

16 y 17 de junio de 2014 - Madrid



# V Jornadas

## Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

GENAEIO - 2014

Programa/Comunicaciones



© Ministerio de Defensa de España

Organizan:



Universidad  
de Granada





V JORNADAS DE ENSEÑANZA Y APRENDIZAJE DE  
LA ESTADÍSTICA Y LA INVESTIGACIÓN OPERATIVA

Madrid,

del 16 al 17 de junio de 2014



**PROGRAMA  
COMUNICACIONES**

**CATÁLOGO GENERAL DE PUBLICACIONES OFICIALES**  
<http://publicacionesoficiales.boe.es/>

Edita:



<http://publicaciones.defensa.gob.es/>

© Autor y editor, 2014

NIPO: 083-14-224-2 (edición papel)

NIPO: 083-14-223-7 (edición en línea)

Depósito Legal: M-27307-2014

Fecha de edición: octubre, 2014

Imprime: Imprenta Ministerio de Defensa

Las opiniones emitidas en esta publicación son exclusiva responsabilidad del autor de la misma.

Los derechos de explotación de esta obra están amparados por la Ley de Propiedad Intelectual. Ninguna de las partes de la misma puede ser reproducida, almacenada ni transmitida en ninguna forma ni por medio alguno, electrónico, mecánico o de grabación, incluido fotocopias, o por cualquier otra forma, sin permiso previo, expreso y por escrito de los titulares del © Copyright.

En esta edición se ha utilizado papel 100% reciclado libre de cloro.



## Presentación

Los días 16 y 17 de junio de 2014 se celebraron las V Jornadas para la Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa en la Escuela de Guerra del Ejército de Tierra en Madrid. Estas jornadas, que forman parte de las actividades que realiza el grupo **GENAEIO**, han sido precedidas por las celebradas en Melilla, en las Palmas de Gran Canaria, en Navarra y en Cádiz en los años 2010, 2011, 2012 y 2013, respectivamente.

Las jornadas pretenden ser un punto de encuentro donde mostrar el estado actual de la enseñanza y de la innovación docente en estadística e investigación operativa y, al mismo tiempo, donde presentar y discutir los últimos avances en estos campos. Se estructuran en base a diversos talleres, cuyo objetivo principal es el aprendizaje y manejo de nuevas técnicas docentes y que a la vez proporcionan recursos y materiales con los que abordar nuestra labor docente desde diferentes perspectivas y entornos de trabajo. Estas jornadas están dirigidas, principalmente, a los profesores universitarios del Área de Estadística e Investigación Operativa.

Para la organización de las jornadas se ha contado con la colaboración de numerosas personas que lo han hecho posible. Destaca el esfuerzo que ha realizado el Ministerio de Defensa, a través de su Secretaría General Técnica, en colaboración con la Universidad de Granada, la Sociedad de Estadística e Investigación Operativa y el grupo GENAEIO. El comité organizador ha estado formado por José Antonio Carrillo Ruíz, Miguel Ángel Sanz Calvo, Ovidio Muñoz Fernández, Antonio Planells del Pozo y Miguel Ángel Daza Arbolí pertenecientes al Área de Información y Ayuda a la Decisión de la Vicesecretaría General Técnica del Ministerio de Defensa.

Este libro contiene el programa, los trabajos de los talleres y los pósters presentados en estas jornadas.

Como coordinadora del grupo GENAEIO deseo agradecer a todos los autores y ponentes, por su trabajo y participación en las jornadas, a los miembros del Comité Organizador, su esfuerzo y dedicación en la organización y planificación de las sesiones, a las entidades organizadoras y colaboradoras, su contribución a la financiación de este evento, y a todos aquellos que de una u otra forma han contribuido a su realización.

M.<sup>a</sup> Jesús García-Ligero Ramírez  
Coordinadora del grupo de trabajo GENAEIO



**Convoca:**

Grupo de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa (GENAEIO)

**Organizan:**

Grupo de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa (GENAEIO)

Secretaría General Técnica del Ministerio de Defensa

Departamento de Estadística e Investigación Operativa de la Universidad de Granada

**Colaboran:**

Sociedad de Estadística e Investigación Operativa (SEIO)





## **Comité Organizador**

Presidente: José Antonio Carrillo Ruíz

Coordinadora GENAEIO: M.<sup>a</sup> Jesús García-Ligero Ramírez

Vocales: Miguel Ángel Sanz Calvo

Ovidio Muñoz Fernández

Antonio Planells del Pozo

Secretario: Miguel Ángel Daza Arbolí



## ÍNDICE

|                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------|----|
| Programación general .....                                                                                                         | 13 |
| Programación detallada .....                                                                                                       | 15 |
| Talleres .....                                                                                                                     | 17 |
| Creación de exámenes <i>online</i> con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave ..... | 17 |
| Problemas numéricos con corrección automática: qué se puede y qué se debe hacer.....                                               | 35 |
| Big data para estadísticos.....                                                                                                    | 53 |
| Resumen.....                                                                                                                       | 53 |
| Pósters .....                                                                                                                      | 71 |
| Herramientas multimedia para la mejora de la enseñanza.....                                                                        | 71 |
| Caso práctico de uso de la función sweave para la elaboración de exámenes usando r y tex.....                                      | 73 |
| Resumen.....                                                                                                                       | 75 |
| Herramienta docente para la creación de actividades individuales con corrección inmediata.....                                     | 77 |
| Estadística e investigación operativa en el proceso de internacionalización de la universidad española .....                       | 79 |
| Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en appes.....                                                     | 81 |
| Comparación de metodologías de evaluación continua .....                                                                           | 83 |
| Material interactivo para el aprendizaje del método simplex dual .....                                                             | 85 |
| Distribuciones de probabilidad con R .....                                                                                         | 87 |
| Enseñando geoestadística con R.....                                                                                                | 89 |
| Desarrollo de un entorno virtual de autoaprendizaje de la estadística .....                                                        | 91 |
| <i>Flippeando</i> con estadística .....                                                                                            | 93 |
| Autores .....                                                                                                                      | 95 |
| Participantes .....                                                                                                                | 98 |



## PROGRAMACIÓN GENERAL

| <b>Hora</b>   | <b>Lunes 16</b>                                   | <b>Hora</b>   | <b>Martes 17</b>                                  |
|---------------|---------------------------------------------------|---------------|---------------------------------------------------|
| 8:30 - 9:00   | Recepción de asistentes y entrega de material     |               |                                                   |
| 9:00 - 9:30   | <b>Inauguración de las jornadas</b>               | 9:00 - 11:00  | <b>Taller 3</b>                                   |
| 9:30 - 11:30  | <b>Taller 1</b>                                   | 11:00 - 11:30 | Pausa café                                        |
| 11:30 - 12:00 | Pausa café                                        | 11:30 - 14:00 | <b>Taller 4</b>                                   |
| 12:00 - 15:00 | <b>Visita</b>                                     | 14:00 - 16:00 | Almuerzo                                          |
| 15:00 - 16:00 | Almuerzo                                          | 16:00 - 17:30 | <b>Sesión 2 de comunicaciones orales/ pósters</b> |
| 16:30 - 18:30 | <b>Taller 2</b>                                   | 17:30 - 19:30 | <b>Reunión del grupo GENAEIO</b>                  |
| 18:30 - 20:30 | <b>Sesión 1 de comunicaciones orales/ pósters</b> |               |                                                   |
|               |                                                   | 21:00         | <b>Acto de clausura (RM Alcázar)</b>              |



## PROGRAMACIÓN DETALLADA

Lunes 16 de junio 2014

**8:30 - 9:00.** Recepción de asistentes y entrega de material.

**9:00 - 9:30. Inauguración de las jornadas.**

Intervienen:

**D. David Javier Santos Sánchez.**

Secretario General Técnico  
(MINISDEF/SEGENTE).

**D.ª M.ª Jesús García-Ligero Ramírez.**

Coordinadora del grupo GENAEIO.

**9:30 - 11:30. Taller 1: «Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave».**

Realiza: **D. Miguel Ángel Daza Arbolí** (Centro de Investigación Operativa de la Defensa, Ministerio de Defensa).

**11:30 - 12:00.** Pausa café.

**12:00 - 15:00.** Visita guiada al Canal de Experiencias Hidrodinámicas de El Pardo (CEHIPAR).

**15:00 - 16:00.** Almuerzo.

**16:30 - 18:30. Taller 2: «Problemas numéricos con corrección automática: qué se puede y qué se debe hacer».**

Realiza: Profesor **D. José Antonio González Alastrué** (Departamento de Estadística e Investigación Operativa, Universidad Politécnica de Cataluña).

**18:30 - 20:30. Sesión 1 de comunicaciones orales/pósters.**

## **Martes 17 de junio 2014**

**9:00 - 11:00. Taller 3: «Big data desde la perspectiva de un estadístico».**

Realiza: **D. Carlos Gil Bellosta** (*Senior Data Scientist* en eBay Inc. Zurich, Suiza).

**11:00 - 11:30.** Pausa café.

**11:30 - 14:00. Taller 4: «Introducción práctica al análisis estadístico y gráfico de grandes volúmenes de datos utilizando R y Hadoop».**

Realiza: **D. Carlos Gil Bellosta** (*Senior Data Scientist* en eBay Inc. Zurich, Suiza).

**14:00 - 16:00.** Almuerzo.

**16:00 - 17:30. Sesión 2 de comunicaciones orales/pósters.**

**17:30 - 19:30. Reunión del grupo GENAEIO.**

**21:00.** Acto de clausura en la Residencia Militar «El Alcázar».



## TALLERES

### Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave

*Carrillo Ruiz, José Antonio;*<sup>1</sup> *Sanz Calvo, Miguel Ángel;*<sup>2</sup> *Muñoz Fernández, Ovidio;*<sup>3</sup> *Planells del Pozo, Antonio;*<sup>4</sup> *Daza Arbolí, Miguel Ángel.*<sup>5</sup>

<sup>1</sup> jcarrillor@et.mde.es, MINISDEF/AIAD.

<sup>2</sup> msanzc@et.mde.es, MINISDEF/UNESOC.

<sup>3</sup> omunfer@et.mde.es, MINISDEF/CIODEF.

<sup>4</sup> aplanells@et.mde.es, MINISDEF/UES.

<sup>5</sup> mdazarb@et.mde.es, MINISDEF/CIODEF.

### Resumen

Utilización de la aplicación LimeSurvey, gestor de encuestas en línea de código abierto, para la realización de un examen tipo test puntuable (incluyendo la opción de penalizar las respuestas erróneas). El alumno conocerá su resultado al finalizar este examen. La aplicación LimeSurvey ofrece la posibilidad de exportar los resultados a varios formatos: CSV, Excel, pdf, spss/pspp, R. De entre ellos nos ocuparemos del que ofrece la posibilidad de exportar los datos a R. La forma en la que exporta los datos a formato R se materializa en dos archivos, uno de tipo CSV que lleva los datos en sí, y otro .r que al ser llamado desde R carga los datos del archivo CSV así como las etiquetas de las opciones de respuesta. Una vez que se dispone de los datos en R lo deseable sería poder generar un informe, imprimible tipo pdf. Para ello haremos uso de la función Sweave de R. Esta función R es capaz de leer de un archivo noweb con extensión .rnw en el que se mezclan los trozos de texto a presentar, en formato LaTeX, y los trozos de comandos y secuencias de R, para ser ejecutado. El resultado que se obtiene al aplicar Sweave es un archivo con extensión .tex en formato LaTeX que lleva los trozos de texto y los resultados de los comandos y secuencias de R (ya ejecutados). Como paso final quedará convertir ese archivo de texto (en LaTeX) en un archivo pdf, para ello bastará con utilizar otra función de R, texi2pdf, aplicada sobre este archivo.

**Palabras clave:** LimeSurvey, R, Sweave, noweb, LaTeX.

**Clasificación AMS:** 62-01, 97-01, 97-04.

## Introducción

Cuando se plantea la realización de un curso, y los de estadística o investigación operativa no están exentos de ello, por lo general, hay que realizar algún tipo de prueba objetiva para valorar los objetivos alcanzados en el mismo. Lo deseable, entonces, sería poder disponer de una herramienta capaz de generar un examen, a ser posible, disponible para ser respondido en modo *online*, que permita corregir y mostrar la corrección al terminar, que incluya la posibilidad de penalizar las respuestas erróneas, y que permita exportar los datos registrados a varios formatos disponibles pero en especial, a uno que permita su posterior tratamiento con el programa R.

El objetivo de este taller será tratar la interacción entre las aplicaciones LimeSurvey, R y LaTeX para poder generar informes de los resultados de los exámenes realizados mediante LimeSurvey. La realización de exámenes con LimeSurvey ya ha sido tratada por otros autores (Gregorio Serrano, 2010), al igual que la combinación R y LaTeX con Sweave que también ha sido tratada por otros autores (Francesc Carmona, 2012). En este trabajo se pretende unir todo el proceso. Desde que se crea un cuestionario (examen) en LimeSurvey hasta llegar a la presentación final de los resultados en pdf escrito con LaTeX, pasando por las operaciones y gráficos necesarios realizados con R.

## Requerimientos

Para la realización de este taller se necesitan los siguientes componentes:

- La aplicación LimeSurvey.
- Un servidor web Apache.
- Una base de datos MySQL.
- R.
- Complementos y funciones de R como Sweave() y texi2pdf().
- Una distribución de LaTeX, como por ejemplo, MikTeX.

## LimeSurvey

El primero de los componentes que necesitamos para este taller es la aplicación LimeSurvey disponible en su página web (<LimeSurvey.org>).

LimeSurvey (anteriormente PHPSurveyor) es una aplicación de código abierto (*open source*)<sup>1</sup> para la aplicación de encuestas en línea, está escrita en PHP y utiliza bases de datos del tipo MySQL, PostgreSQL o MSSQL. Esta aplicación da la posibilidad a usuarios sin conocimientos de programación del desarrollo, publicación y recolección de las respuestas de sus encuestas.

Las encuestas pueden incluir ramificaciones a partir de condiciones, plantillas y diseño personalizado usando un sistema de plantillas web, e incluye utilidades básicas de análisis estadístico para el tratamiento de los resultados obtenidos.

El tipo de acceso a las encuestas puede ser tanto público como controlado estrictamente por las claves que pueden ser utilizadas una sola vez (*tokens*) asignadas a cada persona que participe en la encuesta.

---

<sup>1</sup> Código abierto es la expresión con la que se conoce al *software* distribuido y desarrollado libremente. Se focaliza más en los beneficios prácticos (acceso al código fuente) que en cuestiones éticas o de libertad que tanto se destacan en el *software* libre.

Además, los resultados pueden ser anónimos, separando los datos de los participantes de los datos que proporcionan, inclusive en las encuestas controladas.

## LimeSurvey. Instalación

En el mercado existen numerosos servicios de alojamiento web que ofrecen servicio para LimeSurvey, incluso en la propia página web de LimeSurvey, <<https://www.limeservice.com/en/>>, está disponible un servicio de hospedaje gratuito, con el que poder realizar las primeras pruebas.



Figura 1. Detalle de la página de inicio de LimeSurvey. Servicio de hospedaje

No obstante para realizar una instalación en local o en un servidor necesitamos un servidor web Apache y una base de datos, por ejemplo, MySQL. Actualmente en el mercado es posible disponer de versiones actualizadas de estas herramientas, por separado. Pero la opción más sencilla es realizar una instalación con XAMPP.

XAMPP es un servidor independiente de plataforma, *software* libre, que consiste principalmente en la base de datos MySQL, el servidor web Apache y los intérpretes para lenguajes de *script*: PHP y Perl. El nombre proviene del acrónimo de X (para cualquiera de los diferentes sistemas operativos), Apache, MySQL, PHP, Perl.

El programa está liberado bajo la licencia GNU<sup>2</sup> y actúa como un servidor web libre, fácil de usar y capaz de interpretar páginas dinámicas. Actualmente XAMPP está disponible para Microsoft Windows, GNU/Linux, Solaris y Mac OS X.

Para la instalación en la máquina virtual se ha tomado la versión disponible para Windows.

Aunque también es posible realizar una descarga completa (LimeSurvey + XAMPP) desde la página web de LimeSurvey (<<http://www.limesurvey.org/en/stable-release>>).

<sup>2</sup> La Licencia Pública General de GNU, en inglés *GNU General Public License* (GNU GPL), es la licencia más ampliamente usada en el mundo del *software* y garantiza a los usuarios finales (personas, organizaciones, compañías) la libertad de usar, estudiar, compartir (copiar) y modificar el *software*.



Figura 2. Detalle de la página de XAMPP con las versiones disponibles para los diferentes SS. OO.

Category: Latest stable release

Order by: Default | Name | Date | Hits | [Descending]

| Files:                                                                 |                  |          |                 |
|------------------------------------------------------------------------|------------------|----------|-----------------|
| <a href="#">limesurvey295plus.build149520.zip</a>                      | 2014-05-20 18:44 | 42.85 MB | Downloads: 6418 |
| License: GPLv2 or later                                                |                  |          |                 |
| <a href="#">limesurvey295plus.build149520.tar.gz</a>                   | 2014-05-20 18:44 | 39.58 MB | Downloads: 780  |
| License: GPLv2 or later                                                |                  |          |                 |
| <a href="#">limesurvey295plus.build149520.tar.bz2</a>                  | 2014-05-20 18:44 | 34.05 MB | Downloads: 192  |
| License: GPLv2 or later                                                |                  |          |                 |
| <a href="#">limesurvey295plus.build149520.Tz</a>                       | 2014-05-20 18:44 | 26.42 MB | Downloads: 223  |
| License: GPLv2 or later                                                |                  |          |                 |
| <a href="#">limesurvey295plus.build149520.en.xampp.win32.setup.exe</a> | 2014-05-20 18:44 | 62.48 MB | Downloads: 1481 |
| License: GPLv2 or later                                                |                  |          |                 |

Figura 3. Detalle de la página web de descargas LimeSurvey, seleccionada opción con XAMPP

## LimeSurvey. Configuración

Acceder por primera vez. Para ello hay que indicar en el navegador que desea acceder a la página web que contiene el *script* de administración, si se realizó una instalación estándar esta dirección será algo similar a:

<<http://midominio/limesurvey/admin/admin.php>>.

Aparecerá una pantalla solicitándole su usuario y contraseña, que tendrán como nombre y usuario los que se indicarán por defecto en el archivo *config.php*, credenciales que habitualmente son:

*User: admin*

*Password: password*

Se recuerda la conveniencia de cambiar la contraseña por defecto cuando se acceda por primera vez.

Este usuario administrador (*admin*) tiene privilegios de «Superusuario».

## LimeSurvey. Usuarios

LimeSurvey permite crear usuarios y dotar a cada uno de ellos con una configuración de permisos particularizada. No obstante para este taller se hará uso del usuario *SuperAdmin*, de usuarios *admin* con privilegios para crear encuestas y de usuarios normales, que serán los que realicen las encuestas.

Estos nuevos usuarios administradores solo podrán ver las encuestas que ellos diseñen, así como a los usuarios que ellos den de alta. Es decir, cada uno de estos usuarios administradores puede realizar una gestión particularizada dentro de la plataforma, sin interferir en las encuestas y usuarios de otro administrador (de su mismo nivel).

## LimeSurvey. Encuestas

A la hora de poder gestionar las encuestas, LimeSurvey va mostrando una barra de herramientas cada vez que vamos descendiendo dentro de la organización de una encuesta, hasta llegar a la pregunta, que es el nivel más bajo.

En el primer nivel está el gestor de encuestas.



Figura 4. Detalle de la barra de herramientas del gestor de encuestas

Una vez que hemos seleccionado una encuesta de entre todas las disponibles, LimeSurvey nos mostrará una nueva barra de herramientas, para la encuesta, con nuevas opciones disponibles.

Hemos accedido a una encuesta, ahora el siguiente paso es descender en la orgánica, hasta llegar a la pregunta (el nivel más bajo). Este paso consiste en elegir un Grupo de preguntas o Sección.

En este punto, se debe seleccionar una de las preguntas disponibles (de entre todas las existentes o crear una nueva) para esta Sección o Grupo de preguntas.

Una vez dentro de la pregunta, si esta tiene opciones de respuesta, el siguiente paso será editar las opciones de respuesta. En este caso no se muestra otra barra de herramientas, seguimos en la barra de herramientas de la pregunta seleccionada.

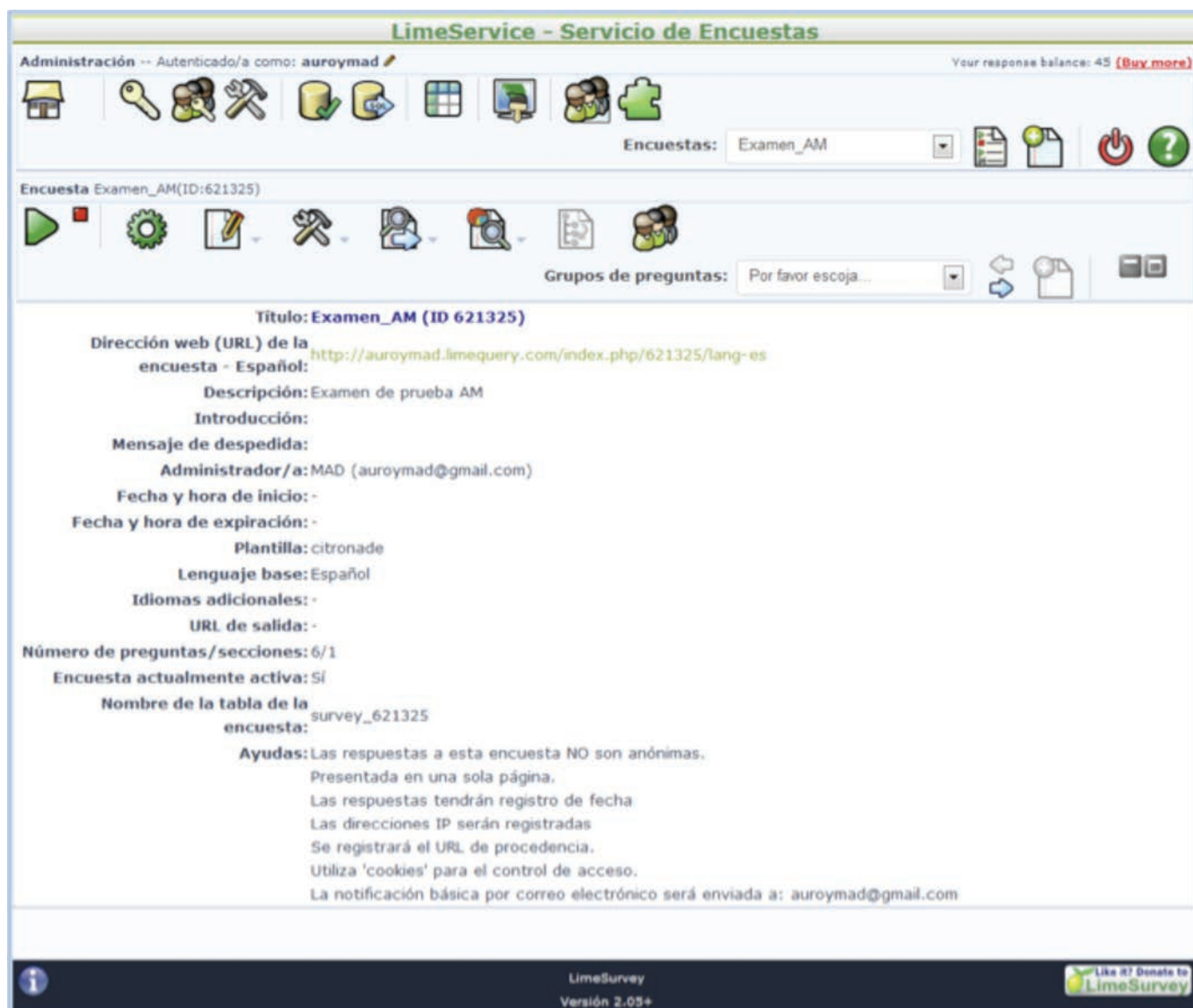


Figura 5. Detalle de la barra de herramientas para una encuesta



Figura 6. Detalle de la barra de herramientas para Grupos de preguntas

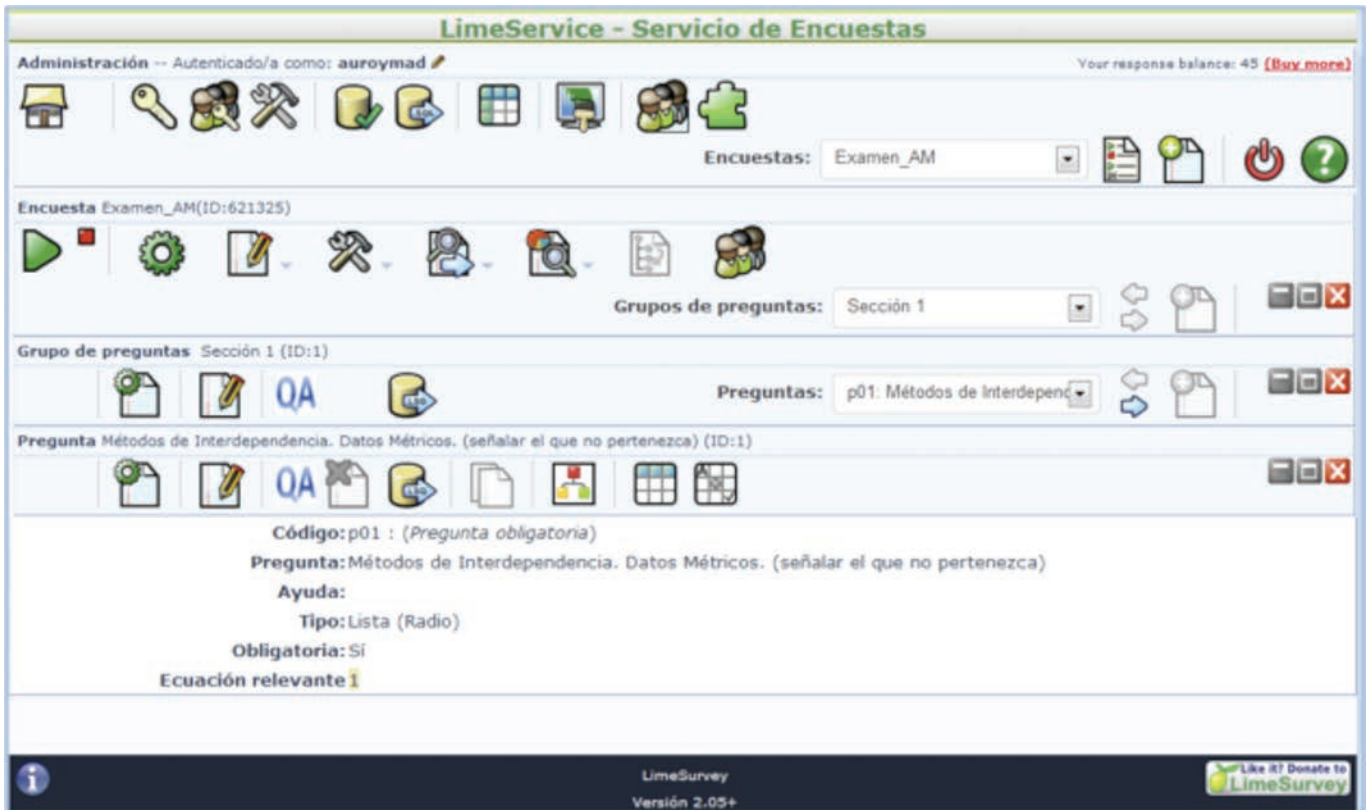


Figura 7. Detalle de la barra de herramientas de una pregunta



Figura 8. Detalle de las opciones de respuesta de una pregunta

A continuación se muestran los tipos de preguntas disponibles en LimeSurvey, agrupados por categorías:

### Pregunta de respuesta única:

- Elegir del 1 al 5
- Lista (radio)
- Lista (desplegable)
- Lista (botones radiales) con comentarios

### Preguntas de máscara:

- Entrada numérica
- Ecuación
- Género
- *Ranking*
- Subir ficheros
- Entrada numérica múltiple
- Fecha/Tiempo
- Mostrar texto
- Selección de idioma
- Sí/No

### Matrices:

- Matriz
- Matriz (elegir del 1 al 10)
- Matriz (números)
- Matriz (textos)
- Matriz de columnas
- Matriz (ampliar, mantener, reducir)
- Matriz (elegir del 1 al 5)
- Matriz (Sí/No/No sé)
- Matriz de doble escala

### Preguntas de opción múltiple:

- Opción múltiple
- Opción múltiple con comentarios

### Preguntas de texto:

- Texto libre (línea)
- Texto libre (párrafo)
- Texto libre (varios párrafos)
- Varios textos cortos

De entre todas las opciones disponibles de tipos de preguntas, para poder elaborar un futuro examen, sobre la base de un cuestionario/encuesta y que permita la calificación de las opciones elegidas, solo se podrán utilizar aquellas en las que LimeSurvey permita puntuar en su opción **Valor de asesoría**, que son:

- Lista (desplegable)
- Lista (radio)
- Lista (botones radiales) con comentarios

En la siguiente figura se muestra, como ejemplo, las opciones de respuesta a una pregunta creada en LimeSurvey con el tipo de pregunta *Lista (radio)*. En esta pregunta se quiere además de puntuar positivamente la opción correcta, puntuar negativamente las opciones incorrectas y no puntuar si se opta por no contestar la pregunta. Como el valor numérico que se puede introducir en la casilla **Valor de asesoría** debe ser un número entero, se ha optado como solución puntuar con un **3** la respuesta correcta, con un **-1** las respuestas incorrectas y con un **0** la opción de dejar sin responder la pregunta. Estos números se han elegido para cumplir con la corrección de aleatoriedad que sugiere descontar los errores divididos por el número de opciones de respues-



ta posibles menos uno. Así pues, como esta pregunta dispone de cuatro opciones de respuesta, cada fallo debe suponer un tercio de una puntuación positiva. Por supuesto al final del proceso habrá que dividir la puntuación obtenida entre tres. Además de adecuar el resultado para que se encuentre en el rango [0,10].



Figura 9. Detalle de las puntuaciones de las opciones de respuesta de una pregunta

Se debe hacer notar que se ha incluido una quinta opción con el literal (en blanco) y que se ha configurado como opción preseleccionada, de tal forma que al acceder a la pregunta esta aparezca marcada y su puntuación sea 0. Además se cumple que al ser la pregunta de respuesta obligatoria por lo menos hay una opción marcada, no dando error y permitiendo continuar con el cuestionario (examen).

### LimeSurvey. Mostrar resultado final del examen al alumno

Hasta este momento se ha configurado LimeSurvey para contabilizar los puntos positivos de los aciertos y los puntos negativos de los errores. Cuando el alumno finalice el examen sería deseable que se le pudiera mostrar una pantalla con su calificación final.

Afortunadamente, para ello LimeSurvey cuenta con un apartado donde especificar esto. Se llama **Evaluaciones** y aparece al seleccionar el menú desplegable de «Propiedades de la encuesta», que se muestra en la figura siguiente:



Figura 10. Opción Evaluaciones del menú Propiedades de la encuesta

Al pulsar sobre la opción Evaluaciones LimeSurvey nos muestra la siguiente pantalla, en la que hemos configurado cómo debe ser la salida en función de la puntuación final alcanzada. En esta la nota final es el resultado de redondear (*round*) a 2 decimales la suma TOTAL de puntos (positivos y negativos) obtenidos en las diversas preguntas, dividido por 15 y multiplicado por 10, para que la nota máxima sea un 10.



Figura 11. Detalle de las Reglas de asesoría creadas para este examen

Después de contestar el examen la plataforma LimeSurvey mostrará al examinado la siguiente pantalla:



Figura 12. Detalle de la pantalla que mostrará LimeSurvey al finalizar el examen

donde nos indica la puntuación obtenida, así como una descripción del límite establecido entre NO APTO y APTO.

LimeSurvey nos da la opción de poder ver en un informe HTML lo que hemos respondido, así como de generar un pdf con ese informe.

Para que la puntuación final obtenida se pueda exportar, se necesita crear una pregunta de tipo ecuación. En este caso se ha creado la pregunta p00, cuyo contenido de la ecuación es la suma de los valores obtenidos en cada una de las preguntas: **sum(p01.value, p02.value, p03.value, p04.value, p05.value)**.

| LimeService - Servicio de Encuestas                                                      |                                                                                                      |
|------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Examen_AM                                                                                |                                                                                                      |
| <a href="#">Exportar a PDF</a>                                                           |                                                                                                      |
| Denominación de la encuesta (ID): Examen_AM (621325)                                     |                                                                                                      |
| Fecha de envío submitdate                                                                | 11/06/2014 12:16:44                                                                                  |
| Dirección IP                                                                             | 193.33.2.114                                                                                         |
| URL de referencia                                                                        | http://auroymad.limequery.com/index.php/admin/questions/sa/answeroptions/surveyid/621325/gid/1/qid/6 |
| Sección 1                                                                                |                                                                                                      |
| Métodos de Interdependencia. Datos Métricos. (señalar el que no pertenezca)              | Análisis Discriminante. [A3]                                                                         |
| Métodos de Dependencia. Variable Dependiente cualitativa. (señalar el que no pertenezca) | Análisis Factorial. [A4]                                                                             |
| Métodos de Interdependencia. (señalar el que no pertenezca)                              | Regresión Logística. [A2]                                                                            |
| Métodos de Dependencia. (señalar el que no pertenezca)                                   | Análisis Cluster. [A3]                                                                               |
| Métodos de dependencia (señale la afirmación correcta).                                  | Suponen las variables en dos grupos (dependientes e independientes). [A4]                            |

Figura 13. Detalle del informe conteniendo las respuestas realizadas por el alumno

The screenshot shows the LimeSurvey administration interface for a specific question. At the top, it says 'LimeService - Servicio de Encuestas'. Below that, there's a navigation bar with icons for home, search, user, settings, and other functions. The main area is divided into sections: 'Administración -- Autenticado/a como: auroymad', 'Encuesta Examen\_AM(ID:621325)', 'Grupo de preguntas Sección 1 (ID:1)', and 'Pregunta {sum(p01.value,p02.value,p03.value,p04.value,p05.value)} (ID:3)'. The question editor shows the code for the question: 'Código:p00 : (Pregunta opcional)', 'Pregunta:sum(p01.value, p02.value, p03.value, p04.value, p05.value)', 'Ayuda:', 'Tipo: Ecuación', 'Obligatoria: No', 'Ecuación relevante 1', and 'Ocultar permanentemente esta pregunta: si'. There are also buttons for 'QA' and 'Exportar a PDF'.

Figura 14. Detalle de la pregunta especial de tipo ecuación, como suma de las puntuaciones de cada pregunta

## LimeSurvey. Resultado a R

Una vez que todos los alumnos hayan contestado la encuesta (examen) llega el momento de recolectar todos esos datos y exportarlos a algún formato especial para ser tratados con posterioridad por algún *software*, como por ejemplo, **R**.

En la siguiente figura se muestran algunos de los formatos de exportación de datos disponibles.

En la figura hemos seleccionado la opción que nos interesa, que se corresponde con exportar los datos y generar un archivo de sintaxis de R (con extensión *.r*). Se descargan dos archivos,

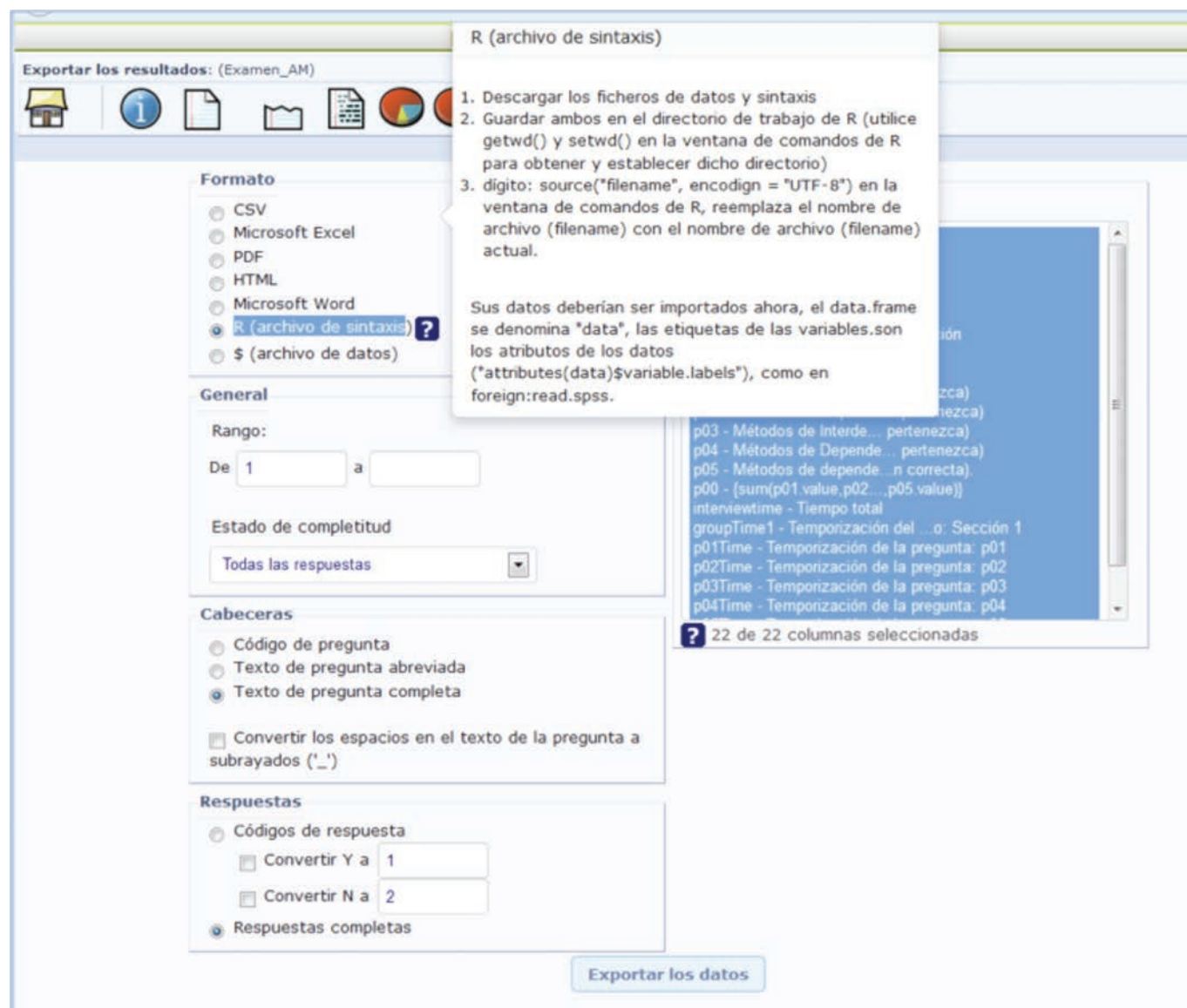


Figura 15. Detalle de la pantalla donde se muestran los distintos tipos de formatos para exportar los datos

uno con los datos en formato CSV y otro con el archivo de sintaxis que además de llamar al fichero de los datos CSV, lleva las etiquetas de las opciones de respuesta de cada una de las preguntas.

## R

R es un lenguaje y entorno de programación para análisis estadístico y gráfico.

Se trata de un proyecto de *software* libre, resultado de la implementación GNU del premiado lenguaje S. R y S-Plus (versión comercial de S) son, probablemente, los dos lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy populares en el campo de la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con finalidades específicas de cálculo o gráfico.

R se distribuye bajo la licencia GNU GPL y está disponible para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux.

## The Comprehensive R Archive Network

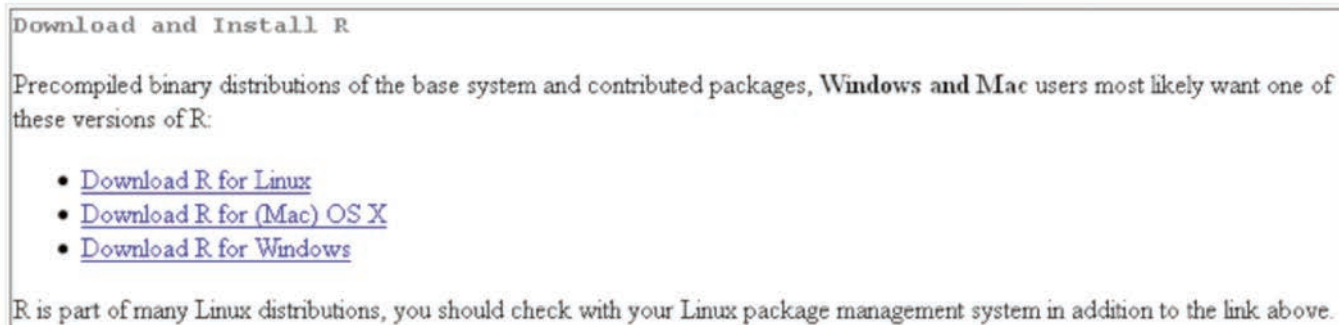


Figura 16. Detalle de la sección Download de <http://cran.es.r-project.org/>

### Sweave()

F. Leisch ha desarrollado un sistema llamado Sweave (Leisch, 2002) que permite incluir código R en documentos de LaTeX, y que funciona de manera similar a Cweb<sup>3</sup>, en lo que se llama un documento de tipo Rnw.

Este tipo de documento se puede procesar de dos maneras distintas:

1. Ejecutar el código R y sustituirlo por su salida para crear un documento LaTeX, que ya se puede compilar para generar un documento final.
2. Extraer todo el código R del documento Rnw, lo que permite ejecutarlo todo seguido y obtener los resultados. De esta forma todo lo que no sea código R será eliminado.

Estas acciones se llevan a cabo con dos funciones de R: 1) **Sweave()** y 2) **Stangle()**.

Como dice el manual oficial de Sweave: «Sweave proporciona un marco flexible para mezclar texto y código R con el objetivo de generar documentos de forma automática». Partiendo de un único archivo fuente que contenga el texto y el código R y que tras el procesamiento con Sweave se obtendrá como resultado texto con formato, el código R utilizado (se puede omitir) y/o los resultados de haber ejecutado el código R (texto, cálculos, gráficos).

El archivo utilizado para este proceso con extensión Rnw es un archivo noweb. Un archivo noweb es un archivo de texto simple que consiste en varios segmentos separados de texto y código de programación. Estos segmentos se llaman trozos (*chunks*). Los hay de dos tipos: trozos de texto y trozos de código.

En la siguiente figura se muestra un extracto del archivo noweb *Examen1.rnw*; en él se han señalado con unos rectángulos las zonas, de ese archivo, que incluyen trozos de código.

<sup>3</sup> CWEB es un sistema de programación de computadoras creado por Donald Knuth y Silvio Levy como un seguimiento al sistema de programación literaria WEB de Knuth, utilizando el lenguaje de programación C (y en menor medida los lenguajes de programación C++ y Java) en lugar de Pascal.

Al igual que WEB, consiste en dos programas principales: CTANGLE, que produce código C compilable de los textos originales, y CWEAVE, que produce documentación imprimible con un formato agradable utilizando TeX.

Estos trozos (*chunks*) de código comienzan con `<<>=` y terminan con `@`.

```

15 \date{Junio 2014}
16
17 \begin{document}
18
19 \maketitle
20 En este documento se muestran, como ejemplo, los resultados obtenidos del examen de prueba realizado en AIAD:
21 \begin{center}
22 Todas las preguntas la nota se ha calculado con la siguiente fórmula
23 \begin{equation}
24 Nota = \frac{Aciertos - \frac{Errores}{opciones-1}}{3}
25 \end{equation}
26 \end{center}
27
28 En primer lugar se muestra un resumen con estadísticos básicos.
29
30 \begin{center}
31 <<fig=FALSE, echo = FALSE>>=
32 source("survey_621325_R_syntax_file.R", encoding="UTF-8")
33 Notas<-as.numeric(data$p00)
34 NumAlumnos<-length(Notas)
35 for (i in 1:length(Notas)) {if (Notas[i]<0) (Notas[i]=0)}
36 for (i in 1:length(Notas)) (Notas[i]=Notas[i]*2/3)
37 Notas<-round(Notas, digits=2)
38 summary(Notas)
39 @
40
41 \end{center}
42 Número de alumnos \Sexpr{NumAlumnos}
43
44 \begin{center}
45 <<fig=TRUE, echo = FALSE>>=
46 boxplot(Notas, ylab="Notas")
47 @
48

```

Figura 17. Extracto del archivo Examen1.rnw, con los trozos de código R resaltados

En el primero de los trozos la opción **fig**=FALSE indica que no se mostrará gráfico alguno, si es generado por el código. La opción **echo**=FALSE indica que no se quiere que en el documento final aparezcan las instrucciones de código R. Lo que sí se mostrarán son los resultados que se obtengan, como pueden ser los que produzca la función de R `summary(Notas)`.

En el segundo de los trozos como lo que se pretende es que se muestre el gráfico (*box-plot*), la opción **fig**=TRUE es la que lo permite. Pero como no se quiere que se muestren las instrucciones que son necesarias para generar dicho gráfico, la opción **echo** está en FALSE.

El resto del texto de este archivo es o bien texto a mostrar, o bien instrucciones de LaTeX para formatear el texto. Nótese que las instrucciones de LaTeX van precedidas de la barra (`\`).

```

15 \date{Junio 2014}
16
17 \begin{document}
18
19 \maketitle
20 En este documento se muestran, como ejemplo, los resultados obtenidos del examen de prueba realizado en AIAD:
21 \begin{center}
22 Todas las preguntas la nota se ha calculado con la siguiente fórmula
23 \begin{equation}
24 Nota = \frac{Aciertos - \frac{Errores}{opciones-1}}{3}
25 \end{equation}
26 \end{center}
27
28 En primer lugar se muestra un resumen con estadísticos básicos.
29
30 \begin{center}
31 \begin{chunk}
32 \begin{Soutput}
33      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
34  0.000  2.000   4.670   4.267   6.665  10.000
35 \end{Soutput}
36 \end{chunk}
37
38 \end{center}
39 Número de alumnos 10
40
41 \begin{center}
42 \includegraphics{Examen1-002}
43
44 \end{center}
45 \newpage
46

```

Figura 18. Extracto del archivo Examen1.tex resultado de ejecutar la función Sweave ("Examen1.rnw")

En la línea 42 del editor pretendemos que en el texto final aparezca el texto «Número de alumnos» seguido del valor *NumAlumnos* que se ha calculado en el *Chunk001*, para ello utilizamos la instrucción LaTeX  $\text{\Sexpr{NumAlumnos}}$ .

En R utilizar la función *Sweave()* de la siguiente forma: *Sweave("Examen1.rnw")*, con lo que se obtendrá un archivo con extensión *.tex* y cuyo extracto se muestra en la figura 18.

Nótese que el primer trozo de código se ha sustituido por su salida y que el segundo trozo de código se ha sustituido por la función LaTeX para incluir un gráfico en el resultado final.

Sweave nombra los gráficos que va generando a lo largo de la ejecución del código. Para nombrarlos usa el nombre del archivo *noweb.rnw* (*Examen1*), un guión, y un número de tres cifras que coincide con el número del trozo (*chunk*) en que ha sido generado, como el gráfico (*box-plot*) se generó en el segundo trozo (002), el nombre para el archivo gráfico será *Examen1-002*.

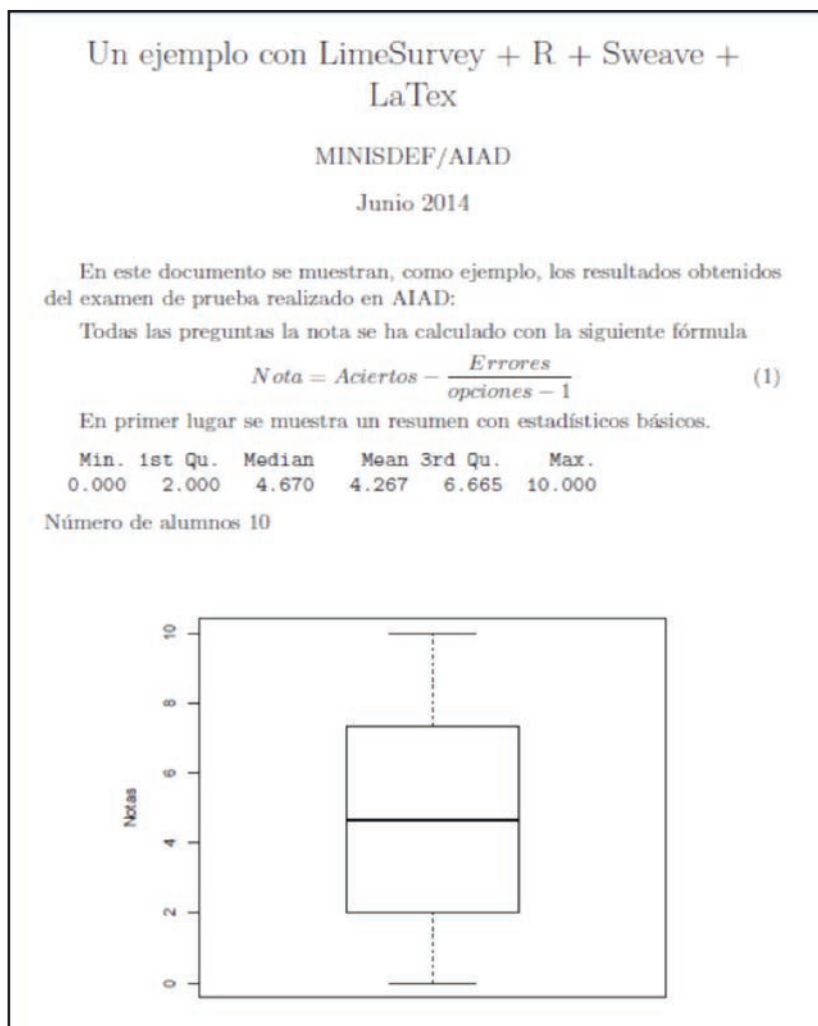


Figura 19. Detalle de un extracto del resultado final en pdf

Por último con el archivo *.tex* obtenido, en este caso, *Ejemplo1.tex*, lo que procede es convertirlo en un archivo con formato pdf. Para ello recurrimos a la función de R *tools::texi2pdf*, así pues, escribiremos *tools::texi2pdf("Examen1.tex")*. Lo que se muestra en la figura 19 es un extracto del pdf obtenido para ese archivo.

Para poder aplicar la función de R `tools::texi2pdf` necesitamos disponer de alguna distribución de *LaTeX*. Para este trabajo hemos utilizado *MiKTeX*.

## MiKTeX

MiKTeX es una distribución TeX/LaTeX para Microsoft Windows que fue desarrollada por Christian Schenk. Las características más apreciables de MiKTeX son su habilidad de actualizarse por sí mismo descargando nuevas versiones de componentes y paquetes instalados previamente, y su fácil proceso de instalación.

La versión actual de MiKTeX es 2.9. Además, tiene características que incluyen MetaPost y pdfTeX y compatibilidad con Windows 7. A partir de la versión 2.7 se incluyó soporte integrado para XeTeX. Desde la versión 2.9 ofrece soporte para ConTeXt Mark IV.

Características:

Las principales características de MiKTeX son las siguientes:

- **Es libre y fácil de instalar.**
- Incluye más de 800 paquetes con tipografías, macros, etc.
- **Tiene un visor propio de archivos dvi** denominado Yap.
- **Su código es abierto.**
- Posee compiladores TeX y LaTeX, convertidores para generar archivos PostScript (.ps), .pdf, .html, etc.; y herramientas para generar bibliografías e índices.
- Posee tres formas de instalación: pequeña, mediana y completa.

Para su configuración e integración con Sweave ver el punto 4 *Configuración de LaTeX* del trabajo *Generación automática de reportes con R y LaTeX* de Mariano Alfonso Morales Rivera, donde explica detalladamente los pasos a realizar tanto para usuarios Linux, como para usuarios Windows.

## Conclusión

En este taller se ha intentado poner de relieve el funcionamiento combinado de varias aplicaciones, muy útiles, distribuidas bajo Licencia Pública General (GNU), como son LimeSurvey, R (Sweave) y LaTeX.

Se ha podido comprobar que juntas hacen posible la creación y gestión de exámenes (encuestas), la extracción de los resultados y la generación de informes con cierto nivel de calidad. Informes que además de incluir texto, pueden incluir resultados y gráficos calculados con toda la potencia que ofrece R. **Todo ello convierte a este conjunto en una muy buena opción para la enseñanza.**

## Referencias

[1] Serrano, G. R. (2010): *Mi primer examen on-line*.

<<http://www.grserrano.es/wp/2010/12/mi-primer-examen-on-line-2/>>.

[2] Carmona, F. (2012): *Generación automática de informes con Sweave y LaTeX*.

<[http://www.ub.edu/stat/docencia/EADB/Manual\\_Sweave.pdf](http://www.ub.edu/stat/docencia/EADB/Manual_Sweave.pdf)>.



[3] Morales Rivera, M. A. (): *Generación automática de reportes con R y LaTeX*.

<[http://cran.r-project.org/doc/contrib/Rivera-Tutorial\\_Sweave.pdf](http://cran.r-project.org/doc/contrib/Rivera-Tutorial_Sweave.pdf)>.

[4] Leisch, F. (2005): *Sweave User Manual (R version 2.7.1)*.

<<http://www.stat.uni-muenchen.de/~leisch/Sweave/Sweave-manual.pdf>>.

[5] Leisch, F. (2002): *Sweave, part I: Mixing R y LaTeX*.

<<http://wiki.leg.ufpr.br/lib/exe/fetch.php/projetos:obesidade:sweave-rnews-2002-3.pdf>>.

[6] Leisch, F. (2002): *Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis*.

<<http://www.stat.uni-muenchen.de/~leisch/Sweave/Sweave-compstat2002.pdf>>.

[7] Gómez Rubio, V. (2010): *Creación de material docente dinámico con LaTeX*.

<[http://mates.albacete.org/ICEAMCM/PDF/v\\_gomezrubio.pdf](http://mates.albacete.org/ICEAMCM/PDF/v_gomezrubio.pdf)>.



## **Problemas numéricos con corrección automática: qué se puede y qué se debe hacer**

*González Alastrué, José Antonio*

jose.a.gonzalez@upc.edu

### **Resumen**

La presente comunicación describe los objetivos, organización, proceso y discusión de un taller orientado a proporcionar a docentes del campo de la estadística y la investigación operativa (y posiblemente de otros campos) una primera noción sobre la construcción de problemas con la plataforma e-status. El taller se estructura en base a cuatro casos que van incrementando el nivel de complejidad, para ir mostrando diversas posibilidades didácticas a la hora de elaborar buenos ejercicios que hagan reflexionar al estudiante.

**Palabras clave:** e-status, plataformas educativas, problemas automáticos.

**Clasificación AMS:** 62-01, 97-01, 97-04.

## Introducción

El objetivo del taller es que el asistente se haga una idea clara sobre cómo se construye un problema con la plataforma e-status. Es un taller del tipo *hands-on*, o «manos a la obra»: es decir, desde el primer momento el participante se ha de involucrar activamente, en vez de darle todo el protagonismo al ponente para que ocupe el tiempo hablando. Por supuesto, dado que es de esperar que en la audiencia haya personas que conocen poco o nada sobre e-status, es necesario dar unas referencias mínimas pero suficientes para proporcionar un contexto.

e-status es un aplicativo que funciona en web, y que se utiliza para poner a disposición de los alumnos (de asignaturas de probabilidad y estadística, generalmente) una serie de problemas que el alumno contesta mediante el navegador, obteniendo la corrección correspondiente de inmediato. Los problemas pueden repetirse tantas veces como sea necesario, puesto que tienen la capacidad de variar sus condiciones aleatoriamente. Al cambiar estas condiciones, las soluciones también son distintas.

El sistema que gobierna la lógica de los problemas es el programa/lenguaje R, bien conocido por los estadísticos. El profesor escribe un código para generar el modelo del problema y para valorar las respuestas de los alumnos. La aplicación se ocupa, entre otros aspectos, de la comunicación en el servidor entre lo que es propiamente el servicio web y R. Puede encontrarse más información en [1] y [2].

La clave de e-status radica en su especialidad. El programa se ha desarrollado desde cero, y sin plantear su integración en entornos más generales, conocidos como LMS (*Learning Management Systems*), de los cuales posiblemente sea el más popular Moodle por el hecho de ser *software* libre. Moodle dispone de algunas herramientas para elaborar cuestiones (o *quizzes*) de tipos muy variados. Sin embargo, la elaboración de preguntas numéricas es limitada y, en opinión del autor, algo farragosa. Existe un paquete para R llamado *exams* que puede integrarse en Moodle, mediante el paquete *Sweave*, y puede utilizarse el lenguaje R para programar la lógica del problema, aunque se trata fundamentalmente de la generación del modelo; la valoración de las respuestas continúa siendo relativamente pobre y compleja. e-status debe incorporar una gestión mínima de alumnos y profesores, asignaturas y grupos, calificaciones y estadísticas de resultados, pero a cambio proporciona agilidad y sencillez de uso al profesor para que este consiga hacer realidad el problema que tiene en la mente.

En las páginas que siguen se va a describir el funcionamiento del taller a partir de los contenidos empleados para el mismo, y que se pueden encontrar en la página web:

<<http://www-eio.upc.es/~josean/genaeio14/problema1.html>>.

## Procedimiento del taller

Previamente a la celebración del taller, se solicitó a los participantes una serie de datos que permitieran efectuar el alta con el perfil de «profesor» (que permite editar problemas) en e-status. Los pocos casos que quedaran sin resolver pueden solucionarse en los minutos iniciales del taller. De esta manera, los participantes pueden entrar en la aplicación inmediatamente. El ponente muestra en el proyector la ubicación del enlace para entrar en el editor, obviando por el momento otras funcionalidades de e-status. También proporciona la dirección web que incluye los contenidos preparados para el taller: se trata de cuatro páginas (*problema1.html*, *problema2.html*, *problema3.html*, *problema4.html*) enlazadas entre sí. Cuatro problemas completos que representan sendas fases de complejidad creciente, y que cada uno de los presentes tiene el cometido de implementar y probar desde el editor.

Como se puede deducir de lo anterior, el taller se ha planteado alejándose de un paradigma muy frecuente: pedir a los asistentes que construyan su trabajo a partir de un tema de su interés particular. En vez de emplear este enfoque, hemos decidido dar un material común a todos, pormenorizadamente detallado para que el participante no tenga más que copiar y pegar una serie de textos en el lugar correspondiente. ¿Creatividad? De acuerdo, sacrificada en aras de la eficiencia (si se dispusiera de más de dos horas, tal vez sería bueno invitar al asistente a experimentar). Hay un acuerdo generalizado, que no se va a poner aquí en duda, de que la implicación personal del alumno es un factor clave para el aprovechamiento de un curso. Sin embargo, el grado de implicación suele ser muy diverso, y es necesario disponer de tiempo sobrado para conseguir resultados satisfactorios en la sesión. Por estas razones se prefiere que este taller procure llevar hasta el final al mayor número posible de asistentes, teniendo en cuenta que tampoco es necesario completar los cuatro ejercicios. Un participante puede sentirse satisfecho si completa solo tres, o dos, problemas, y el resto puede aportar un beneficio marginal, aunque obviamente se le invita a finalizar las tareas en un momento posterior.

La primera pantalla del editor es el organizador de problemas (Figura 20). El ponente explica brevemente que el profesor-autor de problemas tiene la capacidad de abrir las carpetas que necesite para gestionar sus propios problemas, y además puede acceder a los problemas de sus compañeros siempre que estos se lo permitan (la filosofía general es la compartición de materiales, pero sin renunciar al derecho a la privacidad). Lo que el profesor vería en la pantalla son dos estructuras de tipo árbol: a la izquierda puede gestionar el objeto tipo *problema*, y a la derecha el objeto tipo *carpeta* (una operación típica sería mover el problema marcado a la carpeta marcada). Obviamente, al principio no hay ningún objeto creado, solo puede ver tres contenedores genéricos llamados «Carpetas propias», «Externas suscritas» y «Externas no suscritas»; se pide al participante que comience creando una carpeta, y un problema vacío en la misma. Tras ello accede al editor propiamente dicho (Figura 21), y puede empezar a componer el primer ejercicio, que se encuentra estructurado en las Tablas 1 y 2, y que son una réplica de las que se encuentran en la página web del taller.



Figura 20. Organizador de carpetas y problemas

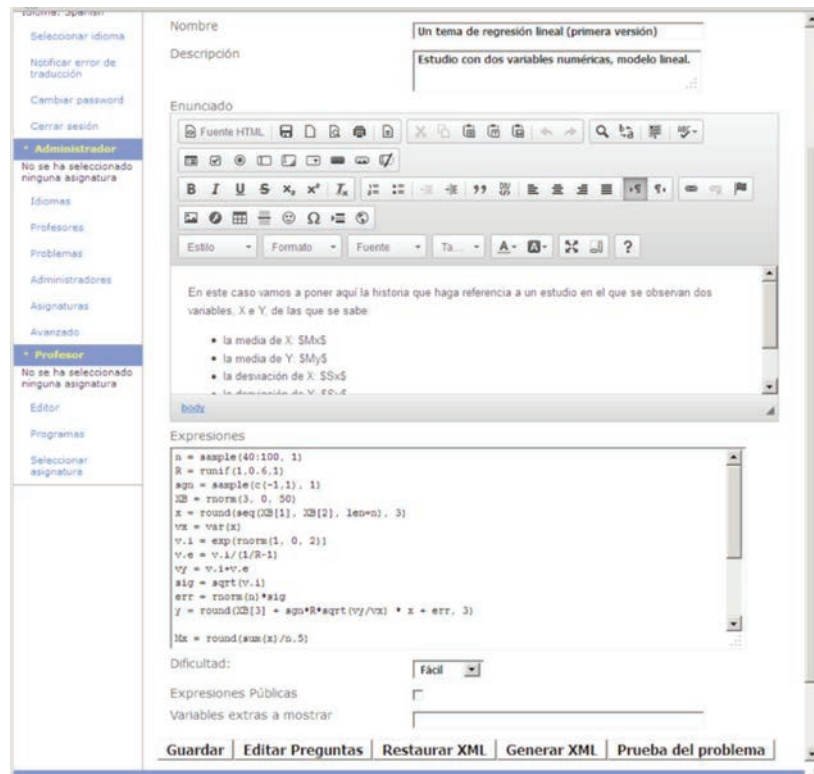


Figura 21. El editor de problemas

| Un tema de regresión lineal (primera versión) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Descripción                                   | Estudio con dos variables numéricas, modelo lineal.                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Enunciado                                     | <p>En este caso vamos a poner aquí la historia que haga referencia a un estudio en el que se observan dos variables, X e Y, de las que se sabe:</p> <ul style="list-style-type: none"> <li>• la media de X: <math>M_x</math></li> <li>• la media de Y: <math>M_y</math></li> <li>• la desviación de X: <math>S_x</math></li> <li>• la desviación de Y: <math>S_y</math></li> <li>• la covariancia de X e Y: <math>V_{xy}</math></li> </ul>                               |
| Modelo R (Expresiones)                        | <pre>n = sample(40:100, 1) R = runif(1,0.6,1) sgn = sample(c(-1,1), 1) XB = rnorm(3, 0, 50) x = round(seq(XB[1], XB[2], len=n), 3) vx = var(x) v.i = exp(rnorm(1, 0, 2)) v.e = v.i/(1/R-1) vy = v.i+v.e sig = sqrt(v.i) err = rnorm(n)*sig y = round(XB[3] + sgn*R*sqrt(vy/vx) * x + err, 3) Mx = round(sum(x)/n,5) My = round(sum(y)/n,5) Vxy = round(cov(x,y),5) Sx = round(sd(x),5) Sy = round(sd(y),5) LM = lm(y ~x) summ = summary(LM) tind = summ\$coef[1,1]</pre> |
| Expresiones públicas                          | Para compartir el modelo con otros autores.                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Variables extras a mostrar                    | (de momento, nada)                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

Tabla 1. Datos principales para la primera versión del ejercicio

Una vez se ha introducido la información contenida en la Tabla 1, se pulsa sobre el botón «Guardar», y seguidamente se elige la opción «Editar preguntas». En la pantalla que aparece inicialmente tendremos una lista vacía: usaremos una opción que aparece en el menú de la izquierda para crear la primera pregunta, con lo que se accede a la página que permite rellenar los diferentes campos que componen una pregunta, cuyo contenido se encuentra en la Tabla 2.

| Preguntas                                |                                                                                                                                                                      |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Enunciado 1                              | Si queremos predecir la Y con la X, utilizando un modelo lineal, ¿cuánto vale el término independiente de la recta estimada (usando el método de mínimos cuadrados)? |
| Solución a la pregunta                   | Tind                                                                                                                                                                 |
| Parámetro de sintaxis                    |                                                                                                                                                                      |
| Sintaxis de la respuesta                 | Real                                                                                                                                                                 |
| Plantilla de expresiones de la respuesta | Error relativo                                                                                                                                                       |
| Parámetro de expresiones                 | 1 (un tanto por ciento)                                                                                                                                              |
| Expresiones de respuesta                 | <pre>if (estatus.error_relativo(solucion_,respuesta_)&lt;=parametro_) {   resultado_&lt;-'1'; }else{   resultado_&lt;-'0'; } (no tocamos nada)</pre>                 |
| Peso                                     | 1                                                                                                                                                                    |
| Ayuda contextual                         | Aquí va un enlace en donde puedes leer más sobre el método:                                                                                                          |
| Link                                     | < <a href="http://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal#Eqnref_12">http://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal#Eqnref_12</a> >                          |
| Intentos                                 | 1                                                                                                                                                                    |
| Mostrar solución                         | (normalmente es lo más sensato)                                                                                                                                      |
| Mensaje correcto                         | ¡Bravo!                                                                                                                                                              |
| Mensaje semicorrecto                     |                                                                                                                                                                      |
| Mensaje incorrecto                       | ¡La próxima vez lo harás mejor!                                                                                                                                      |

Tabla 2. Datos de la pregunta para la primera versión del ejercicio

Con esta información completada, pulsamos sobre el botón «Guardar», con lo que retornamos a la pantalla de preguntas, en las que ya figura una pregunta en la lista, que se identifica gracias al enunciado de la misma, la variable de respuesta y el peso asignado. Podemos retroceder y volver a la pantalla de edición principal con la opción «Editar problema» que aparece en el menú de la izquierda.

El aspecto del problema, tras haber actuado sobre el botón «Prueba del problema», podría ser como el de la Figura 22.

La página del tutorial contiene además algunas ideas más generales, incluidas en la siguiente caja. En primer lugar se menciona el reto de la precisión en la respuesta. Puesto que normalmente las soluciones son números reales, es necesario dar como admisible una respuesta que

esté lo suficientemente cerca del valor exacto que conocemos (cuando sea el caso). En realidad, el criterio a utilizar depende en buena medida del valor que esperamos; quizá no tanto del valor numérico como del tipo. No es lo mismo una probabilidad que una medida cuyo orden de magnitud pueda ser muy variable, y tampoco es lo mismo una probabilidad cualquiera que una probabilidad que sabemos que generalmente va a ser pequeña.

The screenshot shows a web application interface for a statistics problem. On the left, there is a sidebar menu with options like 'Calculadora', 'Ejecuciones', 'Normal', 't-Student', 'Chi-cuadrado', 'F de Fisher', 'Usuario', 'Administrador', and 'Profesores'. The main content area is titled 'Ejecuciones' and displays a time limit of 'Tiempo limite: 44 min 45 sec'. Below this, the problem text reads: 'En este caso vamos a poner aquí la historia que haga referencia a un estudio en el que se observan dos variables, X e Y, de las que se sabe:'. A list of statistics is provided: 'la media de X: 20.46537', 'la media de Y: -28.8853', 'la desviación de X: 8.13213', 'la desviación de Y: 6.64804', and 'la covariancia de X e Y: 53.4786'. The problem question is: '1. Si queremos predecir la Y con la X, utilizando un modelo lineal, ¿cuánto vale el término independiente de la recta estimada (usando el método de mínimos cuadrados)?'. A link is provided: 'Aquí va un enlace en donde puedes leer más sobre el método: [http://es.wikipedia.org/wiki/Regresi%C3%B3n\\_lineal#Eqnref\\_12](http://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal#Eqnref_12)'. There is a text input field for the answer, and buttons for 'Corrección' and 'Editar problema'.

Figura 22. Aspecto de la primera versión del problema

Seguidamente figuran una serie de propuestas, que no son otra cosa más que plantear el campo de acción natural que se abre a continuación, después de crear nuestro primer y simple prototipo. La segunda versión va a abordar algunas de estas cuestiones.

## Comentario

*El criterio del error relativo para corregir una respuesta es muy sensible. ¿Qué pasaría si en un caso la solución estuviera muy cerca de 0? ¿Sería lógico exigirle al menos dos cifras significativas? Por otra parte, sin conocer el valor exacto, el alumno no tiene fácil saber cómo ha de redondear. Tampoco es fácil para el profesor explicar qué criterio se va a utilizar. Otros criterios son más simples, como el error absoluto, pero en cambio puede ser complicado para el profesor definir el umbral de corrección: se podría exigir demasiado o, en el otro extremo, se podrían aceptar como válidas respuestas con errores garrafales.*

## Propuestas

- Probar con varios criterios de corrección
- Probar con diferente número de intentos
- Crear un mensaje orientador (esto no es el término independiente, parece que lo has confundido con la pendiente)
- Añadir en el enunciado un gráfico de la nube de puntos
- Añadir los datos



Un enlace en la parte inferior permite acceder a la página que contiene la segunda versión del ejercicio.

Lo que se plantea es crear un problema nuevo, en vez de añadir algunos cambios a la primera versión, lo cual permitirá al alumno conservar la secuencia de versiones. Para ello recomendamos hacer las siguientes operaciones:

- En la página principal del primer problema, generar una copia que podremos guardar localmente en un fichero XML: «Generar XML».
- Desde el editor, marcar la carpeta de trabajo y elegir la opción «Crear problema».
- Antes de escribir nada sobre los campos de este nuevo problema, pulsar sobre «Restaurar XML», elegir el fichero que se ha generado previamente y aceptar: en este momento tenemos una copia idéntica del primer problema, que procederemos a modificar según las indicaciones que figuran en la Tabla 3.

El texto destacado en color amarillo indica novedades. Por ejemplo, el enunciado es ligeramente diferente. Se le ha dado formato con *tags* de html para crear una tabla, en la que en la parte izquierda figurará el texto propiamente dicho y a la derecha aparecerá un gráfico. Para poder editar explícitamente *tags* de html (o, como en este caso, para poder copiar directamente dicho código), hay que seleccionar en el editor que aparece en la página principal del problema el botón marcado como «Fuente HTML».

| Un tema de regresión lineal (segunda versión) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Descripción                                   | Estudio con dos variables numéricas, modelo lineal.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Enunciado                                     | <p>Incluimos el gráfico:</p> <pre> &lt;table align="left" border="0" cellpadding="1" cellspacing="4"&gt;   &lt;tbody&gt;     &lt;tr&gt;       &lt;td style="width:400px"&gt; En este caso vamos a poner aquí la historia que       haga referencia a un       estudio en el que se observan dos variables, X e Y, de las que se sabe:       &lt;ul&gt;         &lt;li&gt;la media de X: <math>M_x</math>&lt;/li&gt;         &lt;li&gt;la media de Y: <math>M_y</math>&lt;/li&gt;         &lt;li&gt;la desviación de X: <math>S_x</math>&lt;/li&gt;         &lt;li&gt;la desviación de Y: <math>S_y</math>&lt;/li&gt;         &lt;li&gt;la covariancia de X e Y: <math>V_{xy}</math>&lt;/li&gt;       &lt;/ul&gt;       &lt;/td&gt;     &lt;td&gt;\$plos\$&lt;/td&gt;   &lt;/tr&gt; &lt;/tbody&gt; &lt;/table&gt; </pre> |

|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Modelo R (Expresiones)</b>     | <pre> n = sample(40:100, 1) R = runif(1,0.6,1) sgn = sample(c(-1,1), 1) XB = rnorm(3, 0, 50) x = round(seq(XB[1], XB[2], len=n), 3) vx = var(x) v.i = exp(rnorm(1, 0, 2)) v.e = v.i/(1/R-1) vy = v.i+v.e sig = sqrt(v.i) err = rnorm(n)*sig y = round(XB[3] + sgn*R*sqrt(vy/vx) * x + err, 3) Mx = round(sum(x)/n,5) My = round(sum(y)/n,5) Vxy = round(cov(x,y),5) Sx = round(sd(x),5) Sy = round(sd(y),5) LM = lm(y ~x) summ = summary(LM) tind = summ\$coef[1,1] #Atención: plos = ini_imagen(360, 360) par(mar=c(3,3,1,1)) plot(x, y, pch=19) abline(LM, col='grey') fin_imagen() pend = summ\$coef[2,1] </pre> |
| <b>Expresiones públicas</b>       | Para compartir el modelo con otros autores.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>Variables extras a mostrar</b> | x, y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

Tabla 3. Datos principales para la segunda versión del ejercicio

Nuevamente se pulsa sobre el botón «Guardar», y seguidamente se elige la opción «Editar preguntas». Nótese que como se trata de una copia del primer problema ya existe una pregunta, que procedemos a editar y a modificar como se indica en la Tabla 4.

| Preguntas                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Enunciado 1</b>                              | Si queremos predecir la Y con la X, utilizando un modelo lineal, ¿cuánto vale el término independiente de la recta estimada (usando el método de mínimos cuadrados)?                                                                                                                                                                                                                                                                                       |
| <b>Solución a la pregunta</b>                   | Tind                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Parámetro de sintaxis</b>                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Sintaxis de la respuesta</b>                 | Real                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Plantilla de expresiones de la respuesta</b> | Número decimales correctos                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Parámetro de expresiones</b>                 | 2                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Expresiones de respuesta</b>                 | <pre> Adaptamos el código base que aparece: #si entran parametro_ = 4 , el valor absoluto deberá ser igual o menor de 0.00005 (4 decimales correctos) if (estatus.decimales_correctos(solucion_ , respuesta_ , parametro_)) {   resultado_ &lt;- 1 }else{   resultado_ &lt;- 0   if (estatus.decimales_correctos(pend , respuesta_ , parametro_))   {     mensaje_ = 'Parece que estás confundiendo "término independiente" con "pendiente".'   } } </pre> |
| <b>Peso</b>                                     | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Ayuda contextual</b>                         | Aquí va un enlace en donde puedes leer más sobre el método:                                                                                                                                                                                                                                                                                                                                                                                                |

|                      |                                                                                                                                             |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Link                 | < <a href="http://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal#Eqnref_12">http://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal#Eqnref_12</a> > |
| Intentos             | 2                                                                                                                                           |
| Mostrar solución     | (normalmente es lo más sensato)                                                                                                             |
| Mensaje correcto     | ¡Bravo!                                                                                                                                     |
| Mensaje semicorrecto |                                                                                                                                             |
| Mensaje incorrecto   | ¡La próxima vez lo harás mejor!                                                                                                             |

Tabla 4. Datos de la pregunta para la segunda versión del ejercicio

### Comentario

*En este caso se incluye un control (rudimentario) para verificar que el valor proporcionado por el estudiante no sea el coeficiente de la pendiente, un error común. En caso afirmativo, aparecerá un texto destacado. También se podría crear al vuelo el texto que aparece en las cajas «correcto», «semicorrecto» o «incorrecto».*

### Propuestas

- *Introducir otras preguntas relacionadas con el tema, añadiendo las expresiones en R necesarias para generar la solución.*

Como se puede apreciar si se realiza la prueba del problema, se incorpora el gráfico al enunciado, en el que se representa no solo la nube de puntos sino también la recta de regresión (con lo que el estudiante dispone de una valiosa ayuda). También hemos incluido los datos, a los que se puede acceder mediante el enlace «Copiar datos para pegar en otro programa»: este enlace conduce a una pestaña del navegador que posibilita la copia de datos, con facilidades para determinar el símbolo decimal o el carácter que separa las columnas.

Sin embargo, el cambio más importante debe verse en las posibilidades que ofrece el código que acompaña la evaluación de la pregunta. Puesto que el autor del problema puede controlar cualquier aspecto que se pueda implementar en código R, las posibilidades para evaluar flexiblemente y para dar *feedback* adecuado aumentan considerablemente.

La tercera versión del problema amplía el número de preguntas. Con la que se añade se pretende comprobar si el estudiante entiende el concepto de residuos en el modelo lineal y si lo sabe aplicar en casos concretos. Con este fin, se muestran tres gráficos estándar de residuos, aunque solo uno de ellos es el correcto. El reto para el autor de problemas consiste en saber generar los gráficos falsos, que resulten creíbles pero no tanto como para que solamente un ojo experto pueda discernir el auténtico de los demás. Este no es más que un caso particular de pregunta de opción única que, como todo maestro sabe, tiene que contener siempre diversas opciones realistas (al menos una que pueda competir con la solución real), únicamente que con e-status deben crearse mediante un modelo probabilístico lo bastante robusto como para que el riesgo de que aparezca una pregunta demasiado fácil o demasiado difícil sea mínimo.

La tercera versión puede crearse a partir de una copia local generada en XML de la segunda, tal como se procedió la vez anterior; de esta manera conservamos todas las versiones.

| Un tema de regresión lineal (tercera versión) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Descripción                                   | Estudio con dos variables numéricas, modelo lineal.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Enunciado                                     | <pre> &lt;table align="left" border="0" cellpadding="1" cellspacing="4"&gt;   &lt;tbody&gt;     &lt;tr&gt;       &lt;td style="width:400px"&gt; En este caso vamos a poner aquí la historia que       &lt;/td&gt;     &lt;/tr&gt;   &lt;/tbody&gt; &lt;/table&gt; </pre> <p>haga referencia a un estudio en el que se observan dos variables, X e Y, de las que se sabe:</p> <ul style="list-style-type: none"> <li>la media de X: <math>Mx</math></li> <li>la media de Y: <math>My</math></li> <li>la desviación de X: <math>Sx</math></li> <li>la desviación de Y: <math>Sy</math></li> <li>la covariancia de X e Y: <math>Vxy</math></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                           |
| Modelo R (Expresiones)                        | <pre> n = sample(40:100, 1) R = runif(1,0.6,1) sgn = sample(c(-1,1), 1) XB = rnorm(3, 0, 50) x = round(seq(XB[1], XB[2], len=n), 3) vx = var(x) v.i = exp(rnorm(1, 0, 2)) v.e = v.i/(1/R-1) vy = v.i+v.e sig = sqrt(v.i) err = rnorm(n)*sig y = round(XB[3] + sgn*R*sqrt(vy/vx) * x + err, 3) Mx = round(sum(x)/n,5) My = round(sum(y)/n,5) Vxy = round(cov(x,y),5) Sx = round(sd(x),5) Sy = round(sd(y),5) LM = lm(y ~x) summ = summary(LM) tind = summ\$coef[1,1] plos = ini_imagen(360, 360) par(mar=c(3,3,1,1)) plot(x, y, pch=19) abline(LM, col='grey') fin_imagen() pend = summ\$coef[2,1] #Atención: u = sample(1:3, 1) d=0.66 u.2 = c(d^2, d, 1, d^(-1), d^(-2))[u:(u+2)] Rs = array(NA, dim=c(n, 3)) for (i in 1:3) {   j = u.2[i]   if (j==1) { Rs[,i] = summ\$resid }   else {Rs[,i] = sample(qnorm(seq(0.5/n, 1, by=1/n)))* summ\$sigma*j} } ylym=range(Rs) fig = ini_imagen(640,250) opar=par() par(mfrow=c(1,3),mar=c(3.5,2.5,1,0.3)) for (i in 1:3) {   plot(x,Rs[,i], main=i, ylim=ylym, pch=4) } par(opar) fin_imagen() corre = 4-u </pre> |
| Expresiones públicas                          | Para compartir el modelo con otros autores.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Variables extras a mostrar                    | x, y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

Tabla 5. Datos principales para la tercera versión del ejercicio

De nuevo guardamos y procedemos a «Editar preguntas». En esta ocasión volvemos a crear una nueva pregunta, completándola con los contenidos de la Tabla 6. La Figura 23 muestra el aspecto final de la pregunta.

| Preguntas                                |                                                                                                                                                                                    |
|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Enunciado 2                              | Una de estas tres figuras corresponde al diagrama de los residuos para los datos mostrados ajustados con el modelo lineal. Diga con un número de 1 a 3 cuál es la imagen correcta. |
| Nueva pregunta                           | \$fig\$                                                                                                                                                                            |
| Solución a la pregunta                   | Corre                                                                                                                                                                              |
| Parámetro de sintaxis                    |                                                                                                                                                                                    |
| Sintaxis de la respuesta                 | Natural                                                                                                                                                                            |
| Plantilla de expresiones de la respuesta | Exacto                                                                                                                                                                             |
| Parámetro de expresiones                 |                                                                                                                                                                                    |
| Expresiones de respuesta                 | <pre>if (solucion_ == respuesta_) {   resultado_ &lt;- 1 }else{   resultado_ &lt;- 0 }</pre>                                                                                       |
| Peso                                     | 1                                                                                                                                                                                  |
| Ayuda contextual                         |                                                                                                                                                                                    |
| Link                                     |                                                                                                                                                                                    |
| Intentos                                 | 1                                                                                                                                                                                  |
| Mostrar solución                         |                                                                                                                                                                                    |
| Mensaje correcto                         | Excelente                                                                                                                                                                          |
| Mensaje semicorrecto                     |                                                                                                                                                                                    |
| Mensaje incorrecto                       |                                                                                                                                                                                    |

Tabla 6. Datos de la pregunta para la tercera versión del ejercicio

### Comentario

¿Qué hemos hecho? El objetivo consiste en identificar la estructura de los residuos. Se han generado tres gráficos, uno de los cuales es real y los otros no.

*La gracia de la pregunta consiste en saber hacer de trilero: colocar el gráfico correcto en una posición impredecible, y que tampoco ayude la magnitud relativa de la desviación residual.*

Aunque frecuentemente las preguntas que se construyan esperan una respuesta numérica abierta, en e-status es posible preparar preguntas de opción múltiple. Esto será normalmente la única posibilidad si se desea que el estudiante valore la corrección de distintas afirmaciones, para detectar cuál o cuáles son válidas, dado que la técnica para interpretar respuestas textuales abiertas es muy compleja. Por supuesto, las opciones de la pregunta pueden ser también gráficos, como en este caso, expresiones matemáticas o simplemente números.

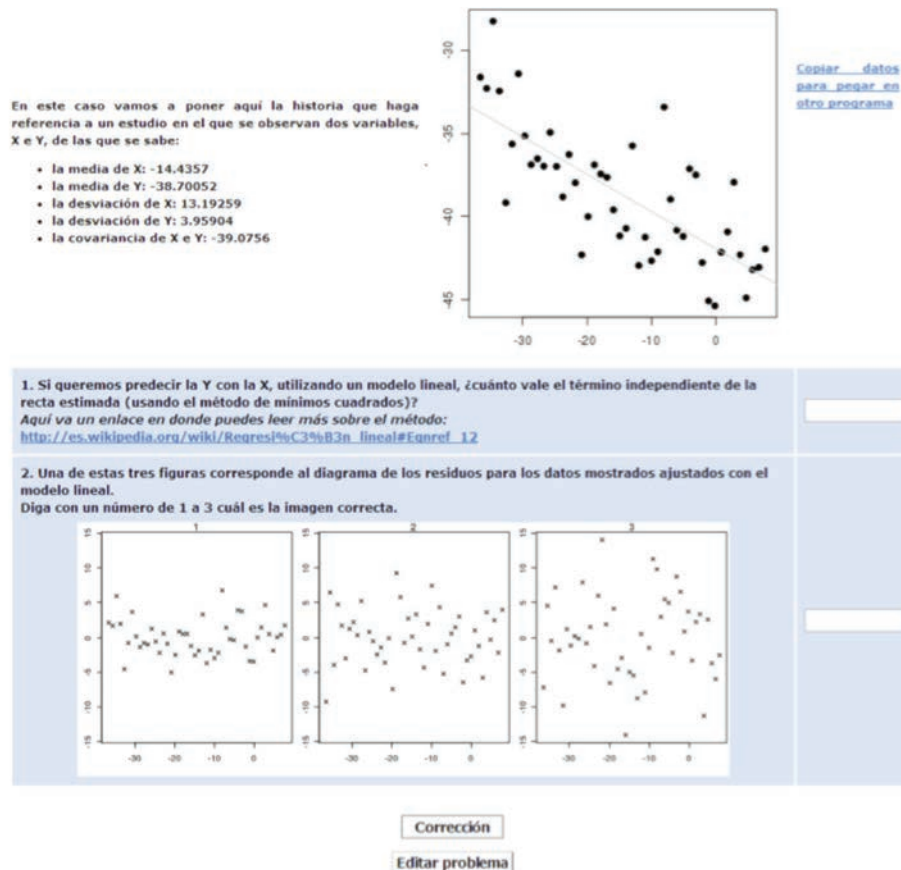


Figura 23. Aspecto de la tercera versión del problema

No hay una forma sencilla de programar preguntas de opción múltiple todavía. Generalmente, el profesor ante la pregunta que quiere plantear elige —además de la correcta— varias posibilidades que le parezcan típicos errores. Después aplica una permutación a las opciones y registra cuál de ellas es la solución. Veamos un ejemplo.

Queremos preguntar: «¿Cuál es el valor que empleamos para un intervalo de confianza del 90% para la media, con una muestra de tamaño  $n$ , y  $\sigma$  desconocida?». Tomamos 4 opciones:

**Op = c(qt(0.95, n-1), qnorm(0.95), qt(0.95, n), qt(0.975, n-1))**

Desordenamos las opciones y anotamos la correcta:

**u = sample(1:4)**

**Ok = which(u==1) # ¿dónde está el 1? Es la opción correcta.**

Preparamos una lista numerada con las opciones, que se mostrará en el enunciado a continuación de la pregunta anterior (como símbolo: \$lista\$):

```
lista = "<ol>"
for (i in 1:4) {
  lista = paste(lista, "<li>", round(Op[u[i]], 4), "</li>")
}
lista = paste(lista, "</ol>")
```

Para esta pregunta la respuesta correcta corresponde a la variable «Ok», y se utiliza la misma plantilla «Exacto», porque se trata de un valor entero y no caben problemas de precisión.

La cuarta versión del problema aborda una técnica que puede ser útil en algunas ocasiones. Consiste en interconectar las diversas preguntas del problema, lo cual permite que el profesor incluya cuestiones cuyo fin es pedir al estudiante que introduzca valores para ser incorporados al problema. Es como cuando un mago invita a un espectador a subir al escenario para participar en primera persona del juego (salvando las distancias).

Antes de continuar, digamos que el código de las preguntas puede utilizar ciertas variables propias de e-status que le permiten acceder a cierto tipo de información, o cambiar el estado del problema. Algunas ya las conocemos:

- `solucion_`: la solución correcta, si está definida.
- `respuesta_`: lo que el alumno ha introducido.
- `resultado_`: la nota de la pregunta.
- `mensaje_`: un texto que aparece tras hacer la corrección de la pregunta.
- `respuestas_`: la lista de todas las respuestas del alumno.
- `resultados_`: las notas para todas las preguntas.
- `chances_`: el número de intentos disponibles.

Además, las variables que se definan tanto en el código principal como en el de las preguntas previas son visibles en el ámbito de cada pregunta (por tanto, hay que procurar no repetir nombres para las variables que deban ser visibles en otros ámbitos). Todo ello permite elaborar este ejercicio: consiste en pedir al alumno que defina una supuesta nueva observación, entrando el valor de X y el de Y. Se le pedirá que este nuevo punto, a unir a un subconjunto determinado de los originales, no sea influyente (ni respecto las X ni respecto las Y). A continuación, se le muestra el gráfico con los puntos seleccionados y el punto del alumno (marcado en color rojo) y se le pide un cálculo para este conjunto, concretamente, un intervalo de confianza para la pendiente.

| Un tema de regresión lineal (cuarta versión) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Descripción                                  | Estudio con dos variables numéricas, modelo lineal.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Enunciado                                    | <pre> &lt;table align="left" border="0" cellpadding="1" cellspacing="4"&gt;   &lt;tbody&gt;     &lt;tr&gt;       &lt;td style="width:400px"&gt; En este caso vamos a poner aquí la historia que haga referencia a un       estudio en el que se observan dos variables, X e Y, de las que se sabe:       &lt;ul&gt;         &lt;li&gt;la media de X: \$Mx\$&lt;/li&gt;         &lt;li&gt;la media de Y: \$My\$&lt;/li&gt;         &lt;li&gt;la desviación de X: \$\$Sx\$&lt;/li&gt;         &lt;li&gt;la desviación de Y: \$\$Sy\$&lt;/li&gt;         &lt;li&gt;la covariancia de X e Y: \$Vxy\$&lt;/li&gt;       &lt;/ul&gt;       &lt;/td&gt;     &lt;td&gt;\$p\$&lt;/td&gt;   &lt;/tr&gt; &lt;/tbody&gt; &lt;/table&gt; &lt;p&gt;&lt;/p&gt; </pre> |

|                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Modelo R</b><br/><b>(Expresiones)</b></p> | <pre> n = sample(40:100, 1) R = runif(1,0.6,1) sgn = sample(c(-1,1), 1) XB = rnorm(3, 0, 50) x = round(seq(XB[1], XB[2], len=n), 3) vx = var(x) v.i = exp(rnorm(1, 0, 2)) v.e = v.i/(1/R-1) vy = v.i+v.e sig = sqrt(v.i) err = rnorm(n)*sig y = round(XB[3] + sgn*R*sqrt(vy/vx) * x + err, 3) Mx = round(sum(x)/n,5) My = round(sum(y)/n,5) Vxy = round(cov(x,y),5) Sx = round(sd(x),5) Sy = round(sd(y),5) LM = lm(y ~x) summ = summary(LM) tind = summ\$coef[1,1] plos = ini_imagen(360, 360) par(mar=c(3,3,1,1)) plot(x, y, pch=19) abline(LM, col='grey') fin_imagen() pend = summ\$coef[2,1] u = sample(1:3, 1) d=0.66 u.2 = c(d^2, d, 1, d^(-1), d^(-2))[u:(u+2)] Rs = array(NA, dim=c(n, 3)) for (i in 1:3) {   j = u.2[i]   if (j==1) { Rs[,i] = summ\$resid }   else {Rs[,i] = sample(qnorm(seq(0.5/n, 1, by=1/n))* summ\$sigma*)} } ylym=range(Rs) fig = ini_imagen(640,250) opar=par() par(mfrow=c(1,3),mar=c(3.5,2.5,1,0.3)) for (i in 1:3) {   plot(x,Rs[,i], main=i, ylim=ylym, pch=4) } par(opar) fin_imagen() corre = 4-u #Atención:lista = seq(from=1, len=9, by=round(n/9)) </pre> |
| <p><b>Expresiones públicas</b></p>              | <p>Para compartir el modelo con otros autores.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <p><b>VARIABLES EXTRAS A MOSTRAR</b></p>        | <p>x, y</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

Tabla 7. Datos principales para la cuarta versión del ejercicio



La Tabla 7 no es muy diferente de la tabla correspondiente a la tercera versión: solo define el subconjunto de los puntos que serán utilizados para la cuarta pregunta. La Tabla 8 contiene la innovación real de este ejercicio (atención: hay que crear dos preguntas).

| Preguntas                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Enunciado 3<br>Nueva pregunta            | Escoja los pares de observaciones de los datos proporcionados correspondientes a las posiciones: \$lista\$. Añada un punto más (le pedimos que su punto siga más o menos la tendencia general), introduciendo primero la coordenada X y luego la Y, separadas por un blanco.                                                                                                                                                                                                                                                                                                              |
| Solución a la pregunta                   | (no relevante)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Parámetro de sintaxis                    | 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Sintaxis de la respuesta                 | Vector                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Plantilla de expresiones de la respuesta | (no relevante)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Parámetro de expresiones                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Expresiones de respuesta                 | <pre> Escriba todo esto: nvo = respuesta_ x2 = c(x[lista], nvo[1]) y2 = c(y[lista], nvo[2]) lm2 = lm(y2 ~ x2) su2 = summary(lm2) sres = lm2\$resid[10]/su2\$sigma mensaje_="" resultado_ = 1 if (abs((nvo[1]-Mx)/Sx) &gt; 3) {   mensaje_ = "La abcisa está muy alejada: este punto sería muy influyente."   resultado_ = 0 } if (abs(sres) &gt; 2) {   mensaje_ = paste(mensaje_, "Parece que tu punto se aleja bastante del resto.")   resultado_ = 0 } else {   plos2 = ini_imagen(300, 300)   par(mar=c(3,3,1,1))   plot(x2, y2, pch=19, col=c(rep(1, 9), 2))   fin_imagen() } </pre> |
| Peso                                     | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Ayuda contextual                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Link                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Intentos                                 | 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Mostrar solución                         | ¡No!                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Mensaje correcto                         | Este es tu nuevo gráfico: <br> \$plos2\$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Mensaje semicorrecto                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Mensaje incorrecto                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Enunciado 4<br>Nueva pregunta            | Con los datos de la pregunta anterior (incluso si la respuesta no fue aceptable), calcule un intervalo de confianza al 95% para la pendiente de la recta $Y = \beta_0 + \beta_1 X$ (separe los dos valores por un espacio, y proporcione por lo menos tres decimales correctos).                                                                                                                                                                                                                                                                                                          |
| Solución a la pregunta                   | (no relevante)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Parámetro de sintaxis                    | 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Sintaxis de la respuesta                 | Vector                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Plantilla de expresiones de la respuesta | (no relevante)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

|                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Parámetro de expresiones | 3 (decimales correctos)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Expresiones de respuesta | <pre> Escriba todo esto: if (!exists("x2")) {   mensaje_ = "Debe responder antes la pregunta previa." } else {   tam = length(x2)   solu = su2\$coef[2,1] + c(-1,1)*qt(0.975, tam-2)*su2\$coef[2,2]   solu = round(solu, 6)   if (respuesta_[1] &gt; respuesta_[2]) respuesta_ = respuesta_[2:1]   resultado_ = 0   len_res = length(respuesta_)   for (i in 1:len_res) {     if (estatus.decimales_correctos(solu[i], respuesta_[i], parametro_)) resultado_ = resultado_ + 1/   len_res   }   if (chances_ &gt; 0) {     msg = "¡Animo!"   } else {     msg = paste("La respuesta correcta es (" , solu[1], " , " , solu[2], ")", sep="")   } } </pre> |
| Peso                     | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Ayuda contextual         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Link                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Intentos                 | 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Mostrar solución         | ¡No!                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Mensaje correcto         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Mensaje semicorrecto     | ¡Qué rabia! Has fallado uno de los dos extremos. \$msg\$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Mensaje incorrecto       | \$msg\$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

Tabla 8. Datos de la pregunta para la cuarta versión del ejercicio

## Comentario

¿Qué hemos hecho? Las preguntas en sí puede que no sean lo más interesante: estaríamos comprobando si el alumno puede imaginar un punto que no sea flagrantemente contrario a la tendencia de la nube representada. También le pedimos un cálculo para estimar por IC la pendiente. Requerirá habilidades de manejo de una herramienta estadística, para seleccionar los valores especificados, realizar el ajuste y obtener el IC.

*Una lectura diferente es ilustrar algunas posibilidades que ofrece la herramienta. Un profesor puede inventarse preguntas que no posean solución única (basta con que haya un procedimiento para valorar la corrección de la respuesta, que se pueda implementar con R).*

*Además, vemos que las preguntas tienen conexión entre ellas (comparten el código, y es posible saber qué se contestó a cierta pregunta y el resultado de dicha evaluación).*

*La consecuencia obvia del aumento de complejidad en la evaluación de las respuestas es: cuidado, aumenta el número de situaciones posibles que hay que prever, y se incrementa el riesgo de cometer errores.*

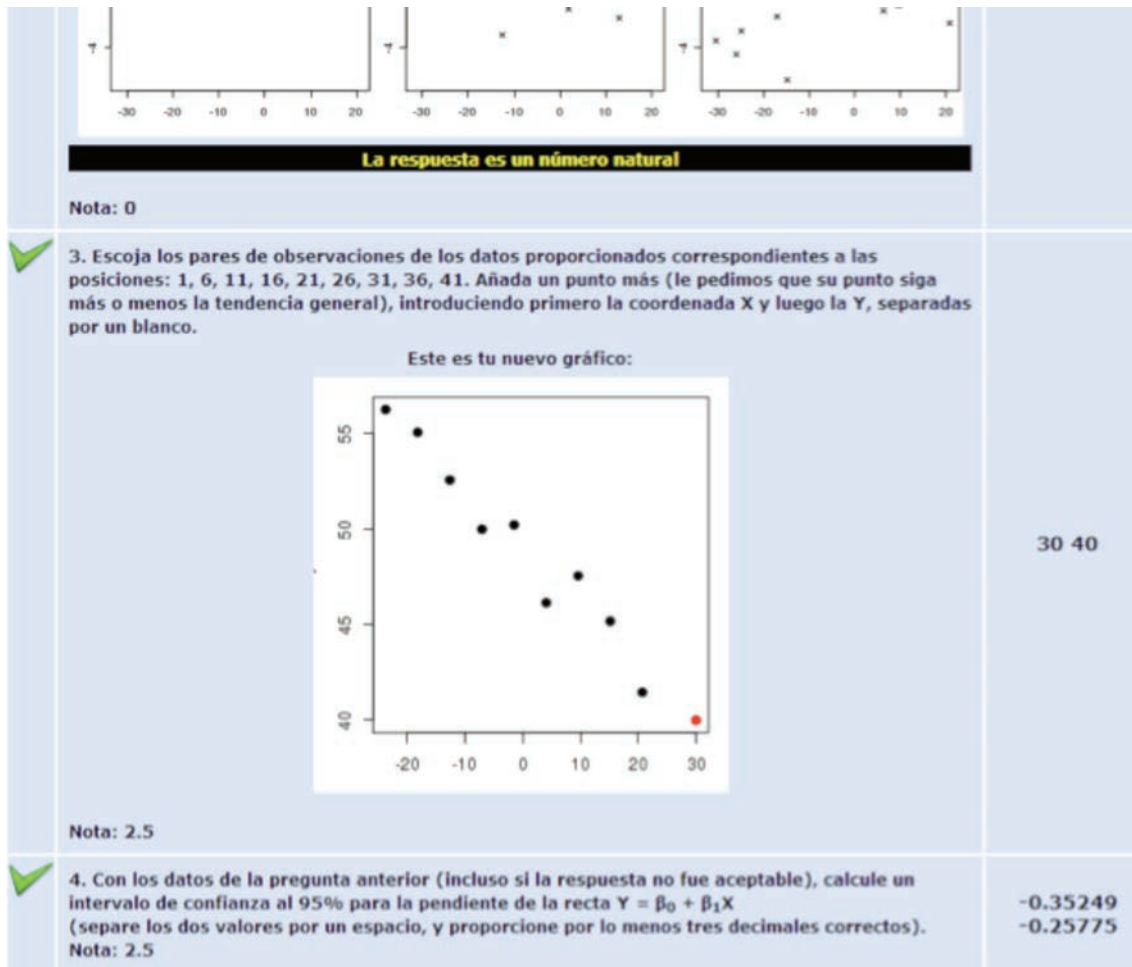


Figura 24. Aspecto de la cuarta versión del problema

La Figura 24 ilustra cómo podría ser un resultado de la solución de estas últimas dos preguntas. El punto introducido (30, 40) es una opción muy razonable —se observa en el gráfico que se muestra tras responder que dicho punto está perfectamente alineado con la tendencia de los puntos seleccionados—. El intervalo consecuente también debe encajar bien con la pendiente de la nube original, si el alumno se ha preocupado por comparar. Por otro lado, si se hubiera introducido un punto *discordante* (hasta lo permitido), el estudiante habría podido comprobar con su propio caso como el nuevo punto influye y altera la magnitud de la pendiente, lo cual también sería un buen ejemplo de *descubrimiento* capaz de asentar mejor el conocimiento.

Con este procedimiento podemos incorporar al problema preguntas que van a exigir al estudiante un esfuerzo diferente del requerido para calcular respuestas únicas, que algunas veces resultan ser mecánicas y previsibles, en cierta manera. También es cierto que se requiere un control más exigente, si se vinculan unas preguntas con otras, más aún si una pregunta depende de una respuesta introducida por el estudiante. Sencillamente, las cosas pueden ir mal, y además es muy conveniente proporcionar explicaciones al estudiante ante respuestas incoherentes.

Por ejemplo: podríamos pedir al estudiante que complete una determinada distribución de probabilidad de una variable discreta. Después le pediremos que halle la variancia de dicha variable. Si las probabilidades resultantes no son positivas o no suman 1 deberíamos notificar el error. Pero si los intentos disponibles para contestar la primera pregunta se agotan antes de responder correctamente, la segunda no se puede contestar, y ambas se puntuarán como incorrectas (aunque el alumno supiera cómo responder a la segunda pregunta).

Puede parecer innecesario decirlo, pero cualquier problema debe ser intensamente validado antes de ponerlo a disposición del alumno; más aún si es elaborado. El coste de rectificar un problema cuando ya ha sido probado por los alumnos es cuantioso, puesto que no es fácil ni recomendable eliminar las resoluciones defectuosas de la base de datos.

## Discusión

El uso de las tecnologías de la información en la docencia de estadística y ramas afines es intenso pero aún no decididamente claro. No es de extrañar, puesto que el sistema actual, que concentra un gran volumen de contenidos en pocas semanas de clase, implica que el profesor sigue tomando el protagonismo porque entiende que es la única manera de cumplir con el compromiso de la guía docente correspondiente. En cambio, las TIC son herramientas para el alumno, a utilizar con el ritmo particular que cada uno sabe que le va bien. Si es que el alumno percibe que la herramienta le aporta algo: no basta con poner los utensilios a su alcance.

El profesor tiene que seleccionar los medios más apropiados para complementar sus clases, y tiene que comprobar que el alumno asocia unos temas con otros. Esta fase de seguimiento es crucial, no basta con saber que en el examen todo saldrá a la luz. No confundamos la evaluación con la calificación. La diferencia está en que tras la evaluación el profesor y los alumnos tienen aún una oportunidad de reconducir el aprendizaje.

Como es normal, los objetivos que persigue el estudiante son variados, y simplificando bastante podríamos decir que son una combinación entre aprobar la asignatura y aprender algo que le pueda ser de provecho en el futuro. Cuando la estadística es una materia relativamente marginal dentro de un plan de estudios, el balance entre estos dos objetivos suele inclinarse llamativamente hacia el primero de ellos. Posiblemente, el profesor se verá obligado a establecer una parte de la calificación en base de los ejercicios resueltos con los recursos TIC empleados, aunque sepa que estos resultados tienen una validez discutible. De lo contrario, hay pocas garantías de que los estudiantes dediquen parte de su tiempo a un esfuerzo que no les reporta un beneficio inmediato (aprender no se percibe así).

El taller que se explica en esta comunicación puede que insista mucho en «lo que se puede hacer», pero indudablemente tiene mucha importancia «lo que se debe hacer», y seguramente más aún. Las posibilidades técnicas de una herramienta al fin y al cabo son medios para llegar a un objetivo, pero lo que guía a un docente cuando diseña la colección de problemas que complementa al resto de materiales es la intención de que el alumno se excite descubriendo conocimientos que quizá él mismo no sospechaba. Dado que el *partenaire* del alumno no es el profesor sino una máquina que responde a un algoritmo, esta no le va a ofrecer una ayuda muy sofisticada o, al menos, será tan rica como la imaginación del autor de los problemas y su habilidad con R le permitan. El buen problema ha de obligar a pensar, al menos al principio (cuando se gasta y pierde la magia, se vuelve mecánico y aburrido), y en esa actividad el alumno interpreta la situación, analiza diversas posibilidades, escoge y descarta procedimientos y, en suma, interviene activamente tratando de utilizar sus recursos para construir nuevo conocimiento.

## Referencias

- [1] González, J. A.; Muñoz, P. (2006): «e-status: an Automatic Web-Based Problem generator - Applications to Statistics», *Computer Applications In Engineering Education*, V14(2), 151-159.
- [2] González, J. A.; Jover, L.; Cobo, E.; Muñoz, P. (2010): «A Web-Based Learning Tool Improves Student Performance In Statistics: A Randomized Masked Trial», *Computers & Education*, V55(2), 704-713.

## **Big data para estadísticos**

*Gil Bellosta, Carlos J.*

cgb@datanalytics.com

### **Resumen**

El *big data* plantea muchas cuestiones de interés estadístico. También es un campo potencial para el desarrollo profesional de los estadísticos. El presente trabajo describe un taller práctico de introducción al *big data* usando Hadoop en el que se plantean y resuelven varios problemas de interés estadístico. Su objetivo último es el de suscitar un debate sobre la pertinencia de ampliar el temario de los estudios de estadística para que incluyan el aprendizaje de técnicas y herramientas para el análisis de grandes bases de datos.

**Palabras clave:** Hadoop, *big data*, *regression*, *logistic regression*, R.

**Clasificación AMS:** 62P30, 68P15.

## Introducción

Este trabajo resume un taller práctico impartido en el seno de las V Jornadas de la GENAEIO. Está dividido en dos partes diferenciadas precedidas de una pequeña introducción que las contextualiza y adecúa al perfil e intereses de los asistentes a las jornadas.

La primera parte de la charla es una introducción al uso del universo del *big data* que hace especial hincapié en Hadoop y termina con un taller en el que resuelve el problema que se ha convertido en el «Hola, mundo» del *big data*: contar registros. La segunda parte replantea esa introducción desde el punto de vista de la estadística y los estadísticos. Termina con otro taller en el que se implementa sobre Hadoop la solución a un problema de regresión lineal y se propone una opción para atacar el problema de la regresión logística.

El objetivo de la charla y los correspondientes talleres es triple. Por un lado, quiere servir de introducción a los asistentes a un conjunto de nuevas tecnologías para el procesamiento y análisis de datos. Por otro, quiere poner de manifiesto la escasa presencia de los estadísticos en el mundo del *big data* en sus diversas dimensiones: académica, laboral, etc. Y, finalmente, invitar a los asistentes, y muy en particular, habida cuenta de su interés en la enseñanza de la estadística y su responsabilidad en la formación de futuras promociones de jóvenes estadísticos, a considerar una renovación del presente currículo estadístico que incorpore elementos que los ayuden en su futuro tanto dentro de la academia como en una carrera profesional fuera de ella.

El presente trabajo sigue la misma estructura que la charla y, por lo tanto, está dividido en cuatro secciones correspondientes a las arriba descritas: los dos talleres precedidos de sus correspondientes introducciones.

## Una introducción al big data

Los matemáticos desarrollan teorías abstractas. Los físicos las utilizan para explicar el comportamiento del mundo. Los ingenieros las convierten en dispositivos y máquinas que permiten transformarlo. Es precisamente así como nació el *big data*.

Lo que conocemos como *big data* nació en el seno de unas cuantas compañías relacionadas con el mundo de Internet. En el desarrollo de su actividad, comenzaron a acumular cantidades cada vez mayores de datos que excedían la capacidad de los sistemas que podían ofrecerles los proveedores habituales. Algunas de estas compañías, conscientes de su capacidad tecnológica, se aprestaron a solucionar el problema por sus propios medios. Desarrollaron nuevas herramientas. No se trató de una evolución de las preexistentes (p. e., las bases de datos tradicionales). Fue una verdadera revolución, comenzando prácticamente de cero. Plantearon soluciones ingenieriles a problemas ingenieriles: los de almacenar y procesar el tipo de datos propios de ese tipo de empresas, es decir, grandes volúmenes, a gran velocidad, de diversa naturaleza, etc.

Las nuevas herramientas, además (y a ello se debe parcialmente su éxito), se concibieron en gran parte dentro del espíritu del *software* libre. Corrían sobre ordenadores baratos, de consumo. Su código se liberó desde un primer momento. Fueron asimiladas y mejoradas, en muchas ocasiones, por empresas de la competencia. Se creó un ecosistema alrededor de ellas.

Estas herramientas y las posibilidades que ofrecían llamaron la atención. Por una parte, la del público en general, que se interesó por cuestiones como la privacidad de los datos y fantaseó con escenarios propios de la ciencia-ficción. También la de empresas y organizaciones de otros sectores distintos del tecnológico que se enfrentaban a problemas similares. Y la de muchos otros, incluida la academia.

La posibilidad de almacenar y procesar grandes datos abrió la posibilidad de analizarlos. Sin embargo, incluso la misma estadística descriptiva, tradicionalmente considerada una disciplina menor, ofrece cierta dificultad técnica en estos contextos. Y mucha más, por supuesto, las cuestiones de tipo inferencial. Por ejemplo, repasando las trazas de actividad de un cliente concreto, ¿es posible inferir algo acerca de sus gustos? ¿O completar su perfil sociodemográfico? Surgen así preguntas de las que la estadística había venido ocupándose tradicionalmente. Así se entiende la curiosidad de los estadísticos por estas nuevas tecnologías; y, a la vez, el renovado interés de sus usuarios por la estadística.

Tanto el presente trabajo como la charla que resume trata no tanto de saciar esa curiosidad sino de alimentarla aún más. Al mismo tiempo, proporcionar un punto de arranque para ulteriores exploraciones.

De todo el universo de herramientas mencionadas más arriba, se ha elegido una, Hadoop, como la más adecuada como introducción al mundo del *big data*. Es posiblemente la más popular de las existentes. De hecho, muchas herramientas de *big data* (p. e., Hive o HBase) son extensiones de Hadoop para resolver problemas concretos. En el caso de Hive, para proporcionar a Hadoop una interfaz SQL. Una primera taxonomía de herramientas de *big data*, de hecho, las clasificaría entre las que corren sobre Hadoop y el resto. Además, Hadoop ha sido la inspiración de otras herramientas posteriores: muchas de ellas fueron desarrolladas para resolver algunas de sus carencias más notables (p. e., Spark).

Hadoop es una herramienta compuesta de dos partes claramente diferenciadas:

- Hadoop como sistema de ficheros.
- Hadoop como plataforma de paralelización de algoritmos.

Estas dos partes están relacionadas con los dos problemas originales que encontraron las compañías pioneras en el *big data*: almacenar datos y procesarlos.

Un sistema de ficheros es una combinación de *hardware* y *software* que permite gestionar ficheros: guardarlos, leerlos, borrarlos, etc. Encuentra espacio sobre el dispositivo físico para nuevos ficheros, los fragmenta de ser necesario, recompone los fragmentos cuando recibe una petición de lectura, etc. Los discos duros, sus predecesores los disquetes, los *pendrives*, etc., son ejemplos de sistemas de ficheros bien conocidos.

Como sistema de ficheros, Hadoop tiene algunas características que lo diferencian de muchos de los más habituales. Estos últimos, típicamente, constan de un único dispositivo y están asociados a una única máquina. Hadoop, sin embargo, es un sistema de ficheros distribuido sobre un clúster de ordenadores. Esto es lo que le permite almacenar ficheros de mayor tamaño que el de cualquier dispositivo físico.

Cuando Hadoop recibe un fichero para su almacenamiento, realiza varias operaciones de entre las cuales, las más importantes son:

- Trocear el fichero en bloques. Nótese que Hadoop está concebido para almacenar ficheros muy voluminosos, que potencialmente exceden la capacidad de almacenamiento de un disco duro tradicional.
- Enviar esos bloques a determinados ordenadores del clúster para que los almacenen de acuerdo con ciertos criterios.
- Realizar copias redundantes: está configurado por defecto (aunque la opción es configurable) para guardar tres copias de cada bloque en máquinas distintas. El objetivo es doble. Por una

parte, para evitar que los datos se pierdan en caso de fallos de *hardware*; por otra, tener más opciones de balanceo de la carga de trabajo.

- Y, obviamente, guardar un registro de qué ordenadores disponen de qué bloques del fichero original para cuando sea necesario recuperarlos y recomponerlos.

Además de las anteriores, también es capaz de realizar otro tipo de tareas, tales como reequilibrar el volumen de datos almacenado por cada clúster para evitar sesgos que afecten al rendimiento, etc. La descripción de estos mecanismos, sin embargo y pese a su interés, excede el alcance del presente trabajo.

Dentro de Hadoop, los ficheros se organizan dentro de una jerarquía de directorios y ficheros (con sus respectivos propietarios y permisos de acceso) de una manera similar a la habitual en sistemas Linux/Unix. Además, como todo sistema de ficheros, Hadoop proporciona herramientas para escribir ficheros sobre él, descargarlos, borrarlos, etc. En particular, Hadoop pone a disposición de sus usuarios una *shell* con los comandos de manipulación de ficheros más habituales: *cp* para copiar, *mv* para mover y renombrar, etc.

El siguiente bloque de código ilustra el uso de una serie de comandos con los que se está operando sobre los ficheros contenidos en Hadoop. Los comandos son análogos a los que se usarían en Linux/Unix con la salvedad de que operan sobre otro sistema de ficheros distinto: primero se crea un nuevo directorio en Hadoop; usando *put* se sube a él un fichero y, finalmente, se le cambia el nombre.

```
hadoop fs -mkdir datos_201405  
  
hadoop fs -put /tmp/misdatos.txt datos_201405/misdatos.txt  
  
hadoop fs -mv datos_201405/misdatos.txt datos_201405/ datos20140525.txt
```

Nótese cómo los comandos *put* y *get* de Hadoop permiten intercambiar ficheros entre Hadoop y el mundo exterior, es decir, el sistema de ficheros anfitrión.

Hadoop sería muy poco práctico si las anteriores fuesen las únicas operaciones que permitiese realizar sobre los datos. Tradicionalmente, fuera de Hadoop y del *big data* en general, cuando un usuario ejecuta un programa que requiere un fichero del disco (p. e., un texto que se quiere editar), el sistema operativo hace una petición al sistema de ficheros para transferirlo a la memoria y poder operar sobre él. Sin embargo, el tamaño de los ficheros para los que Hadoop fue concebido impide que puedan ser procesados de la misma manera.

Un ejemplo típico es el de contar registros: ¿cuántas líneas tiene un determinado fichero? Dependiendo de su tamaño, sería imposible hacer lo siguiente:

```
hadoop fs -cat mifichero.txt | wc -l
```

es decir, extraer una copia fuera de Hadoop y contar el número de líneas utilizando las herramientas habituales (en este caso, de Linux/Unix) para contar líneas.

Tanto por el tamaño potencial de los ficheros como por la naturaleza distribuida de Hadoop, cualquier mecanismo que permita ejecutar programas sobre ficheros tiene que ser paralelo. Así que Hadoop proporciona un mecanismo de paralelización: *mapReduce*.



Otros sistemas de almacenamiento de datos también proporcionan mecanismos propios de manipulación de datos; por ejemplo, las bases de datos relacionales proporcionan un lenguaje, SQL, que permite realizar determinadas operaciones sobre las tablas que almacenan. De hecho, existen extensiones de Hadoop (p. e., Hive) que proporcionan una interfaz basada en SQL (y, por supuesto, construida mediante procesos mapReduce) para operar con datos almacenados en Hadoop. Sin embargo, SQL tiene una serie de limitaciones importantes que mapReduce resuelve:

- No gestiona adecuadamente la paralelización, que en mapReduce es explícita.
- Es muy limitado en términos de las operaciones que pueden realizarse sobre los datos.

En su versión más corriente (y simplificada), mapReduce opera de la siguiente manera:

- El usuario tiene que proporcionar dos programas, el *map* y el *reduce* e indicar sobre qué fichero opera primero (el fichero sobre el que trabaja).
- Hadoop identifica los nodos del clúster en los que se encuentra el fichero y ejecuta copias del *map* sobre dichos nodos en paralelo.

El resultado de esas ejecuciones se envía a otros nodos, los *reducers*, donde se ejecutan también en paralelo copias de *reduce*. El usuario tiene control sobre qué salidas se envían a qué nodos y cuántos *reducers* se ejecutan. El *reduce* toma como entrada la salida de los *maps* y opera sobre ellos.

- El usuario recibe la salida de los *reducers*.

Cualquier proceso de análisis de datos que quiera implementarse sobre Hadoop tiene que atenerse al esquema anterior. Uno de los ejemplos más básicos, el de contar líneas en un fichero se implementaría de la siguiente manera:

- El usuario crea un programa, *map*, que cuenta líneas en un fichero. Su salida es un número entero.
- Hadoop ejecuta *map* sobre los bloques de los que consta el fichero completo en paralelo y envía los valores obtenidos (números enteros) al *reduce*.
- El *reduce*, en este caso, se ejecuta en un único nodo y es únicamente un programa que devuelve la suma de sus entradas.

El taller que se describe en la siguiente sección describe estos pasos con detalle.

### Taller 1: el «Hola, Mundo» del big data

El taller que acompaña la primera parte de la charla consiste en una introducción práctica a Hadoop consistente en la implementación de lo que se ha convertido en, de facto, el «Hola, mundo» del *big data*: contar los niveles de una variable categórica. Es decir, el equivalente a

```
table(mitabla$v)
```

en R o

```
select v, count(*) from mitabla group by 1;
```

en SQL.

El reto consiste en implementar esta operación sobre tablas contenidas en Hadoop.

Hadoop no es una herramienta de fácil instalación. Por eso se ha proporcionado a los asistentes al taller una máquina virtual disponible para su descarga en Internet [1]. Una instalación propiamente dicha de Hadoop corre en clústers de ordenadores. Sin embargo, para facilitar el desarrollo y la enseñanza de esta tecnología, es posible crear instalaciones que simulan clústers sobre una única máquina. La máquina virtual proporcionada encapsula una de tales instalaciones de Hadoop y abstrae a sus usuarios circunstanciales de la no pequeña complejidad de su instalación en sistemas productivos distribuidos.

La máquina virtual es una versión adaptada de otra de Hortonworks [2] que ha sido modificada por el autor para incluir:

- R, como lenguaje de análisis de datos.
- RStudio Server [3], un GUI para R que facilita el trabajo sobre la máquina virtual. De hecho, permite operar sobre ella desde una ventana del navegador abierta desde el ordenador anfitrión y simular por tanto una conexión remota a un clúster de Hadoop. Además, qcon RStudio los usuarios disponen de una plataforma cómoda y probablemente familiar para desarrollar sus programas en R.

El paquete `rnr2` [4] y sus dependencias para interactuar con Hadoop directamente desde R y poder programar en dicho lenguaje los correspondientes *mappers* y *reducers*.

Ha de advertirse que Hadoop es una herramienta desarrollada en Java y que Java es también el lenguaje por defecto para desarrollar aplicaciones sobre dicha plataforma, sean o no de tipo analítico. No obstante, Hadoop dispone de extensiones que permiten utilizar otros lenguajes de programación que han sido aprovechadas por la comunidad de usuarios de R para crear interfaces para Hadoop sobre su herramienta favorita. En este taller se ha utilizado `rnr2` precisamente para poner en valor el conocimiento de R de los asistentes y reducir las barreras de entrada a esta tecnología al no obligarlos a programar en, por ejemplo, Java.

En el taller se trabaja con el fichero `bigiris.txt`, que es una versión recrecida del conjunto de datos iris construido mediante una muestra con repetición de las líneas de este último hasta completar unos cuantos cientos de miles de líneas. Por lo tanto, el fichero contiene cinco columnas; las cuatro primeras son numéricas longitud/anchura del pétalo/sépalo y la última es una etiqueta que identifica el tipo de flor. Nótese además que en un fichero habitual de Hadoop no hay encabezamientos: dado que el fichero se trocea al ser almacenado, no habría garantías de que cada uno de los bloques recibiese esos metadatos.

Una primera solución al problema de contar el número de líneas según el campo *Species* consiste en extraer la tabla de Hadoop, leerla en la memoria local (de la máquina virtual) como una tabla de R y usar entonces los comandos habituales para obtener el resultado (y correspondiente diagrama de barras) deseado:

```
tmp <- read.table(pipe("hadoop fs -cat bigiris.txt"))  
  
barplot(table(tmp$V5))
```

Sin embargo, este subterfugio es a menudo imposible en la práctica. Las tablas típicamente almacenadas en Hadoop no caben en memoria: de otra manera, no estarían almacenadas en Hadoop. Para utilizar Hadoop propiamente es necesario crear un proceso mapReduce.

El más básico que puede plantearse es el siguiente:

```
library(rmr2)

fnam <- "/user/rhadoop/bigiris.txt" # fichero
fif <- make.input.format("csv", sep = "\t") # formato

# función map
my.map <- function(k,v){
  tmp <- table(v$V5)
  keyval(names(tmp), tmp)
}

# proceso mapreduce
res <- from.dfs(mapreduce(
  input = fnam,
  input.format = fif,
  map = my.map
))
```

Primero se carga el paquete `rmr2` para poder acceder a los recursos proporcionados por Hadoop desde R. Las primeras líneas identifican el fichero (de Hadoop) que se desea leer y su formato.

Luego se define la función `map` que realiza los conteos locales. Hadoop trocea el fichero original en bloques y los distribuye entre las diversas máquinas que constituyen el clúster. La función `map` se aplica sobre cada uno de los bloques sin que el usuario pueda asumir nada sobre su ubicación o tamaño.

La función `map`, además, acepta parámetros y devuelve resultados de un formato muy concreto: parejas clave-valor. La clave de la entrada es irrelevante en nuestro ejemplo, pero no el valor. Su salida, sin embargo, tiene la forma valor-frecuencia. Es decir, las claves son los niveles de la variable de interés observados en el bloque en cuestión y los valores su frecuencia de aparición.

En la tercera parte se lanza el proceso en Hadoop. Nótese que en esta ocasión no se ha definido (no es obligatorio) un `reducer`. El resultado es una lista de longitud dos. La primera componente de la lista contiene un vector de claves (los obtenidos por cada uno de los `mappers` sobre cada bloque del fichero) y la segunda componente, una lista de números de la misma longitud: sus frecuencias correspondientes.

El resultado buscado puede obtenerse, por lo tanto, haciendo

```
all.res <- tapply(res$val, res$key, sum)

barplot(all.res)
```

Esta operación agrega las frecuencias correspondientes a las mismas claves. Y es, en esencia, una reducción, un *reduce*. Este paso puede integrarse en el paradigma mapReduce definiendo una función *reduce* de la siguiente manera:

```
my.reduce <- function(k,v){  
  
  keyval(k, sum(v))  
  
}
```

Si entonces se ejecuta

```
res <- from.dfs(mapreduce(  
  input = fnam,  
  input.format = fif,  
  map = my.map,  
  reduce = my.reduce ))  
  
tmp <- res$val  
names(tmp) <- res$key  
  
barplot(tmp, main = "ahora con reducers")
```

Hadoop ejecutará el mismo *mapper* que antes. Pero, después, enrutará las parejas clave-valor con la misma clave a un proceso específico, un *reducer*, que ejecutará la función *reduce*. Cada *reducer* verá todas las parejas clave-valor con una misma clave y solo ellas. Tras sumar las correspondientes frecuencias, generará una nueva combinación de claves-valor. En este caso concreto, cada *reducer* generará una única pareja: la que tiene como clave un determinado nivel y como valor la frecuencia de ocurrencia en el fichero original. Por supuesto, esta frecuencia total es igual a la suma de las frecuencias en los bloques.

Nótese como el *reducer* tiene la misma entrada y salida que el *mapper*: parejas clave-valor. Y cómo la salida del proceso completo es análoga a la del caso anterior. La diferencia es que, ahora, en el resultado final, hay una única clave por nivel.

Internamente, Hadoop ha realizado de manera transparente para el usuario una serie de operaciones:

- Ha distribuido cada función *map* sobre cada uno de los bloques en que estaba partido el fichero original y las ha ejecutado sobre cada uno de ellos.
- Ha capturado las parejas clave-valor generadas por cada uno de los *mappers* y las ha reordenado por clave.
- Ha enviado las que tienen una misma clave a un *reducer*.
- Ha ejecutado la función *reduce* en cada *reduce*.
- Ha devuelto al usuario las parejas clave-valor resultantes.

El hecho de que el formato de entrada y salida de *mappers* y *reducers* esté preespecificado de una manera un tanto inflexible, como parejas clave-valor, tiene ventajas y desventajas. La prin-

principal de entre las últimas es que obliga a utilizar ese tipo de estructura de datos, en ocasiones un tanto antinatural, como formato intermedio para todo tipo de problemas. La ventaja es que es posible concatenar procesos mapReduce y utilizar la salida de uno como entrada para uno posterior. Además de que un paradigma tan amplio como pretende implementar mapReduce no resultaría operativo sin introducir algún tipo de restricción en la interfaz de las funciones, por lo demás, sin otras restricciones, *map* y *reduce*.

Existen implementaciones alternativas que resuelven también el problema anterior. Podría cambiarse la función *map* anterior por

```
my.map <- function(k,v){
  keyval(v$V5, 1)
}
```

que no preagrega. Es decir, que genera multitud de pares, muchos de ellos repetidos: uno para cada fila del fichero original. El resultado obtenido al ejecutar

```
res <- from.dfs(mapreduce(
  input = fnam,
  input.format = fif,
  map = my.map,
  reduce = my.reduce))

tmp <- res$val
names(tmp) <- res$key
barplot(tmp, main = "ahora con reducers y malos mappers")
```

es el mismo que más arriba. Solo que puede argumentarse (y probarse) que el proceso mapReduce así implementado es más ineficiente que el anterior.

Los procesos anteriores pueden generalizarse para realizar otras operaciones básicas. Por ejemplo, sumar campos. Pudiendo sumar y contar, también pueden calcularse medias. Construir tablas de frecuencias es otra de las generalizaciones posibles.

## Big data, estadística y estadísticos

Si se da por buena una definición de *big data* según la cual esta se trata de la disciplina que analiza conjuntos de datos cuyo tamaño excede el tamaño de las herramientas computacionales convencionales, los estadísticos han estado haciendo *big data* desde incluso antes de que a la estadística se la conociese como tal. En efecto, los censos, centrales en la actual función estadística y de los que hay noticia incluso desde la antigüedad, cabrían en la definición anterior. Muchos de los esfuerzos originales de recolección y análisis de datos de los pioneros de la estadística (piénsese en los casos de William Playfair en el siglo XVIII, o de Florence Nightingale durante la guerra de Crimea) supusieron un esfuerzo de recolección, agregación, tabulación y análisis que uno bien puede suponer mucho más sofisticado que los imperantes en sus tiempos.

Sin embargo, algo sucedió en los años veinte y treinta del siglo xx que resumen breve y crudamente Bradley Efron y Persi Diaconis [6]:

«[...] el ordenador está cambiando la teoría de la estadística. Arriba hemos examinado nuevas teorías que han surgido a causa del ordenador. Otro cambio evidente es de los conjuntos de datos enormes que están disponibles a causa de la memoria de los ordenadores. Además, el ordenador permite usar métodos tradicionales para resolver problemas más grandes. El análisis de componentes principales es un buen ejemplo: fue inventado antes de que fuese realmente práctico.»

«Fisher desarrolló una teoría estadística que exprimía al máximo los recursos computacionales disponibles en los años treinta. El objetivo es hacer lo mismo en los ochenta.»

Treinta años después, gran parte de la estadística contemporánea y, en particular, el currículo académico, sigue viviendo al margen de la revolución que trajo el ordenador consigo y, particularmente, de todo el movimiento relacionado con el *big data*.

El autor propone el siguiente dilema: ¿debería la estadística (y su currículo) seguir estudiando y ahondando en los problemas que tradicionalmente y desde los años veinte del siglo xx constituyen su canon? ¿O debería abrirse a ese conjunto de procedimientos y técnicas emergentes que conforman el universo del *big data*?

Existen, por supuesto, contraejemplos que matizan la visión planteada en el párrafo anterior. Por ejemplo, los avances computacionales que subyacen a, por ejemplo, gran parte de los métodos bayesianos modernos. Sería ocioso repasarlos aquí.

Si de entre todas las posibilidades que para la estadística abrieron los ordenadores nos restringimos a las relacionadas con el *big data*, merece la pena revisar *Data Science, Big Data, and Statistics: can we all live together?* [7]. En él, Terry Speed identifica multitud de problemas en los que han trabajado estadísticos durante los últimos veinte años que son, propiamente, *big data* pero sin que necesariamente se reconozcan o etiqueten como tales. En lo que, no obstante, coinciden muchos observadores, incluyendo tanto a Terry Speed como al autor, es en la marginalización de los estadísticos en el *big data*, disciplina que ha sido capturada por investigadores y profesionales procedentes de otras disciplinas: la informática, la física, etc.

Esta situación plantea dos problemas distintos. Por un lado, los estadísticos han quedado (parcialmente y con egregias excepciones) al margen de las oportunidades de desarrollo profesional, de los fondos de investigación, de los proyectos relacionados con el *big data*. Por otra parte, también el *big data* ha perdido la perspectiva enriquecedora que pueden aportar los estadísticos a la disciplina. Desde un punto de vista conceptual, los estadísticos son expertos en el estudio de la variabilidad, los efectos de la multidimensionalidad y la identificación de la causalidad. Por otra parte, existen numerosas técnicas que son patrimonio de la estadística que en ocasiones se infra o mal utilizan en contextos de *big data* precisamente por la escasa presencia de estadísticos trabajando en el área.

Una de las mayores barreras para que los estadísticos puedan acceder a puestos desde los que aportar su visión sobre el análisis de datos es la impericia —siempre en términos relativos con respecto a los graduados en otras disciplinas— técnica de muchos estadísticos. Trabajar efectivamente en entornos *big data* exige un grado de dominio de determinadas herramientas informáticas que excede el que reciben muchos estadísticos durante su formación.

De ahí que uno de los objetivos del presente trabajo sea poner en evidencia esa carencia. Pero, al mismo tiempo, mostrar cómo el dominio de esas herramientas abre a los estadísticos un universo nuevo tanto para la investigación como para su desarrollo profesional fuera de la academia. De ahí que el segundo de los talleres, el que se describe a continuación, explore una serie de técnicas estadísticas habituales desde la perspectiva del *big data* en un intento de unificación de perspectivas.

## Taller 2: dos modelos estadísticos en Hadoop

Este taller plantea implementaciones sobre Hadoop de dos modelos estadísticos muy habituales: la regresión lineal por mínimos cuadrados y la regresión logística.

El primero es sencillo porque la determinación de los coeficientes de una regresión por mínimos cuadrados puede paralelizarse y por lo tanto, reimplementarse fácilmente dentro del paradigma de paralelización, mapReduce, que proporciona Hadoop.

En efecto, el cálculo de los coeficientes asociados al problema de regresión del vector y sobre una matriz de variables predictoras  $X$  está dado por

$$X'X^{-1}X'y$$

y las matrices  $X'X$  y  $X'y$  pueden calcularse por trozos. Por ejemplo, si  $X' = X'_1 X'_2$ , entonces  $X'X = X'_1 X_1 + X'_2 X_2$ . Nótese como tanto  $X'X$  como  $X'y$  tienen un tamaño que no depende del número de filas de  $X$ , que podría ser inmanejable en contextos *big data*, sino el de columnas.

Esta relación puede aprovecharse para construir una solución paralela del proceso de ajuste de una regresión por mínimos cuadrados sobre Hadoop:

- Los *mappers* construyen submatrices  $X'X$  y  $X'y$  sobre los subconjuntos de valores definidos por los bloques originales.
- El *reducer* suma todas las matrices  $X'X$  y  $X'y$  y calcula los coeficientes de acuerdo con la fórmula anterior para devolver los coeficientes de interés.

El *mapper* puede definirse de la siguiente manera:

```
my.map <- function(k,v){
  colnames(v) <- c("a1", "a2", "a3", "x1", "x2", "y")
  model <- lm(y ~ ., data = v)
  x <- qr.X(model$qr)
  xpx <- t(x) %*% x
  xpy <- t(x) %*% v$y
  keyval(1, list(list(xpx = xpx, xpy = xpy)))
}
```

Nótese como la clave es siempre la misma, 1. Eso fuerza a Hadoop a enviar el resultado de todos los *mappers* a un único *reducer*. Nótese también como para recuperar la matriz X en R hay que utilizar la función `qr.X`, que deshace la descomposición QR que internamente utiliza la función `lm`.

El *reduce* es:

```
my.reduce <- function(k,v){
  xpx <- Reduce("+", lapply(v, function(x) x$xpx))
  xpy <- Reduce("+", lapply(v, function(x) x$xpy))
  coefs <- solve(xpx) %*% xpy
  keyval(1, coefs)
}
```

La clave, de nuevo, es irrelevante. El valor asociado es la lista de los coeficientes. Finalmente, el usuario puede correr el programa completo haciendo

```
res <- from.dfs(mapreduce(
  input = fnam,
  input.format = fif,
  map = my.map,
  reduce = my.reduce)
)
```

y obtener finalmente los coeficientes correspondientes de la siguiente manera:

```
t(res$val)
```

Efectivamente, estos coeficientes coinciden con los esperados: es posible en este caso leer los datos completos en memoria y ajustar después el modelo usando el habitual `lm`, para obtener

```
dat <- read.table( pipe("hadoop fs -cat logistic_regression_data.txt"), header = F, sep = "\t")
colnames(dat) <- c("a1", "a2", "a3", "x1", "x2", "y")
coefficients(lm(y~., data = dat))
```



Este primer ejemplo merece una serie de comentarios adicionales. El primero es que, en la práctica, esta implementación solo podría aplicarse a un conjunto limitado de problemas: aquellos en los que la matriz de entrada sea numérica. El segundo es que, de haber predictores categóricos, debido al tratamiento de ellos que hace la función `lm`, es posible que las matrices obtenidas en los distintos *mappers* no sean conformes.

Además, usar regresiones lineales en el ámbito del *big data* tiene poco sentido: se trata de uno de los modelos más rígidos (o parsimoniosos). Esto plantea el siguiente dilema: si los datos, realmente, tienen la estructura que espera el modelo lineal, los coeficientes pueden estimarse mucho más fácilmente ajustando el modelo sobre una muestra mucho menor. Si, por otro lado, no hay motivos para creer que la especificación lineal sea la adecuada, es ocioso tratar de ajustarla.

Sin embargo, el ejemplo anterior tiene gran valor pedagógico: el de la regresión lineal es un problema bien conocido y estudiado, y explota aspectos de él, como la posibilidad de paralelizar el proceso de estimación de los coeficientes, que encajan muy bien en el mecanismo `mapReduce`. Finalmente, exige y pone en valor ciertos conocimientos de álgebra lineal que es necesario poner en valor para lograr dicha paralelización.

El segundo ejemplo es el de una versión de la regresión logística. El algoritmo habitual para el ajuste de una regresión logística no es paralelizable: se trata de una sucesión de ajustes por mínimos cuadrados con pesos. Así que el ajuste global, a diferencia de lo que sucedía con la regresión lineal, no puede realizarse operando individualmente sobre los bloques. No obstante, sí que sería posible paralelizar cada una de las iteraciones encadenando una sucesión de procesos `mapReduce` [8].

En el taller se exploró una opción alternativa inspirada en los modelos que construye el paquete *party* de R [9]. Este paquete permite modelizar un conjunto de datos mediante un árbol cuyas hojas son otros modelos (p. e., regresiones logísticas). Cada uno de los modelos de las hojas se aplica sobre subconjuntos de las observaciones definidos por las bifurcaciones de las ramas del árbol. Las ramas se construyen tratando de maximizar la diferencia entre los modelos que cuelgan de las hojas.

Este procedimiento es sugerente porque al particionar el espacio de observaciones en grupos, los modelos se ajustan sobre subconjuntos de menor tamaño. De un tamaño que podría ser incluso lo suficientemente pequeño como para poder ser procesado en una única máquina.

El procedimiento es también atractivo porque flexibiliza el modelo logístico original, que es un modelo global y en el que todas las observaciones contribuyen al ajuste localizándolo: crea subconjuntos estancos de observaciones dentro de los cuales se ajustan modelos independientes. Los modelos locales tienen con frecuencia un comportamiento superior a modelos de índole global en contextos de *big data* por la heterogeneidad que en ocasiones presentan las poblaciones subyacentes.

El ejemplo siguiente implementa una versión de la regresión logística que fuerza explícitamente la paralelización de acuerdo con los niveles de una serie de variables clave. Esta segmentación de la población base puede tener sentido en muchos contextos prácticos. La idea directora es la de identificar variables que puedan determinar de manera importante el comportamiento de los sujetos de las subpoblaciones que identifican (piénsese, p. e., en sexo y tramo de edad; o provincia y nivel de estudios...).

En el ejemplo siguiente se va a trabajar sobre un conjunto de datos simulados así:

```
gen.logistic.data <- function(n){  
  
  dat <- data.frame(  
  
    a1 = sample(letters[1:10], n, replace = T),  
    a2 = sample(letters[1:2], n, replace = T),  
    a3 = sample(letters[1:3], n, replace = T),  
  
    x1 = exp(rnorm(n) / 10),  
  
    x2 = runif(n)  
  
  )  
  
  a1 <- sort(rnorm(length(levels(dat$a1))))[dat$a1]  
  a2 <- sort(rnorm(length(levels(dat$a2))))[dat$a2]  
  
  prob <- a1 + a2 + 2 * a1 * dat$x1 - a2 * dat$x2  
  prob <- exp(prob)  
  prob <- prob / (1+prob)  
  
  y <- rep(0, n)  
  y[runif(n) < prob] <- 1  
  
  dat$y <- y  
  
  dat  
  }  
  
write.table(gen.logistic.data(1e5), file = "/tmp/logistic_regression_data.txt", col.names = F, row.names = F, sep = "\t")  
  
# en una sesión de ssh se cargan los datos  
  
# en Hadoop así:  
  
# hadoop fs -put /tmp/logistic_regression_data.txt /user/rhadoop/logistic_regression_data.txt
```

Como consecuencia, por construcción, las variables  $a_1$ ,  $a_2$ ,  $x_1$  y  $x_2$  (pero no  $a_3$ ) determinan el comportamiento de  $y$ . Además, existe una interacción entre  $a_1$  y  $x_1$  y entre  $a_2$  y  $x_2$ .

La implementación propuesta divide la población original de acuerdo con los valores de las variables  $a_1$  y  $a_2$  en los *mappers*. Se supone que el usuario sabe de antemano que, efectivamente, estas variables pueden determinar el comportamiento de  $y$ .

Después, en los *reducers*, reconstruye los conjuntos de datos asociados a los valores iguales de las claves y ajusta sobre ellos un modelo logístico.

El resultado de ejecutar

```
fnam <- "/user/rhadoop/logistic_regression_data.txt" # fichero
fif <- make.input.format("csv", sep = "\t") # formato

my.map <- function(k,v){
  colnames(v) <- c("a1", "a2", "a3", "x1", "x2", "y")
  tmp <- split(v, list(v$a1, v$a2))
  keyval(names(tmp), tmp)
}

my.reduce <- function(k,v){
  tmp <- do.call(rbind, v)
  res <- bigglm(y ~ a3 + x1 + x2, data = tmp,
  family=binomial(logit), maxit = 20)
  keyval(k, list(res))
}

res <- from.dfs(mapreduce(
  input = fnam,
  input.format = fif,
  map = my.map,
  reduce = my.reduce)
)
```

son, de nuevo, dos listas. La primera contiene combinaciones de los niveles de las variables *a1* y *a2* y la segunda contiene modelos logísticos —efectivamente, los valores de las combinaciones clave-valor pueden contener objetos más sofisticados que números o listas de números— que corresponden a las observaciones asociadas a dichas claves.

El modelo resultante es, de hecho, una matriz de regresiones logísticas. Y es posible, por ejemplo, crear una función *predict* que permita obtener predicciones sobre nuevos datos así:

```
my.predict <- function(model, new.data){
  my.model <- model$val
  names(my.model) <- model$key

  tmp <- split(new.data, list(new.data$a1, new.data$a2), drop = T)
  tmp <- sapply(names(tmp), function(nm) predict( my.model[[nm]],
  tmp[[nm]], type = "response"))
  do.call(rbind, tmp)
}
```

También es posible estudiar conjuntamente todos los modelos construidos y ver si, efectivamente, los coeficientes obtenidos difieren entre los distintos grupos (o, alternativamente, un único modelo global bastaría). Haciendo

```

coefs <- sapply(1:length(res$key),
function(n){
tmp <- as.data.frame(summary(res$val[[n]])$mat)
tmp$var <- row.names(tmp)
tmp$split <- res$key[n]
tmp
}, simplify = F)
coefs <- do.call(rbind, coefs)
names(coefs) <- c("coef", "lower", "upper", "se", "p", "var", "split")

```

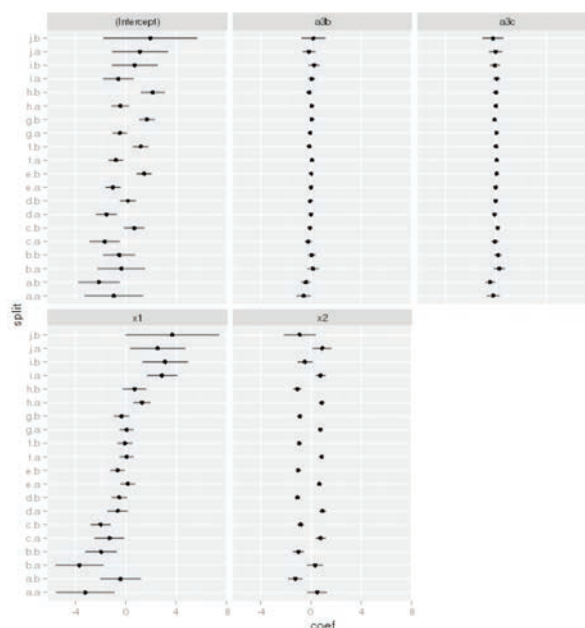
se obtiene información relativa a los coeficientes de cada uno de los submodelos: su valor, intervalos de confianza, etc., para después poder, por ejemplo, representarlos así:

```

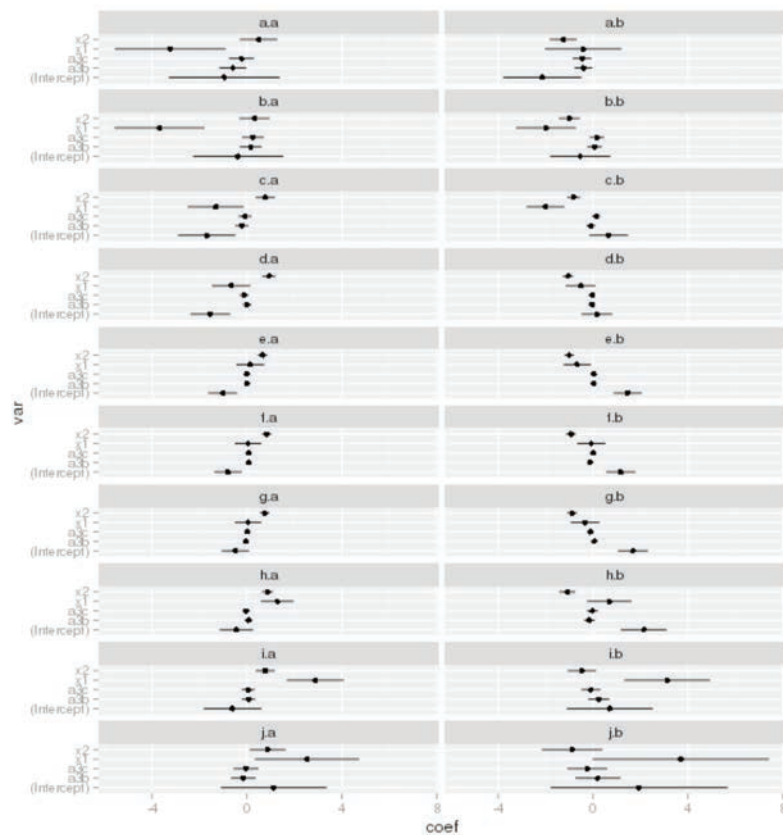
ggplot(coefs, aes(var, coef, ymin = lower, ymax=upper)) +
geom_pointrange() + coord_flip() + facet_wrap(~split, nrow = 10, ncol = 2)
ggplot(coefs, aes(split, coef, ymin = lower, ymax=upper)) +
geom_pointrange() + coord_flip() + facet_wrap(~var)

```

Los gráficos obtenidos son:



y



que muestran como, efectivamente, la variable  $a_3$  tiene poca influencia sobre  $y$  (nula, de hecho, por la manera en que se han simulado los datos) y como, efectivamente, existe una interacción entre  $a_1$  y  $x_1$  y  $a_2$  y  $x_2$ .

## Conclusiones

Este taller ha puesto de manifiesto como herramientas tradicionales de la estadística pueden extenderse y aplicarse en contextos de *big data* (particularmente, sobre conjuntos de datos contenidos en Hadoop). Pero también las dificultades a las que se puede enfrentar un estadístico para trasladar las técnicas tradicionales a estos nuevos entornos.

El *big data* plantea grandes retos, tanto prácticos como teóricos, a los estadísticos. Habrá procedimientos que no puedan extrapolarse, otros sí. A menudo, exigirán reimplementaciones.

En cualquier caso, el *big data* abre nuevos campos y está por ver (y el autor espera que así sea) si las nuevas generaciones de estadísticos contribuirán al desarrollo y expansión de la disciplina (y la profesión) recogiendo el guante.

## Agradecimientos

El autor agradece al Comité Organizador y muy particularmente a María Jesús García Ligeró la invitación para participar en las jornadas.

## Bibliografía y enlaces

- [1] <[http://datanalytics.com/uploads/hortonworks\\_sandbox\\_rstudio.zip](http://datanalytics.com/uploads/hortonworks_sandbox_rstudio.zip)>.
- [2] <<http://hortonworks.com/products/hortonworks-sandbox/>>.
- [3] <<http://www.rstudio.com/products/rstudio/>>.
- [4] <<https://github.com/RevolutionAnalytics/rmr2>>.
- [5] <<http://hadoop.apache.org/docs/r1.2.1/streaming.html>>.
- [6] Efron, B. y Diaconis, P. (1983): *Computer Intensive Methods in Statistics*.
- [7] Speed, T. (2014): *Data Science, Big Data, and Statistics: can we all live together?* <<http://flowingdata.com/2014/07/02/data-science-big-data-and-statistics-all-together-now/>>.
- [8] <<https://github.com/RevolutionAnalytics/rmr2/blob/master/pkg/tests/logistic-regression.R>>.
- [9] Zeileis, A.; Hothorn, T. y Hornik, K. (2008): «Model-Based Recursive Partitioning», *Journal of Computational and Graphical Statistics*, 17(2), 492-514.

## PÓSTERS

### Herramientas multimedia para la mejora de la enseñanza

*Montero Alonso, M. Á.;<sup>1</sup> Lara Aparicio, A.; Mohamed El Maslouhi, M.;  
Martín Ruiz, A.; Pérez Castro, M. Á.; González Jiménez, E.*

<sup>1</sup> mmontero@ugr.es, Universidad de Granada.

#### Resumen

La implantación de los grados y la actualización tecnológica constante en la que estamos inmersos exige la implementación de nuevas estrategias docentes y metodologías, así como disponer de herramientas y tecnología que permita llevarlas a cabo, constituyendo un elemento cotidiano e imprescindible de la sociedad en que vivimos y donde el sistema de enseñanza–aprendizaje ha de ser revisado constantemente y su mejora demanda una renovación y actualización continuas. Presentamos un compendio de herramientas que hemos elaborado dentro del Proyecto de Innovación Docente «Elaboración de materiales multimedia para los estudiantes de Ciencias Sociales» y que está disponible en la web <<http://eues.ugr.es/proyectoccss>>.

**Palabras clave:** materiales multimedia, Proyecto de Innovación Docente.

**Clasificación AMS:** 97D40, 97U80.



## Herramientas multimedia para la mejora de la enseñanza

Montero-Alonso, M. Á.; Lara-Aparicio, A.; Mohamed El-Maslouhi, M.  
Martín-Ruiz, A.; Pérez-Castro, M. Á.; González-Jiménez, E.

\*Universidad de Granada (Campus de Melilla)

### Introducción:

La implantación de los Grados y la actualización tecnológica constante en la que estamos inmersos exige la implementación de nuevas estrategias docentes y metodologías, así como disponer de herramientas y tecnología que permita llevarlas a cabo, constituyendo un elemento cotidiano e imprescindible de la sociedad en que vivimos y donde el sistema de enseñanza-aprendizaje ha de ser revisado constantemente y su mejora demanda una renovación y actualización continua.

### Objetivos:

En éste proyecto multidisciplinar, los componentes aportan sus conocimientos para realizar materiales multimedia para asignaturas de estadística, matemáticas y economía que permita tanto a los docentes que los utilicen como a los estudiantes enseñar-aprender de manera más intuitiva, con técnicas actuales y de forma innovadora para muchos profesores, pero usuales para la mayoría de los estudiantes y donde los objetivos están enfocados a profesores y alumnos.

### Material y Métodos:

En el presente proyecto se han tenido en cuenta los distintos subgrupos de trabajo formados y los materiales que se han creado, adaptando las necesidades de profesores y alumnos.

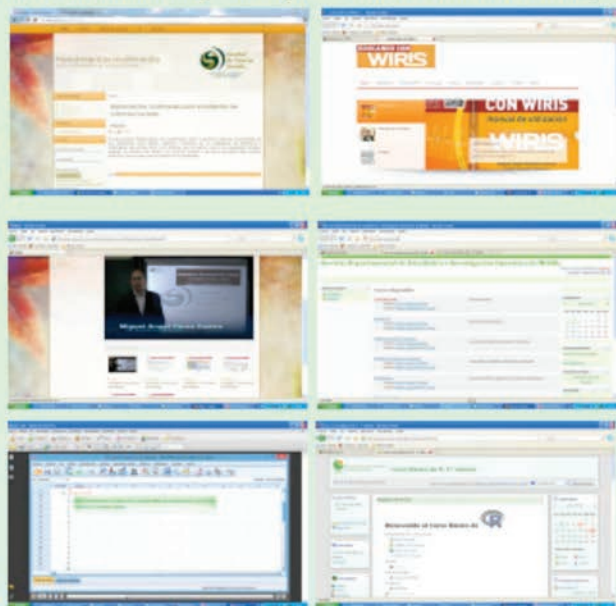
### Resultados:

Los productos o recursos generados en este proyecto están a disposición de toda la comunidad universitaria de forma libre, promoviendo la creación de un espacio de buenas prácticas en la innovación docente en el área de las ciencias sociales. Todo el material general está en la web <http://eues.ugr.es/proyectoccss>

### Bibliografía:

Montero-Alonso, M.A.; Pérez-Castro, M.A.; González-Jiménez, E. y Roldán-Nofuentes, J.A. (2013). Herramientas multimedia para los estudiantes de Ciencias Sociales, En: *X Foro Internacional Sobre la Evaluación de la Calidad de la Investigación y la Educación Superior (FECIES)*, Granada.

Quesada Vázquez, I. y Montero-Alonso, M.A. (2009). Estrategias de enseñanza-aprendizaje para los alumnos de Ciencias Sociales. En: *1<sup>as</sup> Jornadas Andaluzas de Innovación Docente Universitaria*. Córdoba.



### Conclusiones:

La implantación de este proyecto de innovación docente pretende un aumento en la calidad de la enseñanza donde los estudiantes podrán abordar las asignaturas con una mayor fluidez y un mejor conocimiento de la filosofía que subyace con la creación del Espacio Europeo de Educación Superior.



## Caso práctico de uso de la función sweave para la elaboración de exámenes usando r y tex

*Hernández Martín, Zenaida;<sup>1</sup> San Martín Pérez, Montserrat<sup>2</sup>*

<sup>1</sup> zenaida.hernandez@unirioja.es, Universidad de La Rioja.

<sup>2</sup> montse.sanmartin@unirioja.es, Universidad de La Rioja.

### Resumen

Cuando se plantea un examen de estadística, habitualmente se tienen los enunciados por un lado y las soluciones por otro. Además, generar varios exámenes similares con diferentes datos, supone la repetición de los mismos cálculos.

En este trabajo se muestra, de una forma sencilla y muy resumida, cómo utilizar la función Sweave() de R para integrar los resultados estadísticos generados con R en documentos de texto creados con TeX; en particular en el caso de la generación automática de exámenes.

El documento final incluye, no solo el enunciado de un examen (con datos generados aleatoriamente), sino también las soluciones a las preguntas planteadas, para esos datos.

Este sistema permite crear una base de datos de preguntas genéricas, que luego pueden ser utilizadas para la generación de exámenes.

El objetivo del trabajo no es elaborar un manual, hay varios y bastante bien explicados, sino proporcionar un ejemplo práctico que sirva de modelo para que cualquiera pueda utilizarlo directamente.

**Palabras clave:** Sweave, examen.

**Clasificación AMS:** 97U99.



## ***E-learning* para una estadística demográfica más autónoma y práctica**

Aguilera Morillo, M. C.;<sup>1</sup> Huete Morales, M. D.;<sup>2</sup> Vargas Jiménez, M.;<sup>2</sup> Sánchez Borrego, I. R.;<sup>2</sup> García López, P. A.;<sup>2</sup> Quesada Rubio, J. M.;<sup>2</sup> Carrillo Ballesteros, M. B.<sup>3</sup>

<sup>1</sup> maguiler@est-econ.uc3m.es, Universidad Carlos III de Madrid.

<sup>2</sup> Universidad de Granada.

<sup>3</sup> Universidad de Córdoba.

### **Resumen**

El objetivo de este trabajo es adaptar la enseñanza de la estadística demográfica a los nuevos planes de estudios basados en los créditos ECTS. Haciendo uso de recursos informáticos como el paquete Adobe DreamWeaver, lenguaje HTML y el lenguaje de programación orientado a objetos Visual Basic se ha creado una página web que contiene toda la información teórica asociada a los contenidos de estadística demográfica, así como prácticas automatizadas en las que con solo pulsar un icono, el alumno puede aplicar cada una de las técnicas demográficas estudiadas. Estas prácticas van acompañadas de vídeos *flash* sonoros que permiten un seguimiento más adecuado, favoreciendo el trabajo autónomo y la movilidad del alumnado. Como complemento, se ha establecido un sistema para la evaluación de las competencias que el alumnado va adquiriendo a lo largo del curso, mediante la inclusión de cuestionarios de autoevaluación, usando preguntas con respuestas múltiples, de modo que permitan al profesor evaluar las habilidades y conocimientos adquiridos por el alumno.

**Palabras clave:** *e-learning*, demografía.

**Clasificación AMS:** 62-XX.

## E-LEARNING PARA UNA ESTADÍSTICA DEMOGRÁFICA MÁS AUTÓNOMA Y PRÁCTICA



Departamento de Estadística  
Universidad Carlos III de Madrid



Departamento de Estadística e IO  
Universidad de Granada

*Aguilera-Morillo, M.C. (aguilero@es-cvcc.uca.es), Hincé Morales, M.D., Vargas Jiménez, M., Sánchez-Borrego, I.X., García-López, P.A., Quezada Rubio, J.M. y Carrillo-Balazares, M.B.*



V Jornadas de enseñanza y aprendizaje de la Estadística y la Investigación Operativa  
16-17 Junio 2014  
Madrid, España

El objetivo de este trabajo es adaptar la enseñanza de la Estadística Demográfica a los nuevos planes de estudios basados en los créditos ECTS. Haciendo uso de recursos informáticos como el paquete Adobe DreamWeaver, lenguaje HTML y el lenguaje de programación orientado a objetos Visual Basic se ha creado una página web que contiene toda la información teórica asociada a los contenidos de estadística demográfica, así como prácticas acompañadas en la que con solo pulsar un botón, el alumno puede aplicar cada una de las técnicas demográficas estudiadas. Estas prácticas van acompañadas de vídeos flash sonoros que permiten un seguimiento más adecuado, favoreciendo el trabajo autónomo y la actividad del alumno. Como complemento, se ha establecido un sistema para la evaluación de las competencias que el alumno va adquiriendo a lo largo del curso, mediante la inclusión de cuestionarios de autoevaluación, tests de preguntas con respuestas múltiples, de modo que permita al profesor evaluar las habilidades y conocimientos adquiridos por el alumno.

**Referencias**  
 [1] Snel, R. (1984). M.C. Vargas, M. (2011). Estadística aplicada a un Sistema Informático y Lenguaje Web. Proyecto I+D+i.  
 [2] Womack, J., Murdoch, M., Chittam, L. (2010). The Use of the Modeling Survey and Dash. European Association for Population Studies. European Academic Network. October 16-18, 2010.  
 [3] Hincé, M. (2012). Matemáticas con los gemelos de la familia. Universidad de Granada. Granada.  
 [4] Hincé, M. (2012). Análisis estadístico demográfico. Springer-Verlag. New York.  
 [5] Chittam, L. (1999). Visual Basic. Curso de programación. SA MA. Madrid.

### Descripción del contenido de la web:



- Con objeto de mejorar la Estadística Demográfica dentro del sistema de enseñanza superior, se ha creado una herramienta que facilita y mejora el aprendizaje del alumno e igualmente, ayuda al profesorado que imparte este tipo de materias.
- Puesto que esta asignatura se imparte en estudios de diferentes niveles, en los que a veces, el alumno no puede realizar cuestionarios tradicionales de competencias, el uso de una herramienta que simule estos cuestionarios queda justificado.
- La página webada se compone de una introducción, una serie de instrucciones de uso, secciones básicas sobre Estadística Demográfica y unas prácticas sobre diversos aspectos aplicadas a la Estadística Demográfica.
- En cada una de las prácticas, el usuario dispone de un manual teórico y una aplicación automatizada de la misma mediante programación Visual Basic.



### Organización del contenido de la web:

**Práctica 1: Introducción a la Estadística Demográfica**  
 Objetivo: Conocer los conceptos básicos de la Estadística Demográfica y su importancia en el estudio de la población.

**Práctica 2: Índice de Mortalidad**  
 Objetivo: Conocer el concepto de índice de mortalidad y su aplicación en el estudio de la mortalidad.

**Práctica 3: Diferencia entre las series temporales de mortalidad**  
 Objetivo: Conocer la diferencia entre las series temporales de mortalidad y su aplicación en el estudio de la mortalidad.

**Práctica 4: Esperanza de vida en Chile**  
 Objetivo: Conocer el concepto de esperanza de vida y su aplicación en el estudio de la mortalidad.

**Práctica 5: Método de Aringa**  
 Objetivo: Conocer el método de Aringa y su aplicación en el estudio de la mortalidad.

**Práctica 6: Cuentas de mortalidad queales por la función de Bellman-Polard**  
 Objetivo: Conocer el método de Bellman-Polard y su aplicación en el estudio de la mortalidad.

**Práctica 7: Método Logit de Brass para la proyección de tablas de mortalidad**  
 Objetivo: Conocer el método de Brass y su aplicación en el estudio de la mortalidad.



## Herramienta docente para la creación de actividades individuales con corrección inmediata

*Álvarez-Verdejo, Encarnación;<sup>1</sup> Vera López, Juan Antonio<sup>2</sup>*

<sup>1</sup> encarniav@ugr.es, Universidad de Granada, España.

<sup>2</sup> juanantonio.vera@tud.upct.es, Centro Universitario de la Defensa, San Javier, Murcia.

### Resumen

Con la entrada en vigor del Espacio Europeo de Educación Superior, los criterios de evaluación de las distintas materias que se imparten en las universidades españolas incorporan, en general, una valoración en torno al 30% del trabajo continuo desarrollado por el alumno de forma autónoma durante el curso. Con el fin de evitar que los alumnos copien los trabajos, sería recomendable que los trabajos fuesen individualizados. Pero por otra parte, esta metodología presenta el inconveniente del volumen de trabajo que supone por parte del profesor valorar objetivamente los trabajos. Este problema se acentúa más en centros o facultades con alumnado con características específicas; ocurre en los Centros Universitarios de la Defensa o en Facultades de Ciencias Económicas y Empresariales donde los alumnos compatibilizan sus estudios con otra carrera profesional. El presente trabajo presenta una herramienta para evaluar trabajos individualizados de forma rápida y personalizada.

**Palabras clave:** autoevaluación, aprendizaje autónomo, corrección.

**Clasificación AMS:** 62Q05, 91B82, 62P20.

## Herramienta docente para la creación de actividades individuales con corrección inmediata.

*Álvarez-Verdejo Encarnación, Vera-López Juan Antonio*  
Encarnacion.alvarez@ugr.es

Departamento de Métodos Cuantitativos para la Economía y la Empresa. Universidad de Granada, España  
alvarez@mat.us.es

Centro Universitario de la Defensa, Academia General del Aire,  
 Base Aérea de San Javier Murcia

**Descripción:** Herramienta para la creación de actividades individuales con corrección automática.

**Objetivos:** solventar los inconvenientes que conllevan la evaluación continua, los trabajos personalizados y la limitación de número de ejercicios que puede hacer al alumno en materias de estadística, métodos cuantitativos, cálculo, etc.

**Herramienta docente: Proceso de aprendizaje y uso**  
 Es una herramienta docente capaz de generar recursos docentes individuales y autoevaluables. Esta herramienta está siendo usada en fase piloto en asignaturas del Dpto. de Métodos Cuantitativos para la E. y la E. de la Univ. de Granada, y será en breve implementada en el Centro Universitario de la Defensa, en la Academia General del Aire de San Javier.

**Ejemplo de ejercicio propuesto**

1.- En primer lugar el alumno encontrará la siguiente interfaz:

|                          |                                                     |              |              |
|--------------------------|-----------------------------------------------------|--------------|--------------|
| Técnicas Cuantitativas 1 | Técnicas Cuantitativas 2                            | Asignatura 3 | Asignatura 4 |
| Cases virtuales de apoyo | Temas 1 y 2: Variables estadística unidimensionales |              |              |
| Hacer ejercicios         | Tema 3: ...                                         |              |              |
| Trabajo individual       | Tema 4: ...                                         |              |              |
| Trabajo de ordenador     | Tema 5: ...                                         |              |              |
| Hacer un examen          | Temas 6 y 7: ...                                    |              |              |
|                          | Tema 8: ...                                         |              |              |

2.- Completará la siguiente información:

Complete los siguientes datos

|                           |  |
|---------------------------|--|
| Apellidos                 |  |
| Nombre                    |  |
| Grado                     |  |
| Grupos                    |  |
| DMU                       |  |
| Valor de A                |  |
| Valor de B                |  |
| Valor de C                |  |
| Valor de D                |  |
| Último código de pregunta |  |

**CONTINUAR**

3.- El ejercicio y su planteamiento:

**PROBLEMA 1** El siguiente problema de maximización se resuelve aplicando el método de los números de partición. Complete los valores y responda a las preguntas que se van planteando.

**NOTA:** Las variables de decisión se expresarán con sus signos de desarrollo.

**TABLA DE DATOS DE LA PROBLEMA**

|               | C <sub>1</sub> | C <sub>2</sub> | R <sub>1</sub> | R <sub>2</sub> | R <sub>3</sub> |
|---------------|----------------|----------------|----------------|----------------|----------------|
| OBJETIVO      | 1              | 1              |                |                |                |
| RESTRICCIÓN 1 | 2              | 1              | 1              | 1              | 1              |
| RESTRICCIÓN 2 | 1              | 1              | 1              | 1              | 1              |
| RESTRICCIÓN 3 | 1              | 1              | 1              | 1              | 1              |
| RESTRICCIÓN 4 | 1              | 1              | 1              | 1              | 1              |

Problema 1. Obtenga el resultado de maximizar el valor Z (punto de máxima).

Respuesta: El resultado obtenido es:

Obtención de los valores de las variables de decisión:

Respuesta: El resultado obtenido es:

**RESUMEN DEL TRABAJO**

El alumno puede introducir el resultado de la actividad en el campo de texto que aparece a continuación. Haga clic en el botón de "CONTINUAR" para guardar los datos.

**CONTINUAR**

**TRABAJO INDIVIDUAL 1 - MÉTODOS CUANTITATIVOS 1**

|              |                                                                     |
|--------------|---------------------------------------------------------------------|
| UNIVERSIDAD  | UNIVERSIDAD DE GRANADA                                              |
| DEPARTAMENTO | DEPARTAMENTO DE MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA |
| ASIGNATURA   | MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA                 |
| GRUPO        | GRUPO 1                                                             |
| FECHA        | 16/06/2014                                                          |

**RESUMEN DEL TRABAJO**

El alumno puede introducir el resultado de la actividad en el campo de texto que aparece a continuación. Haga clic en el botón de "CONTINUAR" para guardar los datos.

**TABLA DE DATOS DE LA PROBLEMA**

|               | C <sub>1</sub> | C <sub>2</sub> | R <sub>1</sub> | R <sub>2</sub> | R <sub>3</sub> |
|---------------|----------------|----------------|----------------|----------------|----------------|
| OBJETIVO      | 1              | 1              |                |                |                |
| RESTRICCIÓN 1 | 2              | 1              | 1              | 1              | 1              |
| RESTRICCIÓN 2 | 1              | 1              | 1              | 1              | 1              |
| RESTRICCIÓN 3 | 1              | 1              | 1              | 1              | 1              |
| RESTRICCIÓN 4 | 1              | 1              | 1              | 1              | 1              |

Problema 1. Obtenga el resultado de maximizar el valor Z (punto de máxima).

Respuesta: El resultado obtenido es:

Obtención de los valores de las variables de decisión:

Respuesta: El resultado obtenido es:

**RESUMEN DEL TRABAJO**

El alumno puede introducir el resultado de la actividad en el campo de texto que aparece a continuación. Haga clic en el botón de "CONTINUAR" para guardar los datos.

**CONTINUAR**



Centro Universitario de la Defensa



Universidad de Granada

## **Estadística e investigación operativa en el proceso de internacionalización de la universidad española**

*Meca Martínez, Ana*

ana.meca@umh.es, Universidad Miguel Hernández de Elche (UMH)

### **Resumen**

El Gobierno de España, en coordinación con las Comunidades Autónomas y las universidades, ha desarrollado en los últimos años un gran esfuerzo para realizar la transformación y modernización de sus universidades y, en general de todo el sistema universitario español. Esta estrategia de modernización, conocida como Estrategia Universidad 2015, se ha presentado bajo el lema Internacionalización-Excelencia-Dimensión social y Empleabilidad que representa los tres grandes retos que de forma transversal cubren los objetivos estratégicos en los cuatro grandes ámbitos en que se distribuye la EU2015.

En este trabajo centramos nuestra atención en uno de los tres grandes retos de la EU2015, la internacionalización de la universidad española, y comprobamos con gran satisfacción que la respuesta al impulso lanzado por el Ministerio de Educación ha sido bastante positiva. Iniciamos un viaje por el sistema universitario español deteniéndonos en algunas de las universidades que han comenzado a desarrollar su propio plan de internacionalización. Realizamos una parada en la Universidad Miguel Hernández de Elche (UMH) para analizar las actividades realizadas desde el Vicerrectorado de Relaciones Internacionales. Y terminamos describiendo mi experiencia en la enseñanza y aprendizaje de la asignatura Modelos de decisión estratégica, en inglés, en el Grado de Estadística Empresarial.

**Palabras clave:** estadística, investigación operativa, modelos de decisión estratégica.

**Clasificación AMS:** 97-06, 91-01, 90-06.



## ESTADÍSTICA E INVESTIGACIÓN OPERATIVA EN EL PROCESO DE INTERNACIONALIZACIÓN DE LA UNIVERSIDAD ESPAÑOLA

Año XXIV  
Departamento de Estadística, Matemáticas e Informática, Universidad Miguel Hernández de Elche, Alicante

### CICUE, CRUE y Fundación Universidad.es

- Comisión de Internacionalización y Cooperación de las Universidades Españolas (CICUE): comisión sectorial de la CRUE cuyo objeto de actividad son proyectos y relaciones exteriores (Europa e Iberoamérica). Participa activamente en EUA, CUBH y IAU (UNESCO).
- Actividad de la CRUE en el escenario internacional.
- Asegurar una presencia constante de las Universidades españolas en los principales foros universitarios internacionales (EUA y CUBH).
- Reiniciar el diálogo con Conferencias de Rectores de otros países.
- Participar activamente en la creación del Espacio Iberoamericano del Conocimiento (XIC).
- Promover la movilidad y los intercambios con otras regiones.
- La Fundación Universidad.es: fundación para la proyección internacional de las Universidades españolas.
  - Iniciativa del Gobierno de España, promovida conjuntamente con los gobiernos regionales y las propias Universidades (21 de diciembre de 2008).
  - Ayudar a las Universidades en el desarrollo de sus estrategias de internacionalización así como en la puesta en marcha de programas de cooperación universitaria al extranjero.

### Estrategia Universidad 2015 (EU 2015)

- Transformación y modernización del Sistema Universitario Español: el Gobierno de España, en coordinación con las Comunidades Autónomas y las Universidades (CICUE).
- Estrategia de modernización presentada bajo el lema Internacionalización-Eficiencia-Diversidad social y Empleabilidad.
- Visión para el horizonte 2015: Consolidar un sistema universitario fuerte e internacionalmente atractivo y abierto al mundo para mejorar la competitividad internacional de España.
- Misión: transformación de nuestro sistema universitario en un sistema universitario de calidad reconocida a nivel internacional, supliendo como un sistema de educación superior puente entre las culturas globales de conocimiento.
- Objetivos Principales:
  - Promover la internacionalización del sistema universitario español.
  - Aumentar el atractivo internacional de las universidades.
  - Promover la dimensión y la responsabilidad social global de las universidades en la respuesta a los retos globales de la sociedad.
  - Intensificar la cooperación en educación superior con otras regiones del mundo con una perspectiva a largo plazo.

### Internacionalización

- Signos de la escasa proyección internacional de la Universidad española.
  - Participa muy bajo de alumnos internacionales (3,1 por ciento, mitad iberoamericanas).
  - Incentivos mínimos para traer a España estudiantes españoles en el exterior.
  - Flecha de profesores extranjeros.
- Líneas de acción en materia de internacionalización.
  - Políticas orientadas de captación de talento, gestión que sea flexible y modelo de financiación ad hoc para desarrollarla.
  - Captación de profesores de prestigio internacional (entre el 5 y 10 por ciento).
  - Mejorar el marco normativo regulador de estructuras/dimensión de materias y unidades de grado.
  - Incrementar la movilidad universitaria a través de una línea política lingüística (inglés).
  - Participar en las ferias de conocimiento existentes y atraer a las personas con más talento mediante redes de colaboración internacionalizadas tanto en educación, investigación y transferencia.
  - Capaces el entorno internacional, las universidades deberían recibir recursos NO en función de lo que quieren hacer, SINO de lo que han demostrado hacer bien.

### Algunas Universidades públicas con planes propios de internacionalización

- España tiene actualmente 77 universidades (49 públicas) y alrededor de 1.800.000 alumnos universitarios.
- La mayoría de las universidades públicas han comenzado a desarrollar planes de internacionalización con fondos propios.
- Universidades públicas que destacan en el proceso de internacionalización:
  - Universidad Carlos III (Personas con titulación bilingüe en Licenciatura en Economía y ADE en 1994/1995).
  - Universidad de Coruña, Santiago de Compostela y Vigo.
  - Universidad de Granada, Málaga, Cádiz y Sevilla.
  - Universidad de Alicante, Valencia, Politécnica de Valencia y Miguel Hernández de Elche (UMH).
  - Universidad de Burgos, León, Salamanca, Valladolid.
  - Universidad de Barcelona, Autónoma de Barcelona, Politécnica de Cataluña y Pompeu Fabra.
  - Universidad de Alcalá, Autónoma de Madrid, Complutense de Madrid, Politécnica de Madrid, Rey Juan Carlos y UNED.
  - Universidad de Cantabria, Castilla La Mancha, Extremadura, Illes Balears, La Rioja, Oviedo, País Vasco, Pública de Navarra y Zaragoza (G-9, 1977).

### Plan de internacionalización de la UMH

- Convenios con universidades extranjeras.
- Programa Ciencia sin Fronteras con Brasil.
- Becas CICUE, Hainé, Máster en Salud Pública y Becas Fundación Carolina.
- Club ERASMUS expatriates.
- Programas de movilidad (ERASMUS: Europa, DESTINO: resto de países y SECUE: España); estudiantes, PDI y PAS.
- Programa DESTINO: movilidad de PDI.
- Plan UMH de aprendizaje de inglés.
- Cursos de otros idiomas (Alemán, Francés, Italiano, Español).
- DELE - DIPLOMA DE ESPAÑOL COMO LENGUA EXTRANJERA.
- UMH English Summer School.
- Medidas para potenciar la docencia en inglés.
- Dobles titulaciones en la UMH.
- Programas especiales STUDY ABOARD para estudiantes internacionales de la UMH.

### 11. Medidas para potenciar la docencia en inglés

- Intensificar medidas que ayuden a estudiantes de Grado y Máster a prepararse en las mejores condiciones para la realización de actividades formativas en universidades extranjeras y facilitar a estudiantes oceanográficos su formación en nuestros Centros.
- Además de la docencia impartida por nuestros profesores en instituciones extranjeras (ERASMUS teaching staff), la UMH impulsa la docencia en lengua inglesa mediante el reconocimiento y apoyo a la actividad formativa que desarrolla su propio profesorado en dicho idioma.
  - Asignaturas optativas en lengua inglesa en los diplomados, licenciaturas e ingenierías en proceso de creación, así como en las asignaturas pertenecientes a la materia Oportunidad del Centro en los titulaciones de grado.
  - Asignatura de titulaciones de Máster Universitario impartida por PDI de la UMH.
- Los apoyo para impartir de asignaturas en inglés pueden reforzarse con las ayudas de las Comisiones de Proyectos de Innovación Docente que actualmente convoca esta Universidad con la finalidad de poder publicar el material docente redactado en lengua inglesa en la Open Course Ware (OCW) de la UMH.
- Reconocer la docencia impartida por los profesores de la UMH que hayan realizado un programa de movilidad ERASMUS (teaching staff), así como incrementar las intercambios bilaterales.

### Strategic Decision Models en Estadística

- Curso en impartición en inglés en la Licenciatura de Ciencias y Técnicas Estadísticas durante el curso 2012/2013, continúa en el 2013/2014, y pasa a impartirse en Grado de Estadística Empresarial en 2014/2015.
- Estadística extranjera matriculada en la asignatura: University of Bamnas, Poqae (1), University americana Large Boston, Milano (1), Politécnica de Milán (2) (El número total de alumnos por curso no supera 10).
- Resultados esperados del aprendizaje.
  - Preparación al estudiante las herramientas necesarias para el diseño y análisis cuantitativo problemas de decisión multiperiodo: conflicto y cooperación.
  - Asignatura orientada para que el alumno sea capaz de: (i) conocer las diferentes medidas de decisión, (ii) diseñar entre ellas, (iii) elegir las más adecuadas para la resolución de un análisis concreto, (iv) integrar una adecuada comprensión de los conceptos, así como su correcta utilización en cada caso.
  - Preparar al estudiante según el para la realización de actividades formativas en universidades extranjeras, y facilitar a estudiantes extranjeros su formación en nuestros Centros.

### Metodología, evaluación y resultados

- Material OCW-UMH disponible en web: 4 unidades didácticas con, al menos, una sesión práctica, resumen/lista de palabras clave, bibliografía y relación de problemas propuestos.
- Un caso práctico se analiza y estudia, con la participación activa de los alumnos, mientras se introducen los conceptos en cada unidad. El alumno debe realizar en casa los ejercicios propuestos y entregarlos al profesor en la fecha acordada.
- Proceso de evaluación continua que anima al estudiante a seguir el proceso de aprendizaje.
  - Actividades realizadas en las sesiones teóricas/prácticas (20) participación en clase, entrega de ejercicios propuestos, desarrollo de trabajos en grupo, sesiones seminario/talleres.
  - Trabajo final de carácter práctico (30) se presentará en una sesión oral de media hora por alumno.
  - Los alumnos se evalúan entre sí tanto en los ejercicios en-tregados (realización dada por profesor) como en la presentación oral.
  - Requisito indispensable haber asistido al menos al 80 por ciento de las clases.
- Experiencia muy positiva y gratificante en ambas cursos, con un 100 por cien de aprobados en 2012/2013.

### Referencias

- Cátedra Ibero de Gestión y Política Universitaria. Universidad Politécnica de Madrid. Internacionalización de las Universidades de Madrid. Fundación Universidad-España. Madrid, 2014.
- Hogg G. y Vialda J.M. La internacionalización de las Universidades, una estrategia necesaria. Studia XXI, Fundación Europea Socialidad e Innovación, Madrid, 2011.
- Juarez X. La internacionalización, tarea pendiente de la Universidad española. Notas de Prensa PaC España, 2013.
- Lera M., Bravo M., Sola C., Bergelesos A. y Crespo J. Tercer Cuadernillo Universidad Española.2013. PaC España, 2013.
- Mora A. Strategic Decision Models (Modelos de Decisión Estratégica). Digitalización de material docente en inglés OCW-UMH (REP: innovacionumh2012/OCW04). URL de publicación del material en el OCW-UMH: <http://ocw.umh.es/contenidos/docentes/mora/>.
- Ministerio de Educación. Estrategia de Internacionalización del Sistema Universitario Español (EU 2015), 2011.
- Vicerrectorado de Relaciones Internacionales. Memoria de Actividades realizadas durante el curso 2012/2013. Universidad Miguel Hernández de Elche, 2014.



## Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en appes

Ruiz Castro, J. E.;<sup>1</sup> Aguilera, A. M.;<sup>2</sup> Alonso, F. J.;<sup>3</sup> Escabias, M.;<sup>4</sup> García Montero, J.;<sup>5</sup> Raya Miranda, R.<sup>6</sup>

<sup>1</sup> jeloy@ugr.es, Universidad de Granada.

<sup>2</sup> aaguiler@ugr.es, Universidad de Granada.

<sup>3</sup> falonso@ugr.es, Universidad de Granada.

<sup>4</sup> escabias@ugr.es, Universidad de Granada.

<sup>5</sup> javier.garcia.edu@juntadeandalucia.es, Consejería de Educación, Junta de Andalucía.

<sup>6</sup> rraya@ugr.es, Universidad de Granada.

### Resumen

APPES (APP de EStadística) es una aplicación desarrollada para Android cuyo objetivo principal es el aprendizaje de la estadística. Se ha diseñado un contenido teórico, ejercicios guiados, ejercicios a resolver sin solución directa y vídeos tutoriales.

Este trabajo se centra en la presentación de la metodología y los aspectos técnicos de los vídeos tutoriales de APPES. En estos vídeos se hace hincapié en la aplicabilidad de la estadística. Se presentan los contenidos y objetivos, se introduce el ejemplo y se desarrolla en detalle.

En la elaboración se utilizan distintas técnicas como la focalización, la señalización y la presencia en el vídeo del profesor. Todos los vídeos presentan la opción de ser subtítulos para facilitar el aprendizaje a personas con necesidades educativas especiales.

Para la elaboración se ha contado con un equipo de auriculares y micrófono, *webcam* y tabla digitalizadora. Desde el punto de vista de los programas se han utilizado capturadores de pantalla y vídeo (CamStudio, *open source*), Powerpoint de Microsoft Office y editores de vídeo como Lightworks free o Camtasia Studio.

APPES se presenta en detalle en la web <<http://www.ugr.es/~appes>>.

**Palabras clave:** APPES, vídeos tutoriales, necesidades educativas especiales.

**Clasificación AMS:** 97Q60, 97K40, 97R20.



## DESARROLLO METODOLÓGICO Y ASPECTOS TÉCNICOS DE VÍDEOS TUTORIALES EN APPES




Juan Eloy Ruiz Castro, Ana M. Aguilera del Pino, Francisco J. Alonso Morales, Mannel Escabias Machuca, Javier García Montero, Rocío Raya Miranda  
 jeloy@ugr.es, aaguiler@ugr.es, falonso@ugr.es, escabias@ugr.es, javier.garcia.edu@juntadeandalucia.es, rraya@ugr.es

*Universidad de Granada*

**Objetivo:** APPES es una aplicación de libre acceso que está diseñada para que cualquier persona, con interés en el campo de la estadística, se introduzca y sea capaz de adquirir conocimientos desde un punto de vista teórico y aplicado. Para cada uno de los temas de conocimiento se ha diseñado un contenido teórico, ejercicios guiados, ejercicios a resolver sin solución directa y videos tutoriales. Este trabajo presenta la metodología desarrollada y los aspectos técnicos de los videos tutoriales de APPES.

### Contenidos de APPES

APPES es una aplicación desarrollada para Android cuyo objetivo principal es la transmisión y la enseñanza de la Estadística.

Los temas de conocimientos se presentan desagregados

Para cada tema de conocimiento se ha diseñado contenido teórico y práctico.

- Desarrollos teóricos
- Ejercicios guiados
- Ejercicios propuestos
- Vídeos Tutoriales




### Vídeos tutoriales en APPES

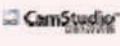
1. **Desarrollo metodológico.** El video tutorial tiene en cuenta los siguientes aspectos:
  - Debe ser atractivo para el usuario.
  - Debe estar compuesto de secuencias con contenido completo.
  - Debe ser diseñado de acuerdo a las características de los potenciales usuarios.
 Y cumplen las siguientes características:
  - El objetivo de cada video tutorial es mostrar la aplicabilidad de los conceptos teóricos a través de ejemplos sencillos y directos.
  - La duración de cada video es de unos 7-8 minutos. Aplicando técnicas de focalización y presencia puntual del profesor.
2. **Aspectos técnicos:** El video tutorial está realizado utilizando los siguientes medios:
  - **Hardware:** CPU, micrófono, auriculares, webcam, tabla digitalizadora










  - **Software:** CamStudio, Snagit 11, Powerpoint, Lightworks free, Camtasia Studio




## ESTRUCTURA DE UN VÍDEO TUTORIAL EN APPES



Presentación



Contenido y objetivos




Introducción del ejemplo



Empleo de técnicas de focalización y señalización



Introducción del profesor en momentos puntuales



Uso de subtítulos para facilitar el aprendizaje a personas con NEE

**REFERENCIAS**

•Ruiz-Castro, J.E., Aguilera, A.M., Escabias, M., García-Montero, J., Raya-Miranda, R. (2013a) Tablets in the learning of descriptive statistics: APPES. INTED 2013 Proceedings CD, 7 pp

•Ruiz-Castro, J.E., Aguilera, A.M., Escabias, M., García-Montero, J., Raya-Miranda, R. (2013b) Estadística en una Tablet: APPES. JVDE 2013

•Ruiz-Castro, J.E., Aguilera, A.M., Escabias, M., García-Montero, J., Raya-Miranda, R. (2014) Can we learn statistics through a Tablet? Yes, we can: APPES. BEIO, 30, 2, 181-198.

**APPES** ©  
 info: <http://www.ugr.es/~appes>  
 e-mail: [appes@ugr.es](mailto:appes@ugr.es)

## Comparación de metodologías de evaluación continua

*Simón de Blas, Clara;<sup>1</sup> Hernández Alonso, Sonia;<sup>2</sup> Montes Díez, Raquel<sup>3</sup>*

<sup>1</sup> clara.simon@urjc.es, URJC.

<sup>2</sup> sonia.hernandez@urjc.es, URJC.


<sup>3</sup> raquel.montes @urjc.es, URJC.

### Resumen

La estructura organizativa de la URJC de profesores en departamentos y escuelas en lugar de facultades, hace posible que la coordinación entre profesores sea más sencilla. Sin embargo, la libertad de cátedra provocaba una creciente demanda por parte del alumnado para la creación de contenidos y materiales unificados por los distintos profesores de diversas áreas de conocimiento. Por ello, el Área de Estadística e I.O. de la URJC creó una comisión para unificar los contenidos generales de la asignatura de estadística para las distintas titulaciones y grados ofertados. Además, los nuevos grados de enseñanza universitaria exigen criterios para la evaluación continua de los alumnos, lo que conlleva una reforma en los métodos tradicionales de enseñanza y una revisión de las metodologías tradicionales de evaluación. En este trabajo se comparan distintas metodologías de evaluación de conocimientos y rendimiento de estudiantes de distintos cursos de estadística en enseñanzas universitarias. El objetivo es diseñar unos criterios de evaluación comunes para los distintos grados de formación universitaria de la URJC en la asignatura de estadística.

**Palabras clave:** metodología enseñanza, diseño de experimentos, herramientas.

**Clasificación AMS:** 62-06, 62-07, 62K15, 97A80.



## COMPARACIÓN DE METODOLOGÍAS DE EVALUACIÓN CONTINUA


### Diseño de criterios de evaluación

**OBJETIVO:** Fijar las herramientas de evaluación continua de los alumnos de Estadística de distintas titulaciones para:


- \* Incrementar la tasa de éxito de egresados
- \* Mejorar el rendimiento académico.
- \* Incrementar el índice de satisfacción .

**Alternativas**


QUESTIONARIOS



HOJAS DE PROBLEMAS PARA ENTREGAR




FOROS



CASOS PRÁCTICOS (DATOS REALES vs SIMULADOS)  
PRÁCTICAS ÚNICAS VS PRÁCTICAS INDIVIDUALIZADAS

PRÁCTICAS CON EL ORDENADOR  
PLANTILLAS (SCRIPTS)



**Caso de estudio**

**CASOS DE ESTUDIO**

- \* 10 GRADOS/TITULACIONES- CIENCIAS AMBIENTALES, INGENIERÍA AMBIENTAL, INFORMÁTICA DE SISTEMAS, INGENIERÍA INFORMÁTICA, INGENIERÍA INFORMÁTICA DEL SOFTWARE, INGENIERÍA INFORMÁTICA DE COMPUTADORES, INGENIERÍA QUÍMICA, INGENIERÍA DE LA ENERGÍA, INGENIERÍA DE TELECOMUNICACIONES, BIOLOGÍA, CIENCIAS EXPERIMENTALES
- \* 2000 ALUMNOS DE LA URJC
- \* CURSOS 2004/2005 HASTA 2013/2014

**DISEÑO DE EXPERIMENTOS**

Prácticas en grupo genéricas

Prácticas ordenador R

Prácticas ordenador SPSS

Hojas de problemas

Nota final de Junio

Prácticas individuales

Cuestionarios campus virtual

Apuntes de programación en R (plantillas de Scripts)

Prácticas ordenador MATLAB

**RESULTADOS**

| Pruebas de los efectos inter-sujetos |                      |         |      |                  |          |      |
|--------------------------------------|----------------------|---------|------|------------------|----------|------|
| Variable dependiente: Nota Junio     |                      |         |      |                  |          |      |
| Origen                               | Suma de cuadrados    | tpo III | df   | Media cuadrática | F        | Sig. |
| Modelo corregido                     | 170,044 <sup>a</sup> |         | 8    | 21,255           | 5,101    | ,000 |
| Intersección                         | 6204,596             |         | 1    | 6204,596         | 1489,122 | ,000 |
| Cuestionarios                        | 24,049               |         | 1    | 24,049           | 5,772    | ,016 |
| Prácticas Personalizadas             | 12,416               |         | 1    | 12,416           | 2,980    | ,085 |
| Prácticas Generales                  | 22,114               |         | 1    | 22,114           | 5,308    | ,021 |
| Hojas problemas                      | 39,454               |         | 1    | 39,454           | 9,472    | ,002 |
| Scripts R                            | 50,458               |         | 1    | 50,458           | 12,110   | ,001 |
| Software                             | 67,995               |         | 3    | 22,665           | 5,440    | ,001 |
| Error                                | 6462,420             |         | 1551 | 4,167            |          |      |
| Total                                | 63661,842            |         | 1560 |                  |          |      |
| Total corregido                      | 6632,464             |         | 1559 |                  |          |      |

a. R cuadrado = ,026 (R cuadrado corregido = ,021)

**CONCLUSIONES**

Mejora del rendimiento académico con

- Prácticas generales
- Hojas de problemas
- Scripts
- MATLAB

Peores rendimientos con

- Cuestionarios del Campus Virtual

Las diferencias entre el uso de diferentes herramientas de evaluación son inferiores a 0.5 puntos en la nota final

**Sugerencias**

- ✓ Práctica semestral aplicada
- ✓ Presentación de los trabajos en grupo en clase
- ✓ Estudio de temas por cuenta propia
- ✓ Foros de discusión
- ✓ Seminarios aplicaciones por titulaciones

Clara Simón de Blas, Raquel Montes, Sonia Hernandez

V Jornadas de la Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

## Material interactivo para el aprendizaje del método simplex dual

*García-Ligero Ramírez, María Jesús;<sup>1</sup> Román Román, Patricia<sup>2</sup>*

<sup>1</sup> mjgarcia@ugr.es, Universidad de Granada.

<sup>2</sup> proman@ugr.es, Universidad de Granada.

### Resumen

En este trabajo se presenta un material docente para la resolución interactiva de un problema de programación lineal mediante el método Simplex Dual.

A partir de la tabla inicial de un problema con solución inicial no factible pero para el que existe una base dual factible, lo primero que se le propone al alumno es comprobar si la solución inicial es óptima. El alumno puede seleccionar dos opciones: Continuar y comprobar o Necesito ayuda. Si selecciona Continuar y comprobar, se le suministra la solución, sin detalles, a la tarea propuesta; con la opción Necesito ayuda se le explica de forma detallada y con animaciones cómo realizar los cálculos necesarios y comprobar las condiciones necesarias para llegar a la respuesta adecuada.

Dado que la solución básica inicial no es óptima pero todos los costes reducidos son no negativos, existe una base dual factible. Así, el siguiente paso que debe llevar a cabo el alumno es aplicar el método Simplex Dual para obtener una nueva solución básica factible. Para ello debe determinar la variable que sale de la base y la que entra y calcular la nueva tabla del problema. De nuevo, para cada una de esas tareas el alumno podrá seleccionar las opciones Continuar y comprobar y Necesito ayuda.

El proceso termina cuando llega a una solución óptima o, en caso de no poder incluir variables en la base, concluye que el problema dual es no acotado y, por tanto, el primal no es factible.

**Palabras clave:** problema de programación lineal, método Simplex Dual, materiales interactivos.

**Clasificación AMS:** C61.

# Material interactivo para el aprendizaje del Método Simplex Dual

García-Ligero Ramírez, M. J. y Román-Hernández, P.  
Departamento de Estadística e Investigación Operativa, Universidad de Granada

**TABLA INICIAL DEL PROBLEMA. COMPROBACIÓN DE LA OPTIMALIDAD DE LA SOLUCIÓN Y DE LA EXISTENCIA DE BASE DUAL FACTIBLE**

**Resolución interactiva del Simplex Dual**

Para que la solución sea óptima debe ser factible. Esto que debe verificarse en el rango de los valores de  $x_1$  y  $x_2$  en el problema original. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea factible. Una vez se haya encontrado la solución factible, se debe comprobar si es óptima. Para ello, se debe verificar si los valores de  $x_1$  y  $x_2$  son los mismos que en la solución óptima. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea óptimo.

**Resolución interactiva del Simplex Dual**

Para que la solución sea óptima debe ser factible. Esto que debe verificarse en el rango de los valores de  $x_1$  y  $x_2$  en el problema original. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea factible. Una vez se haya encontrado la solución factible, se debe comprobar si es óptima. Para ello, se debe verificar si los valores de  $x_1$  y  $x_2$  son los mismos que en la solución óptima. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea óptimo.

**Resolución interactiva del Simplex Dual**

Para que la solución sea óptima debe ser factible. Esto que debe verificarse en el rango de los valores de  $x_1$  y  $x_2$  en el problema original. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea factible. Una vez se haya encontrado la solución factible, se debe comprobar si es óptima. Para ello, se debe verificar si los valores de  $x_1$  y  $x_2$  son los mismos que en la solución óptima. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea óptimo.

**DETERMINACIÓN DE LA VARIABLE QUE ENTRA EN LA BASE Y LA QUE SALE**

**Resolución interactiva del Simplex Dual**

En esta etapa se debe determinar la variable que entra en la base y la que sale. Para ello, se debe calcular el cociente de los valores de la columna de la variable que entra en la base y los valores de la columna de la variable que sale en la base. El menor cociente indica la variable que sale de la base y el mayor cociente indica la variable que entra en la base.

**Resolución interactiva del Simplex Dual**

En esta etapa se debe determinar la variable que entra en la base y la que sale. Para ello, se debe calcular el cociente de los valores de la columna de la variable que entra en la base y los valores de la columna de la variable que sale en la base. El menor cociente indica la variable que sale de la base y el mayor cociente indica la variable que entra en la base.

**CONSTRUCCIÓN DE LA NUEVA TABLA DEL SIMPLEX DUAL**

**Resolución interactiva del Simplex Dual**

En esta etapa se debe construir la nueva tabla del Simplex Dual. Para ello, se debe utilizar los valores de la tabla anterior y los valores de la variable que entra en la base y la variable que sale de la base.

**Resolución interactiva del Simplex Dual**

En esta etapa se debe construir la nueva tabla del Simplex Dual. Para ello, se debe utilizar los valores de la tabla anterior y los valores de la variable que entra en la base y la variable que sale de la base.

**VERIFICACIÓN DE LA OPTIMALIDAD DE LA SOLUCIÓN**

**Resolución interactiva del Simplex Dual**

En esta etapa se debe verificar si la solución es óptima. Para ello, se debe calcular los valores de  $x_1$  y  $x_2$  y comprobar si son los mismos que en la solución óptima. Si no es así, se debe ajustar el valor de  $x_1$  y  $x_2$  para que sea óptimo.

**Una vez evaluada la utilidad de este material con un problema particular, la construcción de este trabajo se plantea como la realización de una aplicación (para estudiantes) para la resolución gráfica interactiva de un problema introducido por el alumno, de forma similar a la aplicación disponible en <http://www.operativa.unizar.es/operativa/operativa.html> para el Método Simplex.**

## Distribuciones de probabilidad con R

*Susi García, Rosario*<sup>1</sup>

<sup>1</sup> rsusi@ucm.es, Universidad Complutense de Madrid.

### Resumen

El objetivo de este trabajo es introducir las distribuciones de probabilidad más conocidas de variables aleatorias discretas y continuas, utilizando sencillas sentencias de R. Se muestran los nombres reservados, en dicho programa, a estas distribuciones que permiten conocer la función de masa o de densidad y la función de distribución en un punto, así como generar valores aleatorios que se ajustan a una distribución concreta y calcular el valor de un cuantil determinado. Para consolidar los conceptos teóricos presentados en el aula se representan gráficamente todas las funciones, trabajando con una misma distribución y distintos parámetros, con funciones de distribución de variables aleatorias discretas y continuas, así como con aproximaciones entre distribuciones.

**Palabras clave:** R, distribuciones de probabilidad, parámetros.

**Clasificación AMS:** 97K60, 00A35.

# Distribuciones de Probabilidad con R

Rosario Susi<sup>(1)</sup>

(1) Dpto. Estadística e Investigación Operativa III, Facultad de Estudios Estadísticos, Universidad Complutense Madrid, Spain

Contacto: rsusi@ucm.es



GRUPO DE INVESTIGACIÓN Y APRENDIZAJE DE LA ESTADÍSTICA Y LA INVESTIGACIÓN OPERATIVA

## Resumen

El objetivo de este trabajo es introducir las distribuciones de probabilidad más conocidas de variables aleatorias discretas y continuas, utilizando sencillas sentencias de R. Se muestran los nombres reservados, en dicho programa, a estas distribuciones que permiten conocer la función de masa o de densidad y la función de distribución en un punto, así como generar valores aleatorios que se ajustan a una distribución concreta y calcular el valor de un cuantil determinado. Para consolidar los conceptos teóricos presentados en el aula se representan gráficamente todas las funciones, trabajando con una misma distribución y distintos parámetros, con funciones de distribución de variables aleatorias discretas y continuas, así como con aproximaciones entre distribuciones.

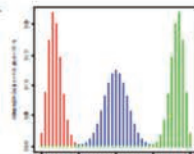
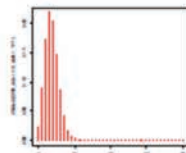


### V.A. DISCRETAS

| Distribución      | Parámetros    | Función de masa          | Función de distribución  | Generación n° aleatorios |
|-------------------|---------------|--------------------------|--------------------------|--------------------------|
| Binomial          | $n, p$        | $dbinom(x, n, p)$        | $pbinom(x, n, p)$        | $rbinom(x, n, p)$        |
| Hipergeométrica   | $KA, N-KA, n$ | $dhyper(x, KA, N-KA, n)$ | $phyper(x, KA, N-KA, n)$ | $rhyper(x, KA, N-KA, n)$ |
| Poisson           | $\lambda$     | $dpois(x, \lambda)$      | $ppois(x, \lambda)$      | $rpois(x, \lambda)$      |
| Binomial Negativa | $r, p$        | $dnbinom(x, r, p)$       | $pnbinom(x, r, p)$       | $rnbinom(x, r, p)$       |
| Geométrica        | $p$           | $dgeom(x, p)$            | $pgeom(x, p)$            | $rgeom(x, p)$            |

#### FUNCIONES DE MASA O PROBABILIDAD: DISTRIBUCIÓN BINOMIAL

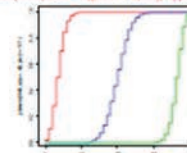
```
> plot(0:40, dbinom(0:40, size=40, prob=1/11), type="l", col="red")
```



```
> plot(0:40, dbinom(0:40, size=40, prob=1/11), type="l", col="red")
> lines(0:40, dbinom(0:40, size=40, prob=1/2), type="l", col="blue")
> lines(0:40, dbinom(0:40, size=40, prob=10/11), type="l", col="green")
```

#### FUNCIONES DE DISTRIBUCIÓN: DISTRIBUCIÓN BINOMIAL

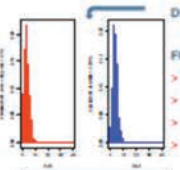
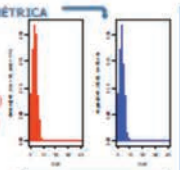
```
> plot(0:40, pbinom(0:40, size=40, prob=1/11), type="l", col="red")
> lines(0:40, pbinom(0:40, size=40, prob=1/2), type="l", col="blue")
> lines(0:40, pbinom(0:40, size=40, prob=10/11), type="l", col="green")
```



#### DISTRIBUCIÓN BINOMIAL & DISTRIBUCIÓN HIPERGEOMÉTRICA

##### FUNCIONES DE MASA

```
> par(mfrow=c(nrow=1, ncol=2))
> plot(0:40, dbinom(0:40, size=40, prob=1/11), type="l", col="red")
> plot(0:40, dhyper(0:40, 10000, 100000, 40), type="l", col="blue")
> par(mfrow=c(nrow=1, ncol=1))
```



#### DISTRIBUCIÓN BINOMIAL & DISTRIBUCIÓN POISSON

##### FUNCIONES DE MASA

```
> par(mfrow=c(nrow=1, ncol=2))
> plot(0:40, dbinom(0:40, size=40, prob=1/11), type="l", col="red")
> plot(0:40, dpois(0:40, lambda=40/11), type="l", col="blue")
> par(mfrow=c(nrow=1, ncol=1))
```

##### FUNCIONES DE DISTRIBUCIÓN

```
> plot(0:40, pbinom(0:40, size=40, prob=1/11), type="l", col="red")
> lines(0:40, ppois(0:40, lambda=40/11), type="l", col="blue")
```

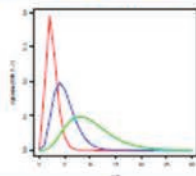
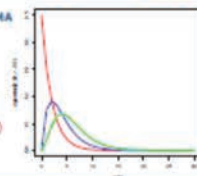
### V.A. CONTINUAS

| Distribución | Parámetros    | Función de densidad     | Generación n° aleatorios | Cálculo del cuantil     |
|--------------|---------------|-------------------------|--------------------------|-------------------------|
| Normal       | $\mu, \sigma$ | $pnorm(x, \mu, \sigma)$ | $rnorm(y, \mu, \sigma)$  | $qnorm(u, \mu, \sigma)$ |
| Uniforme     | $a, b$        | $runif(x, a, b)$        | $runif(y, a, b)$         | $qunif(u, a, b)$        |
| Exponencial  | $\lambda$     | $pexp(x, \lambda)$      | $rexp(y, \lambda)$       | $qexp(u, \lambda)$      |
| Gamma        | $\lambda, r$  | $pgamma(x, r, \lambda)$ | $rgamma(y, r, \lambda)$  | $qgamma(u, r, \lambda)$ |

#### FUNCIÓN DE DENSIDAD: DISTRIBUCIÓN GAMMA

##### CAMBIA EL PRIMER PARÁMETRO

```
> plot(0:30, dgamma(0:30, 1, 0.5), type="l", col="red")
> lines(0:30, dgamma(0:30, 2, 0.5), type="l", col="blue")
> lines(0:30, dgamma(0:30, 3, 0.5), type="l", col="green")
```



#### FUNCIÓN DE DENSIDAD: DISTRIBUCIÓN GAMMA

##### CAMBIA EL SEGUNDO PARÁMETRO

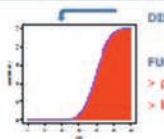
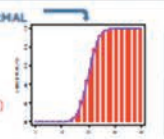
```
> plot(0:30, dgamma(0:30, 5, 2), type="l", col="red")
> lines(0:30, dgamma(0:30, 5, 1), type="l", col="blue")
> lines(0:30, dgamma(0:30, 5, 0.5), type="l", col="green")
```

### APROXIMACIONES DE V.A. DISCRETAS A V.A. CONTINUAS

#### DISTRIBUCIÓN BINOMIAL & DISTRIBUCIÓN NORMAL

##### FUNCIONES DE DISTRIBUCIÓN

```
> plot(0:40, pbinom(0:40, 40, 1/2), type="l", col="red")
> lines(0:40, pnorm(0:40, 20, sqrt(10)), type="l", col="blue")
```



#### DISTRIBUCIÓN BINOMIAL & DISTRIBUCIÓN POISSON

##### FUNCIONES DE DISTRIBUCIÓN

```
> plot(0:60, ppois(0:60, 40), type="l", col="red")
> lines(0:60, pnorm(0:60, 40, sqrt(40)), type="l", col="blue")
```



## Enseñando geoestadística con R

*Militino, Ana;*<sup>1</sup> *Ugarte, María Dolores;*<sup>2</sup> *Goicoa, Tomás;*<sup>3</sup> *Etxeberria, Jaione*<sup>4</sup>

<sup>1</sup> militino@unavarra.es, Universidad Pública de Navarra.

<sup>2</sup> lola@unavarra.es, Universidad Pública de Navarra.

<sup>3</sup> tomas.goicoa@unavarra.es, Universidad Pública de Navarra.

<sup>4</sup> jaione.etxeberrria@unavarra.es, Universidad Pública de Navarra.

### Resumen

El presente trabajo presenta cómo resolver un problema geoestadístico con R utilizando la librería «geoR». Para ilustrarlo se muestran los mapas de las interpolaciones geoestadísticas de temperaturas y precipitaciones del día 9 de junio de 2010. Estos mapas permiten ilustrar como se pueden obtener estimaciones de dichas variables meteorológicas en cualquier día del año y en cualquier punto de Navarra donde no existen pluviómetros que midan dichas variables. Se muestran los variogramas estimados y también los gráficos de validación que ofrece «geoR». También se presentan algunas líneas básicas del código R utilizado.

**Palabras clave:** *kriging*, interpolación, estadística espacial.

**Clasificación AMS:** 62P90.

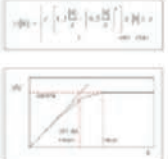
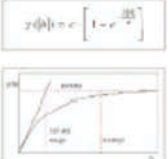


## Enseñando Geoestadística con

**Militino, A.F., Ugarte, M.D., Goicoa, T., Etxebarria, J.**  
Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra  
\*e-mail: militino@unavarra.es

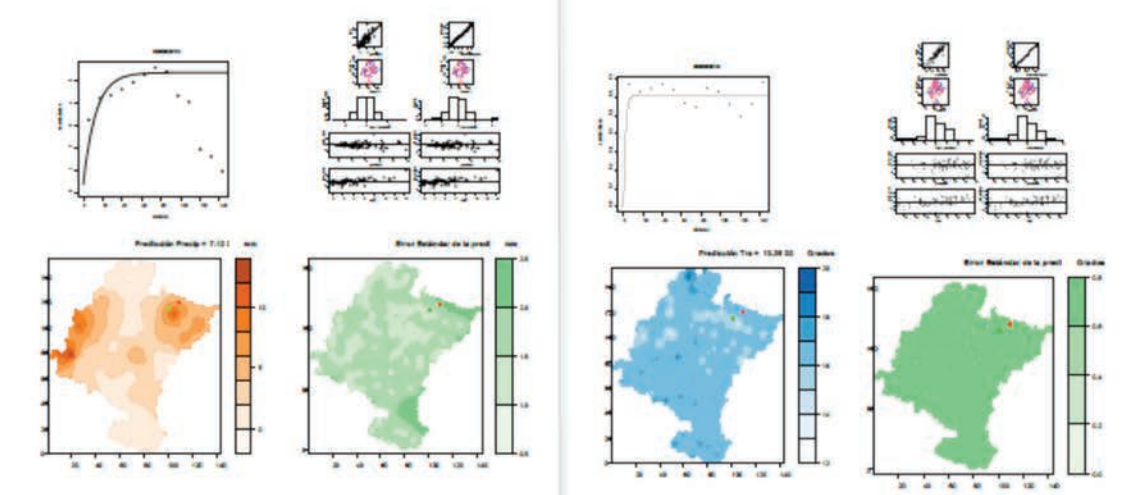
### Objetivo

Enseñar al alumno cómo se realiza la estimación espacial con el paquete `geoR`

| Geoestadística - Kriging Universal                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Kriging Universal                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Es un modelo de regresión lineal con errores correlados (DEPENDIENTES)</p> $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_i \quad i = 1, \dots, n$ <p>A medida <math>x_{1i}</math> es la coordenada <math>x</math>, <math>x_{2i}</math> es la coordenada <math>y</math> y <math>x_{3i}</math> es la coordenada <math>z</math>. En forma matricial</p> $Y = X\beta + \epsilon$ <ul style="list-style-type: none"> <li><math>\bullet</math> <math>Y</math> es el vector respuesta (<math>n \times 1</math>)</li> <li><math>\bullet</math> <math>X</math> es la matriz de diseño (<math>n \times (p+1)</math>)</li> <li><math>\bullet</math> <math>\beta</math> es el vector de coeficientes (<math>(p+1) \times 1</math>)</li> <li><math>\bullet</math> <math>\epsilon</math> es el vector de errores (<math>n \times 1</math>) espacialmente correlados</li> </ul> <p style="font-size: small; text-align: center;">Debido a que el semivariograma empírico no tiene por qué ser definido en negativo, se utiliza en muchos textos. Extraer los más comunes aquí:</p> <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; width: 45%;"> <p><b>Modelo Estérico</b></p> <math display="block">\gamma(h) = \sigma^2 \left( 1 - \frac{3}{2} \left  \frac{h}{c} \right  + \frac{1}{2} \left( \frac{h}{c} \right)^3 \right) \quad  h  \leq c</math>  </div> <div style="border: 1px solid gray; padding: 5px; width: 45%;"> <p><b>Modelo Exponencial</b></p> <math display="block">\gamma(h) = c \left( 1 - e^{-\frac{3h}{c}} \right)</math>  </div> </div> <p style="text-align: center; font-size: small;">a: Rango; c: Meseta</p> <pre style="font-size: x-small; font-family: monospace;">                 ## Datos                 data &lt;- read.csv("datos.txt", as.is=T)                 ## Matriz de diseño                 X &lt;- data[,1:3]                 X &lt;- cbind(1, X)                 ## Vector de respuesta                 Y &lt;- data[,4]                 ## Matriz de covarianza                 gamma &lt;- geoR::variog.empirico(X[,1:3])                 ## Matriz de covarianza teórica                 variog &lt;- geoR::variog.teror(gamma)                 ## Matriz de pesos                 w &lt;- geoR::mat.wts(gamma, variog)                 ## Estimación por mínimos cuadrados                 beta &lt;- geoR::lsq(Y, X)                 ## Estimación de Kriging Universal                 krig &lt;- geoR::krig.ku(X, Y, gamma)                 ## Matriz de varianzas-covarianzas de los errores y par cede de la                 ## variable respuesta. La diagonal es estimada mediante el variograma                 cov &lt;- geoR::mat.cov(krig)                 ## Predicciones                 predict &lt;- predict.ku(krig)                 ## Matriz de varianzas-covarianzas de las predicciones                 cov.pred &lt;- geoR::mat.cov(predict)                 ## Valores predichos                 predict             </pre> | $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$ $y \sim N(\beta X + \epsilon, \Sigma)$ <p>(1) <math>\epsilon \sim N(0, \sigma^2 \Sigma)</math><br/>La matriz <math>\Sigma</math> es la matriz de varianzas-covarianzas de los errores y par cede de la variable respuesta. La diagonal es estimada mediante el variograma</p> <p style="font-size: x-small;"><b>Predicciones</b></p> <pre style="font-size: x-small; font-family: monospace;">                 ## Datos                 data &lt;- read.csv("datos.txt", as.is=T)                 ## Matriz de diseño                 X &lt;- data[,1:3]                 X &lt;- cbind(1, X)                 ## Vector de respuesta                 Y &lt;- data[,4]                 ## Matriz de covarianza                 gamma &lt;- geoR::variog.empirico(X[,1:3])                 ## Matriz de covarianza teórica                 variog &lt;- geoR::variog.teror(gamma)                 ## Matriz de pesos                 w &lt;- geoR::mat.wts(gamma, variog)                 ## Estimación por mínimos cuadrados                 beta &lt;- geoR::lsq(Y, X)                 ## Estimación de Kriging Universal                 krig &lt;- geoR::krig.ku(X, Y, gamma)                 ## Matriz de varianzas-covarianzas de los errores y par cede de la                 ## variable respuesta. La diagonal es estimada mediante el variograma                 cov &lt;- geoR::mat.cov(krig)                 ## Predicciones                 predict &lt;- predict.ku(krig)                 ## Matriz de varianzas-covarianzas de las predicciones                 cov.pred &lt;- geoR::mat.cov(predict)                 ## Valores predichos                 predict             </pre> |

### Interpolación Lluvia el 9-06-2010

### Interpolación Temperatura el 9-06-2010



### Referencias

1. Ugarte MD, Militino AF, Amholt AT (2006). Probability and Statistics with R. Chapman & Hall/CRC.
2. R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
3. Alan T. Amholt (2009). PASWR: PROBABILITY and STATISTICS WITH R. R package version 1.1. <http://CRAN.R-project.org/package=PASWR>
4. Ribeiro P.J. and Diggle, P.J. `geoR`. A package for geostatistical analysis, *R (news)*, 1(2), 15-18

## **Desarrollo de un entorno virtual de autoaprendizaje de la estadística**

*Lara Porras, Ana María;<sup>1</sup> Román Montoya, Yolanda;<sup>2</sup> Parra Ortega, Pedro;<sup>3</sup> Hidalgo Arjona, Manuel;<sup>4</sup> Martín Arias, German<sup>5</sup>*

<sup>1</sup> alara@ugr.es, Departamento de Estadística e I.O. Universidad de Granada.

<sup>2</sup> yroman@ugr.es, Departamento de Estadística e I.O. Universidad de Granada.

<sup>3</sup> parraortega.pedro@gmail.com, desarrollador informático en CeiferIT.

<sup>4</sup> manolohidalgoarjona@gmail.com, estudiante del Grado de Estadística.

<sup>5</sup> germanma\_1@hotmail.com, estudiante del Grado de Estadística.

### **Resumen**

El autoaprendizaje de la estadística es un proceso que requiere de un intenso trabajo por parte del estudiante que puede agilizarse a través del uso de plataformas web adecuadas. Presentamos una plataforma web de libre acceso que incluye tanto material teórico como ejercicios resueltos y guiados con dos programas de computación estadística: SPSS y R.

Dadas las necesidades del proyecto, tras el análisis de las diferentes opciones disponibles en el mercado tecnológico, se ha optado por un sistema de gestión de contenidos (CMS) concreto Wordpress, que permite el desarrollo, manejo y gestión de recursos de la plataforma. Entre las ventajas que aporta, podemos citar:

- Es actualmente la principal plataforma CMS en uso.
- Existe una comunidad que respalda el desarrollo de la plataforma, lo que implica la generación de módulos adicionales instalables para extender funcionalidades necesarias.
- Hay disponible una gran cantidad de información, cursos, tutoriales, etc., para su utilización y mantenimiento.
- El sistema se encuentra alojado en servidores propios que la Universidad de Granada proporciona gratuitamente para el desarrollo de páginas web universitarias.

**Palabras clave:** CMS, Wordpress, R.

**Clasificación AMS:** 6204.

## DESARROLLO DE UN ENTORNO VIRTUAL DE AUTOAPRENDIZAJE DE LA ESTADÍSTICA

El Autoaprendizaje de la Estadística es un proceso que requiere de un intenso trabajo por parte del estudiante que puede agilizarse a través del uso de plataformas webs adecuadas. Presentamos una plataforma web de libre acceso que incluye tanto material teórico como ejercicios resueltos y guiados con dos programas de computación estadística: SPSS y R

Ana M<sup>a</sup> Lara Porras  
Yolanda Román Montoya  
Pedro Parra ortega  
Manuel Hidalgo Arjona  
Germán Martín Arias

**Desarrollo simple de sitios webs** → **Plataforma**

- ❖ Wordpress
- ❖ Drupla
- ❖ Joomla
- ❖ Prestashop
- ❖ Blogger

**CMS:** Aplicación que permite crear la estructura básica para la creación, administración y mantenimiento de contenidos, manejo y gestión de recursos para la plataforma

**Wordpress**

- Soporte fuerte: 10.000 descargas a la hora
- Información para su utilización y mantenimiento: <https://es.wordpress.com>
- Alojado en servidores propios que la UGR proporciona gratuitamente
- La principal plataforma CMS
- <https://es.wordpress.org>

<http://wdb.ugr.es/bioestad>

**Bioestadística**  
Universidad de Granada

**Entorno Virtual**  
DE AUTOAPRENDIZAJE DE LA ESTADÍSTICA

Ver más

Foros

RStudio

IBM SPSS Statistics 20

Localización

Contacto

Twitter

YouTube

REDES SOCIALES

TWITTER

FOROS

Integración de la cuenta youtube.com

PODCAST

VIDEOS TUTORIALES

## ***Flippeando con estadística***

*Serrano Czaia, Isabel;*<sup>1</sup> *Ortega Moreno, Mónica*<sup>2</sup>

<sup>1</sup> iserrano@uhu.es, Universidad de Huelva.

<sup>2</sup> ortegamo@uhu.es, Universidad de Huelva.

### **Resumen**

Cuando impartes docencia de estadística en una titulación de Ciencias Sociales, mantener la atención del alumno e intentar que este muestre interés por la materia se convierte siempre en un reto. Esto ha hecho que los docentes hayan ido buscando nuevas técnicas de enseñanza que despierten y fomenten el interés de los alumnos. En este sentido, las posibilidades que ofrecen las TIC han hecho que estas metodologías aumenten exponencialmente.


Una de estas técnicas emergentes, conocida en algunos foros como Flipped Classroom (la clase al revés), consiste en invertir el proceso de enseñanza. Esta metodología viene siendo aplicada en la modalidad semipresencial de la titulación en la que impartimos docencia.

El profesor proporciona al alumno con tiempo suficiente material (vídeos, PowerPoint, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, eso sí, fuera del aula. El docente prepara una serie de retos o actividades que el alumno deberá desarrollar en el aula, para ello deberá aplicar el conocimiento adquirido previamente. En las clases presenciales se exponen dudas, se debaten opciones óptimas, se presentan resultados... siempre bajo la supervisión, el apoyo y la dirección del profesor.

En este trabajo se describe cómo realizar una Flipped Classroom en las prácticas de estadística con Excel, con alumnos de primer curso.

**Palabras clave:** estadística, aprendizaje interactivo, herramientas *online*.

**Clasificación AMS:** 97U50, 97D40, 97K80.



**UNIVERSIDAD DE HUELVA**

# FLIPPEANDO CON ESTADÍSTICA

Isabel Serrano Czaia ([iserrano@uhu.es](mailto:iserrano@uhu.es)), Mónica Ortega Moreno ([ortegamo@uhu.es](mailto:ortegamo@uhu.es))  
 Depto. Métodos Cuantitativos para la Economía y la Empresa, Estadística e Investigación Operativa. UNIVERSIDAD DE HUELVA

V Jornadas  
Enseñanza y Aprendizaje de la  
Estadística e Investigación Operativa

**GENEIO**

**RESUMEN**

Cuando nuestra docencia de Estadística en una titulación de Ciencias Sociales, mantiene la atención del alumno e intentar que este muestre interés por la materia se convierte siempre en un reto. Esto ha hecho que los docentes hayamos buscado nuevas formas de enseñanza que despierten y fomenten el interés de los alumnos. En este sentido, las posibilidades que ofrecen las TIC han hecho que estas metodologías aumenten exponencialmente. Una de estas técnicas emergentes, conocida en algunos foros como Elipod Classroom (la clave al revés), consiste en invertir el proceso de enseñanza. Esta metodología viene siendo aplicada en la modalidad semi-presencial de la titulación en la que impartimos docencia.

El profesor proporciona al alumno con tiempo suficiente material (vídeos, power point, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, es lo que se llama el aula. El docente prepara una serie de estas actividades que el alumno deberá desarrollar en el aula, para ello deberá aplicar el conocimiento adquirido previamente. En las clases presenciales se exponen dudas, se debaten opciones óptimas, se presentan resultados... siempre bajo la supervisión, el apoyo y la dirección del profesor.

En este trabajo se describe como realizar una Flipped Classroom en las prácticas de Estadística con Excel, con alumnos de primer curso.

**El alumno aprende excel desde cero**

- Entiendo que la metodología de flipped classroom es una metodología innovadora.
- La metodología de flipped classroom es una metodología innovadora.
- La metodología de flipped classroom es una metodología innovadora.
- La metodología de flipped classroom es una metodología innovadora.

**El profesor**

- El profesor prepara una serie de actividades que el alumno deberá desarrollar en el aula.
- El profesor proporciona al alumno con tiempo suficiente material (vídeos, power point, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, es lo que se llama el aula.

**Flipped Classroom**

El alumno

- Desarrolla la actividad de forma autónoma.
- El profesor prepara una serie de actividades que el alumno deberá desarrollar en el aula.
- El profesor proporciona al alumno con tiempo suficiente material (vídeos, power point, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, es lo que se llama el aula.

El profesor

- El profesor prepara una serie de actividades que el alumno deberá desarrollar en el aula.
- El profesor proporciona al alumno con tiempo suficiente material (vídeos, power point, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, es lo que se llama el aula.

**Flipped Classroom**

El alumno

- Desarrolla la actividad de forma autónoma.
- El profesor prepara una serie de actividades que el alumno deberá desarrollar en el aula.
- El profesor proporciona al alumno con tiempo suficiente material (vídeos, power point, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, es lo que se llama el aula.

El profesor

- El profesor prepara una serie de actividades que el alumno deberá desarrollar en el aula.
- El profesor proporciona al alumno con tiempo suficiente material (vídeos, power point, ejemplos, tutoriales...) sobre el tema a tratar y el alumno debe, de forma autónoma o en grupo, comprender y asimilar los nuevos contenidos, es lo que se llama el aula.

## **Autores**

Aguilera, A. M., [aaguiler@ugr.es](mailto:aaguiler@ugr.es), Departamento de Estadística e I.O. Universidad de Granada, «Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en APPES».

Aguilera Morillo, M. C., [maguiler@est-econ.uc3m.es](mailto:maguiler@est-econ.uc3m.es), Universidad Carlos III de Madrid, «*E-learning* para una estadística demográfica más autónoma y práctica».

Alonso, F. J., [falonso@ugr.es](mailto:falonso@ugr.es), Departamento de Estadística e I.O. Universidad de Granada, «Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en APPES».

Álvarez Verdejo, E., [encarniav@ugr.es](mailto:encarniav@ugr.es), Departamento de Métodos Cuantitativos para la Economía y la Empresa, Universidad de Granada, España, «Herramienta docente para la creación de actividades individuales con corrección inmediata».

Carrillo Ruiz, J. A., [jcarrillor@et.mde.es](mailto:jcarrillor@et.mde.es), MINISDEF/AIAD, «Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave».

Carrillo Ballesteros, M. B., Universidad de Córdoba, «*E-learning* para una estadística demográfica más autónoma y práctica».

Daza Arbolí, M. Á., [mdazarb@et.mde.es](mailto:mdazarb@et.mde.es), MINISDEF/CIODEF, «Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave».

Escabias, M., [escabias@ugr.es](mailto:escabias@ugr.es), Departamento de Estadística e I.O. Universidad de Granada, «Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en APPES».

Etxeberria, J., [jaione.etxeberrria@unavarra.es](mailto:jaione.etxeberrria@unavarra.es), Departamento de Estadística e I.O., Universidad Pública de Navarra, «Enseñando geoestadística con R».

García Montero, J., [javier.garcia.edu@juntadeandalucia.es](mailto:javier.garcia.edu@juntadeandalucia.es), Departamento de Ciencias de la Computación, Consejería de Educación, Junta de Andalucía, «Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en APPES».

García-Ligero Ramírez, M. J., [mjgarcia@ugr.es](mailto:mjgarcia@ugr.es), Departamento de Estadística e I.O. Universidad de Granada, «Material interactivo para el aprendizaje del método Simplex Dual».

García López, P. A., Universidad de Granada, «*E-learning* para una estadística demográfica más autónoma y práctica».

Gil Bellosta, C., «Big data desde la perspectiva de un estadístico» e «Introducción práctica al análisis estadístico y gráfico de grandes volúmenes de datos utilizando R y Hadoop».

Goicoa, T., [tomas.goicoa@unavarra.es](mailto:tomas.goicoa@unavarra.es), Departamento de Estadística e I.O., Universidad Pública de Navarra, «Enseñando geoestadística con R».

González Alastrué, J. A., «Problemas numéricos con corrección automática: qué se puede y qué se debe hacer».

González Jiménez, E., Universidad de Granada, «Herramientas multimedia para la mejora de la enseñanza».

Hernández Alonso, S., sonia.hernandez@urjc.es, Departamento de Informática y Estadística, URJC, «Comparación de metodologías de evaluación continua».

Hernández Martín, Z., zenaida.hernandez@unirioja.es, Universidad de La Rioja, «Caso práctico de uso de la función Sweave para la elaboración de exámenes usando R y TeX».

Hidalgo Arjona, M., manolohidalgoarjona@gmail.com, estudiante del Grado de Estadística, «Desarrollo de un entorno virtual de autoaprendizaje de la estadística».

Huete Morales, M. D., Universidad de Granada, «*E-learning* para una estadística demográfica más autónoma y práctica».

Lara Porras, A. M., «Desarrollo de un entorno virtual de autoaprendizaje de la estadística».

Lara Aparicio, A., alara@ugr.es, Departamento de Estadística e I.O. Universidad de Granada, «Herramientas multimedia para la mejora de la enseñanza».

Martín Arias, G., germanma\_1@hotmail.com, estudiante del Grado de Estadística, «Desarrollo de un entorno virtual de autoaprendizaje de la estadística».

Martín Ruiz, A., Universidad de Granada, «Herramientas multimedia para la mejora de la enseñanza».

Meca Martínez, A., ana.meca@umh.es, Departamento de Estadística, Matemáticas e Informática (DEMI), Edificio Torretamarit, Universidad Miguel Hernández de Elche (UMH), «Estadística e investigación operativa en el proceso de internacionalización de la universidad española».

Militino, A., militino@unavarra.es, Departamento de Estadística e I.O., Universidad Pública de Navarra, «Enseñando geoestadística con R».

Mohamed El Maslouhi, M., Universidad de Granada, «Herramientas multimedia para la mejora de la enseñanza».

Montero Alonso, M. Á., mmontero@ugr.es, Universidad de Granada, «Herramientas multimedia para la mejora de la enseñanza».

Montes Díez, R., raquel.montes@urjc.es, Departamento de Informática y Estadística, URJC, «Comparación de metodologías de evaluación continua».

Muñoz Fernández, O., omunfer@et.mde.es, MINISDEF/CIODEF, «Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave».

Ortega Moreno, M., ortegamo@uhu.es, Departamento de Métodos Cuantitativos para la Economía y la Empresa, Estadística e I.O., Facultad de Ciencias del Trabajo, Universidad de Huelva, «*Flippeando* con estadística».

Parra Ortega, P., parraortega.pedro@gmail.com, desarrollador informático en CeiferIT, «Desarrollo de un entorno virtual de autoaprendizaje de la estadística».



Pérez Castro, M. Á., Universidad de Granada, «Herramientas multimedia para la mejora de la enseñanza».

Planells del Pozo, A., aplanells@et.mde.es, MINISDEF/UES, «Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave».

Quesada Rubio, J. M., Universidad de Granada, «*E-learning* para una estadística demográfica más autónoma y práctica».

Raya Miranda, R., rraya@ugr.es, Departamento de Estadística e I.O. Universidad de Granada, «Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en APPES».

Román Montoya, Y., yroman@ugr.es, Departamento de Estadística e I.O. Universidad de Granada, «Desarrollo de un entorno virtual de autoaprendizaje de la estadística».

Román Román, P., proman@ugr.es, Departamento de Estadística e I.O. Universidad de Granada, «Material interactivo para el aprendizaje del método Simplex Dual».

Susi García, R., rsusi@ucm.es, Universidad Complutense de Madrid, «Distribuciones de probabilidad con R».

Ruiz Castro, J. E., jeloy@ugr.es, Departamento de Estadística e I.O. Universidad de Granada, «Desarrollo metodológico y aspectos técnicos de vídeos tutoriales en APPES».

San Martín Pérez, M., montse.sanmartin@unirioja.es, Universidad de La Rioja, «Caso práctico de uso de la función Sweave para la elaboración de exámenes usando R y TeX».

Sánchez Borrego, I. R., Universidad de Granada, «*E-learning* para una estadística demográfica más autónoma y práctica».

Sanz Calvo, M. Á., msanzc@et.mde.es, MINISDEF/UNESOC, «Creación de exámenes *online* con LimeSurvey y obtención de resultados en informes dinámicos pdf (LaTeX + R) con Sweave».

Serrano Czaia, I., iserrano@uhu.es, Departamento de Métodos Cuantitativos para la Economía y la Empresa, Estadística e I.O., Facultad de Ciencias del Trabajo, Universidad de Huelva, «*Flipping* con estadística».

Simón de Blas, C., clara.simon@urjc.es, Departamento de Informática y Estadística, URJC, «Comparación de metodologías de evaluación continua».

Ugarte, M. D., lola@unavarra.es, Departamento de Estadística e I.O., Universidad Pública de Navarra, «Enseñando geoestadística con R».

Vargas Jiménez, M., Universidad de Granada, «*E-learning* para una estadística demográfica más autónoma y práctica».

Vera López, J. A., juanantonio.vera@ cud.upct.es, Centro Universitario de la Defensa, Academia General del Aire, Base Aérea de San Javier, Murcia, «Herramienta docente para la creación de actividades individuales con corrección inmediata».

## Participantes

Aguilar Cortés, Cristóbal; cagucor@ea.mde.es; Ejército del Aire.

Aguilera del Pino, Ana María; aaguiler@ugr.es; Univ. Granada.

Aguilera Morillo, M<sup>a</sup> del Carmen; maguiler@est-econ.uc3m.es; Univ. Carlos III Madrid.

Ahijado del Valle, Luis; lahidel@et.mde.es; Ejército de Tierra.

Alonso Ayuso, Antonio; antonio.alonso@urjc.es; Univ. Rey Juan Carlos (Madrid).

Alonso Sánchez, Alberto; aalosan@ea.mde.es; Ejército del Aire.

Álvarez Verdejo, Encarnación; encarniav@ugr.es; Univ. Granada.

Carrillo Ruiz, José Antonio; jcarrillor@et.mde.es; Vicesecretaría General Técnica.

Castellote Varona, Antonio; acastell@et.mde.es; Ejército de Tierra.

Daza Arbolí, Miguel Ángel; mdazarb@et.mde.es; Vicesecretaría General Técnica.

De Padura y Diéguez, Juan Antonio; jpaduray@et.mde.es; Ejército de Tierra.

Domínguez García-Gil, Juan Ángel; jdomingg@et.mde.es; Ejército de Tierra.

Fernández Militino, Ana; militino@unavarra.es; Univ. Pública Navarra.

Fernández Pérez, Carlos Manuel; cferper@ea.mde.es; Vicesecretaría General Técnica.

García Martín, Rafael; rgarciam@et.mde.es; Ejército de Tierra.

García-Ligero Ramírez, María Jesús; mjgarcia@ugr.es; Univ. Granada.

Gil Bellosta, Carlos; cgb@datanalytics.com; datanalytics.

Gombao Ferrándiz, Juan Carlos; jgombaof@et.mde.es; Ejército de Tierra.

González Alastrué, J.A.; jose.a.gonzalez@upc.edu; Univ. Politècnica de Catalunya.

Hernández Martín, Zenaida; zenaida.hernandez@unirioja.es; Univ. La Rioja.

Jiménez Pascual, Óscar; oscarjim@et.mde.es; Vicesecretaría General Técnica.

Lara Porras, Ana María; alara@ugr.es; Univ. Granada.

Lendínez Lendínez, Pedro Luis; plenlen@et.mde.es; Ejército de Tierra.

López Casado, Paula; paulalc\_76@hotmail.com; Univ. Valladolid (estudiante).

Lopo Márquez, María Ángeles; mlopma1@oc.mde.es; Vicesecretaría General Técnica.

Martín Segura, Reinaldo; rmartin@et.mde.es; SDG. Costes RR. HH.

Martínez Córdoba, Germán; gmarcor@et.mde.es; Vicesecretaría General Técnica.

Meca Martínez, Ana; ana.meca@umh.es; Univ. Miguel Hernández (Elche).

Montero Alonso, Miguel Ángel; mmontero@ugr.es; Univ. Granada (Melilla).

Montes Díez, Raquel; raquel.montes@urjc.es; Univ. Rey Juan Carlos (Madrid).

Muñoz Fernández, Ovidio; omunfer@et.mde.es; Vicesecretaría General Técnica.

Ortega Moreno, Mónica; ortegamo@uhu.es; Univ. de Huelva.

Pastuizaca Fernández, M<sup>a</sup> Nela; nelitmaria@gmail.com; Univ. Politécnica de Valencia.

Planells del Pozo, Antonio; aplanell@et.mde.es; Vicesecretaría General Técnica.

Ramos Méndez, Eduardo; eramos@ccia.uned.es; UNED.

Rivas Serra, Bartolomé; bribser@fn.mde.es; Vicesecretaría General Técnica.

Rivera Galicia, Luis Felipe; luisf.rivera@uah.es; Univ. de Alcalá de Henares.

Ródenas Púa, José Antonio; jrodenas@et.mde.es; Ejército de Tierra.

Román Román, Patricia; proman@ugr.es; Univ. Granada.

Sanz Calvo, Miguel Ángel; msanzc@et.mde.es; Vicesecretaría General Técnica.

Serrano Czaia, Isabel; iserrano@uhu.es; Univ. de Huelva.

Simón de Blas, Clara; clara.simon@urjc.es; Univ. Rey Juan Carlos (Madrid).

Susi García, Rosario; rsusi@ucm.es; Univ. Complutense Madrid.

Vázquez López, Rita; rvazlo1@oc.mde.es; SDG. TIC.

Vera López, Juan Antonio; juanantonio.vera@ cud.upct.es; CUD (AGA-UPCT).

Vinagre Ferrera, Fermín; fvinfer@oc.mde.es; Vicesecretaría General Técnica.