los profesores del módulo)
Observaciones	Los conceptos teóricos, datos y código para el seguimiento del módulo se harán accesibles en documentos en red. Las clases se desarrollarán en aulas de informática provistas de los equipos y programas necesarios, aunque se **recomienda** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 3: MÉTODOS DE REGRESIÓN AVANZADOS PARA LA INVESTIGACIÓN EN CIENCIAS NATURALES CON R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 4 sesiones temáticas. En cada sesión se explican los conceptos estadísticos teóricos ilustrados desde un punto de vista práctico con ejemplos realizados con R, explicando los paquetes y funcionalidades más importantes del software para cada técnica. Además, se irán proponiendo ejercicios cortos a los alumnos para que se practiquen los conocimientos adquiridos.

1. Modelos lineales generalizados (GLM): regresión de Poisson y binomial negativa (4 horas)
2. Modelos lineales generalizados (GLM): regresiones logística y binomial (4 horas)
3. Regresión de Poisson y Binomial Negativa infladas por ceros (6 horas)
4. Ajustes locales y modelos aditivos generalizados (GAM) (6 horas)

MÓDULO 4: ESTADÍSTICA APLICADA A LA INVESTIGACIÓN BIOMÉDICA CON R

PRESENTACIÓN Y OBJETIVOS

El objetivo principal de este módulo es proporcionar a los alumnos los conocimientos de las técnicas estadísticas más utilizadas dentro del campo de las ciencias de la salud, de la epidemiología y de la investigación biomédica en general.

El módulo incluye las técnicas estadísticas más sencillas para analizar **datos clínicos y epidemiológicos**, como son el análisis descriptivo de datos y la inferencia básica, pero la parte más importante está dedicada a los modelos de regresión. Se ha dado especial importancia a las técnicas más usadas en la investigación biomédica, que son la **regresión logística** y **la regresión de Cox**, para analizar **datos de supervivencia**. Los modelos de regresión se explican desde un punto de vista práctico, y se han incluido temas avanzados como: el tratamiento de las **variables de confusión**, el análisis e interpretación de **interacciones**, las estrategias para la **construcción de modelos** de regresión multivariantes y el análisis de la **capacidad predictiva** de los modelos.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Estadística Aplicada a la Investigación Biomédica con R**
Docentes	Jesús Herranz Valera (Bioestadístico IMDEA Alimentación) jesus.herranz@imdea.org
Fechas del Curso	11, 12 y 13 de noviembre de 2015. Horario: 9:00 – 18:00
Inscripciones	Hasta el 10/11/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Instituto IMDEA Alimentación Ctra. Cantoblanco, nº8 Campus UAM – Madrid http://www.alimentacion.imdea.org
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	280€
Créditos	2 ECTS
Requisitos previos	No son necesarios conocimientos de R. Para el máximo aprovechamiento del módulo, son necesarios conocimientos básicos de inferencia estadística (en caso de duda contacten con los profesores del módulo)
Observaciones	Se entregará un manual encuadernado con todas las diapositivas del curso (440 aprox.) y todos los scripts y ficheros de datos Es **imprescindible** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 4: ESTADÍSTICA APLICADA A LA INVESTIGACIÓN BIOMÉDICA CON R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 21 sesiones, con una duración de 1 hora cada una.

En cada sesión se explican los conceptos estadísticos teóricos ilustrados desde un punto de vista práctico con ejemplos realizados con R, explicando los paquetes y funcionalidades más importantes del software para cada técnica. Finalmente, se proponen ejercicios a los alumnos para que puedan practicar los conocimientos adquiridos.

1. Introducción a R
2. Variables y objetos de R
3. Manejo de datos. Ficheros y *dataframes*
4. Gráficos en R
5. Estadística descriptiva y funciones de probabilidad
6. Programación y funciones en R
7. Análisis de tablas de contingencia
8. Inferencia básica con variables continuas
9. Análisis de la varianza
10. Análisis de correlación
11. Regresión lineal simple
12. Regresión lineal múltiple
13. Regresión logística I
14. Regresión logística II
15. Variables de confusión e interacciones
16. Construcción de un modelo de regresión logística
17. Análisis de supervivencia
18. Regresión de Cox I
19. Regresión de Cox II
20. Modelos predictivos. Análisis de curvas ROC
21. Análisis de Medidas Repetidas

MÓDULO 5: MODELOS MIXTOS / JERÁRQUICOS / MULTINIVEL CON R

PRESENTACIÓN Y OBJETIVOS

En este módulo se presentan los **modelos** de regresión **mixtos** (también conocidos como jerárquicos o multinivel), que incluyen factores aleatorios, para las situaciones en los que las unidades de análisis se encuentran agregadas en grupos y las respuestas que se observan en ellas están correlacionadas. Esto sucede por ejemplo, al estudiar distintos barrios dentro de varias ciudades o al preguntarse por alumnos dentro de clases o especies biológicas dentro de familias.

El módulo se desarrollará de forma muy práctica y se centrará en los modelos mixtos lineales, aunque también se expondrá su extensión a modelos generalizados (**GLMM**) y aditivos (**GAMM**). El módulo incluye las estrategias de construcción y diagnóstico de modelos.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Modelos Mixtos / Jerárquicos / Multinivel con R**
Docentes	Javier Seoane (Dpto Ecología, UAM) javier.seoane@uam.es y Carlos Pérez (UAM-Universidad South Bohemia), carlos.perez@uam.es
Fechas del Curso	18, 19 y 20 de noviembre de 2015. Horario: 9:00 – 18:00
Inscripciones	Hasta el 17/11/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Universidad Autónoma de Madrid Ctra. Madrid-Colmenar Viejo (M-607), km 15 Facultad de Ciencias (C/ Francisco Tomás y Valiente, 7) (por confirmar) Madrid http://www.uam.es
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	220€
Créditos	2 ECTS
Requisitos previos	Para el máximo aprovechamiento del módulo, se requiere conocer al menos los rudimentos del lenguaje R y la regresión lineal (en caso de duda contacten con los profesores del módulo)
Observaciones	Los conceptos teóricos, datos y código para el seguimiento del módulo se harán accesibles en documentos en red. Las clases se desarrollarán en aulas de informática provistas de los equipos y programas necesarios, aunque se **recomienda** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 5: MODELOS MIXTOS / JERÁRQUICOS / MULTINIVEL CON R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 2 sesiones temáticas principales y una tercera, más breve, que extiende el uso de los modelos mixtos y combina técnicas de GLM y GAM abordadas en módulos anteriores del curso. En cada sesión se explican los conceptos estadísticos teóricos ilustrados desde un punto de vista práctico con ejemplos realizados con R, explicando los paquetes y funcionalidades más importantes del software para cada técnica. Además, se irán proponiendo ejercicios cortos a los alumnos para que se practiquen los conocimientos adquiridos.

1. Modelos mixtos lineales para datos encajados: incluye construcción y diagnóstico de modelos (10 horas)
2. Modelos mixtos lineales para otros diseños muestrales (factores cruzados y medidas repetidas) (8 horas)
3. Modelos mixtos generalizados (GLMM) y generalizados aditivos (GAMM) (2 horas)

MÓDULO 6: ESTADÍSTICA MULTIVARIANTE CON R

PRESENTACIÓN Y OBJETIVOS

Este módulo tiene por objetivo que los alumnos conozcan los fundamentos y las aplicaciones de las técnicas más utilizadas de análisis multivariante y sepan cómo aplicarlas con R. Estas técnicas no solo tienen interés por sí mismas sino que además son los elementos en los que se basan técnicas más complejas para analizar datos de alta dimensión.

El módulo incluye algunas técnicas descriptivas para datos multivariantes, métodos de reducción de la dimensión y técnicas de clasificación tanto supervisada como no supervisada.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Estadística Multivariante con R**
Docentes	Amparo Baíllo (Dpto. Matemáticas, UAM) amparo.baillo@uam.es y José Ramón Berrendero (Dpto. Matemáticas, UAM), joser.berrendero@uam.es
Fechas del Curso	26 y 27 de noviembre de 2015. Horario: 14:30-19:45
Inscripciones	Hasta el 25/11/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Universidad Autónoma de Madrid Ctra. Madrid-Colmenar Viejo (M-607), km 15 Facultad de Ciencias (C/ Francisco Tomás y Valiente, 7) (por confirmar) Madrid http://www.uam.es
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	110€
Créditos	1 ECTS
Requisitos previos	Para el máximo aprovechamiento del módulo se requieren conocimientos básicos de R y de inferencia estadística (en caso de duda contacten con los profesores del módulo)
Observaciones	Los conceptos teóricos, datos y código para el seguimiento del módulo se harán accesibles en documentos en red. Las clases se desarrollarán en aulas de informática provistas de los equipos y programas necesarios, aunque se **recomienda** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 6: ESTADÍSTICA MULTIVARIANTE CON R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 5 sesiones de 2 horas.

En cada sesión se explican los principales conceptos, tanto estadísticos como relativos al funcionamiento del programa, y se ilustran de manera práctica con ejemplos realizados con R. Finalmente, se proponen ejercicios a los alumnos para que puedan practicar los conocimientos adquiridos.

1. Descripción de datos multivariantes
2. Distribuciones multivariantes
3. Reducción de la dimensión
4. Clasificación no supervisada: análisis de conglomerados
5. Clasificación supervisada: análisis discriminante

MÓDULO 7: TÉCNICAS ESTADÍSTICAS DE DATA MINING CON R

PRESENTACIÓN Y OBJETIVOS

En este módulo se han reunido una serie de técnicas procedentes de distintos campos como el **análisis multivariante, el aprendizaje estadístico, la inteligencia artificial y técnicas de *machine learning***. La mayoría de las técnicas presentadas son técnicas predictivas, cuyo objetivo principal es desarrollar un modelo matemático que permita obtener predicciones en una variable de interés en observaciones no incluidas en la construcción del modelo. Se han seleccionado las **técnicas predictivas más importantes** y cuyo uso está más extendido: regresión lineal y logística, métodos penalizados (LASSO), métodos basados en árboles (CART), support vector machines (SVM), redes neuronales, random forest, etc.

El módulo incluye los aspectos más importantes para la **construcción de modelos predictivos**: pre-procesamiento y descripción básica de los datos, **optimización de los parámetros** involucrados en cada una de las técnicas, **evaluación de la capacidad predictiva** de los modelos **mediante técnicas de remuestreo (validación cruzada y bootstrapping)**, selección de variables que van a formar parte del modelo, comportamiento de cada técnica en **problemas de alta dimensión**, etc.

INSCRIPCIÓN Y DATOS PRÁCTICOS

MÓDULO	**Técnicas Estadísticas de Data Mining con R**
Docentes	Jesús Herranz Valera (Bioestadístico IMDEA Alimentación) jesus.herranz@imdea.org
Fechas del Curso	14, 15, 16 y 17 de diciembre de 2015. Horario: 9:00 – 18:00
Inscripciones	Hasta el 13/12/2015 en http://matriculas.fuam.es/matriculauam/Convocatorias.action
Lugar de celebración	Instituto IMDEA Alimentación Ctra. Cantoblanco, nº8 Campus UAM – Madrid http://www.alimentacion.imdea.org
Nº Plazas	16 plazas. Mínimo de 10 asistentes para impartirse el curso
Precio	420€
Créditos	3 ECTS
Requisitos previos	Para el máximo aprovechamiento del módulo, son necesarios conocimientos de R y de Estadística, incluido modelos de regresión (en caso de duda contacten con los profesores del módulo)
Observaciones	Se entregará un manual encuadernado con todas las diapositivas del curso (540 aprox.) y todos los scripts y ficheros de datos Es **imprescindible** que los participantes traigan su propio **ordenador portátil**, y que se haya instalado el software R previamente

MÓDULO 7: TÉCNICAS ESTADÍSTICAS DE DATA MINING CON R

DESCRIPCIÓN DETALLADA

El módulo está dividido en 21 sesiones, con una duración variable de entre 1 y 2 horas, dependiendo de la importancia del tema.

En cada sesión se explican los conceptos estadísticos teóricos ilustrados desde un punto de vista práctico con ejemplos realizados con R, explicando los paquetes y funcionalidades más importantes del software para cada técnica. Finalmente, se proponen ejercicios a los alumnos para que puedan practicar los conocimientos adquiridos.

1. Introducción. Data Mining. Aprendizaje Supervisado y No Supervisado. Modelos Predictivos. Clasificación y Regresión
2. Nociones de R. Funciones y programación. Tratamiento de datos
3. Análisis descriptivo. Pre-procesamiento de datos
4. Análisis de Componentes Principales (PCA)
5. Análisis Clúster. Métodos Jerárquicos. K-means
6. Análisis Discriminante Lineal (LDA). Clasificador Naïve Bayes
7. Regresión Logística
8. Medidas de Evaluación de Modelos. Medidas de capacidad predictiva. Curvas ROC
9. Técnicas de Evaluación de Modelos. Validación cruzada. Muestras Bootstrap
10. Regresión Lineal. Métodos de Regresión por pasos
11. Regresión sobre Componentes Principales. Mínimos Cuadrados Parciales (PLS)
12. Métodos de Regresión Penalizados. Ridge Regression. LASSO
13. Métodos basados en Vecindad y Núcleos. k-NN
14. Redes Neuronales (NN)
15. Support Vector Machines (SVM)
16. Árboles de Regresión y Clasificación (CART)
17. Multiclasificadores. Bagging. Boosting
18. Random Forest (RF)
19. Selección de Variables. Filter. Wrapper. Recursive Feature Elimination
20. Caso práctico: Comparación de Modelos Predictivos
21. Caso práctico: Evaluación de la Significación Estadística de un Clasificador mediante Test de Permutaciones