

Del 2 al 4 de Junio

Melilla
2010

I Jornadas
Enseñanza y **A**prendizaje
de la **E**stadística e
Investigación **O**perativa

GENAEIO - 2010

Programa
Comunicaciones



Organizan:



Escuela
Universitaria
de Ciencias
Sociales

UNIVERSIDAD DE GRANADA
Campus de Melilla

**Grupo de Enseñanza y
Aprendizaje de la Estadística e
Investigación Operativa**

GENAEIO

I JORNADAS DE ENSEÑANZA Y APRENDIZAJE DE LA ESTADÍSTICA E INVESTIGACIÓN OPERATIVA



Programa Comunicaciones

Junio 2-4, 2010

Melilla

© Ana María Lara Porras

Edita: Ana María Lara Porras

alara@ugr.es

ISBN: 978-84-614-7343-4

Dep. Legal: GR - 438-2011

Presentación

Durante los días 3 y 4 de junio de 2010 se celebrarán en Melilla las I Jornadas sobre la Enseñanza y Aprendizaje de la Estadística e Investigación Operativa que organiza el grupo de trabajo GENAEIO. Es para mi un honor, como coordinadora del grupo, dirigirme a todos los asistentes a estas jornadas que constituyen un foro para la reflexión, el debate, la formación, la información...

Las Jornadas configuran un punto de encuentro entre los profesores para el intercambio de experiencias, la innovación educativa y la reflexión sobre las distintas problemáticas que afectan a la enseñanza y aprendizaje de la EIO. Al mismo tiempo un lugar de encuentro donde abordar los significativos cambios que se están produciendo en la estructura de la Enseñanza Universitaria con la incorporación del Espacio Europeo de Educación Superior.

El presente libro incluye el programa y los trabajos presentados en las I Jornadas de la Enseñanza y Aprendizaje de la Estadística e Investigación Operativa y cuya organización ha sido encargada a la sección departamental del Departamento de Estadística e Investigación Operativa de la Universidad de Granada con sede en Melilla.

Como coordinadora del grupo GENAEIO deseo agradecer la presencia a todos los participantes en las jornadas y felicitar a los ponentes por los trabajos realizados.

Ana María Lara Porras
Coordinadora del grupo de trabajo GENAEIO

Convoca:

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

Organizan:

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

**Escuela Universitaria de Ciencias Sociales
Campus de Melilla. Universidad De Granada**

Colaboran:

Ciudad Autónoma de Melilla

**Vicerrectorado para la
Garantía de la Calidad de la Universidad de Granada**

Consejería de Educación de la Ciudad Autónoma de Melilla

Comité Científico

Ana María Lara Porras
Isabel Quesada Vázquez
Miguel Ángel Montero Alonso

Comité Organizador

Isabel Quesada Vázquez
Ana María Lara Porras
Miguel Ángel Montero Alonso
Juan Antonio Marmolejo Martín

Índice

| | |
|--|------------|
| Programación General | 8 |
| Programación detallada..... | 9 |
| Talleres: | |
| Taller de R (1^a Parte): “Introducción y primeros pasos”..... | 11 |
| Taller de R (2^a Parte): “Estadística Descriptiva Unidimensional y Bidimensional.” | 37 |
| Taller de R (2^a Parte): “Distribuciones e Inferencia”..... | 54 |
| Taller de “Diseño de Herramientas de Autor”..... | 78 |
| Taller de “ Nuevas plataformas virtuales para la docencia” | 95 |
| Taller de “Experiencia de la Enseñanza de la Estadística en las nuevas titulaciones adaptadas al EEES”..... | 108 |
| Autores..... | 121 |
| Participantes..... | 121 |

Programación general

| | Jueves 3 | Viernes 4 |
|--------------|---|--|
| 9:00 | Entrega de documentación | |
| 9:30 | Inauguración | |
| 10:00 | Taller R (1 ^a parte) | Taller de “Diseño de Herramientas de Autor” |
| 11:30 | Café-descanso | Café-descanso |
| 12:00 | Taller R (2 ^a parte) | Taller de “Nuevas plataformas virtuales para la docencia” |
| 14:00 | | |
| 15:00 | | |
| 16:00 | | |
| 17:00 | Visita guiada a la Ciudad Autónoma de Melilla | Taller de “Experiencia de la enseñanza de la Estadística en las nuevas titulaciones adaptadas al EEES” |
| 18:30 | | Café-descanso |
| 19:00 | | Reunión del Grupo GENAEIO |
| 20:00 | | Clausura de las Jornadas |
| 22:00 | | Cena de Clausura |

Programación detallada

Jueves 3 de junio 2010

9:00-9:30. Entrega de documentación.

9:30-10:00. Inauguración de las Jornadas.

Intervienen:

D. Sebastian Sánchez Fernández

Delegado del Rector de la Universidad de Granada.

D. Alejandro Del Canto Bossini

Director de la Escuela Universitaria de Ciencias Sociales de Melilla.

D. Antonio Miranda Montilla

Excmo. Consejero de Educación de la Ciudad Autónoma de Melilla.

Dª. Isabel Quesada Vázquez

Directora de la Sección Departamental de Estadística e Investigación Operativa de la Universidad de Granada en Melilla.

Dª. Ana Mª Lara Porras

Coordinadora del Grupo GENAEIO.

Lugar:

Salón de Grados de la Escuela Universitaria de Ciencias Sociales.

10:00-11:30. Taller de R (1ª Parte): “Introducción y primeros pasos”

Realizan:

D. Luis M. Marín Trechera (Universidad de Cádiz)

D. Antonio Gámez Mellado (Universidad de Cádiz)

11:30-12:00. Café-descanso

12:00-14:00. Taller de R (2ª Parte): “Estadística Descriptiva Unidimensional y Bidimensional. Distribuciones e Inferencia”

Realizan:

D. Luis M. Marín Trechera (Universidad de Cádiz)

D. Antonio Gámez Mellado (Universidad de Cádiz)

14:00. Comida

17:00. Visita guiada a la Ciudad Autónoma de Melilla.

Viernes 4 de junio 2010

10:00-11:30. Taller de “Diseño de Herramientas de Autor”

Realizan:

- D. Luis M. Marín Trechera (Universidad de Cádiz)
D. Antonio Gámez Mellado (Universidad de Cádiz)

11:30_12:00. Café-descanso

12:00-14:00. Taller de “Nuevas plataformas virtuales para la docencia”

Realiza:

- D. Miguel Ángel Daza Arbolí
Centro de Investigación Militar Operativa (Ministerio de Defensa)

14:00. Comida

17:00-18:30 h. Taller de “Experiencia de la Enseñanza de la Estadística en las nuevas titulaciones adaptadas al EEES”

Realiza:

- D. Luis F. Rivera Galicia (Universidad de Alcalá)

18:30-19:00. Café-descanso

19:00. Reunión del Grupo GENAEIO

20:00. Clausura de las Jornadas

22:00. Cena de Clausura

Taller de R

(Introducción y primeros pasos)

Antonio Gámez Mellado¹, Luis Miguel Marín Trechera²

¹antonio.gamez@uca.es, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz

²luis.marin@uca.es, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz
Universidad de Cádiz

Resumen

Esta primera parte del Taller está orientado a aprender a usar R para realizar las tareas estadísticas básicas y, aunque en menor grado, también las de matemáticas. Se ha preparado de forma que sea aplicable en la labor docente del profesor universitario. Una de las ventajas de usar R es su gratuidad, no requiere de licencia al tratarse de software libre, lo que facilita su uso en las aulas de informática de los centros de enseñanza universitaria, así como en los ordenadores personales de los alumnos. Además, y de cara a los centros de enseñanza universitarios, es importante saber que funciona bajo LINUX, aunque también puede funcionar bajo Windows y Macintosh. Además la documentación sobre R está disponible en la red, por lo que tampoco es necesario realizar ningún gasto en material bibliográfico. Es verdad que la mayor parte de la bibliografía está en inglés, pero hay suficiente material en español para poder documentarse. Algunos de estos documentos figuran en la bibliografía del Taller.

Además de la gran variedad de funciones construidas en R, el hecho de que sea programable lo dota de la posibilidad de introducir funciones personalizadas, por lo que presenta una gran adaptabilidad a diferentes campos de estudio. En particular, una de las características de R más interesante es la calidad y facilidad de creación de una gran gama de representaciones gráficas.

Este software estadístico se distribuye bajo la licencia GNU GPL (Licencia Pública General) que es una licencia creada por la Free Software Foundation en los años 80. Esta licencia permite copiar y redistribuir el software, bajo ciertas condiciones, cuyo objetivo es permitir la libre distribución, modificación y uso del software, evitando los intentos de apropiación si no se trasmiten estas mismas libertades a los usuarios.

En esta primera parte del taller abordaremos cuatro bloques de contenidos: La instalación, los primeros pasos con R, el manejo de datos y gráficos, la estadística descriptiva unidimensional y la estadística descriptiva bidimensional.

Palabras clave: Software libre, R, Estadística

Clasificación AMS: 97U50, 97K40, 97K80

1. Introducción. Instalación de R y R-Commander

R puede considerarse como una nueva implementación del lenguaje S, que se desarrolló en 1980 en los laboratorios AT&T Bell. R se desarrolló después de S y S-Plus. Puede encontrarse información sobre S en:

<http://www.insightful.com>

Muchos de los libros y manuales sobre S son útiles para R. El proyecto R fue comenzado en 1995 por Robert Gentleman y Ross Ihaka, del departamento de Estadística de la Universidad de Auckland. La página web oficial del proyecto R es:

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

En esta página, que contiene una gran cantidad de información sobre R, está disponible el software de R para cargarlo en nuestro ordenador, además de otros paquetes complementarios de este software, así como manuales de empleo y diversa documentación.

Aunque R es principalmente usado en Estadística también puede emplearse para otros fines. En el entorno R se hallan implementadas las técnicas estadísticas más usuales. Las más específicas pueden incluirse en forma de biblioteca o paquetes complementarios.

Para instalar el programa R en nuestro ordenador entramos en la página del proyecto R (<http://www.r-project.org>) y desde allí, tras seleccionar el enlace CRAN podemos seleccionar cualquiera de los “mirrors” que se nos ofrecen, por ejemplo el de Argentina, que es el primero que aparece en la lista: <http://mirror.cricyt.edu.ar/r/>

o mejor el de España: <http://cran.es.r-project.org>

En cualquiera de estos sitios de descarga se hallan las instrucciones para instalar R en diversos sistemas operativos.

Para el sistema Windows seguimos los siguientes pasos: En el apartado Download and Install R pulsamos sobre el enlace Windows y luego pulsamos sobre base. En la pantalla siguiente tenemos que descargar el archivo R-2.12.1-win32.exe en un directorio de nuestro ordenador (ver figura 1).

Una vez descargado el archivo anterior, debemos entrar en el directorio y, pulsando sobre el ícono del archivo citado, comenzaremos la instalación de R. Al ejecutar el archivo se instalará el sistema base y los paquetes recomendados. Es conveniente seleccionar todas las opciones que se nos proponen por defecto. ¡Tenga paciencia! La descarga puede requerir bastante tiempo.

Para ejecutarlo se debe pulsar el ícono del programa que aparece en el escritorio, o ir a inicio → archivos de programa → R, para pulsar el ícono correspondiente a este programa. La Ventana de ejecución de R (RGui), presenta el aspecto de la figura 2. Para escribir en dicha ventana, que deberá estar activa, hay que tener en cuenta que R distingue mayúsculas y minúsculas. Teclearemos en primer lugar en la línea de ejecución, que está marcada con el signo > en rojo la orden help(). Una vez introducida la orden **help()** se pulsa la tecla **Enter**. Se muestra on-line la documentación de ayuda de R. A la izquierda aparecen los comandos que se pueden utilizar. Si pulsamos sobre cualquiera de ellos aparece su descripción. Como se puede ver, toda esta ayuda está en inglés.

Para buscar ayuda sobre un comando (por ejemplo mean, que es la palabra en inglés para designar la media) tecleamos en la línea de ejecución, marcada con >:

> help(mean)

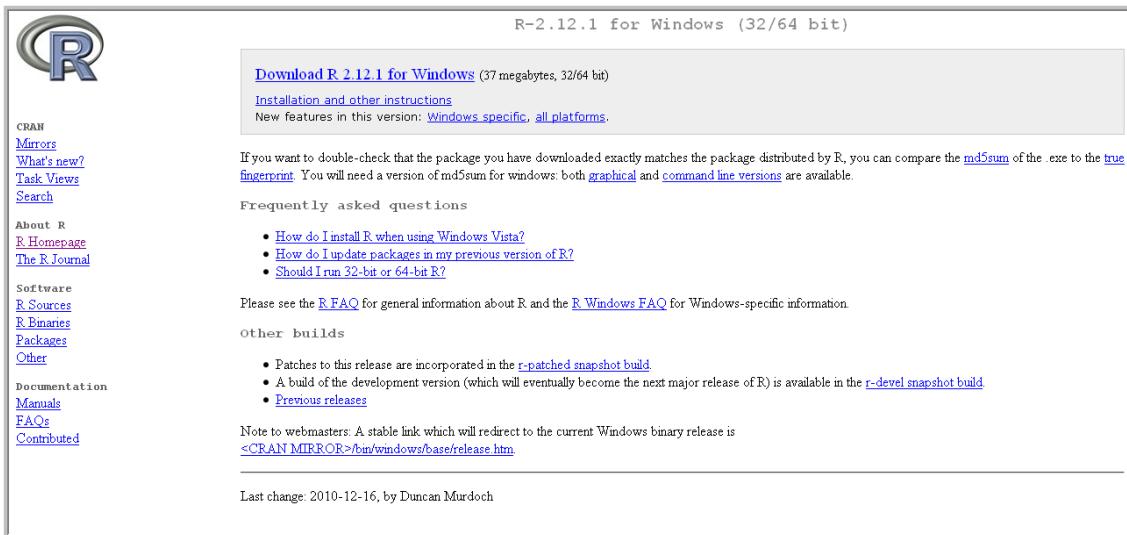


Figura 1: Descarga del programa R

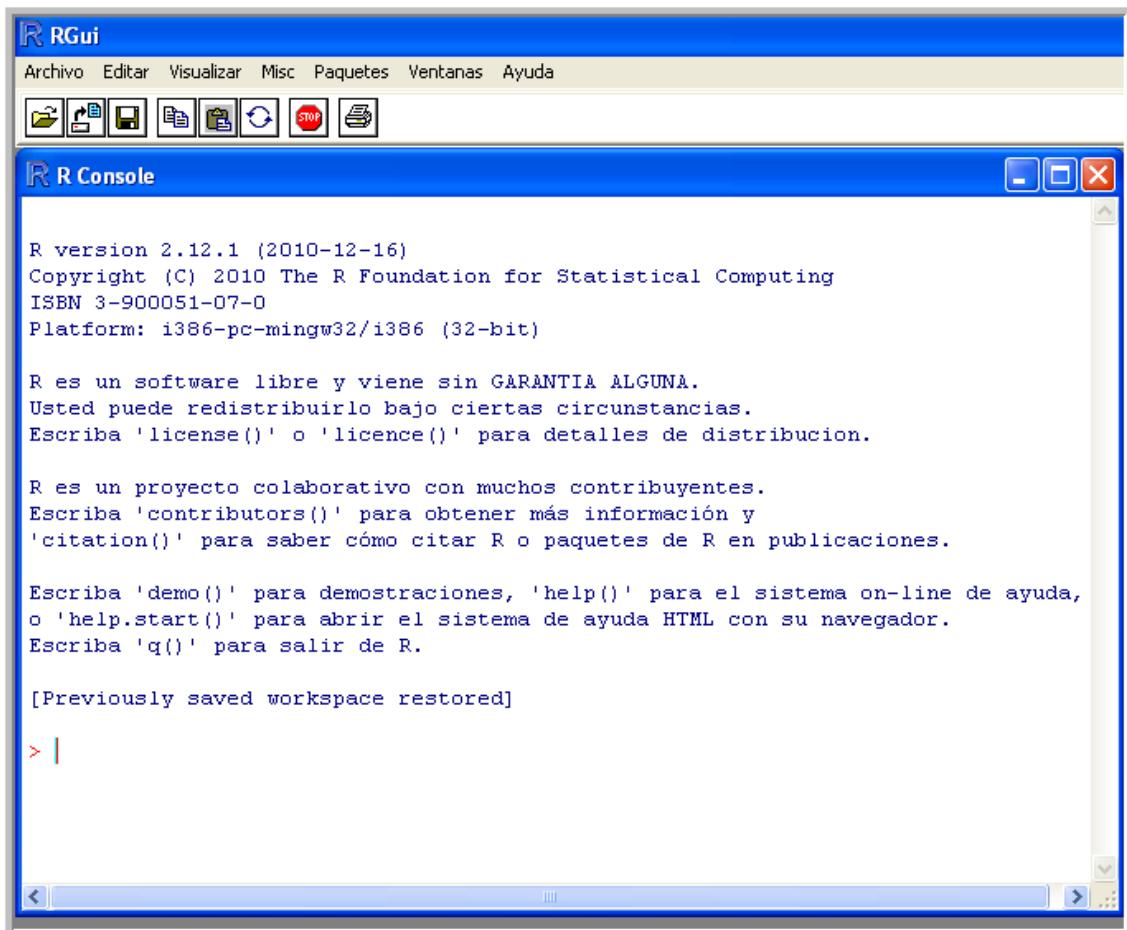


Figura 2: Ventana inicial de ejecución de R

1.1. Primeros pasos con R

A continuación vamos a realizar algunas tareas útiles para comenzar a trabajar con R. Esta presentación inicial tiene como objeto familiarizarse con la programación en R, pero creemos que es más conveniente que los alumnos lleven a cabo estas tareas con el interface gráfico de usuario R-Commander, tal como se explica en la última sección de este tema.

Lo primero que necesitamos es disponer de unos datos. Usamos el fichero Encues.Rdata, de los recursos de este primer bloque. Para ello lo cargamos en R de la forma que se describe a continuación:

- 1) Primero hay que guardar el fichero tema1.Rdata, que contiene dicho fichero de datos, en el directorio de trabajo del programa (si se ha seguido la instalación por defecto este directorio es C:\Archivos de programa\R\R-2.12.1).
- 2) Despues, y desde R, cargaremos este fichero como área de trabajo seleccionando en el menú:
 - Archivo->Cargar área de trabajo-> tema1.Rdata (ver figura 3).
- 3) Para cerciorarnos de que está bien cargado el archivo tecleamos en la línea de ejecución:
 - encues

Nos deben aparecer los datos del fichero que aparecen en la figura 4. Este archivo contiene los resultados de una encuesta en la que se les preguntaba a un grupo de 50 personas su altura en metros, su nivel de estudios, el número de personas que conviven en su domicilio habitual y su peso en kg. Conviene que cerremos la ventana de los datos cuando sigamos trabajando, pues a veces hay problemas por dejar abiertas las ventanas que van surgiendo.

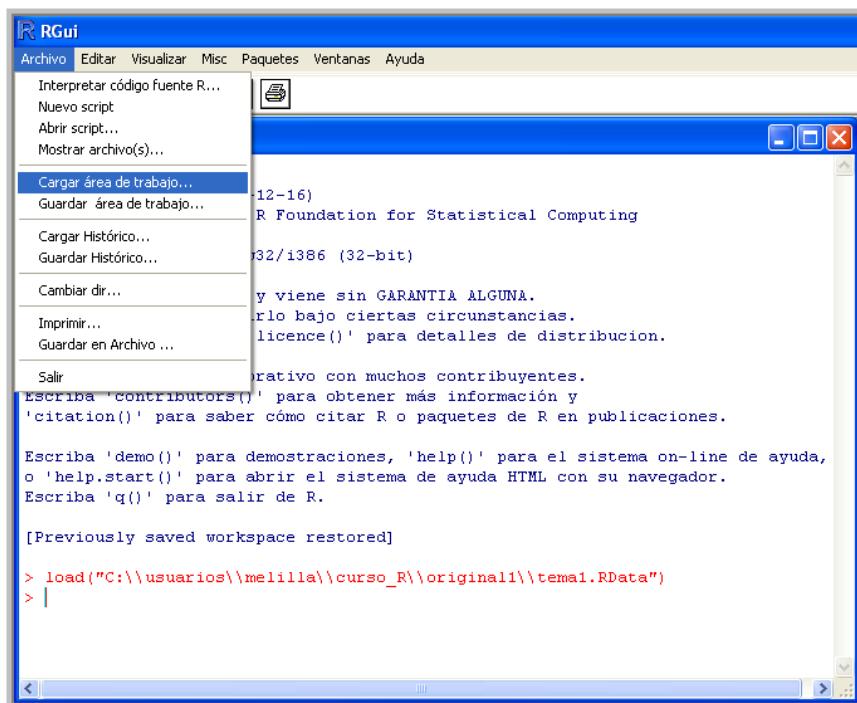
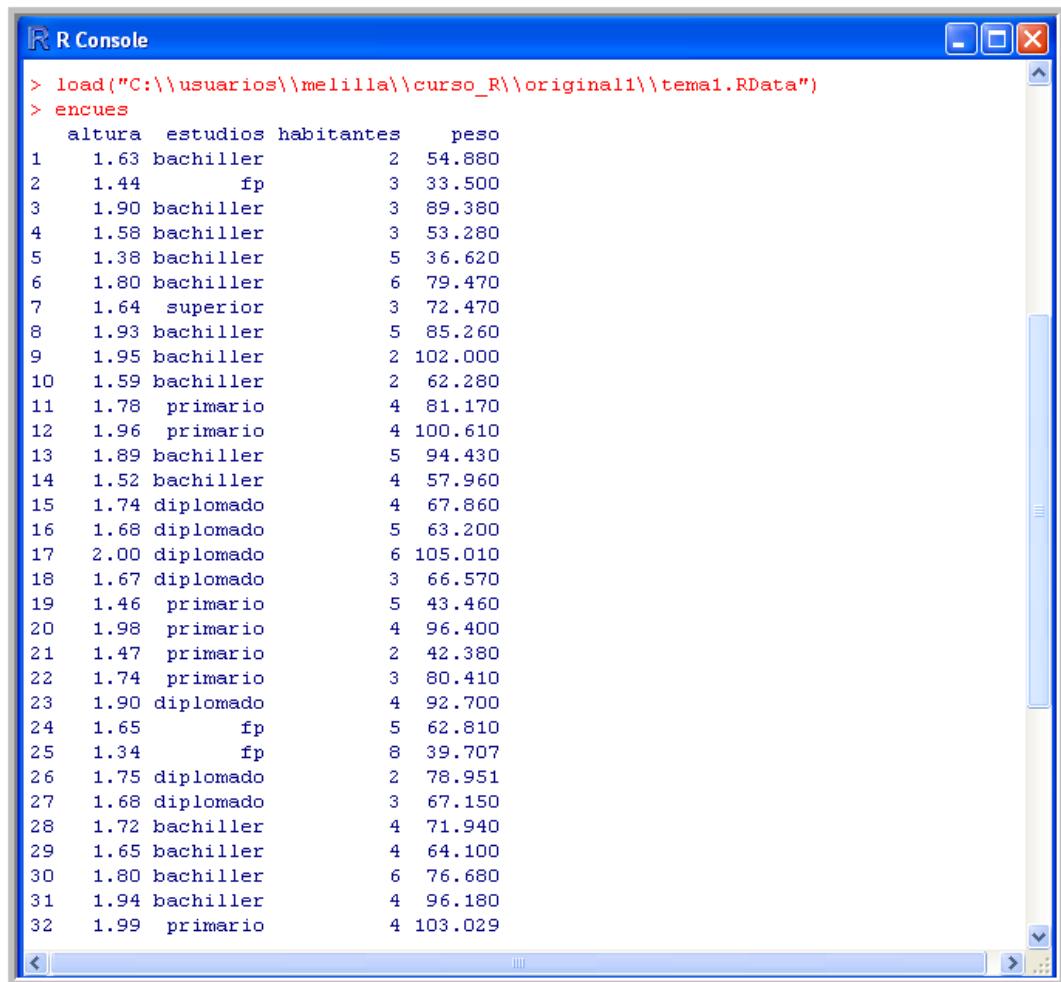


Figura 3: Cargar área de trabajo desde RGui.



The screenshot shows the R Console window with the title 'R Console'. The console displays the following R code and its output:

```

> load("C:\\\\usuarios\\\\melilla\\\\curso_R\\\\original1\\\\tema1.RData")
> encues
  altura estudios habitantes    peso
1   1.63  bachiller          2 54.880
2   1.44      fp            3 33.500
3   1.90  bachiller          3 89.380
4   1.58  bachiller          3 53.280
5   1.38  bachiller          5 36.620
6   1.80  bachiller          6 79.470
7   1.64  superior          3 72.470
8   1.93  bachiller          5 85.260
9   1.95  bachiller          2 102.000
10  1.59  bachiller          2 62.280
11  1.78  primario          4 81.170
12  1.96  primario          4 100.610
13  1.89  bachiller          5 94.430
14  1.52  bachiller          4 57.960
15  1.74  diplomado          4 67.860
16  1.68  diplomado          5 63.200
17  2.00  diplomado          6 105.010
18  1.67  diplomado          3 66.570
19  1.46  primario          5 43.460
20  1.98  primario          4 96.400
21  1.47  primario          2 42.380
22  1.74  primario          3 80.410
23  1.90  diplomado          4 92.700
24  1.65      fp            5 62.810
25  1.34      fp            8 39.707
26  1.75  diplomado          2 78.951
27  1.68  diplomado          3 67.150
28  1.72  bachiller          4 71.940
29  1.65  bachiller          4 64.100
30  1.80  bachiller          6 76.680
31  1.94  bachiller          4 96.180
32  1.99  primario          4 103.029

```

Figura 4: Datos del archivo “encues” en la Consola de R.

Para obtener la tabla de frecuencias de la variable estudios y después realizar el diagrama de barras de la variable estudios tecleamos en la línea de ejecución:

```
> table(encues$estudios)
```

La expresión encues\$estudios designa la variable estudios del fichero encues. No es suficiente con el nombre de la variable, ya que puede haber una variable distinta, pero con el mismo nombre en otro fichero de datos. De esta forma obtenemos la tabla de frecuencias. Si tecleamos la orden barplot(table(encues\$estudios)) se obtiene el diagrama de barras de la variable estudios, como puede verse en la figura 5.

En esta gráfica aparecen en el eje vertical las frecuencias absolutas. Si preferimos que aparezcan las frecuencias relativas la orden que hay que ejecutar es:

```
> barplot(table(encues$estudios)/length(encues$estudios))
```

Para que las barras aparezcan en azul la orden debería ser

```
> barplot(table(encues$estudios),col="blue")
```

Conviene cerrar las ventanas gráficas cuando no sean necesarias.

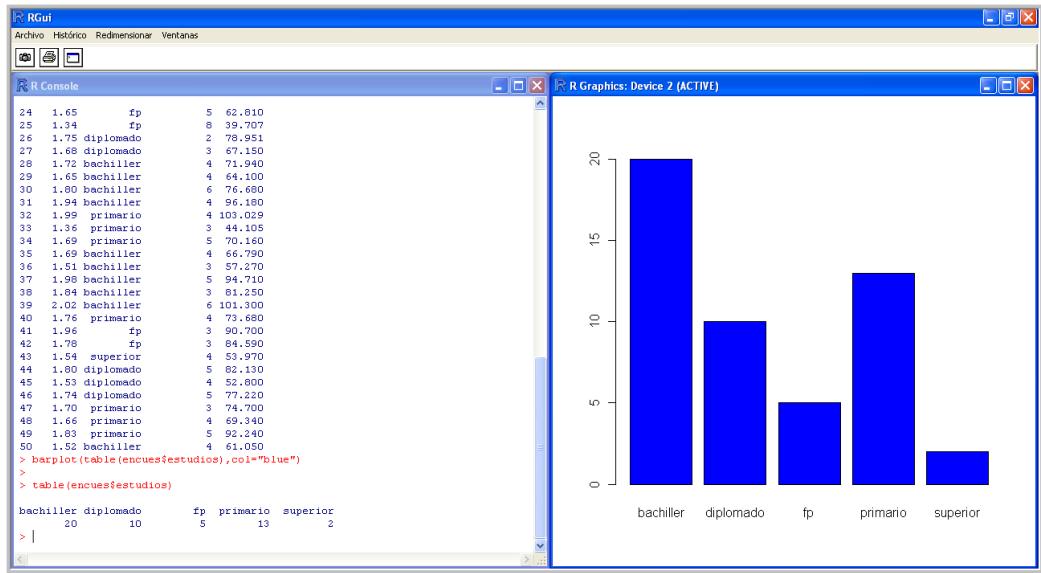


Figura 5: Diagrama de barras de la variable “estudios” del archivo “encues”.

Consideraremos la recta de regresión de la variable peso (dependiente) respecto de la variable altura (independiente):

Para calcular esta recta de regresión tecleamos sucesivamente las siguientes órdenes. No es necesario escribir el comentario que sigue a cada una de ellas. El resultado se ve en la figura 6.

```

> x=encues$altura #asigna el nombre x a la variable altura
> y=encues$peso #asigna el nombre y a la variable peso
> lm(y~x) #calcula los coeficientes de la recta de regresión
> plot(x,y) #dibuja la nube de puntos (x,y)
> abline(-99.26,100.08) #añade la recta de regresión a la nube de puntos

```

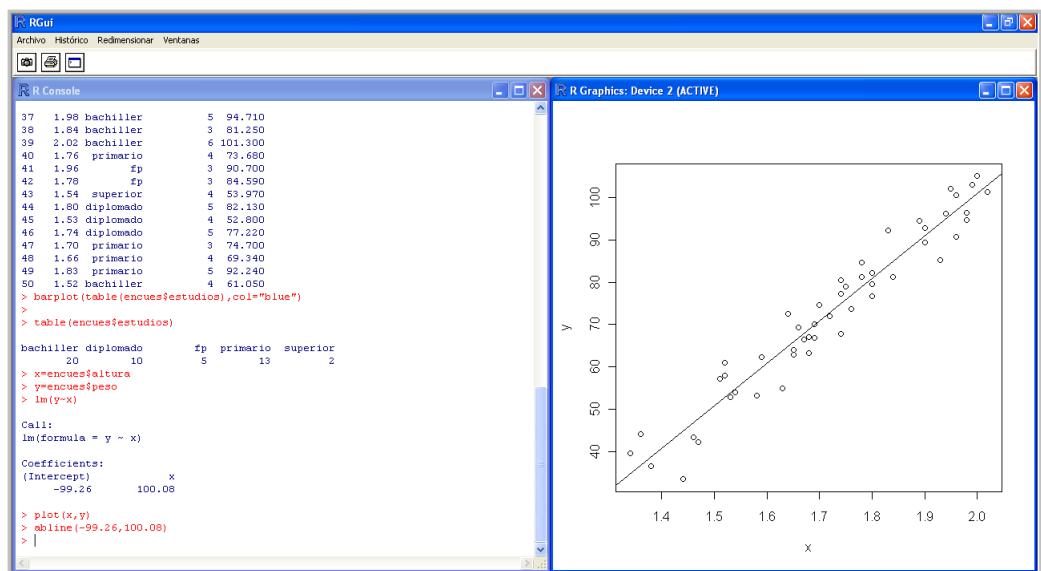


Figura 6: Recta de regresión de la variable peso con respecto a la altura del archivo “encues”.

1.2. Instalación de R-Commander

R-Commander es un manejador de R, que permite realizar muchas de las labores de R, sin necesidad de recordar las órdenes o instrucciones concretas del lenguaje de programación R.

R-Commander se instala por medio de un paquete (en inglés package). El nombre de dicho paquete es Rcmdr y no viene incluido por defecto en R. Para instalarlo es preciso cargar este paquete de la red. Para ello, y dentro de la ventana de ejecución de R, seleccionamos en el menú superior:

Paquetes→ Instalar Paquete(s)

Con esto se nos presentará una lista de “mirrors” desde los que podemos instalar el paquete que seleccionemos. Por ejemplo podemos elegir el primer mirror, Argentina, o mejor el de Spain (Madrid). Tras pulsar aceptar, aparece una lista de paquetes. Tenemos que seleccionar Rcmdr. Cuando se haya cargado nos faltarán aún más paquetes para la completa instalación de Rcmdr, pero se podrán instalar a continuación. A veces el proceso de descarga es algo lento.

1.3. Ejecución de R-Commander

Para ejecutar el paquete instalado hace falta activarlo. Para ello hay que pulsar, desde el menú de R las siguientes órdenes:

Paquetes→ Cargar paquetes.

Se debe seleccionar Rcmdr en la ventana que se despliega. Si este procedimiento presentará alguna dificultad o no funcionara, puede realizarse la misma operación tecleando, en la ventana de comandos de R, la orden library(Rcmdr). Cuando en una sesión de R se ha cerrado R Commander, pero no RGui, y se quiere reabrir R Commander, lo adecuado es usar Commander().

Tras realizar esta tarea nos aparecerá la ventana de ejecución de R. La pantalla presentará un aspecto similar al de la figura 7. Es frecuente que la ventana de R-Commander quede oculta por las restantes ventanas que tengamos abiertas. En esas ocasiones no hay más remedio, para poder seguir trabajando, que abrirla de nuevo desde la barra de tareas de Windows. Puede bastar con pulsar el icono de ventana en R, para ver ambas simultáneamente.

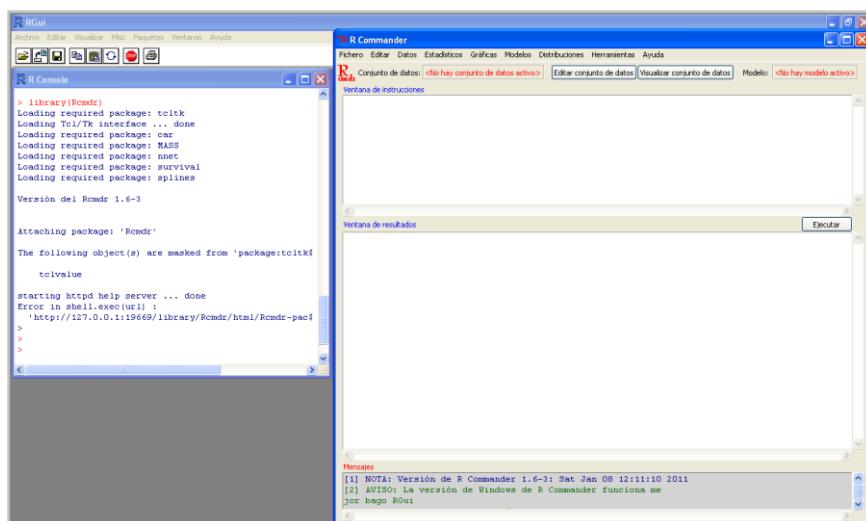


Figura 7: Ventana de inicio de R-Commander.

Ahora ya estamos preparados para trabajar con R-Commander.

El acceso a las funciones de R-Commander se realiza seleccionando, en el menú superior, con el ratón la opción que necesitemos para nuestro trabajo. Los elementos que se pueden elegir son: Fichero, Editar, datos, Estadísticos, etc. En la sección siguiente, así como en el resto del Taller iremos usando estas opciones.

La ventana de R-Commander se divide en tres partes, ventana de instrucciones, ventana de resultados y ventana de mensajes. La ventana superior es la de instrucciones, allí aparecen automáticamente las órdenes que se están ejecutando con R-Commander. En la ventana de resultados se mostrarán estas órdenes y además los resultados que se van obteniendo. Es decir, R-Commander muestra también, en lenguaje de R, las órdenes cuya ejecución se ha solicitado con el sistema de menús de R-Commander.

Además se pueden introducir, manualmente, las órdenes del lenguaje R. Para ello se selecciona dicha orden, simplemente dejando el cursor sobre ella y pulsando el botón Ejecutar, que aparece entre ambas ventanas. La salida de la orden aparecerá en la ventana inferior (la de resultados). Es de destacar que podemos seleccionar, para ejecutarla con el botón Ejecutar, cualquier orden de las que aparecen en la ventana de instrucciones y no solamente la última, como ocurría en RGui. También es posible realizar esta función pulsando el botón derecho del ratón y pulsando Ejecutar en el menú emergente que se ha abierto.

1.4. Ejemplos de uso de R-Commander

Vamos a realizar, a modo de ejemplo, algunas de las tareas que ya realizamos con R. Lo primero que debemos hacer es cargar los datos que vamos a emplear. Para ello pulsamos en:

Datos→Conjunto de datos activo → Seleccionar conjunto de datos activo.

Elegimos “encues”. Este nombre deberá aparecer como conjunto de datos activo. Para que en un momento dado R pueda actuar sobre un conjunto de datos, estos deben estar incluidos en este fichero activo.

Una vez cargado el fichero que vamos a emplear, si deseamos comprobar su contenido podemos pulsar el botón visualizar conjunto de datos. Es conveniente ir cerrando las ventanas que no necesitamos, así que una vez que hayamos visto los datos se debe cerrar la ventana de los datos. La tabla de frecuencias de la variable “estudios” se realiza ahora pulsando:

Estadísticos→ Resúmenes→ Distribución de frecuencia.

Con ello se nos muestra una ventana para seleccionar la variable que deseemos. Como la única variable cualitativa en nuestros datos es estudios, es la única que se nos muestra. Tras pulsar aceptar, surge la ventana de la figura 8. A veces la ventana de R-Commander se oculta debajo de las demás. Si ocurre esto, buscarla en la barra de herramientas de Windows. Se puede apreciar que muestra las órdenes de R en la ventana de instrucciones. Los resultados en la parte inferior. Estos resultados están expresados en frecuencia absoluta y en porcentaje. Observar lo que aparece en las ventanas de instrucciones y de resultados.

Para realizar el diagrama de barras de la variable estudios, seleccionamos:

Gráficas→ Gráfica de Barras

El diagrama de barras que se obtiene es similar al que se puede ver en la figura 5. Ahora tenemos ocasión de manipular las órdenes que han ido apareciendo en la ventana de instrucciones de R-Commander. Si queremos que las barras aparezcan en color azul, añadimos

dentro de la última orden que aparece en la ventana de instrucciones el argumento col="blue", obteniendo:

```
barplot(table(encues$estudios),xlab="estudios",ylab="Frequency",col="blue")
```

Manteniendo el cursor sobre esta orden pulsamos el botón ejecutar. Con esto obtenemos un diagrama de barras con las barras en color azul, como el de la figura 5.

Las gráficas aparecen siempre en la ventana de ejecución de RGui. A veces no se ven inmediatamente, bien porque quedan ocultadas por otras ventanas abiertas, bien porque RGui no se ha ejecutado a pantalla completa.

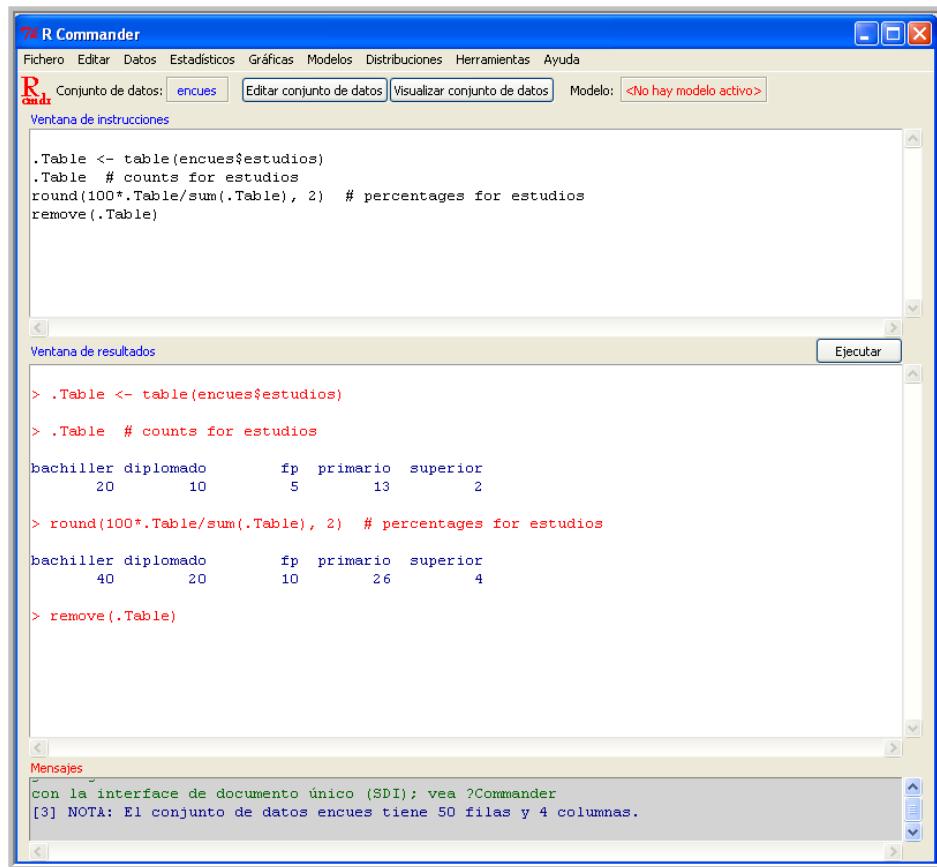


Figura 8: Ventana de R-Commander. Tabla de frecuencias de la variable “estudios”.

Vamos a representar con R-Commander la recta de regresión de la variable peso (dependiente) respecto de la variable altura (independiente). Para calcular la recta de regresión seleccionamos:

Estadísticos-> Ajuste de Modelos->Regresión lineal

Nos aparece la ventana de la figura 9, donde están seleccionadas las variables necesarias.

Aparecen los resultados que pueden verse en la figura 10, donde se incluye el análisis de la regresión. Hemos resaltado en negrilla los coeficientes de la recta de regresión y también el coeficiente de Correlación lineal.



Figura 9: Ventana de selección de variables para la regresión lineal.

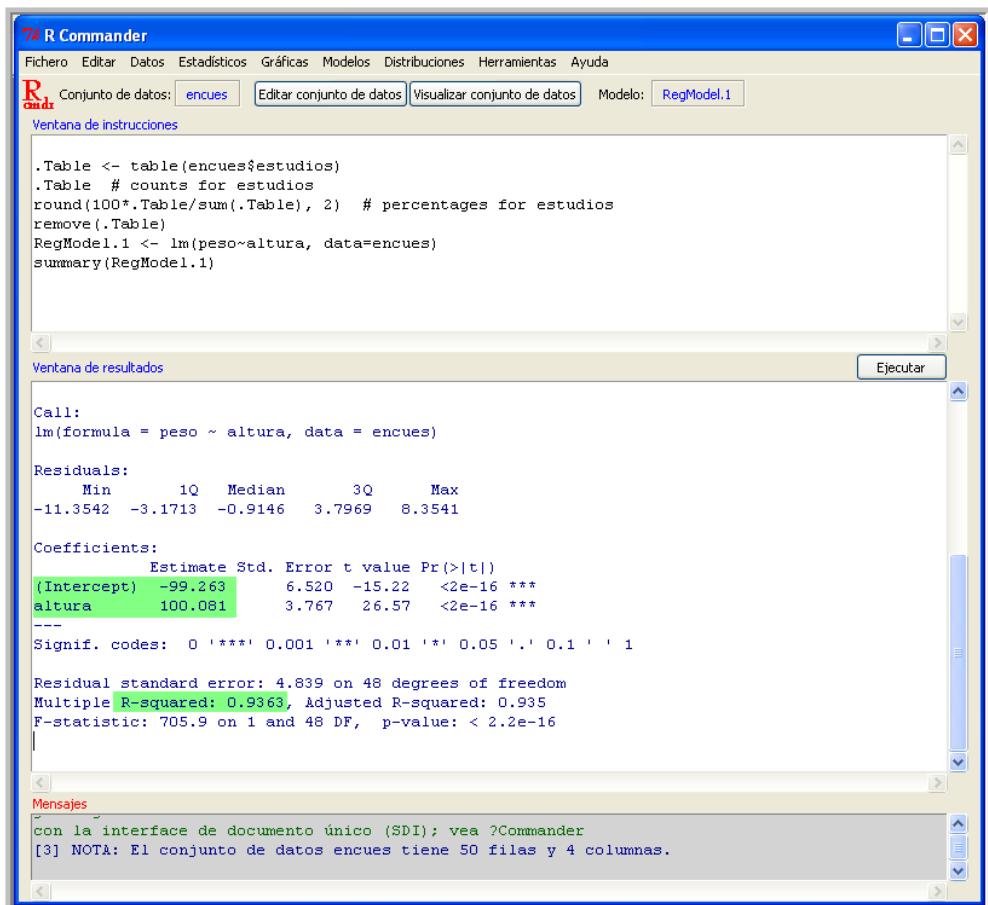


Figura 10: Regresión lineal de la variable peso frente a la variable altura del archivo “encues”.

Por lo tanto la recta de regresión resulta: $\text{peso} = -99.263 + 100.081 * \text{altura}$.

Como el coeficiente de determinación es $R^2 = 0.9363$, bastante cercano a 1, concluimos que la recta hallada ajusta bastante bien la nube de puntos (x, y).

En cuánto a la representación gráfica de los datos y de la recta de regresión, similar a la de la figura 6, puede obtenerse pulsando en los menús de R-Commander:

Gráficas → Diagrama de Dispersión.

2. Manejo de datos y representaciones gráficas

Los datos son la materia prima de cualquier análisis estadístico. Para manejarlos, con R o con R-Commander, es necesario que estén registrados en un fichero con el formato propio de R. En la primera parte de esta sección trataremos tres formas posibles de realizar esta operación con R-Commander: introducir directamente los datos, importarlos desde un fichero externo y utilizar ficheros de datos ya incluidos en un paquete de R.

2.1. Introducción directa de datos

Esta opción suele emplearse cuando el número de datos que queremos introducir no es demasiado grande. Para ilustrar la forma de introducir directamente datos en R-Commander editaremos un fichero con los valores de las cuatro primeras líneas del fichero encues, que hemos manejado en la sección anterior.

Los valores son los que aparecen en la figura 4. Para realizar esta operación, desde el menú superior de R-Commander, pulsamos Datos → Nuevo Conjunto de Datos. Se nos pide el nombre del fichero donde queremos guardar los datos. Introducimos el nombre prueba. Se abre una ventana como la de la figura 11 y se introducen los datos que necesiten.

Pulsamos en el nombre de la primera variable, que se llama de momento var1. Escribimos en su lugar altura y seleccionamos el tipo numeric, ya que la variable altura contiene números. Pulsar la tecla Enter.

Cuando hayamos terminado la introducción de los cuatro primeros datos de esta variable, continuamos con la variable estudios. Pulsamos en esta ocasión sobre el nombre de la segunda variable, ahora var2, y la cambiamos por el nombre de nuestra segunda variable, estudios. Seguimos los mismos pasos que en el caso anterior, con la única diferencia de que, como la variable estudios es cualitativa, no podemos seleccionar la opción numeric, sino que elegimos la opción character, que es la adecuada para este tipo de variables. Cuando acabemos de introducir los cuatro primeros datos de esta variable, el editor de datos presentará el aspecto de la figura 11.

| R Editor de datos | | | | | |
|-------------------|--------|---------------|------|------|------|
| | altura | estudios | var3 | var4 | var5 |
| 1 | 1.54 | bachiller | | | |
| 2 | 1.75 | universitario | | | |
| 3 | 1.84 | fp | | | |
| 4 | 1.77 | superior | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | |
| 19 | | | | | |

Figura 11: Editor de datos de R-Commander.

Ahora vamos a guardar los datos que ya hemos introducido. Para ello pulsamos Archivo → Cerrar (también puede hacerse cerrando la ventana de los datos). Con ello nos aparece el fichero prueba como conjunto de datos activo. Si queremos seguir añadiendo las dos columnas que quedan, habitantes y peso, debemos pulsar el botón Editar Conjunto de Datos para que aparezca de nuevo la ventana que acabamos de cerrar, donde podemos seguir introduciendo los datos restantes hasta completar nuestra tarea. Cuando acabemos guardamos el fichero completo pulsando de nuevo Archivo→ Cerrar, o cerrando la ventana.

Esta operación de edición (pulsando Editar Conjunto de Datos) es válida para modificar cualquier fichero de datos que tengamos en ese momento disponible en nuestra sesión de trabajo. Para saber cuales son esos ficheros podemos pulsar en el botón en el que aparece el nombre del fichero de datos activos. Si no hay fichero de datos activo pulsar en el botón que aparece en este mismo lugar, pero que contiene la frase No hay archivo de datos activo. Cuando aparezcan los nombres de estos ficheros seleccionamos el que queramos usar.

Hay distintas formas de manipular los datos de un fichero para modificarlos o crear nuevas variables a partir de ellos. Si entramos en Datos → Modificar variables del conjunto de datos activo, se despliega una ventana donde se pueden seleccionar diversas opciones, que iremos utilizando a lo largo del taller conforme vayan siendo necesarias para nuestro trabajo.

De momento, y para mostrar un ejemplo, vamos a seleccionar el fichero encues, si no estuviera ya seleccionado, como conjunto de datos activo. Posteriormente seleccionamos desde el menú de R Commander Datos → Modificar variables del conjunto de datos activo → Segmentar variable numérica. Esta opción puede ser interesante para construir una tabla de frecuencias de una variable numérica, por ejemplo la variable altura. En la ventana que se despliega rellenamos las opciones tal como aparecen en la figura 12. Con ello se creará la variable clasesaltura, que presentará el aspecto que puede verse en la figura 13.



Figura 12: Segmentar una variable.

En R-Commander no hay ningún procedimiento específico, desde el menú, para construir la tabla de frecuencias para una variable numérica como altura. Pero la variable clasesaltura, que acabamos de crear, nos suministra una forma de hacerlo, pues ha asignado cada elemento de la variable altura a una clase de las que resultan al dividir todos los datos en cinco intervalos de igual amplitud.

| | altura | estudios | habitantes | peso | clasealtura |
|----|--------|-----------|------------|---------|-------------|
| 1 | 1.63 | bachiller | 2 | 54.880 | (1.61,1.75] |
| 2 | 1.44 | fp | 3 | 33.500 | (1.34,1.48] |
| 3 | 1.90 | bachiller | 3 | 89.380 | (1.88,2.02] |
| 4 | 1.58 | bachiller | 3 | 53.280 | (1.48,1.61] |
| 5 | 1.38 | bachiller | 5 | 36.620 | (1.34,1.48] |
| 6 | 1.80 | bachiller | 6 | 79.470 | (1.75,1.88] |
| 7 | 1.64 | superior | 3 | 72.470 | (1.61,1.75] |
| 8 | 1.93 | bachiller | 5 | 85.260 | (1.88,2.02] |
| 9 | 1.95 | bachiller | 2 | 102.000 | (1.88,2.02] |
| 10 | 1.59 | bachiller | 2 | 62.280 | (1.48,1.61] |
| 11 | 1.78 | primario | 4 | 81.170 | (1.75,1.88] |
| 12 | 1.96 | primario | 4 | 100.610 | (1.88,2.02] |
| 13 | 1.89 | bachiller | 5 | 94.430 | (1.88,2.02] |
| 14 | 1.52 | bachiller | 4 | 57.960 | (1.48,1.61] |
| 15 | 1.74 | diplomado | 4 | 67.860 | (1.61,1.75] |
| 16 | 1.68 | diplomado | 5 | 63.200 | (1.61,1.75] |
| 17 | 2.00 | diplomado | 6 | 105.010 | (1.88,2.02] |
| 18 | 1.67 | diplomado | 3 | 66.570 | (1.61,1.75] |
| 19 | 1.46 | primario | 5 | 43.460 | (1.34,1.48] |
| 20 | 1.98 | primario | 4 | 96.400 | (1.88,2.02] |
| 21 | 1.47 | primario | 2 | 42.380 | (1.34,1.48] |
| 22 | 1.74 | primario | 3 | 80.410 | (1.61,1.75] |
| 23 | 1.90 | diplomado | 4 | 92.700 | (1.88,2.02] |
| 24 | 1.65 | fp | 5 | 62.810 | (1.61,1.75] |
| 25 | 1.34 | fp | 8 | 39.707 | (1.34,1.48] |
| 26 | 1.75 | diplomado | 2 | 78.951 | (1.75,1.88] |
| 27 | 1.68 | diplomado | 3 | 67.150 | (1.61,1.75] |
| 28 | 1.72 | bachiller | 4 | 71.940 | (1.61,1.75] |
| 29 | 1.65 | bachiller | 4 | 64.100 | (1.61,1.75] |
| 30 | 1.80 | bachiller | 6 | 76.680 | (1.75,1.88] |

Figura 13: Variable segmentada “clasealtura”.

Para construir dicha tabla de frecuencias pulsamos Estadísticos → Resúmenes → Distribución de frecuencias y seleccionamos la variable *clasesaltura*. Mirando ahora la ventana de resultados de R-Commander hallaremos los datos que nos permiten construir la siguiente tabla de frecuencias para la variable altura:

| Tabla de frecuencias para la variable altura | | |
|---|---------------------|-------------|
| Clases | Frecuencia absoluta | Porcentajes |
| (1.34, 1.48] | 6 | 12% |
| (1.48, 1.61] | 7 | 14% |
| (1.61, 1.75] | 15 | 30% |
| (1.75, 1.88] | 9 | 18% |
| (1.88, 2.02] | 13 | 26% |

Tabla 1. Tabla de frecuencias de la variable altura.

Si ahora deseamos borrar la nueva variable, para dejar el fichero como estaba, pulsamos Datos → Modificar variables del conjunto de datos activo → Eliminar variable del conjunto de datos. En la ventana que se despliega seleccionamos la variable que queramos eliminar. En este caso *clasesaltura*.

También puede construirse una tabla de frecuencias para una variable numérica, usando la información que nos suministra el histograma de frecuencias realizado con R Commander. Pero este procedimiento lo trataremos en esta misma sección dentro del apartado de representaciones gráficas.

2.2. Importar datos desde un fichero externo

Seleccionamos Datos → Importar datos. En la ventana que se despliega observamos que es posible importar datos en diferentes formatos. Aquí importaremos a R un fichero de datos de SPSS (familiar.sav), de Excel (familiar.xls) y otro de texto (familiar.txt).

Para importar datos desde un fichero SPSS (.sav) seguimos la serie de instrucciones Datos→ Importar datos →Desde datos SPSS. Se abre una ventana como de la figura 14. En el nombre del conjunto de datos introducimos el nombre que deseemos para el nuevo fichero, por ejemplo famspss. Los demás campos pueden dejarse tal como están.



Figura 14. Importando datos desde SPSS.

A continuación se nos pedirá, desde una ventana de Windows, el nombre del fichero de SPSS que se desee importar. Buscando el directorio en el que se encuentre este fichero, elegiremos familiar.sav. Si el programa SPSS no está instalado en nuestro ordenador, no podemos ver el contenido de este último fichero. Ahora lo podemos ver desde R, porque aparece como archivo de datos activo. Pulsando el botón Visualizar conjunto de datos veremos ahora el contenido de dicho fichero. Observamos que aparecen una serie de datos marcados con NA. Son los datos que faltan. Los ficheros de datos de R han de ser del tipo llamado data.frame, es decir que todas las variables deben tener el mismo número de datos. Si hay datos en blanco han de rellenarse con NA.

Para importar datos desde un fichero Excel (.xls) seguimos la misma serie de instrucciones, Datos→ Importar datos →Desde Excel, Access oDbase. Como nombre de los datos podemos elegir famexcel. El fichero de datos de Excel que hay que importar es familiar.xls. Pulsando el botón Visualizar conjunto de datos veremos ahora el contenido de dicho fichero.

Para importar datos desde un fichero de texto (.txt) pulsamos sucesivamente la serie de instrucciones Datos→ Importar datos →Desde archivo de texto o portapapeles. Como nuevo nombre para los datos podemos elegir famtxt. El fichero de datos que vamos a importar es familiar.txt.

Es frecuente introducir en R nuestros propios datos por medio de un archivo de texto, porque la introducción directa de datos en R, que hemos tratado previamente, es más lenta y, como ya se ha indicado, no es cómoda cuando hay que manejar una gran cantidad de datos. Así qué a menudo escribiremos nuestros nuevos datos en un archivo de texto. En ese caso es importante recordar que el fichero ha de ser del tipo data.frame (todas las variables con el mismo número de datos y NA en los lugares donde faltan datos)

2.3. Utilización de datos de paquetes de R

Los paquetes de R suelen contener ficheros de datos adecuados para las funciones incluidas en él, y que pueden servir como material para trabajar con las funciones propias de esos paquetes o de cualquier otro. También los podemos emplear para proponer trabajos a nuestros alumnos, con la ventaja de que incluyen información (aunque en inglés) sobre los datos que contienen.

Algunos estos ficheros ya están disponibles en nuestra sesión de trabajo, los de paquetes ya cargados. Para acceder a estos ficheros seleccionamos, desde el menú de R-Commander Datos → Conjunto de datos en paquetes. (ver la figura 15).

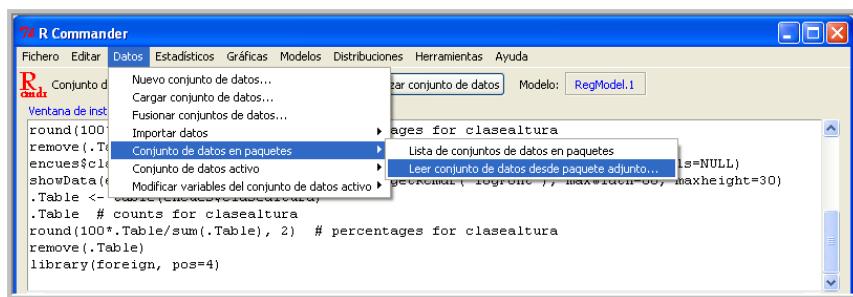


Figura 15: Menú R-Commander “Leer conjunto de datos en paquete adjunto”.

Si elegimos la primera opción nos aparece una lista de los archivos cargados con una breve descripción de su contenido. La lista aparece en una ventana de RGUi. La segunda opción nos permitirá cargar los ficheros de datos de dichos paquetes. En esta ocasión vamos a abrir el fichero Seatbelts del paquete datasets, así que pulsamos en Leer conjunto de datos desde paquete adjunto. La ventana que se nos abre debe quedar como la de la figura 16. Una vez aceptadas estas opciones tenemos disponible el fichero Seatbelts (para localizar este fichero se debe tener en cuenta que en la lista de ficheros aparecen dos secuencias alfabéticas sucesivas). Posteriormente, usaremos este fichero para introducir algunas de las representaciones gráficas que describimos en la segunda parte de esta sección y, además, aprovecharemos la ocasión para extraer información sobre los datos que contiene.

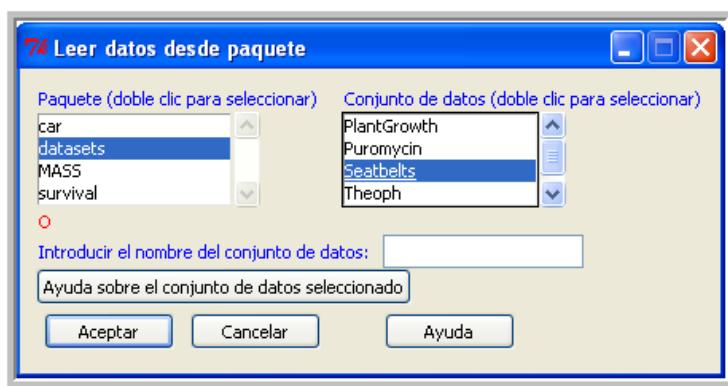


Figura 16: Menú R-Commander “Leer conjunto de datos en paquete adjunto”.

Cuando los datos que vamos a usar no están en los paquetes cargados debemos instalar y cargar primero el paquete en el que está incluido, actuando seguidamente como se ha hecho para cargar el fichero Seatbelts.

2.4. Representación gráfica de datos

Vamos a hacer algunas modificaciones al diagrama de barras de la figura 5 que nos permitan obtener un nuevo diagrama de barras de la misma variable tal como el de la figura 17. Para ello es necesario modificar la orden

```
barplot(table(encues$estudios), xlab="estudios", ylab="Frequency")
```

Esta orden se genera desde R-Commander añadiendo una serie de parámetros. Si queremos informarnos adecuadamente sobre estos parámetros, entramos en la ayuda de R desde RGui. Pulsamos Ayuda y Funciones R (texto). La orden de la que deseamos obtener más información es barplot, así que introducimos esa palabra en la ventana que aparece. Se despliega entonces la ventana representada en la figura 18.

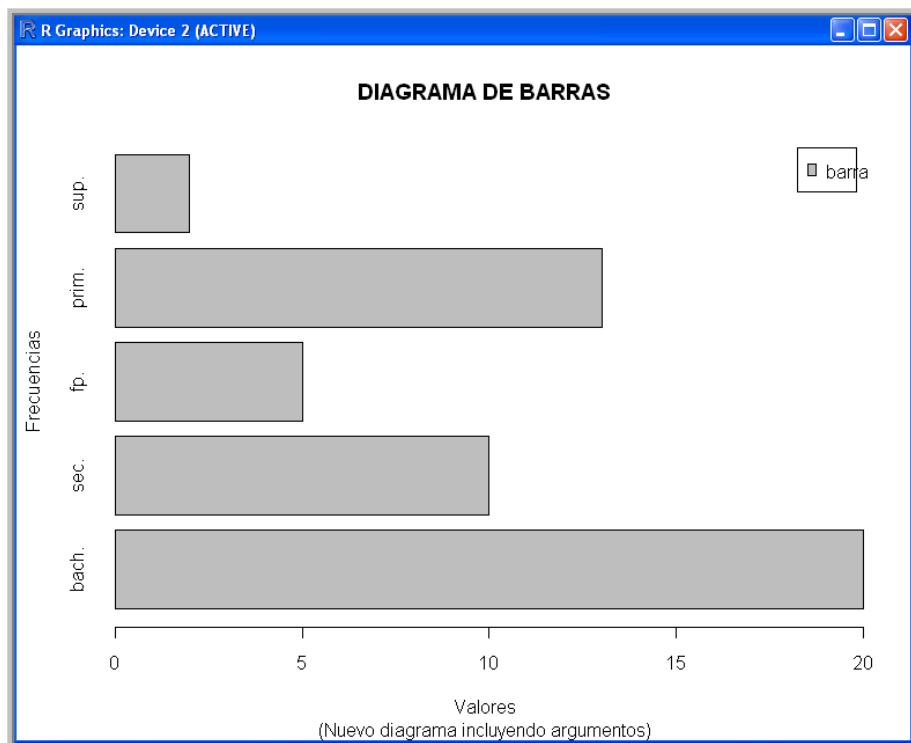


Figura 17: Diagrama de barras de la variable “estudios”.

En la ventana de la figura 18, se describe la orden barplot y se especifican los valores que pueden incluirse como argumentos. Algunos de ellos los hemos incorporados en los ejercicios que siguen, para tener ocasión de perfeccionar el diagrama de barras que suministra R-Comander usando sólo el menú desplegable.

Merece la pena destacar el apartado Examples de esta ayuda, ya que los ejemplos suelen ser ejecutables. Copiando el ejemplo que aparece y pegándolo en RGui, o en R-commander, podemos ejecutarlo. Si se quiere ejecutar el ejemplo orden a orden para observar los detalles de programación, o al menos poder ver con tranquilidad cada uno de los gráficos que aparecen, es más fácil hacerlo desde R-Commander.

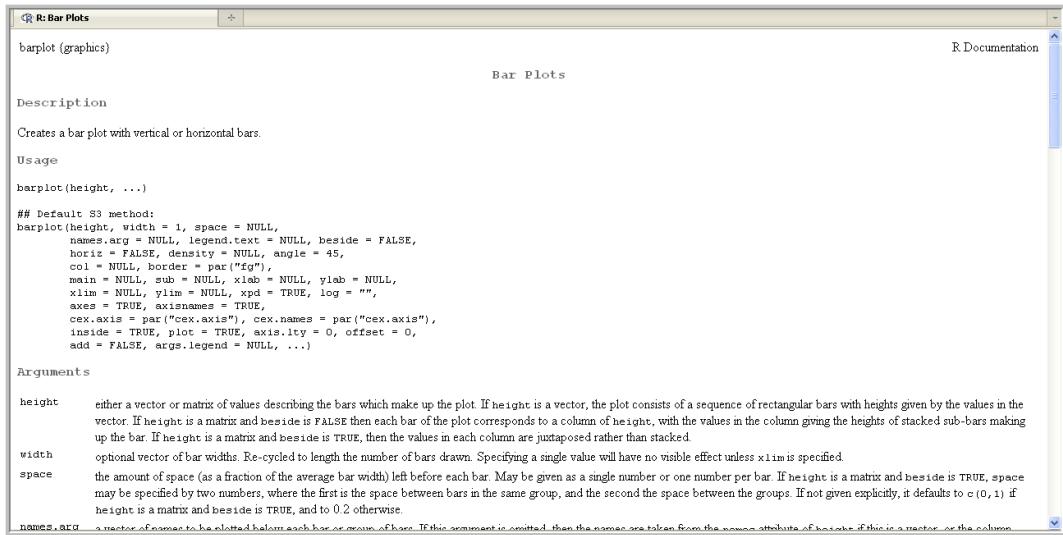


Figura 18: Ayuda de la orden barplot.

Para obtener el diagrama de barras de la figura 17, se debe introducir la siguiente línea en R-Commander (copiar todo seguido en un único renglón):

```
barplot(table(encues$estudios), xlab="Valores", ylab="Frecuencias",
names.arg = c("bach.", "sec.", "fp.", "prim.", "sup."),
legend.text = "barra", main ="DIAGRAMA DE BARRAS",
sub = "(Nuevo diagrama incluyendo argumentos)", horiz=TRUE)
```

Otra opción es teclear desde R:

```
{barplot(table(encues$estudios), xlab="Valores", ylab="Frecuencias",
+ names.arg = c("bach.", "sec.", "fp.", "prim.", "sup."), legend.text = "barra",
+ main ="DIAGRAMA DE BARRAS", sub = "(Nuevo diagrama incluyendo argumentos)",
+ horiz=TRUE)}
```

Para hacer un diagrama de sectores hay que seleccionar:

Gráficas→Gráfica de Sectores

Se selecciona la variable estudios. Obtendremos la gráfica de la figura 19. Para copiar una figura en el portapapeles se pulsa sobre la ventana en la que aparece con el botón derecho del ratón, seleccionando copiar como metafile o como bitmap.

Para hacer un histograma para la variable altura, seguimos desde el menú de R-Commander:

Gráficas→Histograma

Seleccionamos la variable altura y, por ejemplo, recuento de frecuencias. Emerge una ventana en RGui con un histograma de frecuencias de la variable altura. En la ventana de instrucciones de R Commander aparece:

```
Hist(encues$altura, breaks="Sturges", col="darkgray")
```

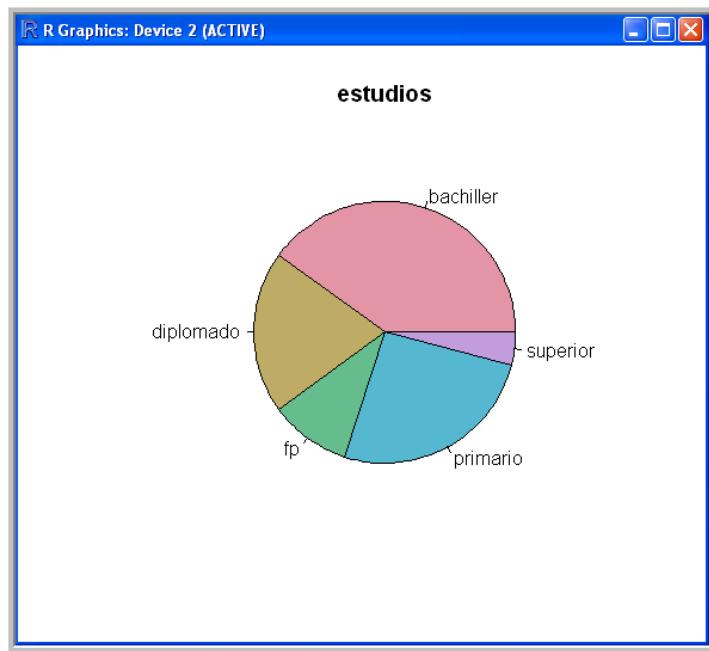


Figura 19: Diagrama de sectores de la variable “estudios”.

Cerramos la ventana de la gráfica. Si ahora ejecutamos, desde R-Commander, la orden:

```
datostabla=hist(encues$altura, breaks="Sturges", col="darkgray")
```

y después introducimos y ejecutamos desde la ventana de R-Commander:

```
datostabla
```

Obtendremos de nuevo el histograma de la figura 20, pero además una serie de valores en la ventana de resultados (obsérvese que la función hist está escrita ahora con h minúscula). Usaremos los valores obtenidos para construir una tabla de frecuencias de la variable altura. Destacamos a continuación los datos útiles para construir esta tabla de frecuencias:

```
$breaks
```

```
[1] 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1
```

```
$counts
```

```
[1] 3 3 7 11 11 5 9 1
```

```
$mids
```

```
[1] 1.35 1.45 1.55 1.65 1.75 1.85 1.95 2.05
```

La tabla de frecuencias de la variable altura es la siguiente:

| Intervalo | Marca de clase | Frecuencia absoluta |
|-----------|----------------|---------------------|
| (1.3,1.4] | 1.35 | 3 |
| (1.4,1.5] | 1.45 | 3 |
| (1.5,1.6] | 1.55 | 7 |
| (1.6,1.7] | 1.65 | 11 |

| | | |
|-----------|------|----|
| (1.7,1.8] | 1.75 | 11 |
| (18,1.9] | 1.85 | 5 |
| (1.9,2.0] | 1.95 | 9 |
| (3.0,2.1] | 2.05 | 1 |

Tabla 2: Tabla de frecuencias de la variable altura.

Si queremos una tabla de frecuencias similar a la que habíamos representado en el tema anterior usamos (introduciéndola en una única línea) la siguiente orden.

```
datostabla=hist(encues$altura, breaks=c(1.34,1.48,1.61,1.75,1.88,2.02), col="darkgray")
```

Con los resultados de esta orden formamos la tabla 3.

| Tabla de frecuencias para la variable altura | | |
|---|---------------------|-------------|
| Clases | Frecuencia absoluta | Porcentajes |
| (1.34, 1.48] | 6 | 12% |
| (1.48, 1.61] | 7 | 14% |
| (1.61, 1.75] | 16 | 32% |
| (1.75, 1.88] | 8 | 16% |
| (1.88, 2.02] | 13 | 26% |

Tabla 3: Tabla de frecuencias de la variable altura.

Se observa que la tabla 3 no es exactamente igual que la tabla 1 que aparece en el apartado 2.1. El motivo es que aquí el dato 1.75, que queda en la frontera entre dos clases se ha asignado (correctamente) a la clase (1.61, 1.75]. En la tabla 1 de la sección 2.1, donde se había seguido el procedimiento de segmentar la variable, se asignaba este dato a la clase siguiente.

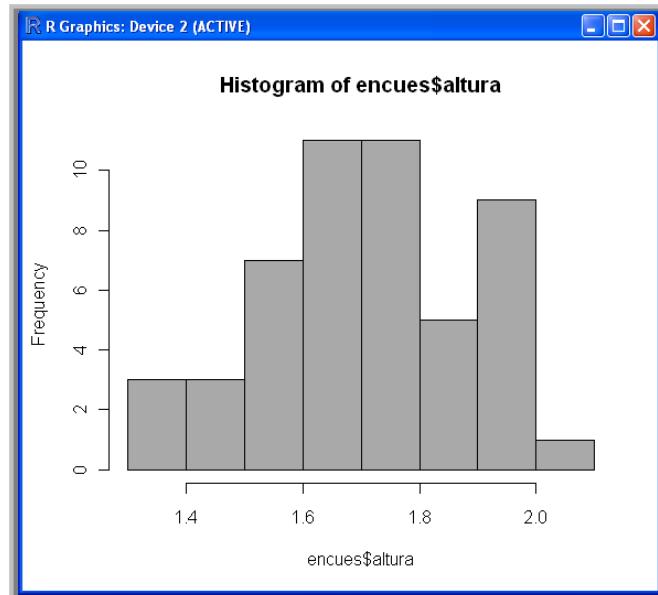


Figura 20: Histograma de la variable “altura”.

El gráfico secuencial es apropiado para representar datos que aparecen en un cierto orden dentro de la variable. Por ejemplo, los valores de las temperaturas medias mensuales de una ciudad durante varios años. Los valores suelen aparecer tal como se han recogido, es decir, en orden cronológico. Este tipo de datos se conoce con el nombre de series temporales.

La gráfica de la figura 21 representa la variable drivers del fichero Seatbelts, perteneciente al paquete datasets. Tendremos que seleccionar este fichero (Seatbelts) como conjunto de datos activo. Como ya se había abierto en la subsección 2.3, este fichero ya debería estar cargado.

Esta gráfica se ha obtenido mediante la secuencia:

Gráficas → Gráfica secuencial

Y seleccionando la variable drivers. Se observa en la figura 21 una serie de segmentos verticales, cada uno de los cuales mide, el valor del dato de drivers que aparece en el fichero en el orden indicado por el eje horizontal de la figura (hay 192 datos).

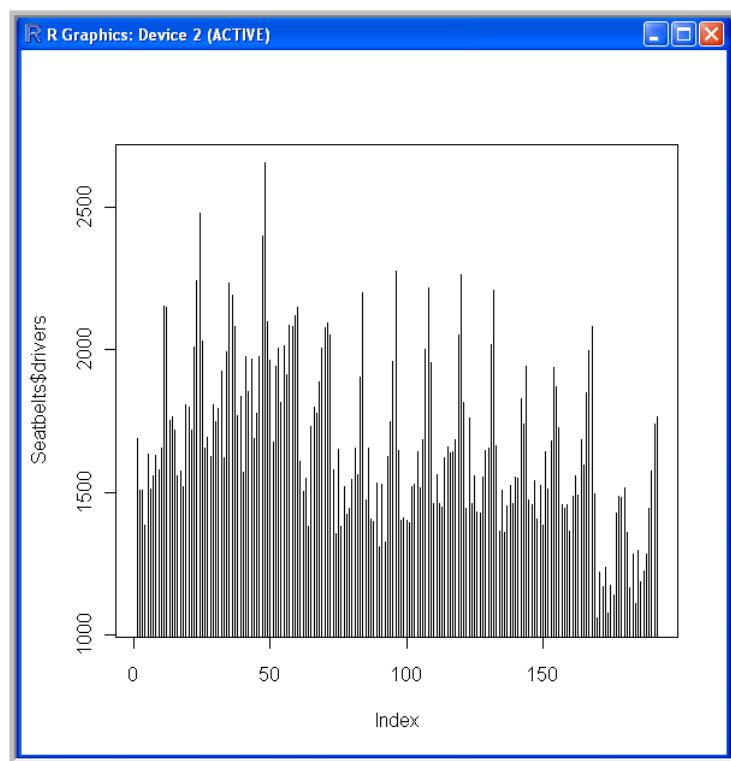


Figura 21. Gráfica secuencial de la variable “drivers” del fichero seatbelts.

Si deseamos información sobre la variable representada podemos seguir las instrucciones:

Datos → Conjunto de datos activo → Ayuda sobre el conjunto de datos activos

Con ello vemos la ventana de la figura 22, que nos da información sobre los datos del fichero Seatbelts. Esta información nos ha permitido saber que los datos del fichero se refieren a una investigación realizada en Gran Bretaña sobre los viajeros muertos o con heridas graves en accidentes de Tráfico. Los datos corresponden al periodo de tiempo 1969 - 1984. Son 16 años, es decir 192 meses. La variable Drivers recoge exactamente el número de conductores muertos o gravemente heridos en cada uno de estos 192 meses. Hay otras variables. Por ejemplo, la primera DriversKilled es el número de conductores muertos. La última variable (law) nos

muestra que en los últimos 22 meses se había impuesto la ley de obligatoriedad de uso del cinturón de seguridad.

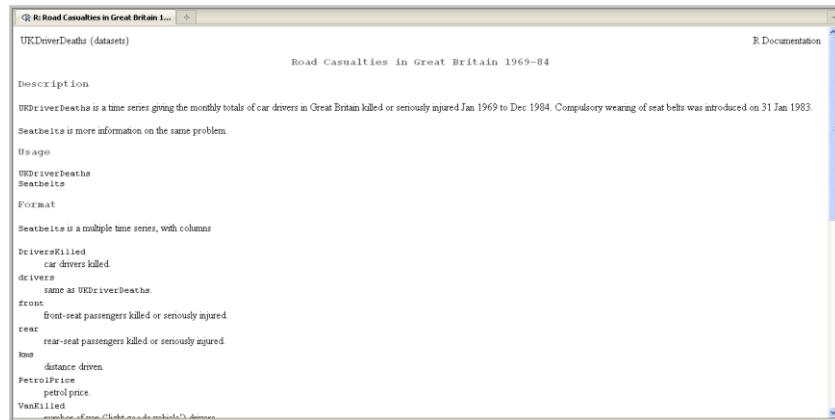


Figura 22. Información sobre el fichero seatbelts.

Vamos a modificar algo la última gráfica de la figura 21. Cambiamos, en la ventana de instrucciones donde aparece:

```
plot(Seatbelts$drivers, type="h")
type="h" por type="l". Ejecutando la nueva orden:
plot(Seatbelts$drivers, type="l")
```

Y aparece la gráfica de la figura 23.

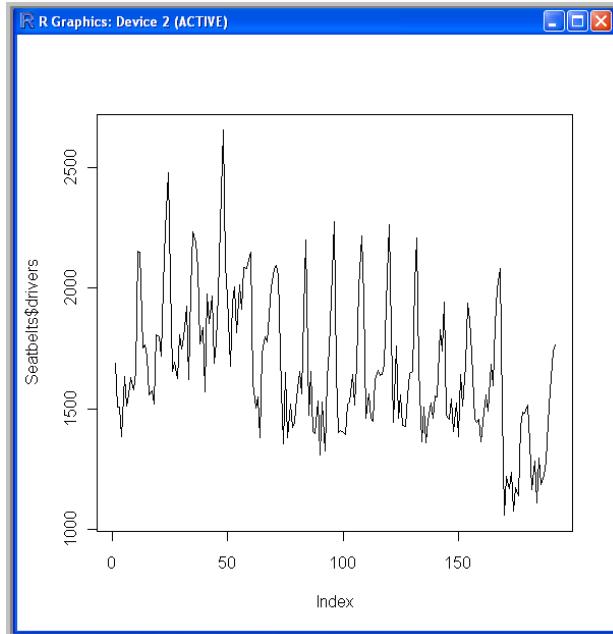


Figura 23. Gráfica de serie temporal de la variable “drivers” del fichero seatbelts.

Esta gráfica secuencial representa la poligonal que une los extremos de los segmentos de la gráfica de la figura 21. Aunque las figuras 21 y 23, representan la misma información sobre la variable drivers, parece que esta última figura permite verla con más claridad. Puede detectarse bastante información sobre los datos. Es interesante destacar este hecho, pues la gráfica nos da información que es casi imposible de conseguir observando los datos en bruto. Por ejemplo:

- Permite observar una pauta de periodicidad en los datos.
- Se observan 16 picos, correspondientes a los 16 años. De estos valores, los más altos, coinciden aproximadamente con los meses numerados con múltiplos de 12. Posiblemente corresponden a los meses de Diciembre.
- En los primeros años el número de accidentes graves ha ido aumentando.
- Luego viene un periodo algo menos dramático, unos 9 años, donde el número de estos accidentes ha sido menor, conservándose más o menos estable (salvo la periodicidad).
- Un hecho destacable es el importante descenso de la siniestralidad que se observa en el último periodo, coincidiendo con la puesta en vigor de la ley de obligatoriedad del cinturón de seguridad.

Este ejemplo nos parece extraordinariamente adecuado para que nuestros alumnos aprendan a interpretar las gráficas y perciban el papel que pueden llegar a jugar para desvelar la información que puede estar escondida dentro de un conjunto de datos numéricos.

Los diagramas de cajas, conocidos también como diagramas de cajas y bigotes. Estas gráficas se construyen a partir de los datos de la variable resaltando los cuartiles y la mediana de la muestra.

Para hallar el diagrama de caja de la variable peso, se selecciona el fichero encues como conjunto de datos activo. Desde el menú de R-Commander seleccionamos Gráficas → Diagrama de caja y en la ventana emergente seleccionamos la variable peso. Se obtiene con ello la gráfica de la figura 24.

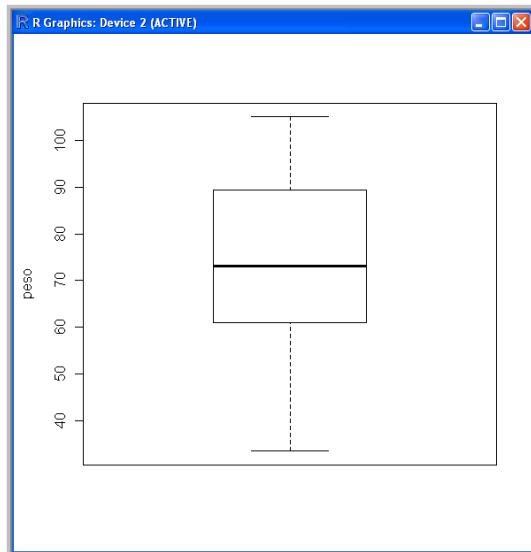


Figura 24. Gráfica de cajas y bigotes de la variable “peso” del fichero encues.

La longitud de la caja (zona rectangular) abarca desde el primer cuartil hasta el tercer cuartil. En el interior de la caja está marcada una línea a la altura de la mediana. Las líneas verticales por arriba y por abajo de la caja se conocen con el nombre de "bigotes".

El extremo del bigote superior es la menor de las siguientes medidas:

- a) el primer cuartil más una vez y media la altura de la caja (rango intercuartílico)
- b) la observación mayor de la muestra.

Análogamente, el extremo del bigote inferior es la mayor de las siguientes medidas:

- a) el primer cuartil menos una vez y media la altura de la caja (rango intercuartílico).
- b) la observación menor de la muestra.

Los datos que sobresalen de los bigotes se califican como datos Anómalos u Outliers, como puede verse en la figura 25, donde se ha introducido en la variable peso el dato 140.

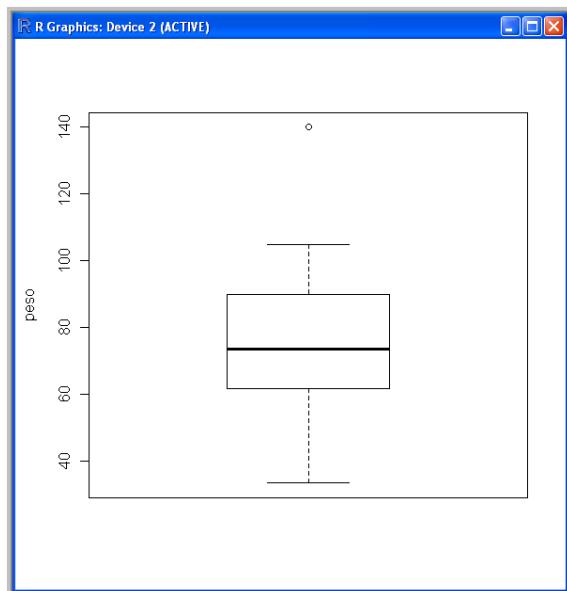


Figura 25. Gráfica de cajas y bigotes de la variable “peso” del fichero encues con outliers.

El diagrama de caja de la figura 25 se ha realizado añadiendo en el fichero encues el valor 140 en la variable peso. Este dato es demasiado grande con relación al resto y se detecta como outlier, puesto que queda por encima del bigote superior.

Si no hemos tenido la precaución de guardar el fichero encues con un nuevo nombre al añadir el nuevo dato, el fichero encues queda ahora modificado. Para devolverlo a su estado primitivo conviene borrar la fila 51. Esto puede hacerse pulsando desde el menú del interface de R-Commander:

Datos → Conjunto de datos activo → Borrar fila(s) del conjunto de datos activo.

3. Conclusiones y agradecimientos

Hoy día el uso de programas estadísticos se ha convertido en una rutina a la hora de analizar los datos de las investigaciones en la mayoría de las disciplinas científicas. Los Estadísticos nos encontramos con frecuencia con peticiones de ayuda por parte de autores y directores de investigaciones que manifiestan tener problemas con el uso de estos programas, o, peor aún, con la interpretación de los resultados y los gráficos que han devuelto estos programas.

Es una práctica común en algunas disciplinas científicas confiar el análisis de datos a paquetes estadísticos comerciales, de los que se extraen más estadísticos que información. Es imposible obtener respuestas a medida, porque estos programas proporcionan técnicas y gráficas a la medida que el programador decidió en su día. Algunos programas se han diseñado para que sea el investigador el que decida qué respuestas quiere. Estos programas, como por ejemplo el software estadístico R, son además gratuitos.

En este Taller se ha pretendido proporcionar una introducción al programa R, que representa una alternativa a programas estadísticos de uso más extendido. A través de la utilización de este software libre se ha ido poniendo de manifiesto cómo es el propio autor del trabajo quien toma las decisiones con respecto a qué estrategias, gráficos y técnicas utilizar, así como sobre el formato de presentación que prefiere para sus resultados y sus gráficos. Es sencillamente imposible utilizar R sin criterio. Se debe comenzar por saber lo que se quiere antes de dar un sólo clic con el ratón sobre un menú en un programa estadístico.

Creemos además que el programa R proporciona todas las opciones necesarias para utilizarlo en la práctica docente universitaria. Aunque obtener resultados es más costoso inicialmente, a la larga, los resultados son mucho más concretos, adaptables y modificables. Debemos tener en cuenta que R no es sólo un programa de cálculo estadístico, sino que es un lenguaje de programación estadístico, con el que podemos implementar cualquier procedimiento o técnica estadística, por novedosa que ésta sea. Además, en la actualidad existen más de 2000 librerías o paquetes disponibles para casi cualquier propósito.

Creemos que R es un entorno de programación en el que se han incluido rutinas que permiten de forma fácil y directa la realización de multitud de cálculos estadísticos; también la confección de gráficas e imágenes de una sencillez y una potencia equiparable, o superior, a la de otros programas comerciales. Por todo ello, os animamos a que utilicéis el software libre R tanto en el ámbito docente como en el de la investigación.

Queremos agradecer a los organizadores de las I Jornadas de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa, celebradas en Melilla en junio de 2010, que hayan pensado en nosotros para el desarrollo de este Taller de R. Queremos también agradecer a los integrantes del Grupo de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa (GENAEIO) de la SEIO por su participación y su especial motivación en el desarrollo de este Taller. Esperamos y deseamos que el desarrollo del mismo haya resultado interesante y productivo para todos.

4. Bibliografía y documentación

- [1] Introducción a R: <http://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- [2] Manual resumido de R: <http://cran.r-project.org/doc/contrib/curso-R.Diaz-Uriarte.pdf>
- [3] Diverso material docente: <http://www.ub.es/stat/docencia/EADB/Curso basico de R.htm>
- [4] Para principiantes: http://es.geocities.com/r_vaquerizo/Manual_R3.htm

[5] Introducción al uso del R Commander:

[http://geminis.dma.ulpgc.es/profesores/personal/asp/
Documentacion/Manual_R_commander.pdf](http://geminis.dma.ulpgc.es/profesores/personal/asp/Documentacion/Manual_R_commander.pdf)

[6] R para principiantes:

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

[7] Amplio manual sobre gráficos estadísticos con R:

<http://cran.r-project.org/doc/contrib/grafi3.pdf>

[8] Curso básico de R:

http://www.ub.es/stat/docencia/EADB/Curso_basico_de_R.pdf

[9] Introducción a R:

http://www.cesma.usb.ve/~lbravo/co5314/MANUAL_Rb.pdf

Acerca de los autores



Luis Miguel Marín Trechera es Profesor Titular de Estadística e Investigación Operativa en la Universidad de Cádiz. Antes de dedicarse a la docencia ocupó puestos de responsabilidad en la Administración Pública y la empresa privada.

Sus líneas de investigación están centradas en técnicas de Inteligencia Artificial en Análisis de Datos, aplicación de Técnicas Estadísticas en problemas de ingeniería y utilización de Nuevas Tecnologías y Metodologías en Educación, habiendo publicado numerosas aportaciones en revistas, congresos y libros.

Tiene amplia experiencia en el uso de Internet como herramienta educativa y en formación del profesorado universitario y preuniversitario durante los últimos 15 años, a través su participación como ponente en diversos cursos y planes de desarrollo, tanto a nivel nacional como internacional.

Colabora con la AGAE (Agencia Andaluza de Evaluación de la Calidad y Acreditación Universitaria) como evaluador de proyectos de innovación docente y del campus andaluz virtual.

Actualmente es Coordinador de Innovación Docente de la Escuela Superior de Ingeniería de Cádiz.



Antonio Gámez Mellado es Profesor Titular de Escuela Universitaria del Departamento de Estadística e Investigación Operativa de la Universidad de Cádiz. Licenciado en Matemáticas, especialidad Estadística e Investigación Operativa por la Universidad de Granada, sus líneas de investigación son la Enseñanza y el Aprendizaje de la Estadística e I.O. y la fiabilidad de sistemas complejos. Tiene una amplia experiencia docente tanto en Bachillerato como en la Universidad. Ha publicado diversos libros de carácter docente en el ámbito de la enseñanza de la Estadística en la Ingeniería y en el ámbito de las Matemáticas Aplicadas a las Ciencias Sociales. Actualmente es el ponente de la Universidad de Cádiz (UCA) de la materia “Matemáticas Aplicadas a las Ciencias Sociales II”. Ha recibido

multitud de premios y distinciones en proyectos de innovación docente en la Universidad de Cádiz y en la Unidad para la Calidad de las Universidades Andaluzas. Ha coordinado diversos proyectos de innovación docente en la UCA. En la actualidad es evaluador externo de la Agencia Andaluza de Evaluación de la Calidad y Acreditación Universitaria (AGAE), para proyectos de innovación docente, y ha evaluado multitud de proyectos de innovación docente en las Universidades Andaluzas.

Participa como investigador principal en el Proyecto de Excelencia “Re-evalúa” y en otros proyectos I+D+I como EvalCOMIX, EvalHIDA, FLEXO, etc. que han sido financiados por el Ministerio de Educación, el Ministerio de Industria, Turismo y Comercio, la Consejería de Economía, Innovación y Ciencia de la Junta de Andalucía, etc. Es investigador principal en diversos proyectos dentro del Programa de Cooperación Interuniversitaria e Investigación Científica (PCI) en Jordania, Cuba, etc.

Ha impartido cursos de formación en especialización en el uso de herramientas informáticas para la docencia en el Aula Virtual, cursos de formación para el profesorado universitario en distintas universidades españolas y diversas actividades formativas en cursos de verano en universidades españolas. Ha impartido y coordinado planes de desarrollo para el Personal Docente e Investigador (PDI) en la Universidad de Cádiz y en la Universidad de Cantabria.

Ha dictado multitud de conferencias relacionadas con las Matemáticas y la Estadística en diversos ámbitos. Ha impartido cursos de formación en especialización en diversas áreas de Matemáticas dentro del proyecto “Thales-CICA-Web”. Ha sido evaluador interno en la UCA y evaluador externo en diversos cursos del Campus Andaluz Virtual (CAV).

Es autor de multitud de comunicaciones en congresos nacionales e internacionales. En la actualidad es tutor académico de la Organización de Estados Iberoamericanos (OEI) en los cursos de especialización en formación matemática en Nandutí (Paraguay).

Taller de R

(Estadística Descriptiva Unidimensional y Bidimensional)

Antonio Gámez Mellado¹, Luis Miguel Marín Trechera²

¹*antonio.gamez@uca.es*, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz

²*luis.marin@uca.es*, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz
Universidad de Cádiz

Resumen

Esta segunda parte del Taller está orientado a aprender a usar R para realizar el cálculo de los parámetros estadísticos básicos de estadística descriptiva unidimensional. También se muestra como implementar diversos modelos de ajuste y el cálculo de parámetros de estadística descriptiva bidimensional.

Palabras clave: Software libre, R, Estadística

Clasificación AMS: 97U50, 97K40, 97K80

1. Estadística descriptiva unidimensional

Con R-Commander pueden calcularse los parámetros estadísticos más usuales (media, varianza, desviación, cuartiles, etc.). El manejo de R-Commander es sencillo, y suele ser suficiente para los estudios básicos. En la primera parte de esta sección tratamos estos casos. Sin embargo hay otros parámetros algo menos usados que no son accesibles desde R-Commander, pero que pueden programarse desde R. El cálculo de los parámetros estadísticos básicos mediante commandos de R se tratará, más adelante. Además existe la posibilidad de realizar estas tareas por medio de programas, o funciones, realizados en lenguaje R. Al final de esta sección proponemos algunas funciones, que sirven para hallar diversos parámetros y para hacer tablas de frecuencias más detalladas que las obtenidas en las secciones anteriores. Para realizar programas es necesario conocer y manejar el lenguaje R. No obstante no es necesario programar en R para emplearlas. Así que podemos usar estas funciones, a modo de caja negra, sin más que ejecutar el programa e incluir los datos precisos en sus argumentos.

Para obtener un resumen de los parámetros estadísticos de un conjunto de datos es necesario que aparezca como Conjunto de datos activo. Por ejemplo si el fichero activo es encues, pulsando desde R-Commander las órdenes Estadísticos→Resúmenes→Conjunto de datos activo se obtiene la ventana de la figura 1, que contiene información sobre las variables del fichero. En esta figura observamos que para las variables cuantitativas se han registrado los valores mínimos y máximos, el primer y tercer cuartil, la mediana (o segundo cuartil) y la media. Para la única variable cualitativa (estudios) se obtiene la frecuencia absoluta de cada uno de los valores que contiene. También se especificaría, si se diera el caso, el número de valores ausentes (NA) para cada variable que contenga valores desconocidos.

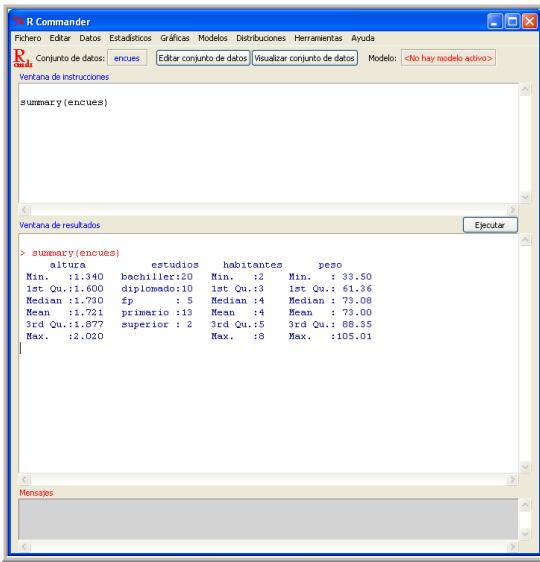


Figura 1: Resumen estadístico de las variables del archivo “encues”.

Otra opción, que trata las variables numéricas una a una es seguir la serie de comandos Estadísticos→Resúmenes→Resúmenes numéricos.

Con esta secuencia se despliega una ventana tal como la de la figura 2 que permite obtener media, cuasidesviación, mínimo, primer cuartil, mediana, tercer cuartil y máximo de la variable altura. Podemos seleccionar otros cuantiles (percentiles), e incluso obtener estos parámetros para los subgrupos definidos por la variable estudios.

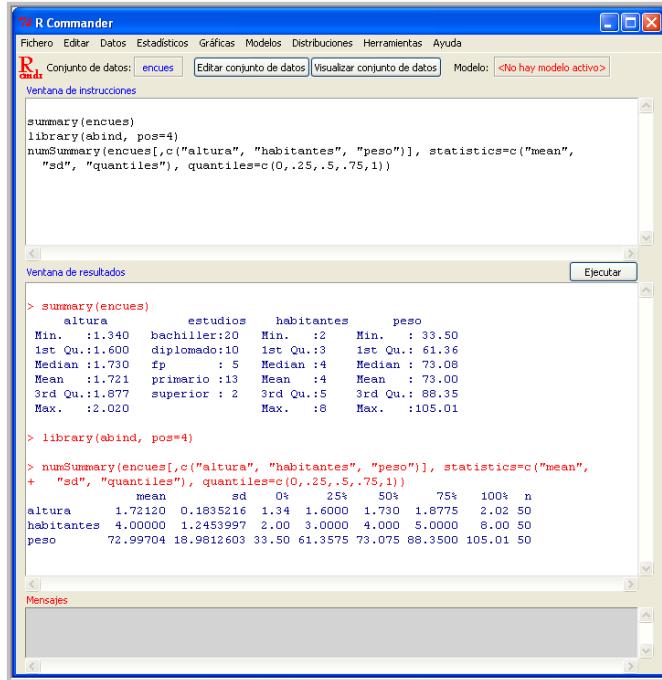


Figura 2: Resúmenes numéricos de las variables del archivo “encues”

Veamos como calculamos las medias de la variable altura clasificada según el nivel de estudios. En este caso la ventana de selección tendrá el aspecto de la figura 3. El grupo se ha seleccionado pulsando en el botón Resumir por grupos.

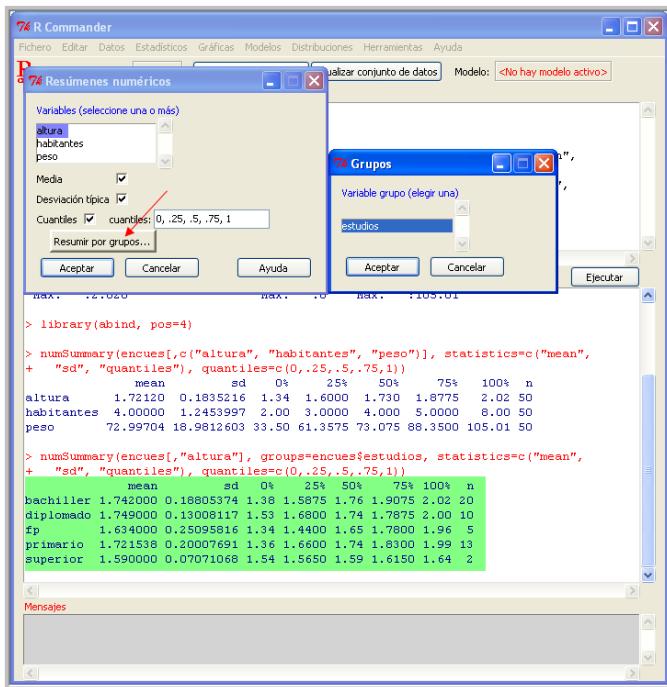


Figura 3: Resumen numérico de la variable altura según subgrupo “nivel de estudios”.

El resultado, que aparece en la ventana de instrucciones de R-Commander de la figura 3, especificando la media y el número de datos de la variable altura según el nivel de estudios, es el siguiente:

| | medias | n |
|-----------|----------|----|
| bachiller | 1.742000 | 20 |
| diplomado | 1.749000 | 10 |
| fp | 1.634000 | 5 |
| primario | 1.721538 | 13 |
| superior | 1.590000 | 2 |

Tabla 1: Medias marginales de la variable altura según “nivel de estudios”.

Los comandos necesarios para calcular los parámetros estadísticos más usuales son:

max = máximo min = mínimo range = extremos mean = media
median = mediana var = cuasivarianza quantile = percentiles

Este último comando (`quantile`) calcula por defecto los percentiles 0%, 25%, 50%, 75%, 100% que corresponden a los valores mínimo, primer cuartil, mediana, tercer cuartil y máximo.

Las órdenes necesarias para calcular los parámetros de asimetría (skewness) y curtosis (kurtosis) no son accesibles con R-Comander, ni tampoco desde el paquete básico de R, así que hay que incorporar paquetes complementarios que las incluyan.

Veamos como puede conseguirse esto. En primer lugar hay que localizar los paquetes que contienen estos comandos. Para ello, desde la pestaña de ayuda de Rgui, seleccionamos **Buscar ayuda**: Se nos despliega una ventana, en la que se introduce la palabra de la que se deseé obtener información. Por ejemplo skewness (asimetría). Después de aceptar, nos aparece la ventana de la figura 4, que contiene los documentos en los que aparece esta palabra. Encontramos varios paquetes que contienen este comando. Por ejemplo los paquetes: "agricolae" y "e1071". Instalamos y cargamos cualquiera de ellos. Si estuvieran ya instalados sólo tendríamos que cargar dicho paquete.

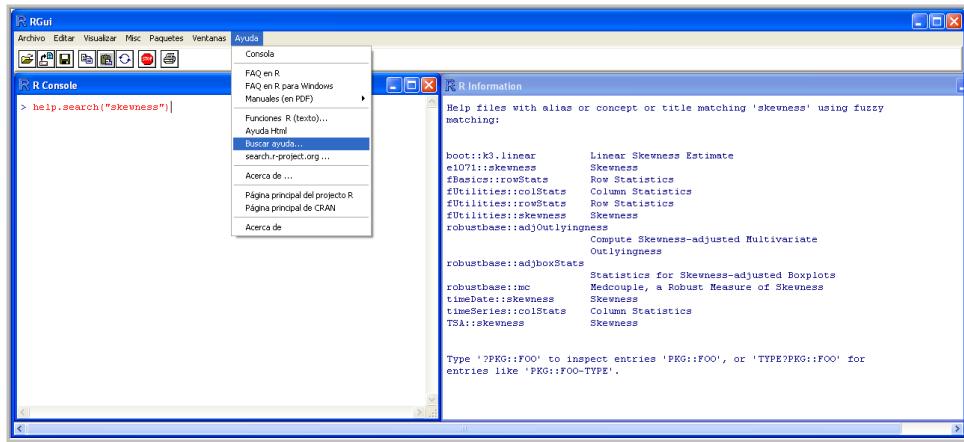


Figura 4: Información sobre “skewness”en la ayuda de RGui.

Para ver la información sobre este parámetro de asimetría (skewness) entramos, por ejemplo, en el paquete “agricolae”. Para ello, tras pulsar la pestaña de ayuda y seleccionar en el submenú que aparece Ayuda Html, seleccionamos sucesivamente los enlaces packages y agricolae, con los no aparece la información sobre el contenido de este paquete. En la lista de comandos aparecen skewness y kurtosis, así que con este paquete podemos calcular los coeficientes de asimetría y curtosis. Pulsando el enlace skewness nos encontramos en una página web donde se describe el comando, su uso y la fórmula que se emplea para calcularlo.

Para calcular los coeficientes de asimetría y curtosis de la variable altura del fichero encues, debemos ejecutar las órdenes que siguen:

```
skewness(encues$altura)
[1] -0.2133998
kurtosis(encues$altura)
[1] -0.749942
```

2. Cálculo de parámetros estadísticos creando funciones en R

A continuación se definen una serie de comandos o funciones personales que permiten calcular la media, cuasivarianza, varianza, desviación típica y cuasidesviación de una variable y que se evaluarán especificando explícitamente las operaciones necesarias en cada caso.

```
media=function(x){sum(x)/length(x)}
varianza=function(x){sum((x-media(x))^2)/length(x)}
```

```

cuasivarianza=function(x){ sum((x-media(x))^2)/(length(x)-1)}
varianza2=function(x){ cuasivarianza(x)*(length(x)-1)/length(x)}
desvtip=function(x){ sqrt(varianza(x)) }
cuasdesv=function(x){ sqrt(cuasivarianza(x)) }

```

El comando `sum` calcula la suma de todos los valores de la variable, `length` el número de datos, `sqrt(x)` calcula la raíz cuadrada de cada uno de los elementos de `x`, si `x` es un vector. Naturalmente si `x` es un número calcula la raíz cuadrada de su valor.

Si queremos hallar la media, la cuasivarianza, la varianza, desviación típica o la cuasidesviación de la variable altura del fichero encues tecleamos primero el código anterior, para definir las funciones. Después las evaluamos introduciendo las siguientes líneas:

```

x=encues$altura
media(x)# equivale a mean(x)
cuasivarianza(x) # equivale a var(x)
varianza(x) # o varianza2(x)
desvtip(x)
cuasdesv(x)# equivale a sd(x)

```

También se pueden definir los parámetros estadísticos, así como realizar muchas otras tareas, por medio de programas realizados en R. Por ejemplo, el siguiente programa permite obtener la media, la varianza, el máximo y el mínimo de un conjunto de variables

```

f = function(z,t)
{
  datos = list(z,t)
  medias = lapply(datos,mean) # lapply aplica una función sobre una lista
  varianzas = lapply(datos,var)
  maximos = lapply(datos,max)
  minimos = lapply(datos,min)
  for(i in 1:length(datos)) # Se realizan las siguientes instrucciones para cada una de las
    # variables de la lista de datos
  {
    cat("Distribución ",i,".", "\n") # La función cat es para visualizar cosas
    cat("media: ",medias[[i]], "varianza:", varianzas[[i]],
        "maximo:", maximos[[i]], "minimo:", minimos[[i]], "\n")
    cat("-----", "\n")
  }
}

```

Para ejecutar la función hay que dar previamente los valores a `z` y `t`.

```

z=encues$altura
t=encues$peso
f(z,t)

```

Naturalmente el número de argumentos de `f` podría ampliarse. Por ejemplo:

```
f = function(z,t,h)
{
  datos = list(z,t,h)
  medias = lapply(datos,mean) # lapply aplica una función sobre una lista
  varianzas = lapply(datos,var)
  maximos = lapply(datos,max)
  minimos = lapply(datos,min)
  for(i in 1:length(datos))
  {
    cat("Distribución ",i,":","\n") # La función cat es para visualizar cosas
    cat("media: ",medias[[i]], "varianza:", varianzas[[i]],
        "máximo:", maximos[[i]], "mínimo:", minimos[[i]], "\n")
    cat("-----","\n")# "\n" añade una línea
  }
}
```

Después de introducir el código de la función, para ejecutarla hay que dar previamente los valores de z, t y h. Como en este caso z y t ya deben estar definidos anteriormente, sólo queda añadir h:

```
h=encues$habitantes
f(z,t,h)
```

Obtendremos la siguiente salida en la ventana de instrucciones:

```
Distribución 1 :
media: 1.7212  varianza: 0.03368016  máximo: 2.02  mínimo: 1.34
-----
Distribución 2 :
media: 72.99704  varianza: 360.2882  máximo: 105.01  mínimo: 33.5
-----
Distribución 3 :
media: 4  varianza: 1.551020  máximo: 8  mínimo: 2
-----
```

Estos resultados se han obtenido directamente de la ventana de R-Commander, como puede verse en la figura 5.

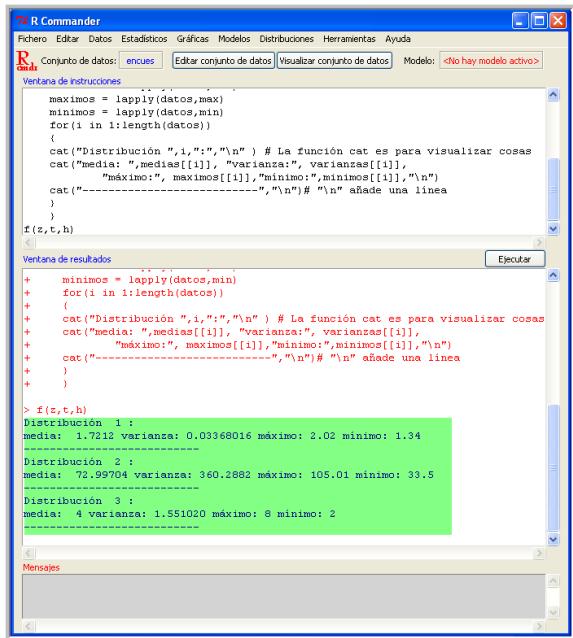


Figura 5: Parámetros estadísticos de “altura”, “peso” y “habitantes” creando funciones en R.

3. Función para realizar la tabla de frecuencias en R

El siguiente programa calcula los valores necesarios para construir una tabla de frecuencias de una variable numérica y presenta en pantalla estos resultados en dos formas: En forma de lista de valores y en forma de matriz.

Esta última opción suministra los mismos valores pero presenta un aspecto más cuidado y ordenado, y por tanto la información es más clara. Este programa, como cualquier otro se puede utilizar sin entenderlo en absoluto. Sin embargo, si se desea avanzar en la programación R es recomendable analizar con detalle cada una de sus líneas e incluso realizar algunas modificaciones en dicho programa.

Programa para obtener una tabla de frecuencias de una variable numérica con R

```

tabla=function(x)
{
  # Cálculo de valores
  nuint=1 + 3.3 * log10(length(x))
  nuint
  pt=pretty(x,n=nuint)
  # pretty(x) calcula los intervalos con un procedimiento interno de R
  ma=pt
  uno=pt[1]
  dos=min(x, na.rm = TRUE)
  if(uno==dos) {pr=min(x)-(pt[2]-pt[1])
  ma=c(pr,pt)}
  a=cut(x,breaks=ma)
  l=length(ma)
  exti=ma[c(1:l-1)]# extremos inferiores
  exts=ma[c(2:l)]# extremos superiores
```

```

marclas=(exti+exts)/2 # marcas de clase
f=table(a) # frecuencias absolutas
xcor=na.omit(x)
nn=sum(f)
fr=f/nn # frecuencias relativas
p=100*fr # porcentajes
fa=cumsum(f) # frecuencias absolutas acumuladas
fra=cumsum(fr) # frecuencias relativas acumuladas
pa=cumsum(p) # porcentajes acumulados
# Lista de valores
cat("-----","\n")
cat("DATOS PARA REALIZAR UNA TABLA DE FRECUENCIAS","\n")
cat("-----","\n")
cat("intervalos:",levels(a),"\n")
cat("marcas de clase:", marclas," \n")
cat("frec. absolutas:",f," \n")
cat("frec. relativas:",fr," \n")
cat("porcentajes:", p," \n")
cat("frec. abs. acum.:",fa," \n")
cat("frec. rel. acum.:",fra," \n")
cat("porcent. acum.:",pa," \n")
cat(" ", "\n")
cat(" ", "\n")
# Tabla de valores. Se define la matriz lista, detallando el nombre de sus columnas.
lista=matrix(c(levels(a),marclas,f,fr,p,fa,fra,pa),nrow=length(f),ncol=8)
colnames(lista)=c(" intervalo", "Marc.de.Cl", "Frec.Ab.", "Frec.Rel",
"Porcent.", "F.R.Acum", "F.A.Acum", "Porcent.Acum")
rownames(lista)=rownames(lista, do.NULL = FALSE, prefix = "int. ")
cat("-----","\n")
cat("TABLA DE FRECUENCIAS CORRESPONDIENTE:","\n")
cat("-----","\n")
lista
}

```

Fin del programa para obtener una tabla de frecuencias de una variable numérica con R.

Ejecutar primero la función **tabla** copiando previamente el código anterior en la ventana de instrucciones de R-Comander. Aplicarla después a los datos de una variable numérica tecleando, por ejemplo:

```
tabla(encues$altura)
```

O bien escribir las instrucciones siguientes:

```
z=encues$altura
```

```
tabla(z)
```

Los datos que se obtienen en la ventana de resultados de R-Commander para construir la tabla de frecuencias de la variable altura del archivo encues pueden verse en la figura 6.

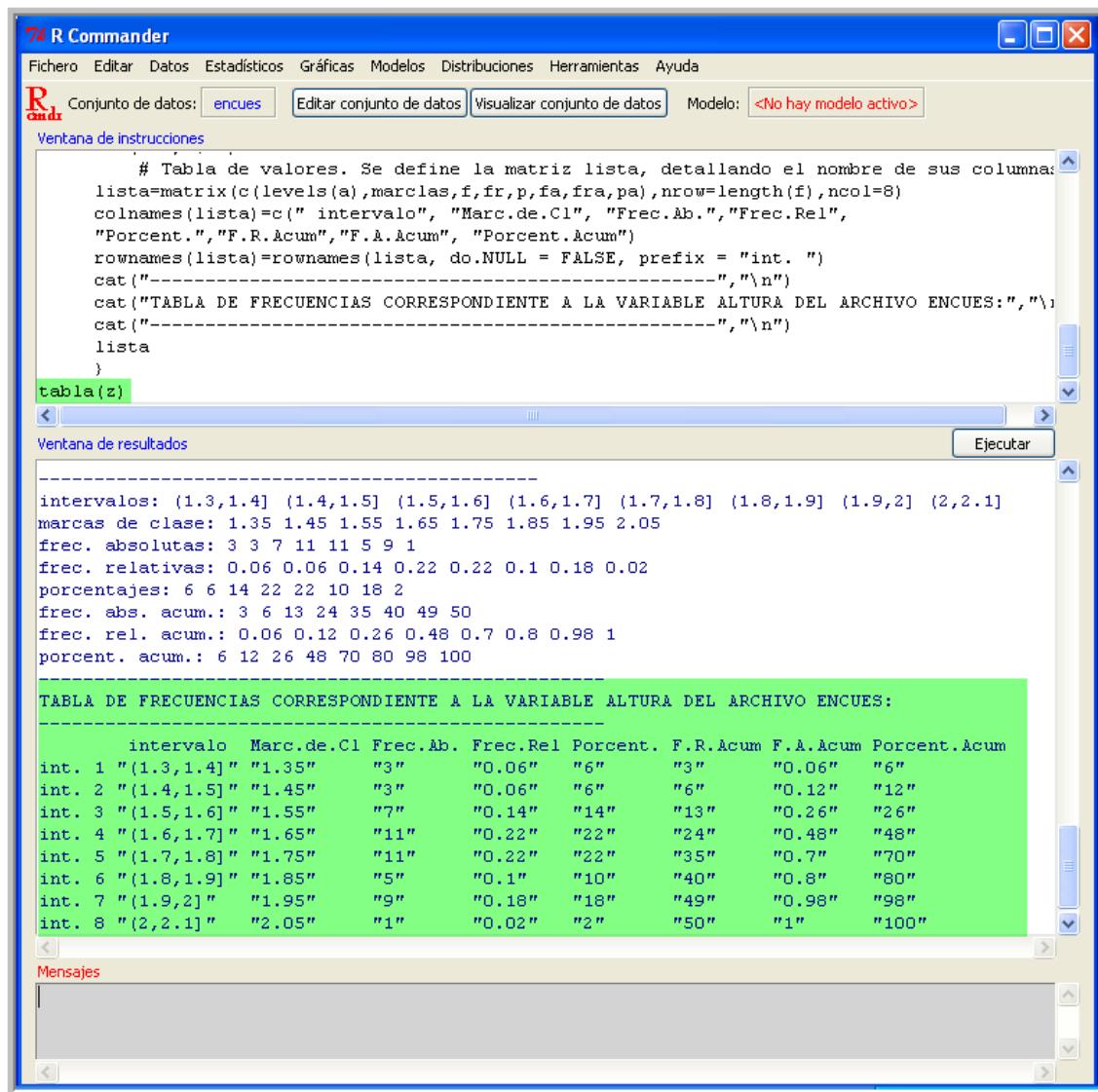


Figura 6: Tabla de frecuencias de la variable altura del archivo encues.

Otras opciones:

Para obtener otro número de intervalos hace falta intercambiar la orden `a=cut(x, pretty(x))` por otra, en la que se especifique el número de intervalos que deseamos hacer. Por ejemplo: `a=cut(x, breaks=6)`.

Si quisieramos especificar la posición de los extremos de los intervalos podemos teclear este otro comando:

`a=cut(x, breaks=c(1,1.5,2.2))` Se definirán dos intervalos de desigual amplitud. Desde luego los valores de los extremos de los intervalos deben ser apropiados para los datos. Es conveniente informarse de los valores mínimo y máximo de nuestra variable. Esto podemos hacerlo mediante la orden `range(x)`

4. Estadística descriptiva bidimensional

En algunas ocasiones estudiamos más de un carácter de una población, por lo que consideramos datos de varias variables. Podemos tratar cada una de ellas independientemente, por ejemplo aplicando las técnicas de estadística descriptiva unidimensional. Pero a menudo nos interesa también ver si existe algún tipo de relación entre los caracteres en estudio. Por ejemplo, si consideramos la edad y la altura de un conjunto de niños podremos observar que, por promedio, su altura va creciendo con la edad. Es decir observaremos una relación entre ambas variables.

En las secciones que describiremos a continuación trataremos del uso del programa R en la aplicación de las técnicas de la estadística descriptiva que nos ayudan a estudiar las relaciones que pueden aparecer entre dos variables. Lo mismo que en el caso de una variable esta información la recogemos por medio de parámetros estadísticos obtenidos a partir de las muestras, o también por medio de tablas y de representaciones gráficas que resumen la información sobre dos o más variables, y que nos descubren hechos sobre los caracteres en estudio que no es posible extraer de los datos numéricos en bruto, ni del estudio estadístico de cada una de las variables por separado.

5. Gráficos de dispersión

Cuando las variables son ambas numéricas, una de estas técnicas bivariantes consiste simplemente en representar en el plano los puntos correspondientes a la variable bidimensional en estudio. Usaremos el fichero Notas obtenido de Manual_Rb.pdf (pág. 339 y sig.) de la Bibliografía Comentada.

Por ejemplo, vamos a representar simultáneamente las variables matA y matC. Para ello, entramos en R-Commander. Una vez que se ha declarado Notas como fichero de datos activo, seguimos desde el menú las instrucciones: Gráficas → Gráfica XY y seleccionamos las variables matA y matC en la ventana emergente, y se obtendrá como resultado la nube de puntos de la figura 7.

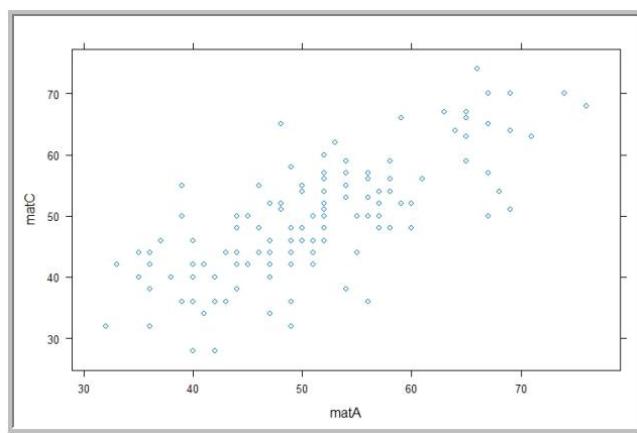


Figura 7: Nube de puntos de las variables matA y matC.

Con ello obtenemos la nube de puntos de la figura 7, en la que observamos que los puntos no se distribuyen por igual por todo el plano, sino que tienen una cierta tendencia a situarse en los alrededores de una de las diagonales. Esto puede interpretarse como que hay alguna relación entre las variables, es decir que no son del todo independientes. Además esta relación puede

concretarse en el siguiente hecho: conforme aumenta el índice de aptitud para las matemáticas (matA) tiende a aumentar el grado de conocimiento en dicha materia (matC).

También es posible representar las distintas gráficas según los niveles de un factor. En este caso la selección en la ventana es la que aparece en la figura 8, donde aparecen seleccionadas las variables anteriores y el factor (o grupo) sexo.

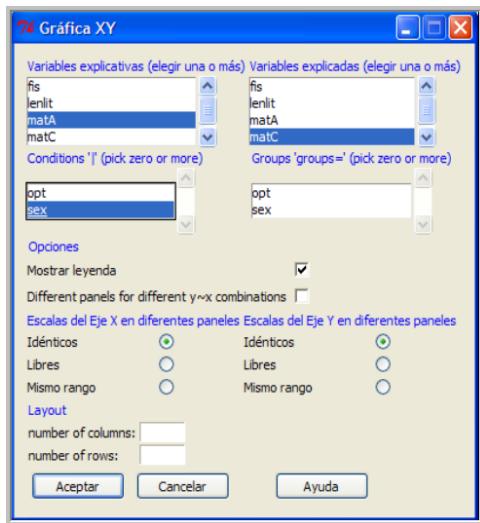


Figura 8: Ventana de selección de variables y factor o grupo “sexo”.

Tras pulsar en el botón Aceptar, se muestran las nubes de puntos, que se observan en la figura 9. Se puede observar que estas nubes de puntos muestran un grado de relación entre ambas variables similar para ambos sexos.

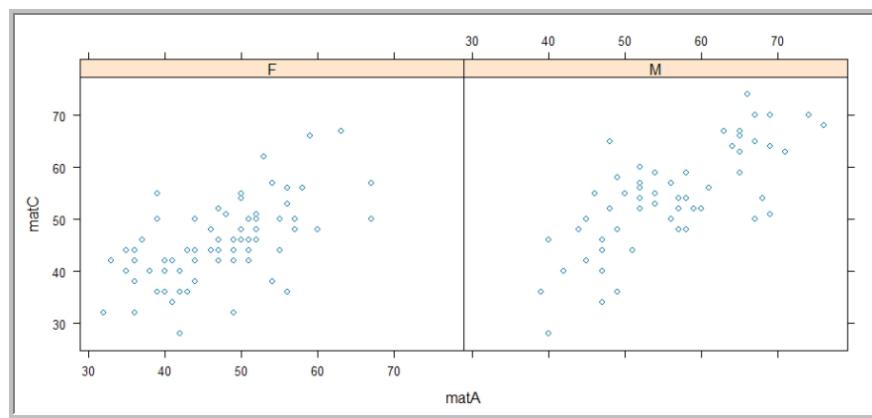


Figura 9: Nubes de puntos de las variables matA y matC según el factor sexo.

6. Covarianza y coeficiente de correlación lineal

Estos parámetros también nos dan una medida de la importancia de la dependencia entre dos variables. Para hallar el coeficiente de correlación lineal entre dos variables, por ejemplo para las variables fis y matC, pulsamos en el menú de R-Commander: Estadísticos → Resúmenes → Matriz de correlaciones. Se seleccionan las dos variables anteriores fis y matC (para seleccionar la segunda variable hay que pulsar la tecla Alt Gr o mantener pulsada la tecla Ctrl).

El resultado que se obtiene en la ventana de R-Commander es la siguiente matriz:

| | fis | matC |
|------|-----------|-----------|
| fis | 1.0000000 | 0.6876345 |
| matC | 0.6876345 | 1.0000000 |

El coeficiente de correlación lineal de ambas variables resulta 0.6876345. El valor 1 es el coeficiente de correlación de las variables fis y matC consigo mismas, que como es lógico toma el valor máximo posible, 1, de este coeficiente.

Si deseamos el valor de la cuasivarianza de ambas variables, sólo tenemos que cambiar el comando cor usado en R-Commander para hallar la matriz de correlaciones por el comando cov. Así resultará:

```
cov(Notas[,c("fis","matC")], use="complete.obs")
```

La salida que se obtiene en la ventana de resultados de R-Commander es la matriz siguiente:

| | fis | matC |
|------|----------|----------|
| fis | 81.79892 | 61.96565 |
| matC | 61.96565 | 99.27481 |

7. Rectas de regresión

Para calcular la recta de regresión de una variable con respecto a otra, por ejemplo matC (como variable dependiente) y matA (variable independiente o explicativa), seleccionamos en R-Commander: Estadísticos → Ajuste de Modelos → Regresión Lineal, rellenando la ventana emergente con las variables indicadas, obtenemos los resultados que se pueden ver en la figura 10.

```
> summary(RegModel.1)

Call:
lm(formula = matC ~ matA, data = Notas)

Residuals:
    Min      1Q  Median      3Q     Max 
-17.0344 -4.0122 -0.1095  5.0764 18.0542 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.41421   3.11646   3.342  0.00109 ** 
matA        0.76107   0.05962 12.765 < 2e-16 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 6.663 on 130 degrees of freedom
Multiple R-squared: 0.5562, Adjusted R-squared: 0.5528 
F-statistic: 162.9 on 1 and 130 DF,  p-value: < 2.2e-16
```

Figura 10: Regresión lineal de la variable matC con respecto a matA.

De esta información destacamos que la recta de regresión de la variable matC con respecto a la variable matA tiene la expresión siguiente:

$$\text{matC} = 0.76107 \times \text{matA} + 10.41421$$

Y el coeficiente de determinación R^2 es igual a 0.5562.

También pueden ser útiles las siguientes funciones:

```
predict.lm(RegModel.1, data.frame(matA=50))
predict(RegModel.1)
residuals(RegModel.1)
coef(RegModel.1)
```

La primera línea nos da el valor de la ordenada de la recta de regresión que corresponde a un valor de matA (en el ejemplo matA=50).

$$\text{matC}=0.76107 \times \text{matA} + 10.41421 = 0.76107 \times 50 + 10.41421 = 48.468$$

En la segunda predict nos suministra la ordenada de la recta de regresión para todos los valores de la variable independiente (matA) contenidos en el fichero de datos.

La instrucción residuals nos devuelve los errores de regresión (o residuos) es decir la diferencia entre los valores de la variable dependiente contenidos en los datos (matC) y los obtenidos con la orden predict.

Por último coef(RegModel.1) nos da ordenada en el origen y la pendiente de la recta de regresión.

A veces es conveniente almacenar en el mismo archivo algunas de las salidas de los comandos junto con las variables ya contenidas en dicho archivo. El conjunto de órdenes siguiente nos va a incorporar dentro de un mismo fichero, que en este caso llamaremos Nmas, los datos contenidos en Notas junto con dos variables más, que van a ser los valores predichos por la regresión (fis ~lenlit) y los residuales o errores cometidos por está:

```
RegModel.2=lm(formula = fis ~lenlit, data = Notas)
summary(RegModel.2)
pred=predict(RegModel.2)
res=residuals(RegModel.2)
Nmas=cbind(Notas,pred,res)
```

Para representar la recta de regresión de la variable “fis” con respecto a “lenlit”, con su nube de puntos desde R-Commander se sigue la cadena de órdenes: Gráficas → Diagrama de Dispersión, con ello nos aparece una ventana que tendremos que seleccionar en la forma que aparece en la figura 11. La recta de regresión resultante es la que aparece en la figura 12.

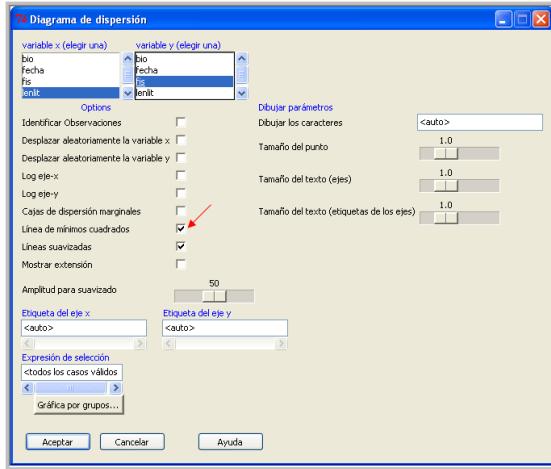


Figura 11: Ventana de selección recta de regresión.

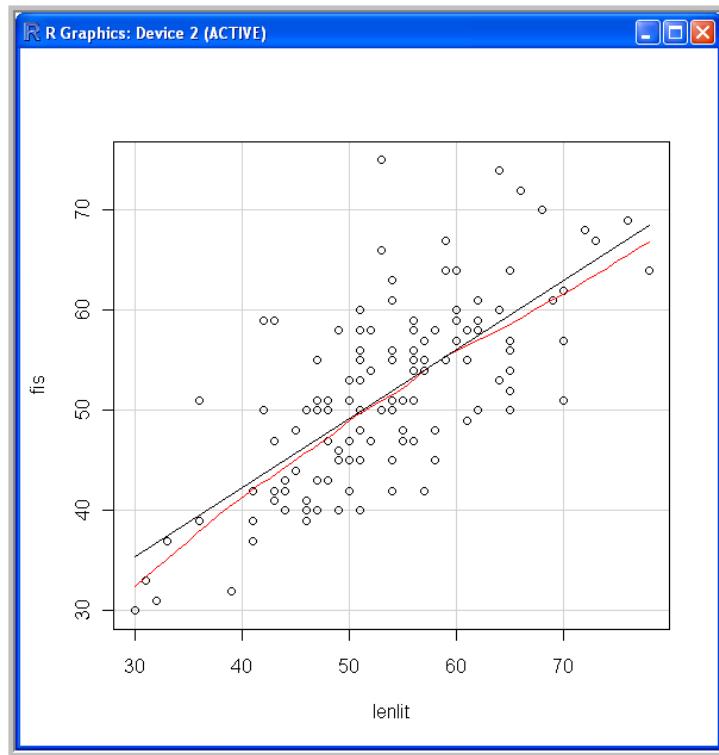


Figura 12: Recta de regresión de la variable lenit con respecto a fis, con la nube de puntos.

Como vemos una cierta curvatura en la línea suavizada podemos intentar un ajuste a alguna función que nos parezca más adecuada que una recta

8. Ajuste de otros modelos lineales

Veamos ahora como podríamos ajustar la nube de puntos de la figura 12 con un polinomio de segundo grado.

Para ello seleccionamos en el menú de R-Commander:

Estadísticos → Ajuste de modelos → Modelo lineal

Con ello se nos abre una ventana que rellenaremos como aparece en la figura 13.

Los resultados del ajuste parabólico se pueden ver en la figura 14.

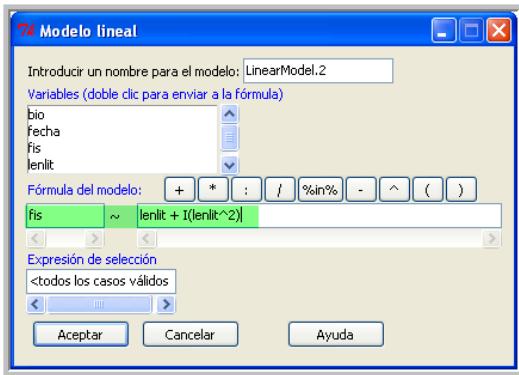


Figura 13: Ventana de selección del modelo de regresión parabólico.

```

Call:
lm(formula = fis ~ lenlit + I(lenlit^2), data = Notas)

Residuals:
    Min      1Q  Median      3Q     Max 
-12.3809 -5.0514 -0.4889  3.4497 23.3287 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.407204  13.412482   0.105  0.9166    
lenlit      1.200346  0.502512   2.389  0.0184 *  
I(lenlit^2) -0.004754  0.004650  -1.022  0.3085    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 6.489 on 129 degrees of freedom
Multiple R-squared: 0.4932, Adjusted R-squared: 0.4853 
F-statistic: 62.76 on 2 and 129 DF,  p-value: < 2.2e-16

```

Figura 14: Datos del ajuste parabólico de la variable “fis” frente a “linlit”.

La función de ajuste resultante es aproximadamente:

$$\text{fis} = 1.407 + 1.2 \text{ lenlit} + 0.005 \text{ lenlit}^2$$

En la tabla 5 se indican las fórmulas que se emplean en R para ajustar otros modelos:

| Modelo | Expresión | Fórmula en R |
|---------------------|------------------|-------------------|
| recta | $a+bX$ | X |
| recta sin constante | aX | $X-1$ |
| Cuadrático | $a+bX+cX^2$ | $X+I(X^2)$ |
| Cúbico | $a+bX+cX^2+dX^3$ | $X+I(X^2)+I(X^3)$ |
| logarítmico | $a+b\log(X)$ | $\log(X)$ |
| Inverso | $a+(b/X)$ | $I((1/X))$ |

Tabla 2. Fórmulas para distintos tipos de ajustes.

9. Conclusiones y agradecimientos

Hoy día el uso de programas estadísticos se ha convertido en una rutina a la hora de analizar los datos de las investigaciones en la mayoría de las disciplinas científicas. Los Estadísticos nos encontramos con frecuencia con peticiones de ayuda por parte de autores y directores de investigaciones que manifiestan tener problemas con el uso de estos programas, o, peor aún, con la interpretación de los resultados y los gráficos que han devuelto estos programas.

Es una práctica común en algunas disciplinas científicas confiar el análisis de datos a paquetes estadísticos comerciales, de los que se extraen más estadísticos que información. Es imposible obtener respuestas a medida, porque estos programas proporcionan técnicas y gráficas a la medida que el programador decidió en su día. Algunos programas se han diseñado para que sea el investigador el que decida qué respuestas quiere. Estos programas, como por ejemplo el software estadístico R, son además gratuitos.

En este Taller se ha pretendido proporcionar una introducción al programa R, que representa una alternativa a programas estadísticos de uso más extendido. A través de la utilización de este software libre se ha ido poniendo de manifiesto cómo es el propio autor del trabajo quien toma las decisiones con respecto a qué estrategias, gráficos y técnicas utilizar, así como sobre el formato de presentación que prefiere para sus resultados y sus gráficos. Es sencillamente imposible utilizar R sin criterio. Se debe comenzar por saber lo que se quiere antes de dar un sólo clic con el ratón sobre un menú en un programa estadístico.

Creemos además que el programa R proporciona todas las opciones necesarias para utilizarlo en la práctica docente universitaria. Aunque obtener resultados es más costoso inicialmente, a la larga, los resultados son mucho más concretos, adaptables y modificables. Debemos tener en cuenta que R no es sólo un programa de cálculo estadístico, sino que es un lenguaje de programación estadístico, con el que podemos implementar cualquier procedimiento o técnica estadística, por novedosa que ésta sea. Además, en la actualidad existen más de 2000 librerías o paquetes disponibles para casi cualquier propósito.

Creemos que R es un entorno de programación en el que se han incluido rutinas que permiten de forma fácil y directa la realización de multitud de cálculos estadísticos; también la confección de gráficas e imágenes de una sencillez y una potencia equiparable, o superior, a la de otros programas comerciales. Por todo ello, os animamos a que utilicéis el software libre R tanto en el ámbito docente como en el de la investigación.

Queremos agradecer a los organizadores de las I Jornadas de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa, celebradas en Melilla en junio de 2010, que hayan pensado en nosotros para el desarrollo de este Taller de R. Queremos también agradecer a los integrantes del Grupo de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa (GENAEIO) de la SEIO por su participación y su especial motivación en el desarrollo de este Taller. Esperamos y deseamos que el desarrollo del mismo haya resultado interesante y productivo para todos.

10. Bibliografía y documentación

- [1] Introducción a R: <http://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- [2] Manual resumido de R: <http://cran.r-project.org/doc/contrib/curso-R.Diaz-Uriarte.pdf>
- [3] Diverso material docente: <http://www.ub.es/stat/docencia/EADB/Curso basico de R.htm>
- [4] Para principiantes: http://es.geocities.com/r_vaquerizo/Manual_R3.htm

[5] Introducción al uso del R Commander:

[http://geminis.dma.ulpgc.es/profesores/personal/asp/
Documentacion/Manual_R_commander.pdf](http://geminis.dma.ulpgc.es/profesores/personal/asp/Documentacion/Manual_R_commander.pdf)

[6] R para principiantes: http://cran.r-project.org/doc/contrib/rdebuts_es.pdf

[7] Amplio manual sobre gráficos estadísticos con R:

<http://cran.r-project.org/doc/contrib/grafi3.pdf>

[8] Curso básico de R: http://www.ub.es/stat/docencia/EADB/Curso_basico_de_R.pdf

[9] Introducción a R: http://www.cesma.usb.ve/~lbravo/co5314/MANUAL_Rb.pdf

Taller de R (Distribuciones e Inferencia)

Antonio Gámez Mellado¹, Luis Miguel Marín Trechera²

¹antonio.gamez@uca.es, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz
²luis.marin@uca.es, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz
 Universidad de Cádiz

Resumen

En esta tercera parte del Taller de R se muestra el manejo de distribuciones estadísticas usando R, así como una breve introducción a la inferencia estadística con el software libre R. Asimismo vamos a estudiar en esta última parte del Taller algunas de las distribuciones de probabilidad más usuales, y abordaremos también la forma de resolver algunos problemas sencillos relacionados con la inferencia estadística usando R.

Palabras clave: Software libre, R, Estadística

Clasificación AMS: 97U50, 97K40, 97K80

1. La distribución Binomial.

El manejo de la distribución Binomial con R-Commander lo vamos a ilustrar con el siguiente ejemplo:

Considerando la variable aleatoria X que sigue una distribución $B(5, 0.3)$, se pide:

- a) Represente gráficamente sus funciones de probabilidad y de distribución.
- b) Calcule los valores de la función de densidad y de la función de distribución.
- c) Calcule el cuantil correspondiente a 0.5.
- d) Calcule las probabilidades siguientes: $P(X=3)$, $P(X \leq 3)$, $P(X \geq 3)$, $P(X < 3)$, $P(X > 3)$
- e) Genere dos muestras con 7 elementos cada una, procedentes de esta variable.

a) Para representar gráficamente sus funciones de probabilidad y de distribución, seleccionamos sucesivamente desde el menú principal de R-Commander las opciones:

Distribuciones → Distribuciones Discretas → Distribución binomial → Gráficas de la distribución binomial

Esta selección puede verse en la figura 1. A continuación, aparece una ventana emergente que se completa tal como aparece en la figura 2.

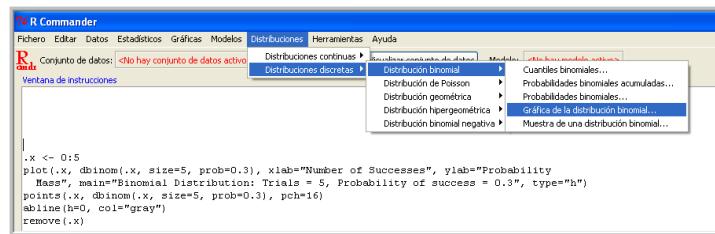


Figura 1: Menú de selección para gráficas de la distribución binomial.

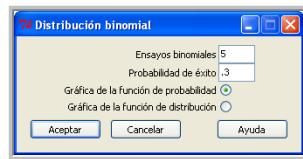
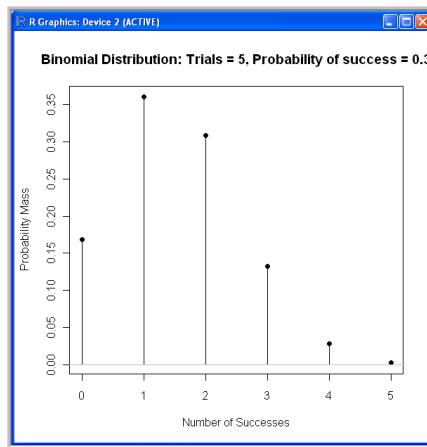


Figura 2: Ventana de selección de la función de probabilidad binomial.

De esta forma obtenemos la representación gráfica de la función de probabilidad de la distribución $B(5, 0.3)$ que se puede ver en la figura 3. Las órdenes generadas por R-Commander son las siguientes:

```
.x <- 0:5
plot(.x, dbinom(.x, size=5, prob=0.3), xlab="Number of Successes",
      ylab = "Probability Mass", main="Binomial Distribution: Trials = 5,
      Probability of success = 0.3", type="h")
points(.x, dbinom(.x, size=5, prob=0.3), pch=16)
abline(h=0, col="gray")
remove(.x)
```

Figura 3: Función de probabilidad de la Binomial $B(5,0.3)$.

Para representar la función de distribución basta seleccionar en el mismo menú anterior de la figura 3 de R-Commander, la segunda opción (Gráfica de la función de distribución), y se obtiene la gráfica de la figura 4.

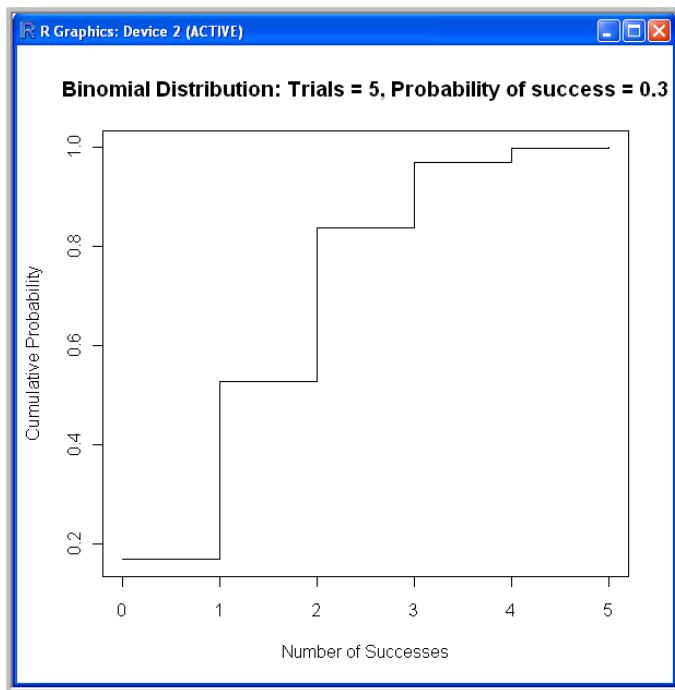


Figura 4: Función de distribución de la Binomial B(5,0.3).

b) Para calcular los valores de la función de densidad y de la función de distribución se seleccionan las opciones del menú de R-Commander:

Distribuciones → Distribuciones Discretas → Distribución binomial → Probabilidades binomiales.

Se obtienen los siguientes valores para la función de probabilidad.

0 0.16807
 1 0.36015
 2 0.30870
 3 0.13230
 4 0.02835
 5 0.00243

Los valores de la función de distribución siguiendo las siguientes órdenes del menú o las siguientes instrucciones desde la ventana de instrucciones:

Distribuciones → Distribuciones Discretas → Distribución binomial → Probabilidades binomiales acumuladas

La ventana emergente que aparece se rellena tal y como aparece en la figura 5.

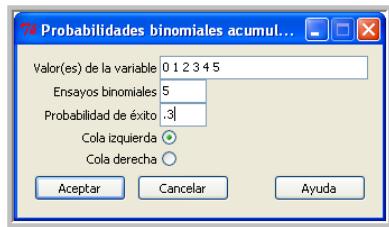


Figura 5: Valores de la función de Distribución de la binomial.

La instrucción y los valores que se generan en la ventana de salida son los siguientes:

```
pbinom(c(0,1,2,3,4,5), size=5, prob=0.3, lower.tail=TRUE)
[1] 0.16807 0.52822 0.83692 0.96922 0.99757 1.00000
```

c) Para calcular el valor del cuantil correspondiente a 0.5 se sigue la siguiente secuencia en el menú de R-Commander:

Distribuciones → Distribuciones Discretas → Distribución binomial → Cuantiles binomiales

Se rellena la ventana emergente tal y como aparece en la figura 6. De esta forma se obtiene lo siguiente:

```
qbinom(c(0.5), size=5, prob=0.3, lower.tail=TRUE)
[1] 1
```

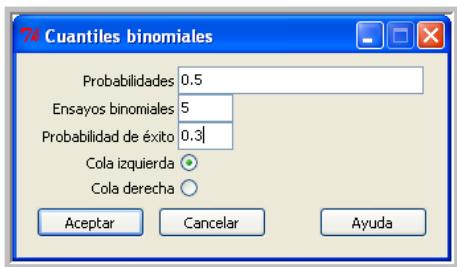


Figura 6: Ventana para el cálculo de cuantiles.

En realidad tenemos que $P(X \leq 0) = 0.16807 = P(X < 1) < 0.5 < P(X < 2) = P(X \leq 1) = 0.52822$.

d) Para calcular las probabilidades $P(X=3)$, $P(X \leq 3)$, $P(X \geq 3)$, $P(X < 3)$, $P(X > 3)$ tenemos:

Para cada uno de estos valores detallamos la secuencia que se selecciona en el menú, la orden generada y/o el resultado obtenido por R-Commander:

- $P(X=3)$:

Distribuciones → Distribuciones Discretas → Distribución binomial → Probabilidades binomiales.

Se obtienen todos los valores de la función de probabilidad. Se selecciona el valor deseado:

3 0.13230

- $P(X \leq 3)$:

Distribuciones → Distribuciones Discretas → Distribución binomial → Probabilidades binomiales acumuladas (cola izquierda)

```
pbinom(c(3), size=5, prob=0.3, lower.tail=TRUE)
```

```
[1] 0.96922
```

- $P(X \geq 3)$:

Distribuciones → Distribuciones Discretas → Distribución binomial → Probabilidades binomiales acumuladas (cola derecha y valor 2)

```
> pbinom(c(2), size=5, prob=0.3, lower.tail=FALSE)
```

```
[1] 0.16308
```

- $P(X < 3) = P(X \leq 2)$

```
0.83692
```

- $P(X > 3) = P(X \geq 4)$:

Distribuciones → Distribuciones Discretas → Distribución binomial → Probabilidades binomiales acumuladas (cola derecha y valor 3) y se obtiene como resultado:

```
0.03078
```

e) Para generar dos muestras de 7 elementos de esta variable aleatoria, seleccionamos: Distribuciones → Distribuciones Discretas → Distribución binomial → Muestra de una distribución binomial, y rellenando la ventana emergente tal como la de la figura 47, las muestras aparecen en el fichero activo que se denomina Muestras_binomiales. Visualizando los datos que se han obtenido en esta ocasión, obtenemos las muestras especificadas en la figura 8. Los números de la primera fila corresponden a la primera muestra. En la primera prueba se han obtenido dos éxitos, en la segunda prueba se ha obtenido un único éxito, en la tercera 3 éxitos, etc.

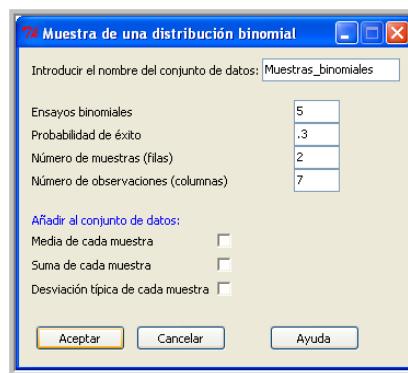


Figura 7: Ventana de selección de muestras para la Distribución B(5, 0.3).

| | obs1 | obs2 | obs3 | obs4 | obs5 | obs6 | obs7 |
|---------|------|------|------|------|------|------|------|
| sample1 | 2 | 1 | 3 | 2 | 1 | 0 | 2 |
| sample2 | 2 | 1 | 1 | 3 | 3 | 1 | 0 |

Figura 8: Datos de las dos muestras de tamaño 7 de la distribución B(5, 0.3).

Es de notar que las muestras son diferentes para cada ejecución del programa, salvo que se inicie el programa con la misma semilla. Esto se realiza con la orden `set.seed(n)`, donde n debe ser un entero concreto cada vez que se desee repetir la misma secuencia de números aleatorios.

Por ejemplo, Para generar dos secuencias diferentes se puede ejecutar dos veces:

```
a<-rbinom(5*1, size=5, prob=0.3)
```

```
a
```

Para generar dos secuencias iguales podemos ejecutar dos veces

```
set.seed(12)
```

```
a<-rbinom(5*1, size=5, prob=0.3)
```

```
a
```

Cambiando la semilla 12 por otro valor se puede volver a cambiar la secuencia.

2. Teorema Central del Límite y su simulación

El Teorema Central del Límite nos asegura que si una variable aleatoria X tiene media μ y varianza finita σ^2 , la distribución de las medias de sus muestras de tamaño n tiende hacia una $N(\mu, \sigma/\sqrt{n})$ conforme aumenta el tamaño de la muestra.

Distribución de la media muestral

Dada una variable aleatoria X, de media μ y varianza finita σ^2 , seleccionando distintas muestras de n elementos y hallando sus respectivas medias, obtendremos, por regla general, un resultado diferente para cada una de estas medias muestrales, debido a la influencia que tiene el azar en la selección de cada muestra. Por tanto la media de las muestras de tamaño n es otra variable aleatoria. Denotemos esta variable aleatoria “media muestral” por X . ¿Qué distribución sigue esta variable aleatoria? No hay una respuesta genérica ya que depende de las distribuciones de partida. Pero si el número de elementos de la muestra es amplio, el Teorema Central del Límite nos da una solución aproximada para la distribución del estadístico media muestral: puesto que nos asegura que si una variable aleatoria X tiene media μ y varianza finita σ^2 , la distribución de las medias de sus muestras de tamaño n tiende hacia una $N(\mu, \sigma/\sqrt{n})$ conforme aumenta el tamaño de la muestra.

Es decir que la distribución de la variable media muestral se va aproximando a la distribución Normal al ir creciendo el valor de n.

Simulación del Teorema Central del Límite usando R

Para visualizar este teorema, usando muestras simuladas, podemos seguir los pasos:

Paso 1: Generar N muestras de n elementos cada una que sigan la función de distribución de la variable X.

Paso 2: Calcular las medias de cada una de estas N muestras.

Paso 3: Considerando la muestra M, de N elementos, formada con las N medias obtenidas en el Paso 2, calcular:

3.a) La media de M (la media de las medias muestrales de X).

3.b) La cuasidesviación típica de M , ya que la cuasivarianza muestral es un estimador insesgado de la varianza poblacional.

3.c) Realizar una tabla de frecuencias y un histograma de frecuencias para la variable M .

Paso 4 (Comprobación):

4.a) Se comparará el valor obtenido en 3.a) con μ .

4.b) Se comparará el valor obtenido en 3.b) con σ/\sqrt{n} .

4.c) Se comparará la tabla de frecuencias con las probabilidades correspondientes a los intervalos de la tabla que corresponderían a la distribución $N(\mu, \sigma/\sqrt{n})$ indicada en la tesis del teorema. Igualmente se comparará el histograma de frecuencias obtenido en 3.c) con la representación gráfica de la distribución teórica $N(\mu, \sigma/\sqrt{n})$.

Se podrá apreciar un buen acuerdo entre la simulación y el enunciado del teorema si se tienen las siguientes precauciones: El valor de N (número de muestras generadas) debe ser lo bastante grande para permitir dibujar un histograma de frecuencias razonable. Es recomendable que el número de elementos, n , de cada muestra sea al menos 30, salvo que la distribución de partida sea normal, en cuyo caso n puede tomar cualquier valor puesto que la media de una normal se distribuye como una normal exacta.

Vamos a proponer tres ejemplos. En todos ellos las cinco primeras líneas están realizadas usando R-Commander con el procedimiento siguiente:

Distribuciones → Distribuciones Discretas → Distribución binomial → Muestra de una distribución binomial. Las dos últimas son para realizar la gráfica que permite comparar el histograma de frecuencias de la variable media muestral con la gráfica de la función de densidad prevista en cada caso por el Teorema Central del Límite.

Ejemplo 1: Simulación de 1000 muestras de 5 elementos de la $B(5,0.15)$. (Ver figura 9).

```
Muestras_binomiales<-as.data.frame(matrix(rbinom(1000*5, size=5, prob=0.15), ncol=5))
rownames(Muestras_binomiales)<- paste("sample", 1:1000, sep="")
colnames(Muestras_binomiales)<- paste("obs", 1:5, sep="")
Muestras_binomiales$mean <- rowMeans(Muestras_binomiales[,1:5])
Hist(Muestras_binomiales$mean, prob=T, scale="frequency", breaks="Sturges")
curve(dnorm(x,mean=0.75, sd=0.3570714),from=-0.5,to=1.925, add=T)
```

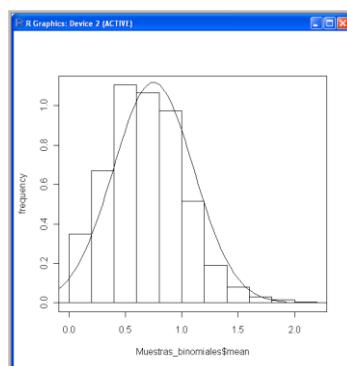


Figura 9: Gráfica correspondiente al Ejemplo1.

Ejemplo 2: Simulación de 10000 muestras de 30 elementos de la $B(5,0.15)$. (Ver figura 10).

```
Muestras_binomiales<-as.data.frame(matrix(rbinom(10000*30,size=5,prob=0.15),ncol=30))
rownames(Muestras_binomiales)<- paste("sample", 1:10000, sep="")
colnames(Muestras_binomiales)<- paste("obs", 1:30, sep="")
Muestras_binomiales$mean <- rowMeans(Muestras_binomiales[,1:30])
Hist(Muestras_binomiales$mean, prob=T, scale="frequency", breaks="Sturges")
curve(dnorm(x,mean=0.75,sd= 0.1457738), from=-0.5, to=1.925, add=T)
```

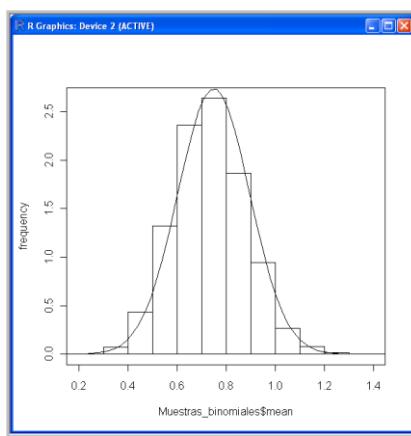


Figura 10: Gráfica correspondiente al ejemplo 2.

Ejemplo 3: Simulación de 10000 muestras de 300 elementos de la $B(5,0.15)$. (Ver figura 11).

```
Muestras_binomiales<-as.data.frame(matrix(rbinom(10000*300,size=5,prob=0.15), ncol=300))
rownames(Muestras_binomiales)<- paste("sample", 1:10000, sep="")
colnames(Muestras_binomiales)<- paste("obs", 1:300, sep="")
Muestras_binomiales$mean <- rowMeans(Muestras_binomiales[,1:300])
Hist(Muestras_binomiales$mean, prob=T, scale="frequency", breaks="Sturges")
curve(dnorm(x,mean=0.75, sd=0.046098),from=0.5,to=0.95, add=T)
```

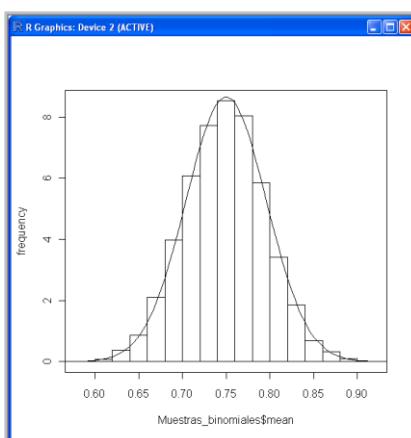


Figura 11: Gráfica correspondiente al ejemplo 3

Este último caso, el de la figura 11, es el más ajustado de los tres, ya que el histograma está realizado con 10000 valores y el tamaño de la muestra de la que se hayan las medias es bastante grande (300 elementos). Calculando la media y la desviación típica de la variable que contiene la media de cada muestra (Muestras_binomiales\$mean), resulta lo siguiente:

```
mean(Muestras_binomiales$mean)
[1] 0.7507283
sd(Muestras_binomiales$mean)
[1] 0.04633999
```

Los valores teóricos suministrados por el Teorema Central del Límite para este caso serían:

$$\mu = np = 5 \times 0.15 = 0.75$$

$$\sigma = \sqrt{npq} / \sqrt{N} = \sqrt{5 \times 0.15 \times 0.85} / \sqrt{300} = 0.046098$$

Así que hay un buen acuerdo entre la distribución $N(\mu, \sigma / \sqrt{n}) = N(0.75, 0.046098)$ y los valores de las medias de las muestras de 300 elementos de la $B(5, 0.15)$ del ejemplo 3.

3. La distribución Normal

Mostraremos a continuación cómo se puede manejar R-Commander para obtener distintos elementos de la distribución normal a través de ejemplos de la distribución $X \sim N(50, 5)$.

Ejemplo 4: Dibujar las funciones de densidad y de distribución de la distribución $N(50, 5)$.

Para la gráfica de la función de densidad comenzamos seleccionando desde el menú de R-Commander: Distribuciones → Distribuciones Continuas → Distribución Normal → Graficas de la distribución Normal, rellenando la ventana emergente tal como la de la figura 12.

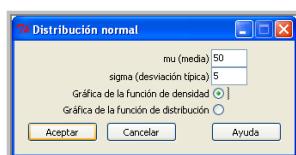


Figura 12: Selección de la gráfica de la función de densidad de la normal.

Para la gráfica de la función de distribución seguimos el mismo proceso seleccionando en la misma ventana de la figura 12 la segunda opción.

Ejemplo 5: Hallar la probabilidad de que la variable X tome un valor menor o igual que 45.

Elegimos en el menú de R-Commander: Distribuciones → Distribuciones Continuas → Distribución Normal → Probabilidades Normales. Introducimos los datos como en la figura 13.

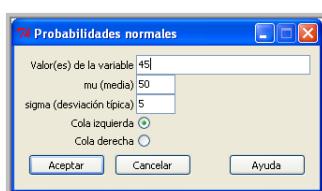


Figura 13: Selección de la probabilidad de la función de distribución de la normal.

El resultado obtenido es 0.1586553

Ejemplo 6: Hallar la probabilidad de que la variable X tome un valor mayor que 57. Seleccionar en este caso la cola derecha como en la figura 14.

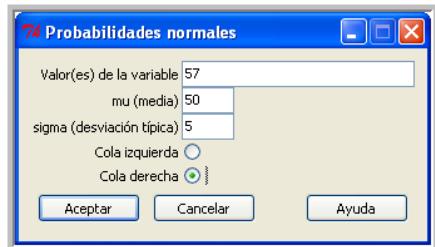


Figura 14: Función de distribución de la normal (cola derecha).

El resultado que obtenemos es el siguiente:

> pnorm(c(57), mean=50, sd=5, lower.tail=FALSE)

$$P(X < 57) = P(X \leq 57) = 0.08075666.$$

Ejemplo 7: Calcule la probabilidad siguiente: $P(53 < X < 60)$.

Calculamos la función de distribución en los valores 53 y 60, tal como se ve en la figura 15.

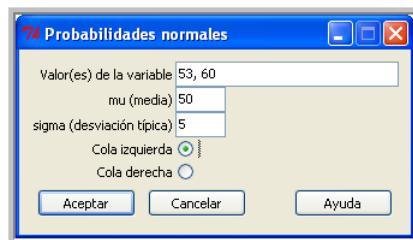


Figura 15: Probabilidad de varios valores de la distribución normal.

Resultan los valores 0.7257469 y 0.9772499, por lo que obtenemos:

$$P(53 < X < 60) = P(X < 60) - P(X < 53) = 0.9772499 - 0.7257469 = 0.251503$$

Ejemplo 8: Genera 10000 muestras de 1 elemento cada una de una variable $X \sim N(50, 5)$. Calcula la frecuencia relativa del suceso $A = \{X / 53 < X < 60\}$ y comparar el resultado con la probabilidad teórica obtenida en el ejemplo 7.

Seleccionamos Distribuciones → Distribuciones Continuas → Distribución Normal → Muestra de una distribución normal. Se rellena la ventana como en la figura 16. En este caso la muestra generada aparecerá en la primera columna.

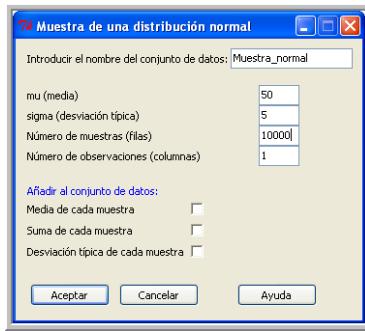


Figura 16: Generación de 10000 muestras de una distribución N(50, 5).

Para contar los elementos de A recurrimos al procedimiento Datos → Modificar variables del conjunto de datos activos → Recodificar variables, rellenando la ventana emergente como en la figura 17.



Figura 17: Ventana emergente para contar los elementos de A.

Para contar el número de "síes" pulsamos en el menú de R-Commader: Estadísticos → Resúmenes → Distribución de frecuencias, obteniendo lo siguiente:

```
.Table <- table(Muestras_normales$var_cod)
Table # counts for varcod
no si
7479 2521
100*.Table/sum(.Table) # percentages for varcod
no   si
74.79 25.21
remove(.Table)
```

El valor teórico obtenido en el ejemplo 7 es 0.2515. El valor experimental, es decir el obtenido con 10000 muestras de 1 elemento cada una, resulta 0.2521. El resultado, es aleatorio, así que puede haber ligeras variaciones al ir cambiando la muestra generada.

Ejemplo 9: Calcular el valor de A, B, C tales que $P(X \leq A) = 0.20$, $P(X \leq B) = 0.40$, $P(X \leq C) = 0.95$.

Entrando en Distribuciones → Distribuciones Continuas → Distribución Normal →Cuantiles normales y rellenando la ventana como la de la figura 18.

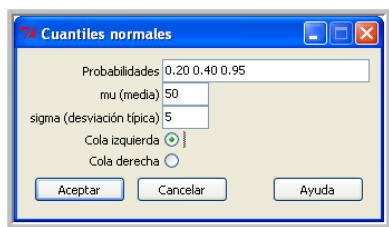


Figura 18: Ventana cuantiles de una distribución normal.

Obtenemos lo siguiente:

```
> qnorm(c(0.20,0.40,0.95), mean=50, sd=5, lower.tail=TRUE)
[1] 45.79189 48.73326 58.22427
```

Es decir, obtenemos que $A = 45.79189$, $B = 48.73326$, $C = 58.22427$.

Ejemplo 10: Representar en el mismo gráfico las funciones de densidad de las distribuciones $N(0.75,0.3)$, $N(0.75,0.5)$, $N(0.75,0.7)$.

Estas representaciones nos pueden servir para ver la influencia que tiene el valor de la desviación típica en la representación gráfica de la función de densidad.

Se pueden ejecutar las siguientes líneas de programa en la ventana de instrucciones de R-Commander, y se obtiene el gráfico de la figura 19:

```
curve(dnorm(x,mean=0.75, sd=0.3), from=-1.25,to=2.75)
curve(dnorm(x,mean=0.75, sd=0.5), from=-1.25,to=2.75, add=T,col="red")
curve(dnorm(x,mean=0.75, sd=0.7), from=-1.25,to=2.75, add=T,col="blue")
```

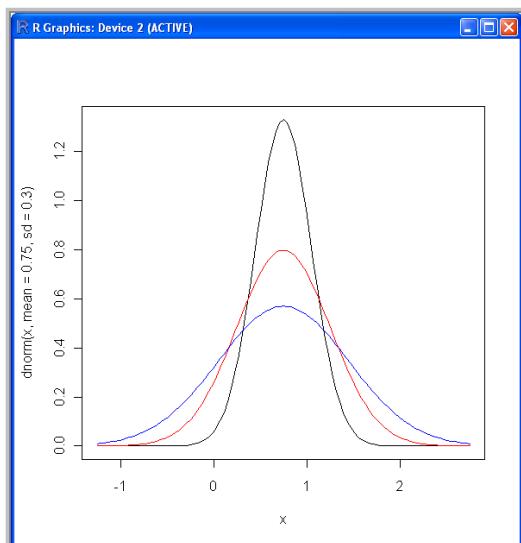


Figura 19: Ventana distribuciones normales, $N(0.75, 0.3)$, $N(0.75, 0.5)$ y $N(0.75, 0.7)$.

4. La distribución uniforme. Las funciones de distribución en R

En los apartados previos hemos trabajado con R-Commander. Sin embargo es bastante sencillo trabajar desde la consola de RGui. R dispone de una gran colección de funciones de distribución.

Vamos a ver las funciones de R relacionadas con la distribución uniforme U(2,5):

```
dunif(x, min=2, max=5) # Valor de la función de densidad de la U(min,max) en x.  
punif(q, min=2, max=5) # Valor de su función de distribución en q.  
qunif(p, min=2,max=5) # Función de distribución inversa o cuantiles.  
runif(n, min=2,max=5) # Generación de una muestra de n elementos de una U(min,max).
```

De manera similar se pueden manejar las restantes funciones. Todas ellas siguen el mismo convenio de iniciales: d, p, q, r, seguido del nombre de la distribución. El primer argumento siempre tiene el mismo significado para todas ellas:

x, q vectores de cuantiles
p vector de probabilidades
n número de observaciones a generar

En la siguiente tabla hacemos un resumen de estas funciones:

Distribuciones de probabilidad

| Nombre en R | Distribución | Parámetros | Por defecto |
|-------------|--------------------|-----------------|-------------|
| unif | Uniforme | min, max | 0 , 1 |
| norm | Normal | mean, sd | 0 , 1 |
| Inorm | Log Normal | meanlog, sdlog | 0 , 1 |
| t | t-Student | df | |
| cauchy | Cauchy | location, scale | 0 , 1 |
| exp | Exponencial | rate | |
| beta | Beta | shape1, shape2 | |
| f | F –Fisher-Snedecor | df1, df2 | |
| gamma | Gamma | shape,scale | - , 1 |
| weibull | Weibull | shape, scale | - , 1 |
| logis | Logística | location, scale | 0 , 1 |
| chisq | Chi-cuadrado | df | |
| binom | Binomial | size, prob | |
| pois | Poisson | lambda | |
| nbinom | Binomial negativa | size, prob | |
| geom | Geométrica | prob | |

| | | | |
|-------|-----------------|---------|--|
| hyper | Hipergeométrica | m, n, k | |
|-------|-----------------|---------|--|

En relación con las funciones de distribución pueden tener interés las siguientes consideraciones:

Se pueden obtener los valores mínimo, primer cuartil, mediana, tercer cuartil, máximo de la muestra con la función fivenum():

```
fivenum(rnorm(1000))
```

La instrucción sample(x,size,replace=TRUE,prob) obtiene una muestra de tamaño “size” de x, con o sin reemplazamiento (replace=TRUE, replace=FALSE), pudiendo tener los elementos de x probabilidades distintas a la uniforme, que es la distribución que se supone por defecto).

Ejemplo 11:

```
x<-c(1:10)
```

```
sample(x,3)
```

```
[1] 9 5 7
```

```
sample(x) # Permutaciones de los elementos de x
```

```
[1] 5 2 6 1 4 7 3 10 9 8
```

```
y<-sample(5:15,5)
```

```
y
```

```
[1] 5 7 12 8 15
```

Muestra idénticamente distribuida:

```
sample (1:3,replace=TRUE,10)
```

```
[1] 3 2 3 2 3 3 3 1 3 3
```

Muestras con asignación de probabilidad a cada elemento, P(1)=0.1, P(2)=0.4, P(3)=0.5:

```
sample (c(1,2,3),prob=c(0.1,0.4,0.5),replace=TRUE, 20)
```

```
[1] 1 2 3 3 2 3 3 3 2 3 3 3 2 1 2 2 1
```

5. Inferencia Estadística

En esta sección describiremos algunos aspectos relacionados con la inferencia estadística.

5.1. La distribución muestral de la media. Intervalos de confianza para la media

En esta primera parte trataremos algunas consideraciones teóricas. El cálculo práctico de un intervalo de confianza y la realización de tests de hipótesis a partir de una muestra concreta se hará en el siguiente apartado.

La distribución de la media de las muestras de tamaño n de una variable aleatoria de media μ y desviación típica σ que siga una distribución normal es $N(\mu, \sigma/\sqrt{n})$ independientemente del tamaño n de la muestra. Si la distribución de dicha variable aleatoria no fuera normal, en virtud del Teorema Central del Límite, la distribución de la media de las muestras de tamaño n se va aproximando a la distribución $N(\mu, \sigma/\sqrt{n})$ según aumenta el tamaño de la muestra.

De ello se deduce que la variable $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ sigue una distribución $N(0,1)$ si la distribución de partida fuera normal. Si no fuera así, se aproximaría a la $N(0,1)$ si el tamaño de la muestra fuera suficientemente grande. A efectos prácticos suele considerarse suficiente que n sea mayor o igual que 30 ($n \geq 30$). Del citado teorema se deduce que si la distribución de partida es normal se cumple:

$$P[|(\bar{X} - \mu)/(\sigma/\sqrt{n})| < z_{\alpha/2}] = 1 - \alpha$$

o equivalentemente

$$P[\bar{X} - z_{1-\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{X} + z_{1-\alpha/2}(\sigma/\sqrt{n})] = 1 - \alpha$$

Si la distribución de la muestra no es normal entonces las dos igualdades anteriores se cumplen aproximadamente si $n \geq 30$.

El intervalo $(\bar{X} - z_{1-\alpha/2}(\sigma/\sqrt{n}), \bar{X} + z_{1-\alpha/2}(\sigma/\sqrt{n}))$ se llama intervalo de confianza para la media al $100 \times (1 - \alpha)\%$. Esto significa que se espera que este intervalo contenga el valor de μ en el $100 \times (1 - \alpha)\%$ de las muestras generadas. Suelen emplearse intervalos de confianza, obtenidos a partir de una única muestra, para acotar el valor de μ . Sin embargo, al ser la muestra aleatoria, no es seguro que el intervalo de confianza contenga el verdadero valor de μ . No sabemos, ni podemos saber, si la muestra que usamos pertenece al $100 \times (1 - \alpha)\%$ muestras que contienen el valor de μ , o al $100 \times \alpha\%$ de muestras que no lo contienen.

El siguiente programa de R nos permite generar 50 muestras aleatorias de 30 elementos cada una, procedentes de una distribución $N(3,1)$ y representar gráficamente los respectivos intervalos de confianza junto con la media teórica, que es 3.

```
Muestras_normales<-as.data.frame(matrix(rnorm(50*30,mean=3,sd=1), ncol=30))
rownames(Muestras_normales)<- paste("sample", 1:50, sep="")
colnames(Muestras_normales)<- paste("obs", 1:30, sep="")
Muestras_normales$mean <- rowMeans(Muestras_normales[,1:30])
amp=alfa=0.05; amp=qnorm(1-alfa/2)*1/sqrt(30)
matplot(rbind(Muestras_normales$mean-amp,Muestras_normales$mean+amp),
        rbind(1:50,1:50),type="l",lty=1)
abline(v=3) # draw line for media verdadera.
```

Para generar las muestras y la correspondiente media se ha usado el menú de R-Commander con la selección de la figura 20 (las tres primeras líneas del programa). En cambio para las tres últimas órdenes se ha introducido el código directamente en la ventana de comandos.

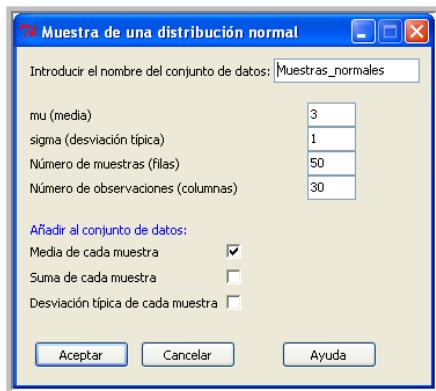


Figura 20: Selección para generar 50 muestras de 30 elementos de una $N(3,1)$.

Con este programa se ha generado la gráfica de la figura 21, donde puede observarse que algunos de estos intervalos no contienen el valor de la media.

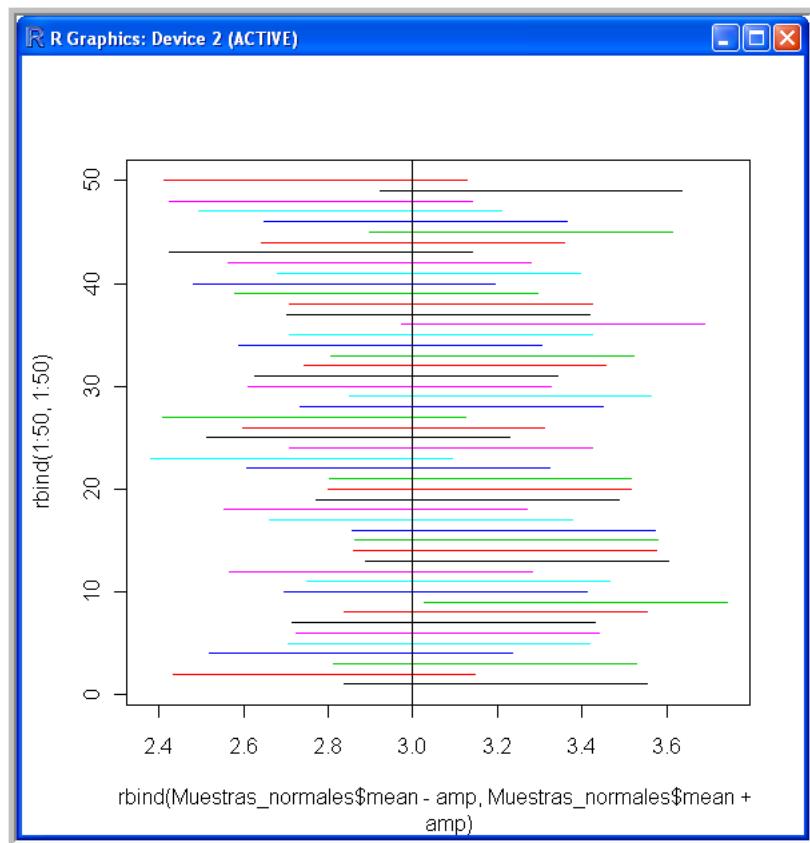


Figura 21: Representación de intervalos de confianza para 50 muestras de una $N(3,1)$.

5.2. Intervalo de confianza y test de hipótesis para la media a partir de una muestra

Cuando los intervalos de confianza para la media se estiman a partir de una muestra no suele conocerse el valor de σ de la distribución de la que procede la muestra. Por este motivo no puede utilizarse la distribución $N(0,1)$ de la variable $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$ para hallar el intervalo de confianza $(\bar{X} - z_{1-\alpha/2}(\sigma / \sqrt{n}), \bar{X} + z_{1-\alpha/2}(\sigma / \sqrt{n}))$ para μ , ya que no se conoce σ . Por eso en un caso práctico, y si la distribución en estudio es normal, se usa casi siempre el siguiente teorema:

Si X es una variable aleatoria $N(\mu, \sigma)$

$$T = \frac{\bar{X} - \mu}{S_c / \sqrt{n}}$$

se distribuye según una t de Student con $n-1$ grados de libertad, siendo \bar{X} la media de la muestra y S_c su cuasidesviación.

Si la variable aleatoria X no sigue una distribución normal, este resultado puede aplicarse a efectos prácticos si el número de elementos de la muestra es mayor o igual que 60.

Usando el teorema anterior y siguiendo un procedimiento similar al empleado para calcular el intervalo $(\bar{X} - z_{1-\alpha/2}(\sigma / \sqrt{n}), \bar{X} + z_{1-\alpha/2}(\sigma / \sqrt{n}))$ donde se usa la distribución $N(0,1)$ de $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$, se obtiene que, si no se conoce σ , la expresión de un intervalo de confianza para la media poblacional μ con una confianza $100(1-\alpha)\%$ es:

$$(\bar{X} - t_{1-\alpha/2}(S_c / \sqrt{n}), \bar{X} + t_{1-\alpha/2}(S_c / \sqrt{n}))$$

Estos intervalos de confianza pueden obtenerse directamente con R Commander entrando desde el menú:

Estadísticos → Medias → Test t para una muestra

Veámoslo con un ejemplo:

Ejemplo 12: Calculemos un intervalo de confianza al 95%, para la media de una población normal a partir de la siguiente muestra.

6.411, 4.324, 5.2825, 3.2689, 3.705, 4.9736, 1.8977, 4.4681, 3.7961, 4.9666

Se introducen los datos previamente en la variable var1 del fichero ejemplo. La ventana emergente resultante del procedimiento citado se rellena como la de la figura 22.

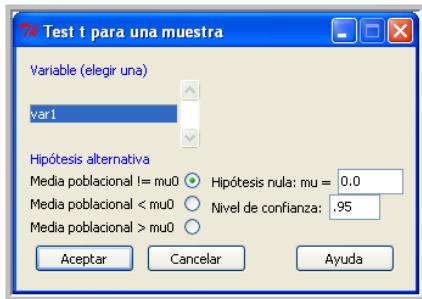


Figura 22: Obtención de intervalos de confianza y test de hipótesis para μ .

Con ello R-Commander presenta la siguiente salida en la ventana de resultados:

```
data: ejemplo$var1
t = 11.0091, df = 9, p-value = 1.599e-06
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
3.423865 5.194835
sample estimates:
mean of x
4.30935
```

En conclusión: el intervalo de confianza al 95% para la media poblacional, basado en la muestra dada en el ejemplo 12, es el siguiente:

(3.423865, 5.194835)

Se especifica a continuación cómo se haría este cálculo a mano usando, por ejemplo, una tabla de la distribución t-Student con 9 grados de libertad.

En primer lugar obtendríamos la media y la cuasidesviación de la muestra, que resultan 4.30941 y 1.2378 respectivamente. El valor de $t_{1-\alpha/2} = t_{1-0.05/2} = t_{0.975}$ se ha calculado con la condición: $P(t < t_{0.975}) = 0.975$. Por tanto, mirando en la tabla de la t de Student con 9 grados de libertad obtenemos que su valor es: $t_{0.975} = 2.2622$.

Sustituyendo en la expresión del intervalo de confianza anterior los valores obtenidos y teniendo en cuenta que el número de elementos de la muestra es 10, obtenemos el siguiente intervalo de confianza para la media poblacional:

$$(4.3094 - 2.2622 \times (1.2378/\sqrt{10}), 4.3094 + 2.2622 \times (1.2378/\sqrt{10})) = (3.4239, 5.1949).$$

5.3. Contraste de hipótesis para la media

Un contraste o test de hipótesis para la media es un criterio para decidir si, conociendo los datos de una muestra, se puede suponer que un número dado es razonablemente adecuado como valor de la media poblacional μ de la distribución de la que procede. Por ejemplo, ¿es razonable aceptar la hipótesis $\mu=0$, es decir que la media verdadera o poblacional de la distribución de la que procede la muestra del ejemplo anterior sea 0? A primera vista parece que no. La respuesta a esta pregunta puede basarse en el intervalo de confianza (3.423865, 5.194835) para la media. Como el valor 0 no está en el intervalo de confianza respecto de la media se rechaza que la

media poblacional tome el valor $\mu=0$ con una confianza del 95%. No obstante también puede darse la respuesta usando la primera parte de la salida de R-Commander dada en el anterior apartado, donde habíamos rellenado en la casilla correspondiente a la hipótesis nula, $\mu=0$. Ver la figura 22.

En efecto la información requerida está contenida en las siguientes líneas:

$t = 11.0091$, $df = 9$, $p\text{-value} = 1.599e-06$

Los datos que nos ofrece esta línea son los siguientes:

$$T = \frac{\bar{X} - \mu}{S_c / \sqrt{n}} = \frac{4.3094 - 0}{1.2378 / \sqrt{10}} = 11.009.$$

df (degrees of freedom) = grados de libertad = $n - 1 = 10 - 1 = 9$

$p\text{-value} = P(|T| > 11.009) = 2 \times (1 - \text{TDist}(11.009; 9)) = 1.5989 \times 10^{-6}$

La respuesta a la pregunta: ¿Es $\mu=0$ con una confianza del 95%?, se basa en el $p\text{-value}$. Si $p\text{-value}$ es mayor que α no se rechaza la hipótesis nula, en caso contrario si se rechazaría. Es decir si queremos aceptar o rechazar la hipótesis nula, $\mu=0$, hacemos el siguiente razonamiento. Como $p\text{-value} = 1.5989 \times 10^{-6} \leq \alpha=0.05$ se rechaza la hipótesis de partida de que la media poblacional sea 0.

En cambio, si propusieramos $\mu=4$ como valor de la media poblacional, deberíamos introducir el valor $\mu=4$ en la ventana del test de hipótesis (ver la figura 22). En este caso, en la ventana de resultados de R-Commander nos suministraría la siguiente información:

$t = 0.7903$, $df = 9$, $p\text{-value} = 0.4497$.

Como la respuesta al contraste de hipótesis se basa en el $p\text{-value}$, y éste es mayor que α , no se rechaza la hipótesis nula ($\mu=4$). El $p\text{-value}$ es ahora $0.4497 > \alpha=0.05$ y no se rechaza la hipótesis de partida de que la media poblacional sea 4. También puede darse la misma respuesta basándose en el intervalo de confianza. El intervalo de confianza debe contener el valor 4 para que esta hipótesis sea aceptable. En este caso el intervalo de confianza es (3.423865, 5.194835), que contiene el valor 4. Lo que acabamos de realizar es un contraste bilateral. Para realizar contrastes unilaterales ($\mu>4$, $\mu<4$) seleccionaríamos la opción adecuada en la ventana de selección que podemos ver en la figura 22.

5.4. Estimación de una proporción. Contraste de hipótesis para una proporción

Con frecuencia se requiere estimar el valor de una proporción o porcentaje a partir de los datos de una muestra. Esta situación se presenta, por ejemplo, cuando la variable aleatoria es la respuesta a una pregunta con sólo dos respuestas posibles. Así ocurre cuando se realiza un referéndum, aunque también podría ser algún tipo de encuesta de tipo político, social, religioso, etc. Lo importante es que la variable aleatoria sólo puede tomar dos valores, y se pueda modelizar con una distribución de Bernoulli.

Ejemplo 13: Usando la muestra registrada en el archivo “propor.rda” halle un intervalo de confianza al 95% para la proporción p de “síes”. Decida, al 95% de confianza, si el valor 0.7 es aceptable para dicha proporción.

Elegimos el fichero propor.rda como fichero de datos activo. A continuación seleccionamos en el menú de R Commander:

Estadísticos → Proporciones → Test de proporciones para una muestra

La ventana emergente que se nos ofrece la rellenamos en la forma indicada en la figura 23.

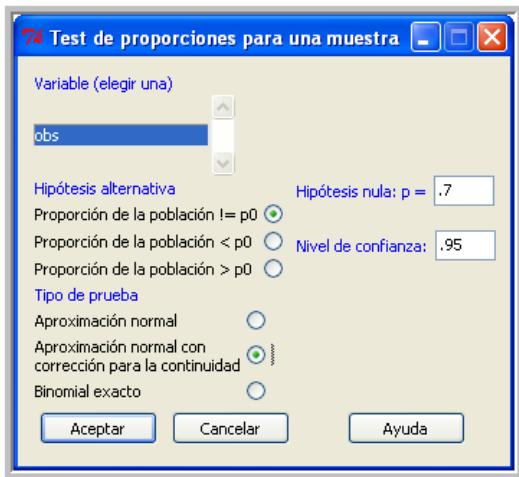


Figura 23: Estimación de una proporción.

La proporción, p , de “síes” es la media de la distribución de Bernoulli asociada ($si=1$, $no=0$). Si no tenemos otra información que la contenida en la muestra, tomariamos como estimación puntual de esta proporción, p , 0.77, que es la proporción de “síes”, y la media, de nuestra muestra.

Para responder las cuestiones formuladas en este ejemplo actuamos de la siguiente forma:

Como la muestra es amplia (100 elementos) y la media de la distribución, de Bernoulli en este caso, sigue según el teorema central del límite una distribución normal, se ha seleccionado la opción aproximación normal con corrección para la continuidad, porque el tamaño de la muestra es suficientemente amplio y está opción suministra mejor aproximación que la primera. Sin embargo la opción de usar la binomial es exacta, y es la única recomendable si el tamaño de la muestra es pequeño.

La salida del programa, en la que hemos destacado los resultados relevantes para nuestro problema, es la siguiente:

```
.Table
obs
  si  no
  77 23
> prop.test(rbind(.Table), alternative='two.sided', p=0.7, conf.level=.95, correct=TRUE)
  1-sample proportions test with continuity correction
  data: rbind(.Table), null probability 0.7
  X-squared = 2.0119, df = 1, p-value = 0.1561
  alternative hypothesis: true p is not equal to 0.7
```

95 percent confidence interval:

0.673059 0.845785

sample estimates:

p

0.77

Como el p-value = $0.1561 > 0.05$ encontramos que el valor 0.7 es aceptable como valor de la proporción de “síes”.

El intervalo de confianza, al 95%, para esta proporción p es (0.673059, 0.845785) que contiene el valor 0.7 propuesto. Está sería otra forma de decidir que el valor 0.7 es aceptable para la proporción de “síes”.

Si usamos, con R-Comander, la versión sin corrección de continuidad (correct=False) obtenemos para el intervalo de confianza para la proporción: (0.6784562, 0.8415673).

Usando la distribución exacta (elegir la opción tercera, binomial exacta) el intervalo de confianza para la proporción resulta el siguiente: (0.6751413, 0.8482684).

5.5. Programa en R para la representación gráfica de un intervalo de confianza para una proporción

Generamos 50 muestras de la Bernoulli con 100 elementos cada una. La idea del programa que describimos en este apartado es representar un intervalo de confianza para la proporción para cada una de estas muestras con el objeto de observar su aleatoriedad.

Las 50 muestras y su medias las generamos desde el menu de R Commander siguiendo el procedimiento Distribuciones → Distribuciones discretas → Distribución Binomial → Muestras de la distribución binomial, con lo que aparece una ventana que rellenamos como la de la figura 24.

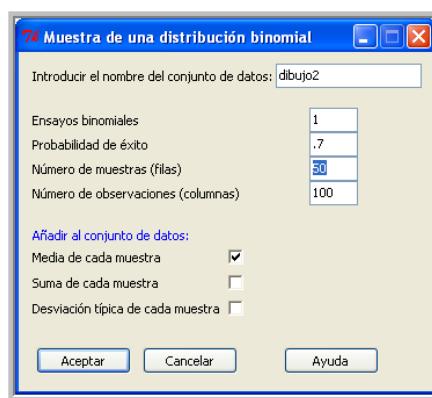


Figura 24: Generación de muestras de $Be(p)$, equivalente a $B(1,p)$.

Este procedimiento genera las siguientes líneas de programa.

```
dibujo2 <- as.data.frame(matrix(rbinom(50*100, size=1, prob=0.7), ncol=100))
```

```
rownames(dibujo2) <- paste("sample", 1:50, sep="")
colnames(dibujo2) <- paste("obs", 1:100, sep="")
dibujo2$mean <- rowMeans(dibujo2[,1:100])
```

Se genera un fichero que se sitúa como conjunto de datos activo y que contiene 50 muestras de una distribución de Bernoulli junto con sus medias, o estimaciones de p correspondientes a cada muestra.

A continuación pulsaremos las líneas de programa que siguen:

```
alfa=0.05; amp=qnorm(1-alfa/2)*sqrt(dibujo2$mean*(1-dibujo2$mean)/100)
matplot(rbind(dibujo2$mean-amp,dibujo2$mean+amp),rbind(1:50,1:50),type="l", lty=1)
abline(v=0.7) # draw line for media verdadera.
```

Como en el caso del apartado 5.1, aparece un gráfico que, en este caso, recoge los 50 intervalos de confianza para la proporción de la población obtenidos a partir de cada una de estas 50 muestras. Se observa en la figura 25 que, como suele ocurrir, unos pocos de estos intervalos no contienen la proporción verdadera. También puede ocurrir que en algunas simulaciones, como la de la figura 26, todos los intervalos contengan a la proporción verdadera.

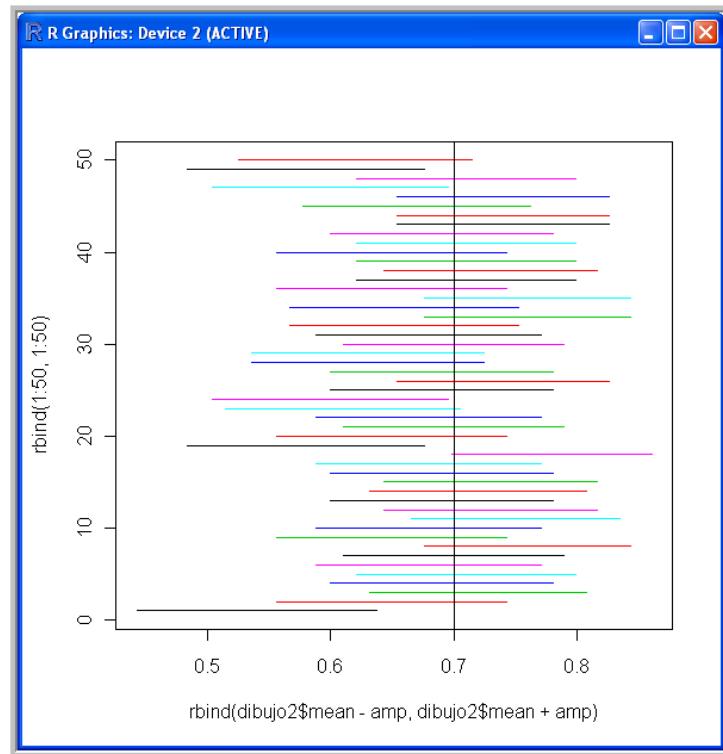


Figura 25: Intervalos de confianza para la proporción para 50 muestras. Simulación 1.

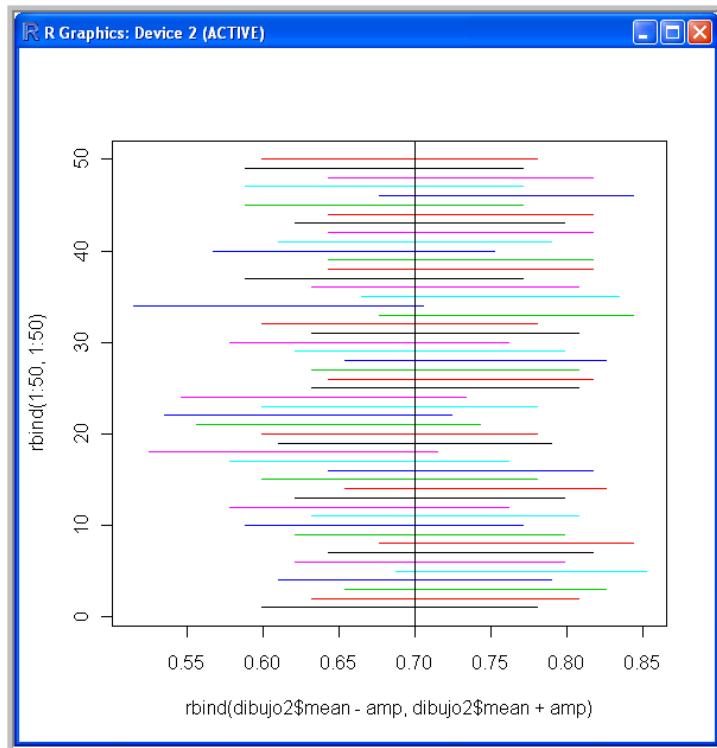


Figura 26: Intervalos de confianza para la proporción para 50 muestras. Simulación 2.

6. Conclusiones y agradecimientos

Hoy día el uso de programas estadísticos se ha convertido en una rutina a la hora de analizar los datos de las investigaciones en la mayoría de las disciplinas científicas. Los Estadísticos nos encontramos con frecuencia con peticiones de ayuda por parte de autores y directores de investigaciones que manifiestan tener problemas con el uso de estos programas, o, peor aún, con la interpretación de los resultados y los gráficos que han devuelto estos programas.

Es una práctica común en algunas disciplinas científicas confiar el análisis de datos a paquetes estadísticos comerciales, de los que se extraen más estadísticos que información. Es imposible obtener respuestas a medida, porque estos programas proporcionan técnicas y gráficas a la medida que el programador decidió en su día. Algunos programas se han diseñado para que sea el investigador el que decida qué respuestas quiere. Estos programas, como por ejemplo el software estadístico R, son además gratuitos.

En este Taller se ha pretendido proporcionar una introducción al programa R, que representa una alternativa a programas estadísticos de uso más extendido. A través de la utilización de este software libre se ha ido poniendo de manifiesto cómo es el propio autor del trabajo quien toma las decisiones con respecto a qué estrategias, gráficos y técnicas utilizar, así como sobre el formato de presentación que prefiere para sus resultados y sus gráficos. Es sencillamente imposible utilizar R sin criterio. Se debe comenzar por saber lo que se quiere antes de dar un sólo clic con el ratón sobre un menú en un programa estadístico.

Creemos además que el programa R proporciona todas las opciones necesarias para utilizarlo en la práctica docente universitaria. Aunque obtener resultados es más costoso

inicialmente, a la larga, los resultados son mucho más concretos, adaptables y modificables. Debemos tener en cuenta que R no es sólo un programa de cálculo estadístico, sino que es un lenguaje de programación estadístico, con el que podemos implementar cualquier procedimiento o técnica estadística, por novedosa que ésta sea. Además, en la actualidad existen más de 2000 librerías o paquetes disponibles para casi cualquier propósito.

Creemos que R es un entorno de programación en el que se han incluido rutinas que permiten de forma fácil y directa la realización de multitud de cálculos estadísticos; también la confección de gráficas e imágenes de una sencillez y una potencia equiparable, o superior, a la de otros programas comerciales. Por todo ello, os animamos a que utilicéis el software libre R tanto en el ámbito docente como en el de la investigación.

Queremos agradecer a los organizadores de las I Jornadas de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa, celebradas en Melilla en junio de 2010, que hayan pensado en nosotros para el desarrollo de este Taller de R. Queremos también agradecer a los integrantes del Grupo de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa (GENAEIO) de la SEIO por su participación y su especial motivación en el desarrollo de este Taller. Esperamos y deseamos que el desarrollo del mismo haya resultado interesante y productivo para todos.

7. Bibliografía y documentación

- [1] Introducción a R: <http://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- [2] Manual resumido de R: <http://cran.r-project.org/doc/contrib/curso-R.Diaz-Uriarte.pdf>
- [3] Diverso material docente: <http://www.ub.es/stat/docencia/EADB/Curso basico de R.htm>
- [4] Para principiantes: http://es.geocities.com/r_vaquerizo/Manual_R3.htm
- [5] Introducción al uso del R Commander:
[http://geminis.dma.ulpgc.es/profesores/personal/asp/
Documentacion/Manual R commander.pdf](http://geminis.dma.ulpgc.es/profesores/personal/asp/Documentacion/Manual_R_commander.pdf)
- [6] R para principiantes: http://cran.r-project.org/doc/contrib/rdebuts_es.pdf
- [7] Amplio manual sobre gráficos estadísticos con R:
<http://cran.r-project.org/doc/contrib/grafi3.pdf>
- [8] Curso básico de R: <http://www.ub.es/stat/docencia/EADB/Curso basico de R.pdf>
- [9] Introducción a R: http://www.cesma.usb.ve/~lbravo/co5314/MANUAL_Rb.pdf

Herramientas de autor para la creación de contenidos para la Docencia de la Estadística

Luis Miguel Marín Trechera¹, Antonio Gámez Mellado²

¹luis.marin@uca.es, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz

²antonio.gamez@uca.es, Escuela Superior de Ingeniería de Cádiz, C/Chile s/n 11002 Cádiz
Universidad de Cádiz

Resumen

Internet es en la actualidad una poderosa herramienta docente que permite poner a disposición de los estudiantes una gran cantidad de material interactivo para potenciar los aprendizajes. Por ese motivo, los profesores se ven cada vez más apremiados a dominar herramientas que les permitan crear este tipo de materiales.

En el presente trabajo se realiza una breve introducción a dos de esas herramientas, mostrando ejemplos de utilización de las mismas en el ámbito de la docencia de la Estadística e Investigación Operativa.

Las herramientas seleccionadas son eXe y Hot Potatoes. Con la primera de ellas es posible crear de una manera intuitiva material didáctico estructurado en unidades de aprendizaje. La segunda está especializada en la creación de cuestionarios que faciliten la evaluación y/o autoevaluación de los estudiantes.

Palabras clave: Herramientas de autor, eXe learning, Hot Potatoes, Enseñanza de la Estadística

Clasificación AMS: 97U50, 97K40, 97K80

1. Introducción

Internet ha supuesto una revolución en prácticamente todos los ámbitos de nuestra vida. Las relaciones sociales, las transacciones comerciales, la búsqueda de información, por poner unos ejemplos, se realizan ahora de una manera totalmente distinta a como se hacían hace solamente una década. Y el ámbito educativo no es ajeno a todos estos cambios. Antes al contrario, podemos afirmar que en el campo de la educación es uno de los más afectados por los cambios, y son muchas las iniciativas que se han tomado en los distintos niveles educativos para potenciar el uso de las nuevas tecnologías. Y no se trata de una moda pasajera. Internet ha entrado de lleno en la educación, y lo ha hecho para quedarse.

No en vano el uso de las nuevas tecnologías es mayor entre las generaciones más jóvenes. Los profesores se ven obligados a enfrentarse a generaciones que han asumido la tecnología de manera natural, son lo que se ha denominado *nativos digitales*, y los cambios siguen produciéndose. En los últimos años se ha pasado sucesivamente de alumnos sin ordenador, a alumnos con un amplio arsenal de dispositivos informáticos, que luego fueron reemplazados por alumnos con conexiones cada vez más veloces y éstos por alumnos que emplean varias horas al día a sumergirse en las redes sociales.

Estos cambios tecnológicos, que han tenido una respuesta por parte de las instituciones universitarias con la puesta en marcha de las aulas virtuales, unidos a la gran transformación que ha supuesto en el ámbito universitario el proceso de Bolonia y el EEES, están obligando al profesorado a realizar un proceso de adaptación. Muchos profesores se sienten impulsados a aumentar su nivel competencial en lo referente al diseño y desarrollo de materiales educativos.

El desarrollo de material educativo debe realizarse llegando a un equilibrio entre los recursos tecnológicos utilizados, los objetivos educativos y los hábitos de los estudiantes. Tal y como afirma Jakob Nielsen en su último trabajo [1] “Resulta peligroso asumir que los estudiantes son expertos tecnólogos”. La tecnología les resulta natural, se han educado utilizando, pero les disgustan los interfaces complicados, y suelen abandonar los sitios web ignorando a menudo la información que buscaban si no la encuentran de manera inmediata.

Son muchas las herramientas informáticas existentes para facilitar estas labores de creación de material educativo interactivo multimedia. Tras haber realizado un proceso de análisis de diversas herramientas se han seleccionado dos, basándonos en la facilidad de uso y en las características de los materiales generados. Las herramientas seleccionadas son Hot Potatoes y eXe (E-learning Xhtml Editor). Hot potatoes está especializada en la creación de cuestionarios que faciliten la evaluación y/o autoevaluación de los estudiantes. Con la eXe es posible crear de una manera intuitiva material didáctico estructurado en unidades de aprendizaje.

En este trabajo se realiza una breve descripción de estas dos herramientas, mostrando ejemplos de utilización de las mismas en el ámbito de la docencia de la Estadística e Investigación Operativa.

2. Hot Potatoes

Hot Potatoes es un conjunto de seis herramientas de autor, desarrollado por el equipo de Investigación y Desarrollo del “Humanities Computing and Media Centre” de la Universidad de Victoria (Canadá). Esta aplicación es gratuita para su uso en instituciones educativas sin ánimo de lucro y está disponible en la dirección web del proyecto [2].

Estas herramientas permiten elaborar ejercicios interactivos basados en páginas Web de seis tipos básicos. La interactividad de los ejercicios se consigue mediante JavaScript y HTML dinámico, pero no es necesario tener conocimientos de JavaScript o HTML para poder utilizar estas aplicaciones. El sistema funciona de un modo bastante sencillo e intuitivo, de modo que el programa se encarga de generar páginas Web a partir de los datos proporcionados por el usuario (enunciados de las preguntas, textos, respuestas, etc.). No obstante, los programas están diseñados para que casi todos los elementos de las páginas puedan ser personalizados por los autores. De este modo, sabiendo un poco de los lenguajes HTML y JavaScript, se pueden hacer casi todos los cambios que se deseen en los ejercicios o en el formato de las páginas Web.

Como se ha comentado anteriormente, Hot Potatoes es un conjunto de seis herramientas (o “patatas”). Cada una de estas herramientas permite crear un tipo de ejercicio diferente. Las herramientas son las siguientes:

- **JQuiz:** Ejercicios basados en preguntas.
- **JCloze:** Ejercicios de llenar huecos.
- **JMatch:** Ejercicios de emparejamiento.
- **JMix:** Ejercicios para desordenar palabras o frases.
- **JCross:** Crucigramas.

- **The Masher:** Para crear unidades enlazadas a partir de diferentes ejercicios.

Para elaborar un ejercicio podemos distinguir tres fases o pasos:

Introducción de los datos: En esta fase se introduce el enunciado de las preguntas, se incluyen las propuestas de respuestas alternativas en los casos oportunos, las posibles indicaciones explicativas ante respuestas erróneas, etc.

Configuración del formato de salida: En esta fase se pueden indicar unas instrucciones generales para la realización del ejercicio así como modificar el texto de los mensajes, avisos y botones. Del mismo modo se puede configurar la apariencia de la página web, seleccionando el tipo de letra, colores, etc. También se pueden modificar parámetros generales del cuestionario, como el número de intentos permitidos, limitación de tiempo, barajar el orden de las preguntas y respuestas, etc.

Creación de las páginas Web: El último paso del proceso es exportar los ejercicios, generándose la página html correspondiente que puede ser subida a cualquier sitio web. En este paso se puede seleccionar en algunos casos si se desea usar las opciones de arrastrar y soltar. También es posible realizar la exportación en formato SCORM. Además de la creación de la página html es importante guardar el archivo en el formato propio de hot potatoes, ya que de este modo se podrán realizar modificaciones posteriores. Cada uno de los módulos tiene un formato específico, como puede verse en la tabla 1: jqz para JQuiz, jcw para JCross, jcl para JCloze, jmt para JMatch, jmx para JMIX y jms para The Masher. Algunos LMS como Moodle son capaces de interpretar directamente estos ficheros.

Tabla 1: Extensión de los archivos según el tipo de herramienta

| Tipo de herramienta | Extensión del archivo |
|---------------------|-----------------------|
| JQuiz | *.jqz |
| JCross | *.jcw |
| JCloze | *.jcl |
| JMatch | *.jmt |
| JMix | *.jmx |
| The Masher | *.jms |

3. Ejemplos de utilización de las distintas Herramientas Hot Potatoes

A continuación se da una breve descripción de cada una de las herramientas, mostrando ejemplos de utilización en el ámbito de la estadística:

3.1. JQuiz (ejercicios basados en preguntas)

JQuiz es una herramienta que permite elaborar ejercicios basados en preguntas. Cada test puede estar formado por un número ilimitado de preguntas. Se pueden formular cuatro tipos de preguntas:

- **Pregunta de respuestas múltiples:** los estudiantes deben seleccionar la respuesta correcta de entre las propuestas. Tras cada selección se muestran las indicaciones

específicas para dicha respuesta incluidas por el profesor al crear el ejercicio, explicándole porque está bien o mal. Si la respuesta es incorrecta, el estudiante puede seguir seleccionando opciones hasta dar con la respuesta correcta. La puntuación de cada pregunta dependerá del número de intentos realizados antes de acertar la pregunta. Una vez que se ha elegido una respuesta correcta el estudiante puede seguir pulsando en los botones para ver las indicaciones de las demás respuestas sin que por ello se le penalice. En la figura 1 puede verse un ejemplo de página creada con JQuiz que muestra una pregunta de tipo verdadero-falso, como caso particular de pregunta de elección múltiple.

Cuestionario Autoevaluación ANOVA

Cuestionario Hot Potatoes

Escoge la respuesta correcta para cada pregunta, haciendo click sobre la letra correspondiente.

1 / 7 => Mostrar todas las preguntas

En la tabla ANOVA se representa la idea de que la varianza se puede descomponer en las distintas fuentes que la originan. (Responda si es verdadera o falsa la afirmación anterior)

A. ? Falsa

B. ? Verdadera

Figura 1. Ejemplo de actividad creada con JQuiz

- **Pregunta de respuestas cortas:** los estudiantes deben escribir una respuesta en una casilla de texto en la página, y pulsar en el botón Comprobar para ver si es correcta. Si la respuesta está en la lista de respuestas previstas (sea correcta o no) se muestran las indicaciones para dicha respuesta. La puntuación obtenida depende del número de intentos realizados por el estudiante antes de dar la respuesta correcta. También se puede incluir un botón Pista, para proporcionar al estudiante una letra de la respuesta correcta; la utilización del botón Pista conlleva una penalización en la puntuación.
- **Pregunta híbrida:** es una combinación de una pregunta de respuestas múltiples y una pregunta de respuesta corta. En este tipo de preguntas, al estudiante se le presenta una casilla de texto para que escriba la respuesta. Si después de un número

de intentos prefijado no se ha dado con la respuesta correcta, la pregunta cambia a otra de respuestas múltiples para que sea más fácil para el estudiante.

- **Pregunta de multi-selección:** en este tipo de preguntas el estudiante ha de seleccionar todas las respuestas correctas de entre las propuestas. En la figura 2 se muestra una pregunta de este tipo.

The screenshot shows a question from a self-assessment quiz on ANOVA. The title is "Cuestionario Autoevaluación ANOVA" and the section is "Cuestionario Hot Potatoes". A green box contains the instruction: "Escoge la respuesta correcta para cada pregunta, haciendo click sobre la letra correspondiente." Below this, a navigation bar shows "3 / 7" and buttons for "Mostrar todas las preguntas", "<=", and "=>". The question text reads: "Para poder construir el test ANOVA tenemos que asumir que se cumplen las siguientes hipótesis (señale todas las que sean correctas)". The options are listed as follows:

- A. Heterogeneidad
- B. Llinealidad
- C. Homogeneidad
- D. Normalidad
- E. Homocedasticidad
- F. Dependencia
- G. Independencia

Figura 2: Pregunta de multi-selección creada con JQuiz.

3.2. JCloze (ejercicios de llenar huecos)

JCloze se utiliza para elaborar ejercicios de completar las palabras que faltan en un texto. La idea de este tipo de ejercicios es que el estudiante complete todas las respuestas antes de comprobar el resultado. Cuando se han completado todos los huecos, el estudiante pulsará sobre el botón Comprobar para ver las respuestas. Las respuestas correctas serán introducidas en el texto; las respuestas incorrectas se dejarán en las casillas de texto, para que puedan corregirse. Cuando el estudiante comprueba una respuesta que no es totalmente correcta, se incurre en una penalización. Por tanto, la puntuación final dependerá del número de veces que se ha comprobado la respuesta antes de dar con la correcta.

En los ejercicios JCloze, se puede incluir un botón Pista que proporcionará al estudiante una letra de la respuesta correcta (en función de la posición del cursor). También se puede incluir una pista específica para cada hueco.

En la figura 3 se muestra un ejemplo de página creada con JCloze. En este caso se muestra una página con dos huecos, en un cuestionario sobre medidas estadísticas.

The screenshot shows a Cloze exercise page with a blue header bar containing a 'Volver al índice' link. The main title is 'Medidas Estadísticas' and the section title is 'Ejercicio de completar frases'. Below this, there is a text input field with the placeholder 'Escribe en los espacios en blanco las palabras adecuadas.' followed by a horizontal line with two empty text boxes labeled 'dispersión' and 'posición'. In the main content area, there is a sentence: 'La media es una medida de posición central, mientras que la varianza es una medida de dispersión.' Below the sentence is a 'Verificar' button. At the bottom of the page is another 'Volver al índice' link.

Figura 3. Ejemplo de Página creada con JCloze.

3.3. JMatch (ejercicios de emparejamientos)

JMatch se utiliza para crear ejercicios de emparejamiento. Básicamente esto quiere decir que se disponen una serie de elementos a un lado y cada uno de ellos ha de emparejarse con los elementos dispuestos en el lado opuesto.

Los ejercicios generados con JMatch pueden ser de dos tipos: modelo estándar y modelo arrastrar y soltar. El modelo estándar utiliza una lista desplegable de elementos, siendo el formato a utilizar cuando los elementos de la derecha son solo de tipo texto y haya más de siete u ocho elementos. Si los elementos son gráficos debe utilizarse el modelo arrastrar y soltar. En la figura 4 se muestra un ejemplo de ejercicio de emparejamiento creado con JMatch. En este caso, al tratarse de emparejar textos con imágenes se ha optado por la opción de arrastrar y soltar.

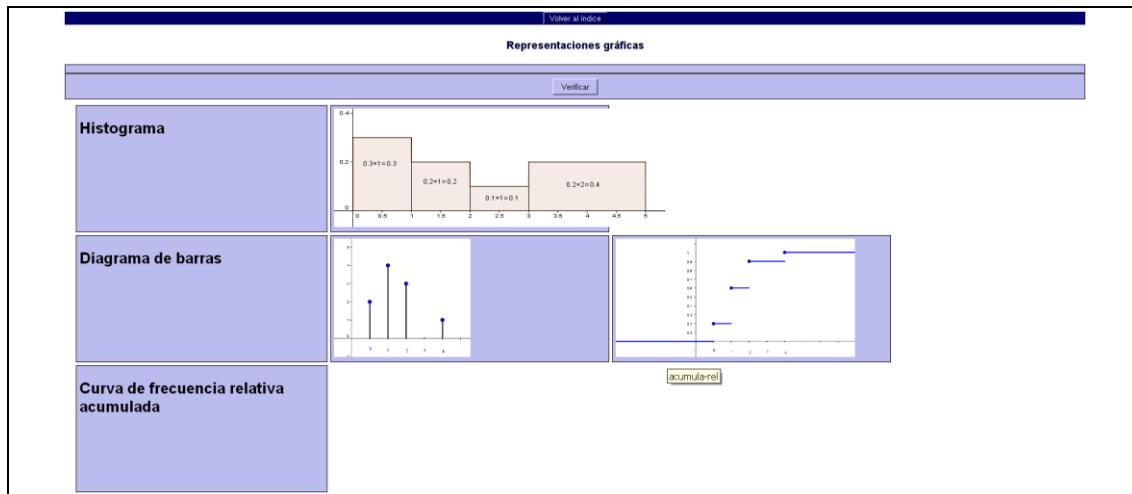


Figura 4. Ejemplo de Página creada con JMatch.

3.4. JMIX (ejercicios para desordenar palabras o frases)

JMix se utiliza para elaborar ejercicios de ordenar palabras o frases. Puede ordenar las palabras de una frase, o las letras de una palabra. Como JMatch, JMIX puede presentarse en dos modalidades: modelo estándar, y modelo arrastrar y soltar. En el modelo estándar hay que ir construyendo la frase por orden, pulsando sobre cada una de las palabras, mientras que en el otro modelo cada bloque puede ser arrastrado y soltado en la posición de la frase que se desee. La figura 5 muestra un ejemplo de página creada con JMIX usando el modelo arrastrar y soltar.

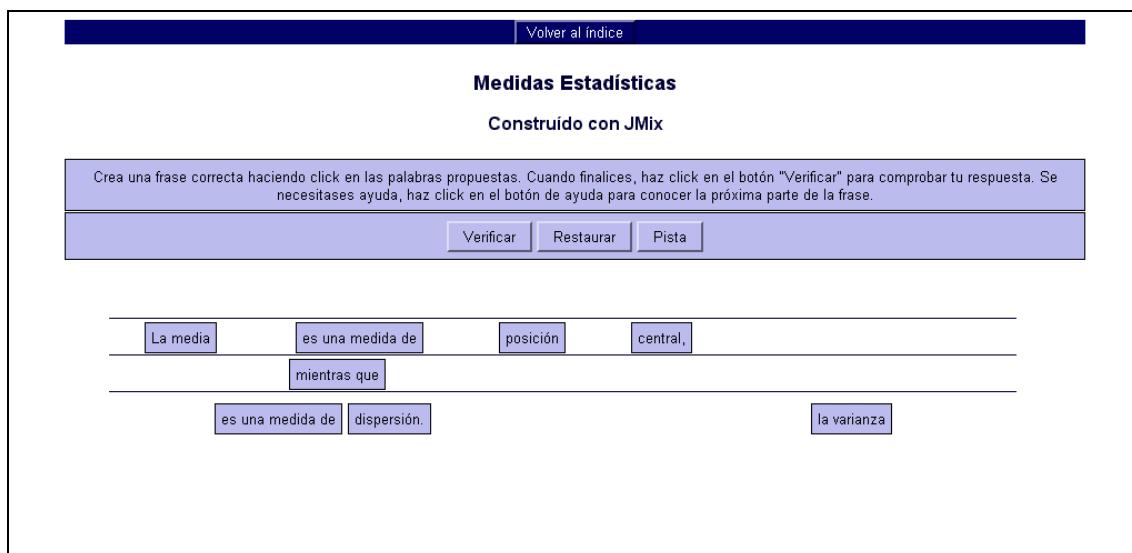


Figura 5. Ejemplo de ejercicio creado con JMIX

3.5. JCross (crucigramas)

JCross se utiliza para realizar crucigramas. Hay dos fases en la elaboración de un ejercicio de crucigrama; primero deberá introducir las letras en la cuadrícula, para después añadir las definiciones. La figura 6 ilustra el proceso de creación del crucigrama con JCross. En la figura 7 puede verse como quedaría la página creada en el proceso anterior.

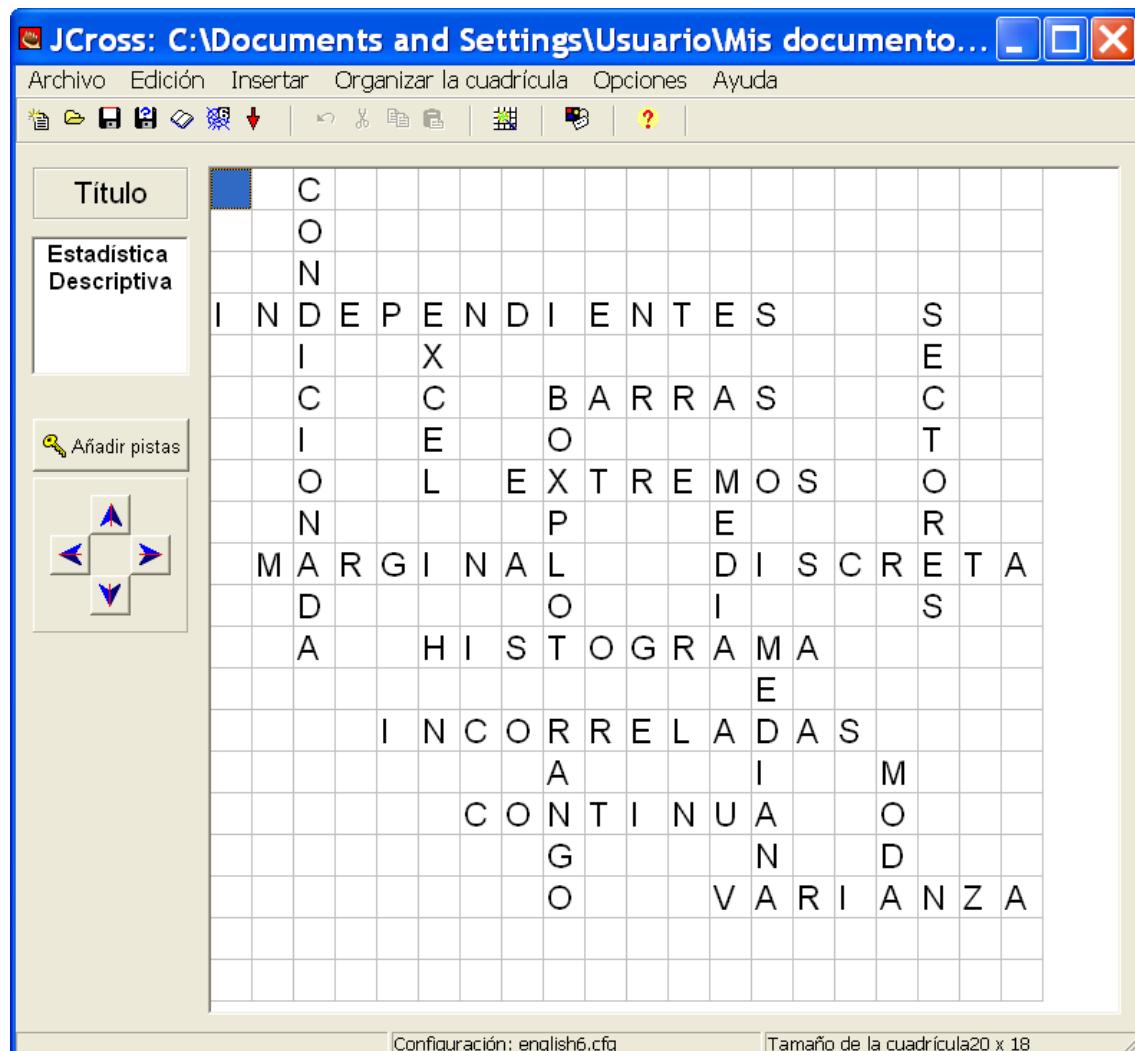


Figura 6. Proceso de creación de un crucigrama con JCross.



Figura 7. Página creada con JCross.

Las páginas creadas con JCross están especialmente indicadas para ejercitarse a los alumnos al manejo adecuado de las definiciones de los distintos conceptos.

3.6. The Masher (para crear unidades enlazadas a partir de diferentes ejercicios)

The Masher es un tipo de aplicación diferente de las otras patatas de Hot Potatoes. Sirve para enlazar distintos ejercicios creados con las otras herramientas. Se genera una página de índice y en cada página aparece la opción de volver al índice o pasar al ejercicio siguiente. Las páginas enlazadas no pueden estar generadas con herramientas distintas, creándose de esta manera una sucesión de actividades de los distintos tipos.

3.7 Integración de Hot Potatoes en Moodle

Las páginas generadas con Hot Potatoes pueden enlazarse desde cualquier página web. Las puntuaciones mostradas pueden servir a los estudiantes para la autoevaluación de sus conocimientos. Es de especial utilidad el uso conjunto de Hot Potatoes con un LMS (Learning Management System) que lo soporte, como Moodle. De este modo puede almacenarse las calificaciones obtenidas por cada alumno, con lo cual los ejercicios pueden ser usados por el profesor con fines de evaluación. Moodle es capaz de entender tanto los ficheros html como los específicos de cada una de las herramientas (*.jqz, *.jwq,...).

4. eXe Learning

El proyecto eXe Learning está financiado por el Gobierno de Nueva Zelanda y coordinado por la University of Auckland, The Auckland University of Technology, y Tairawhiti Polytechnic. En el proyecto también participan un amplio grupo de colaboradores de todo el mundo. La herramienta eXe proporciona un editor multiplataforma de recursos educativos Open Source (Código Abierto), que permite construir contenido web didáctico sin necesidad de tener conocimientos de lenguajes de edición y marcado como XML ó HTML. El programa puede descargarse de la página web del proyecto [3].

Según la nomenclatura de eXe, con esta herramienta creamos proyectos que están constituidos por unidades que se estructuran jerárquicamente formando un árbol de contenidos. El proyecto resultante es una colección de páginas web con un formato visual común, mostrándose en la parte izquierda de la pantalla el índice con el árbol de contenidos para facilitar la navegación entre las distintas páginas. En la parte derecha de cada página hay una amplia zona donde se muestra el contenido correspondiente. La Figura 8 muestra un ejemplo de una página de un proyecto creado con eXe. A la izquierda aparece el índice, que irá desplegando las subsecciones al ir mostrando cada sección, y a la derecha las páginas de contenidos.

La zona de contenidos estará formada por una o varias unidades más pequeñas, a las que se puede dotar de una utilidad específica. Para ello eXe cuenta con una colección de idevices, es decir, herramientas capaces de introducir una caja de contenidos dotándolos de una apariencia visual propia (con un ícono identificativo específico). Cada idevice presenta características propias, y es posible crear otros nuevos. En general, cada idevice suele mostrar un editor web bastante intuitivo, con lo que es fácil introducir textos, imágenes, fórmulas, etc. encargándose luego el programa de integrarlo en la página con su formato correspondiente. El proceso de creación de un proyecto es, de este modo, bastante rápido y sencillo.

El aspecto visual final mostrado es bastante elegante y uniforme, al aplicarse a todas las páginas del proyecto un mismo estilo seleccionado. Este estilo, sin embargo, puede modificarse, eligiendo otro de un catálogo de estilos predefinidos. También es posible crear un estilo propio.

Además, se pueden definir unos parámetros generales comunes a todo el proyecto, como un título común que aparecerá en todas las páginas, que puede incluir una imagen de fondo. También pueden especificarse una serie de metadatos siguiendo el modelo establecido por la DCMI (Dublín Core Matadata Initiative) [4]. De este modo se facilita la clasificación apropiada de estos recursos por buscadores inteligentes.

Los distintos tipos de idevices existentes pueden clasificarse según la utilidad que tienen de la siguiente manera:

idevices de presentación de contenidos: Permiten insertar textos o gráficos, enmarcándolo y resaltándolo con un ícono y un título. Los idevices de este tipo son los denominados *Objetivos*, *Pre-conocimiento* y *Actividad*. El idevice *Texto Libre* también entraría en esta categoría, aunque en este caso el texto no se resalta de ninguna manera especial. La figura 8 muestra un ejemplo de una página constituida por 3 idevices de presentación de contenidos. La figura 9 muestra una página con los idevices Objetivos y Pre-conocimiento. La figura 10 muestra esa misma página generada con un estilo visual distinto.

Tema 1

Objetivos

1.- Introducción a la Estadística.

2.- Conceptos básicos

3.- Tablas de Frecuencias

4.- Representaciones gráficas

Actividad 2 para entregar

3.- Medidas Estadísticas,

- 3.1.- Medidas de posición central**
- 3.2.- Medidas de posición no central.
- 3.3.- Medidas de dispersión.
- 3.4.- Características de forma

3.1.- Medidas de posición central

Las medidas de posición central nos detectan una serie de puntos alrededor de los cuales se agrupan de alguna manera el resto de observaciones.

Media

La más importante de las medidas de posición central es la media aritmética, que denotaremos como \bar{X}

Es el centro de gravedad de los datos. Si todos los valores observados hubiesen tomado un mismo valor constante de modo que la suma total fuese la misma que la obtenida, ese valor constante sería la media aritmética. Su fórmula de cálculo es:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k f_i x_i$$

Cuando los datos no son de naturaleza aditiva, sino multiplicativa, se utiliza la **media geométrica**. En otros casos, cuando las frecuencias mantienen una relación inversa respecto a los valores de la variable, se utiliza la **media armónica**.

Aunque existan estas otras medias, cuando especifiquemos el término **Media** sin indicar más, nos referiremos a la media aritmética. También es común referirse a la media aritmética como el valor esperado de la variable, usando la notación $E[X]$

Podemos destacar la siguiente propiedad de la media aritmética: Si realizamos una transformación lineal de los datos, a media de los datos transformados coincide con la transformación de la media. Esto es:

$$y_i = a + b x_i \Rightarrow \bar{Y} = a + b \bar{X}$$

Lo que usando la notación de esperanzas puede indicarse como:

$$y_i = a + b x_i \Rightarrow E[Y] = a + b E[X]$$

Esta propiedad la cumple la media aritmética pero no las otras medias ni, por lo general, la mayoría del resto de medidas estadísticas. También hay que destacar que la propiedad anterior no se cumple si la transformación de los datos no es lineal.

Mediana

Definimos la mediana como aquel valor que hace que la mitad de las observaciones sean menores o iguales a él y la otra mitad mayor o igual que él.

Si el número total de observaciones es n , y ordenamos los datos de menor a mayor, la mediana será:

- Si n es impar, el valor que ocupe el lugar $(n+1)/2$
- Si n es par, el punto medio entre los valores que ocupen las posiciones $n/2$ y $n/2+1$

Moda

Es el valor que presenta una mayor frecuencia. Si es único se dice que la distribución es unimodal, si no lo es, se le dice multimodal.

« Anterior | Siguiente »

Figura 8. Ejemplo de proyecto creado con eXe.

| | |
|--|---|
| <p>Tema 1</p> <p>Objetivos</p> <p>1.- Introducción a la Estadística.</p> <p>2.- Conceptos básicos</p> <p>3.- Tablas de Frecuencias</p> <p>4.- Representaciones gráficas</p> <p>Actividad 2 para entregar</p> <p>3.- Medidas Estadísticas.</p> | <p>Objetivos</p> <p> Competencias y Objetivos</p> <p>Al terminar esta lección, deberá ser capaz de:</p> <ul style="list-style-type: none"> • Comprender los distintos usos de la Estadística: Distinguir entre un estudio con fines descriptivos u otro cuya finalidad sea realizar una inferencia, relacionando los conceptos de población y muestra. • Identificar los distintos tipos de datos (cuándo una característica de la población es un atributo o una variables discreta o continua, datos temporales, escalas ordinales,...). • Agrupar los datos de una manera conveniente, para una más fácil presentación de los datos, identificando los distintos tipos de elementos que aparecen en las mismas (frecuencias, marcas de clase, alturas, ...). • Realizar la representación gráfica más apropiada para los datos. • Resolver problemas que impliquen el cálculo de distintas medidas estadísticas de los datos. <p> Conocimientos Previos</p> <p>Este es el primer tema del curso. Los requisitos previos necesarios para poder adquirir las competencias indicadas son manejar con soltura los conceptos correspondientes a los contenidos desarrollados en las asignaturas de Matemáticas de Secundaria y Bachillerato. Aunque en dichos contenidos aparecen conceptos de Estadística, es frecuente que el primer contacto de los alumnos con la materia se produzca en esta asignatura. Esto puede conllevar la realización de un pequeño esfuerzo adicional.</p> |
|--|---|

Figura 9. Página de un proyecto creado con eXe con dos idevices de presentación de contenidos.

| | |
|--|---|
| <p>Tema 1</p> <p>Objetivos</p> <p>1.- Introducción a la Estadística.</p> <p>2.- Conceptos básicos</p> <p>3.- Tablas de Frecuencias</p> <p>4.- Representaciones gráficas</p> <p>Actividad 2 para entregar</p> <p>3.- Medidas Estadísticas.</p> | <p>Objetivos</p> <p> Competencias y Objetivos</p> <p>Al terminar esta lección, deberá ser capaz de:</p> <ul style="list-style-type: none"> • Comprender los distintos usos de la Estadística: Distinguir entre un estudio con fines descriptivos u otro cuya finalidad sea realizar una inferencia, relacionando los conceptos de población y muestra. • Identificar los distintos tipos de datos (cuándo una característica de la población es un atributo o una variables discreta o continua, datos temporales, escalas ordinales,...). • Agrupar los datos de una manera conveniente, para una más fácil presentación de los datos, identificando los distintos tipos de elementos que aparecen en las mismas (frecuencias, marcas de clase, alturas, ...). • Realizar la representación gráfica más apropiada para los datos. • Resolver problemas que impliquen el cálculo de distintas medidas estadísticas de los datos. <p> Conocimientos Previos</p> <p>Este es el primer tema del curso. Los requisitos previos necesarios para poder adquirir las competencias indicadas son manejar con soltura los conceptos correspondientes a los contenidos desarrollados en las asignaturas de Matemáticas de Secundaria y Bachillerato. Aunque en dichos contenidos aparecen conceptos de Estadística, es frecuente que el primer contacto de los alumnos con la materia se produzca en esta asignatura. Esto puede conllevar la realización de un pequeño esfuerzo adicional.</p> |
|--|---|

Figura 10. La página anterior en la que se ha modificado el estilo visual.

idevices para incluir actividades: Son similares a los anteriores pero incluyen uno o varios botones que permiten mostrar información adicional. Entrarían en este tipo los idevices *Actividad de lectura*, *Caso de Estudio* y *Reflexión*. En la figura 11 se muestra cómo se muestra nueva información al pulsar el botón.

The screenshot shows a digital activity interface. On the left, there is a vertical navigation menu with sections like 'Tema 1', 'Objetivos', '1.- Introducción a la Estadística', '2.- Conceptos básicos', '3.- Tablas de Frecuencias', '4.- Representaciones gráficas', 'Actividad 4.1', 'Actividad 4.1.a', 'Actividad 4.1.b' (which is highlighted), 'Actividad 4.1.c', 'Actividad 4.1.d', 'Actividad 4.1.e', 'Actividad 4.1.f', 'Actividad 4.2', 'Actividad 2 para entregar', and '3.- Medidas Estadísticas'. Below this is a 'Solución' section with a 'Click aquí' button.

The main content area is titled 'Actividad 4.1.b' and contains the following text: 'Supongamos que los datos de la tabla representan el número de hijos que tienen los empleados de una empresa.' Below this is a table:

| x_i | n_i |
|-------|-------|
| 0 | 2 |
| 1 | 4 |
| 2 | 3 |
| 4 | 1 |
| Suma | 10 |

Below the table is a bullet point: 'b) Realice la representación gráfica más apropiada para las frecuencias relativas'

Figura 11. Ejemplo de Actividad.

This screenshot shows a 'Solución' section of an activity. It contains the same bullet point as Figure 11: 'b) Realice la representación gráfica más apropiada para las frecuencias relativas' and a 'Click aquí' button. Below this is a text block: 'Si representamos gráficamente los puntos (x_i, f_i) y hacemos el diagrama de barras, nos quedaría:' followed by a bar chart. The x-axis is labeled with values 0, 1, 2, 3, 4. The y-axis has tick marks at -1, 0, and 1. Blue dots representing the data points from the table are plotted at each integer value on the x-axis.

Figura 12. Ejemplo de Actividad. Al pulsar el botón se muestra nueva información

idevices para incluir preguntas y cuestionarios: Entrarían en este tipo los idevices *Actividad de espacios en blanco*, *Pregunta elección múltiple*, *Pregunta selección múltiple*, *Pregunta verdadero falso* y *Examen SCORM*. Cada uno de estos idevices permite incluir una o varias preguntas de un mismo tipo. El examen SCORM permite crear un cuestionario con preguntas de distintos tipos. La figura 13 muestra un ejemplo de Actividad de espacios en blanco. La figura 14 muestra un ejemplo de Pregunta de elección múltiple.

Tema 1
Objetivos
1.- Introducción a la Estadística.
2.- Conceptos básicos
3.- Tablas de Frecuencias
4.- Representaciones gráficas
Actividad 2 para entregar
3.- Medidas Estadísticas.

Actividad 1

Actividad de Espacios en Blanco

Rellene los espacios en blanco

Según la definición de _____, la Estadística es la ciencia encargada de registrar hechos de interés para el Estado.

En cambio, para Fisher la Estadística es la rama de la matemática aplicada cuyo fin consiste en aplicar conceptos _____ a datos procedentes de observaciones.

Estas dos definiciones nos llevan a distinguir dos grandes ramas dentro de la Estadística:

Estadística _____. Tiene por objeto recoger, describir y analizar las características de una población, tratando de poner de manifiesto las regularidades y estructura del conjunto de la población, no ocupándose de los comportamientos individuales.

_____**Estadística**. Parte de la estadística que se encarga de generalizar los resultados obtenidos a partir de muestras a toda la población.

Enviar

« Anterior | Siguiente »

Figura 13. Ejemplo de Actividad de espacios en blanco.

Tema 1
Objetivos
1.- Introducción a la Estadística.
2.- Conceptos básicos
Actividad 2
3.- Tablas de Frecuencias
4.- Representaciones gráficas
Actividad 2 para entregar
3.- Medidas Estadísticas.

Actividad 2

Preguntas

El conjunto sobre el que se va a realizar una investigación se denomina

Muestra
 Población

Los atributos son caracteres

cuantitativos discretos
 cualitativos
 cuantitativos continuos

Respuestas propuestas

« Anterior | Siguiente »

Figura 14. Ejemplo de Pregunta de elección múltiple.

idevices para agregar elementos externos: Artículo Wiki, RSS, Sitio externo del Web. Conviene indicar que estos elementos externos no se incluyen de manera dinámica, sino que lo que se incluye es una copia estática de la página web, RSS o artículo de la Wikipedia tal y como esté en el momento de la creación de la página, pudiendo incluso modificarse el contenido de las mismas en la copia. En la figura 15 se muestra un ejemplo de inserción de un artículo wiki.

Fisher

Biografías

- Fisher**
- Laplace
- Pearson

Ronald Fisher

De Wikipedia, la enciclopedia libre
Saltar a [navegación](#), [búsqueda](#)

Ronald Aylmer Fisher, (n. 17 de febrero de 1890 – m. 29 de julio de 1962) [científico](#), [matemático](#), [estadístico](#), [biólogo evolutivo](#) y [genetista](#) inglés. Fisher realizó muchos avances en la estadística, siendo una de sus más importantes contribuciones, la [inferencia estadística](#) creada por él en 1920.

Contenido

- [1 Biografía académica](#)
- [1.1 Estadística](#)
- [1.2 Genética de poblaciones y teoría evolutiva](#)
- [1.3 Eugenesia](#)
- [2 Referencias](#)
- [3 Bibliografía](#)
- [3.1 Selección de artículos](#)
- [3.2 Libros](#)
- [3.3 Biografías](#)
- [4 Enlaces externos](#)

Biografía académica

Fisher nació en [East Finchley, Londres](#). En 1909 la escasez de sus recursos económicos y su extraordinaria capacidad académica le valieron una beca para cubrir su estancia en el [Gomville and Caius College](#) de la [Universidad de Cambridge](#), donde obtuvo su graduación en matemáticas en 1913. Dos años antes, se había convertido en uno de los fundadores más activos de la Sociedad de Eugenesia de la Universidad de Cambridge, junto con [John Maynard Keynes](#), R. C. Punnett y Horace Darwin, hijo de Charles Darwin.

Durante la guerra, Fisher atravesó momentos de extrema carestía económica. A pesar de las dificultades, comenzó a escribir reseñas de libros para la [Eugenic Review](#) e incrementó gradualmente su interés en el trabajo genético y estadístico. Publicó varios artículos sobre [biometría](#), incluyendo el célebre *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*, que inauguró la fundación de la llamada genética biométrica e introdujo la metodología del [análisis de varianza](#), considerablemente superior a la de la correlación. El artículo mostraba que la herencia de rasgos, mensurables por valores reales, los valores de variables continuas, era consistente con los principios mendelianos.



Ronald Aylmer Fisher.

Estadística

En 1919 Fisher empezó a trabajar en la [Rothamsted Experimental Station](#) ([Harpenden, Hertfordshire](#), Inglaterra). Allí comenzó el estudio de una extensa colección de datos, cuyos resultados fueron publicados bajo el título general de *Studies in Crop Variation*. Durante los siguientes siete años, se dedicó al estudio pionero de los principios del diseño de experimentos (*The Design of Experiments*, 1935), elaboró sus trabajos sobre el [análisis de varianza](#) y comenzó a prestar una atención especial a las ventajas metodológicas de la computación de datos (*Statistical Methods for Research Workers*, 1925).

Figura 15. Ejemplo de inserción de un artículo Wiki.

idevices para adjuntar otros elementos multimedia: Los idevices de este tipo son *Galería de imágenes*, *Imagen ampliada* y *Applet de Java*. Los dos primeros permiten mostrar varias imágenes en una galería o ampliar partes de una imagen, siendo más usadas en proyectos correspondientes a otras ramas de conocimiento. Sin embargo la última opción es especialmente interesante en el ámbito de la Estadística, ya que permite integrar elementos interactivos en el proyecto. En particular permite integrar applets creados con GeoGebra. En la figura 16 se muestra la inclusión de un applet creado con GeoGebra que permite calcular interactivamente los valores de la distribución normal estándar.

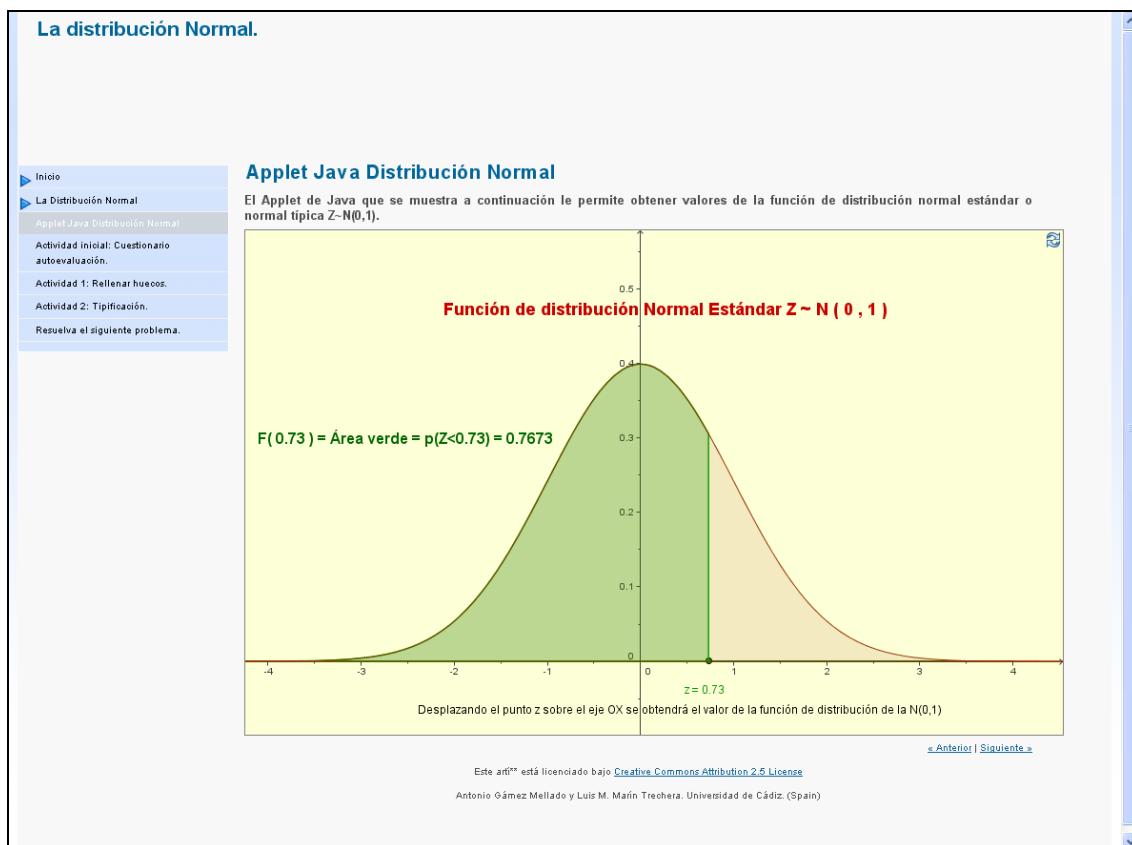


Figura 16. Ejemplo de inserción de un Applet de Java

5. Conclusiones

Las herramientas mostradas permiten con poco esfuerzo crear materiales didácticos interactivos para su uso a través de Internet. Además de servir para proporcionar a los estudiantes información para facilitarles la adquisición de contenidos permiten dotarles de medios afectivos para realizar una autoevaluación de las capacidades adquiridas. Además, los materiales generados están dotados de interactividad, lo que les proporciona un valor añadido.

Ambas herramientas permiten además exportar los materiales en formato SCORM, lo que garantiza su uso en distintas plataformas LMS. Es precisamente al integrar estas herramientas con un LMS cuando se les puede extraer mayores niveles de rendimiento.

6. Agradecimientos

Queremos agradecer a los organizadores de las I Jornadas de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa, celebradas en Melilla en junio de 2010, así como a los integrantes del Grupo de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa (GENAEIO) de la SEIO, por las facilidades dadas para la celebración de las mismas y por el interés mostrado.

7. Bibliografía

- [1] Nielsen, J. (2010). *College Students on the Web: Usability Guidelines for Creating Compelling Websites for College Students. Some new estimates for distribution functions.* Nielsen Norman Group Report. <http://www.nngroup.com/reports/students/>
- [2] Página principal de Hot Potatoes. <http://hotpot.uvic.ca/>
- [3] Página principal del proyecto eXe. <http://exelarning.org/wiki>
- [4] Dublin Core Metadata Initiative. <http://www.dublincore.org/>

Nuevas plataformas virtuales para la docencia

Daza Arbolí, Miguel Ángel¹

¹mdazarb@et.mde.es, CIODEF (MINISDEF)

Resumen

Este trabajo pretende ser una guía de cómo configurarse uno, personalmente, su propio sitio de Moodle, paso a paso. Instalando un servidor Apache, una Base de Datos MySQL y realizando los primeros ajustes sobre Moodle en cuanto a Idioma, Apariencia, Calendario. Dejando para un taller posterior como crear cursos y desarrollarlos.

Palabras clave: Moodle, Plataforma Virtual Educativa, Instalación.

Clasificación AMS: 97C70, 97D40.

1. Introducción

Lo más habitual es que nuestro conocimiento de Moodle sea como una plataforma virtual educativa ya montada, bien sea en el servidor de la Universidad, Instituto, Colegio, Empresa, etc... Pues bien, el objetivo que persigue este taller es mostrar lo fácil que puede resultar instalar un servidor de Moodle propio. En nuestro propio ordenador, con el que poder realizar todas las facilidades que proporciona Moodle. Crear nuestros propios cursos, nuestros recursos, y si fuera necesario, exportarlos al servidor de nuestra institución.

2. Primeros Pasos.

En primer lugar lo que debemos hacer es visitar la página web de Moodle (www.moodle.org). Una vez allí dirigirnos al apartado de descargas. Buscamos la descarga que mejor se ajuste a nuestras condiciones de ordenador y sistema operativo.

Continuaremos la explicación con el supuesto de que elegimos como sistema operativo Windows y la versión 1.97 de Moodle.

Después de realizar la descarga del archivo, archivo que viene comprimido en formato zip, lo único que debemos hacer es descomprimir este archivo sobre una carpeta (la que deseemos).

En la imagen de la derecha se muestra la carpeta resultante al descomprimir el archivo ZIP.

La carpeta se compone de tres archivos y una carpeta.

A destacar los ejecutables START MOODLE y STOP MOODLE que son los que posibilitarán que se ejecute o detenga el servidor apache y la base de datos MySQL.

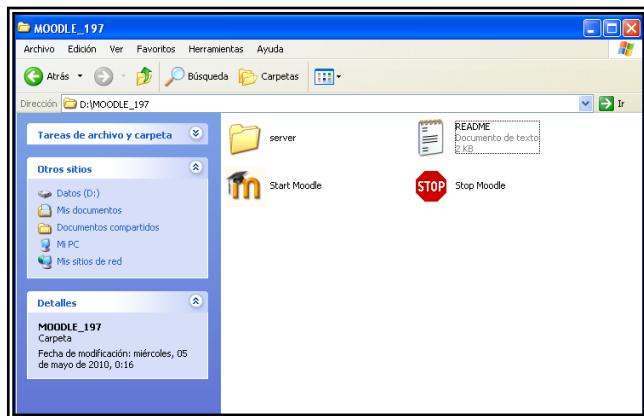


Figura 1.- Carpeta con el contenido archivo ZIP

3. Habilitando Servidor Apache y Base de Datos MySQL.

Para el funcionamiento correcto de Moodle, necesitamos instalar una base de datos (en MySQL) y un servidor de páginas WEB como Apache.

Bien, para iniciar la activación del servidor Apache y de la Base de Datos MySQL, sólo debemos ejecutar el archivo START MOODLE, con lo que se iniciará el proceso de instalación del servidor **Apache HTTP Server**, solicitando al usuario permiso para instalarlo y ejecutarlo. Posteriormente solicitará permisos para instalar la Base de Datos MySQL.



Figura 2.- Aviso Servidor Apache



Figura 3.- Aviso Base de Datos MySQL

Por defecto, en la instalación sobre nuestro equipo portátil, la dirección del sitio Moodle será <http://localhost>.

Al poner esa dirección en un navegador la primera página que se cargará será <http://localhost/install.php>.

4. Instalando y configurando Moodle.

La página <http://localhost/install.php> nos muestra el comienzo de la instalación. Aparece un desplegable para que seleccionemos el idioma que seguiremos durante la instalación. En nuestro caso elegimos Español-Internacional (es) y la página además nos muestra la siguiente información:

*Esta versión de **Moodle Windows Installer** incluye las aplicaciones necesarias para que **Moodle** funcione en su ordenador, principalmente:*

- Apache 2.2.11
- PHP 5.2.9
- MySQL 5.1.33 (Community Server)

Pulsamos el botón SIGUIENTE y la siguiente página que nos muestra el proceso de instalación son los resultados de una serie de comprobaciones que realiza para verificar si Moodle funcionará correctamente o no.

| | Correcto |
|-------------------------------------|----------|
| Versión PHP | Correcto |
| Inicio automático de sesión | Correcto |
| Magic Quotes Run Time | Correcto |
| Manejo Inseguro de Ajustes Globales | Correcto |
| Modo Seguro (Safe Mode) | Correcto |
| Subidas de archivos | Correcto |
| Versión GD | Correcto |
| Límite de memoria | Correcto |

Figura 4.- Comprobaciones previas.

Figura 5.- Direcciones Web y Directorios

Pulsamos el botón SIGUIENTE y la siguiente página nos muestra la dirección del servidor y de las carpetas donde se guardarán los datos. Estas direcciones se pueden cambiar y ajustar a nuestras necesidades.

Si estamos conformes pulsamos el botón SIGUIENTE.

La nueva página que aparece nos solicita algunos datos para terminar de configurar la Base de Datos y el acceso a ésta. Entre ellos el NOMBRE de la Base de Datos y una Contraseña para el “súper usuario” llamado “root”.

moodle

Instalación

Ahora necesita configurar los ajustes de la base de datos donde se almacenarán la mayoría de los datos de Moodle. El instalador creará la base de datos con los ajustes especificados más abajo.

Tipo: fijado a "mysql" por el instalador
Host: fijado a "localhost" por el instalador
Nombre: nombre de la base de datos, e.g., moodle
Usuario: el valor por defecto es "root"
Contraseña: contraseña de la base de datos
Prefijo de Tablas: prefijo opcional utilizado por todos los nombres de las tablas

Tipo: mysql
Servidor: localhost
Base de datos: moodle
Usuario: root
Contraseña: _____
Prefijo de tablas: mdl_

< Anterior Siguiente >

Figura 6.- Datos SUPER-USUARIO

Completamos los datos necesarios y pulsamos el botón SIGUIENTE.

Comprobaciones del Servidor

| Nombre | Información | Informe | Estado |
|---------------|-------------|--|--------|
| unicode | | ● es obligatorio que esté instalado/activado | OK |
| database | mysql | ● versión 4.1.16 es obligatoria y está ejecutando 5.1.33 | OK |
| php | | ● versión 4.3.0 es obligatoria y está ejecutando 5.2.9 | OK |
| php_extension | iconv | ● es recomendable que esté instalado/activado | OK |
| php_extension | mbstring | ● es recomendable que esté instalado/activado | OK |
| php_extension | curl | ● es recomendable que esté instalado/activado | OK |
| php_extension | openssl | ● es recomendable que esté instalado/activado | OK |
| php_extension | tokenizer | ● es recomendable que esté instalado/activado | OK |
| php_extension | xmlrpc | ● es recomendable que esté instalado/activado | OK |
| php_extension | ctype | ● es recomendable que esté instalado/activado | OK |
| php_extension | pcntl | ● es obligatorio que esté instalado/activado | OK |

< Anterior Siguiente >

La siguiente página muestra los resultados de unas comprobaciones realizadas al sistema:

Pulsamos el botón SIGUIENTE.

Figura 7.- Comprobaciones del Servidor

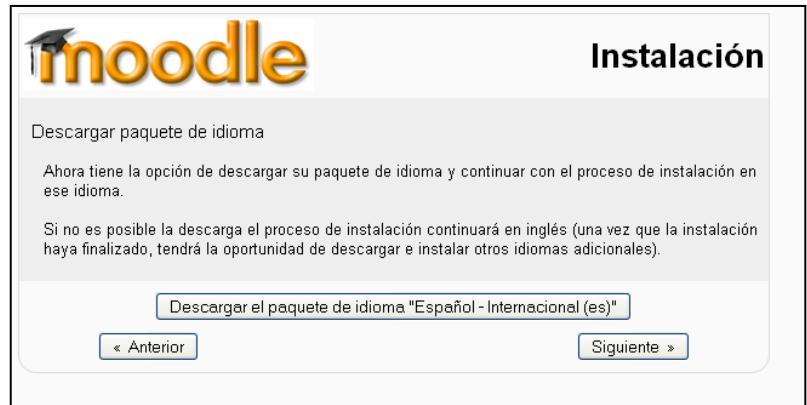


Figura 8.- Descarga paquete Idioma

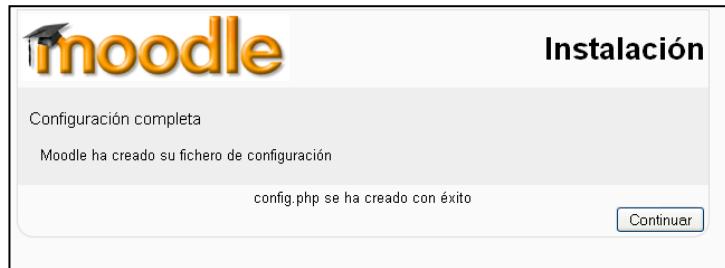


Figura 9.- Creación archivo CONFIG.PHP

Pulsando el botón SIGUIENTE, el proceso de instalación nos informa que se ha completado con éxito la creación del fichero CONFIG.PHP.

Pulsando el botón CONTINUAR.

Una vez terminada la creación del archivo de configuración, el proceso de instalación nos cuenta la licencia de distribución que posee Moodle, preguntándonos si hemos entendido los términos de la misma.

Después de contestar, afirmativamente, continuamos con el proceso.

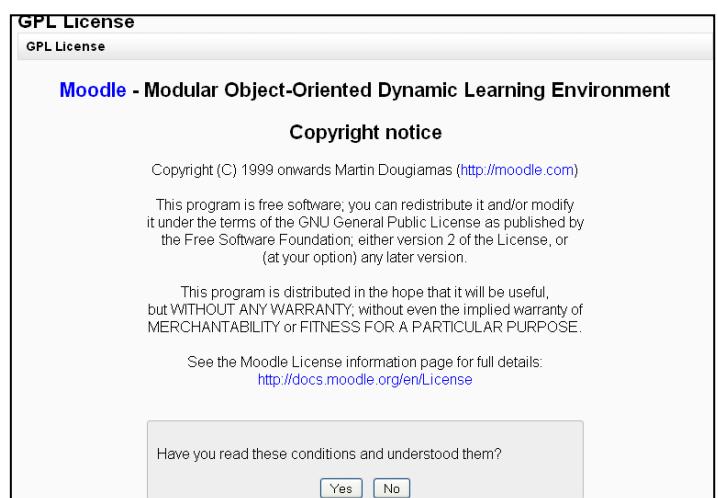


Figura 10.- Licencia de Distribución

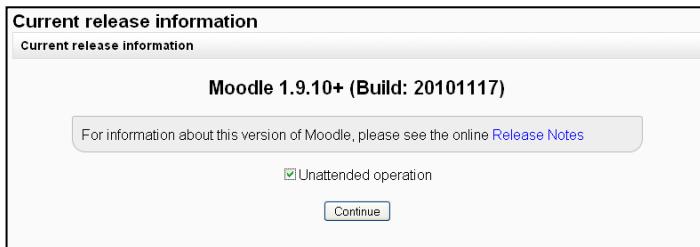


Figura 11.- Versión de Moodle que se instalará

Al marcar esta opción lo que conseguimos es que el proceso de instalación no se detenga después de la creación de cada una de las, innumerables, tablas que necesita crear en la Base de Datos MySQL, antes definida.

La siguiente página nos indica que versión de Moodle se instalará.

Pero lo más relevante de esta página es la posibilidad de marcar "Unattended operation" (operación desatendida).

This screenshot shows the 'Setup administrator account' page. It has a 'General' tab selected. A note at the top says: 'On this page you should configure your main administrator account which will have complete control over the site. Make sure you give it a secure username and password as well as a valid email address. You can create more admin accounts later on.' The form fields include:

- Username*: admin
- New password*: [redacted] Urmask
- Force password change:
- First name*: Admin
- Surname*: User
- Email address*: [redacted]
- Email display: Allow everyone to see my email address
- Email activated: This email address is enabled
- City/town*: [redacted]
- Select a country*: Select a country...
- Timezone: Server's local time
- Preferred language: English (en)
- Description: [redacted]

Figura 12.- Datos ADMINISTRADOR Moodle

This screenshot shows the 'Edit profile' page for the administrator account. It includes sections for:

- Picture of**: Shows a placeholder for a profile picture with a 'New picture' button and a 'Picture description' input field.
- Interests**: A text area for listing interests.
- Optional**: A section containing a 'Show Advanced' link and an 'Update profile' button.

At the bottom, a red note says: 'There are required fields in this form marked*.'

Figura 13.- Datos del ADMINISTRADOR Moodle

Para terminar pulsar el botón UPDATE PROFILE, (actualizar perfil).

The settings shown below were added during your last Moodle upgrade. Make any changes necessary to the defaults and then click the "Save changes" button at the bottom of this page.

New settings - Front Page settings

Full site name:

Short name for site (eg single word):

Front Page Description:
summary:
Trebuchet, 10pt, Long,
Path:
This description of the site will be displayed on the front page.

New settings - Manage authentication

Self registration: Default: Disable
If an authentication plugin, such as email-based self-registration, is selected, then it enables potential users to register themselves and create accounts. This results in the possibility of spammers creating accounts in order to use forum posts, blog entries etc. for spam. To avoid this risk, self-registration should be disabled or limited by Allowed email domains setting.

Figura 14.- Ajuste página principal SITIO

Lo siguiente que se nos pide es el **nombre completo de nuestro sitio** (Full site name). Un **nombre abreviado** para el sitio (Short name for site). Una descripción para la página principal (Front Page Description).

El desplegable nos da la opción de habilitar, o no, el que los nuevos usuarios se den de alta ellos mismos a través de e-mail.

Al final el resultado es:

Cursos MAD

You are logged in as Admin User (Logout) English (en) Turn editing on

Site Administration □

- Notifications
- Users
- Courses
- Grades
- Location
- Language
- Modules
- Security
- Appearance
- Front Page
- Server
- Networking
- Reports
- Miscellaneous

Available Courses

No courses in this category

Cursos MAD

Calendar □ December 2010

| Sun | Mon | Tue | Wed | Thu | Fri | Sat |
|-----|-----|-----|-----|-----|-----|-----|
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | 29 | 30 | 31 | |

You are logged in as Admin User (Logout)

moodle

Figura 15.- Vista página principal del sitio

5. Ajuste del Idioma.

El siguiente paso es cargar los archivos de los idiomas que deseemos, en nuestro caso nos interesa el idioma Español-Internacional.

The screenshot shows the Moodle Site Administration interface under the Language section. On the left, there's a sidebar with various administration links like Notifications, Users, Courses, etc. The main area has a heading 'Language packs' with a note: 'Because Moodle can not connect to download.moodle.org, we are unable to do language pack installation automatically. Please download the appropriate zip file(s) from the list below, copy them to your C:\nMOODLE\server\moodledata\lang\ directory and unzip them manually.' Below this, there are two tabs: 'Installed language packs' (which currently only lists 'English (en)') and 'Available language packs'. The 'Available language packs' tab lists many languages with their names in their respective scripts and download links. The 'Español - Internacional (es)' entry is highlighted with a red border.

| Available language packs | Action |
|-------------------------------------|--------------------------|
| Afrikaans | Download |
| አማርኛ | Download |
| Беларуская | Download |
| Български | Download |
| Bosanski | Download |
| Català | Download |
| Čeština | Download |
| Dansk | Download |
| Deutsch - Du | Download |
| Deutsch | Download |
| Ελληνικά | Download |
| English | Download |
| Español - Argentina | Download |
| Español - España | Download |
| Español - Mexico | Download |
| Español - Internacional (es) | Download |

Figura 16.- Cargar paquetes de idiomas

Una vez descargado nos aparecerá en el apartado de paquetes de lenguajes instalados (Installed language packs).

This screenshot shows the same Moodle Site Administration interface as Figure 16, but now the 'Installed language packs' tab is selected, showing both 'English (en)' and 'Español - Internacional (es)'. The rest of the interface and the available language list are identical to Figure 16.

| Installed language packs |
|------------------------------|
| English (en) |
| Español - Internacional (es) |

Figura 17.- Paquetes de Idiomas cargados

Una vez descargado el Idioma, lo seleccionamos.

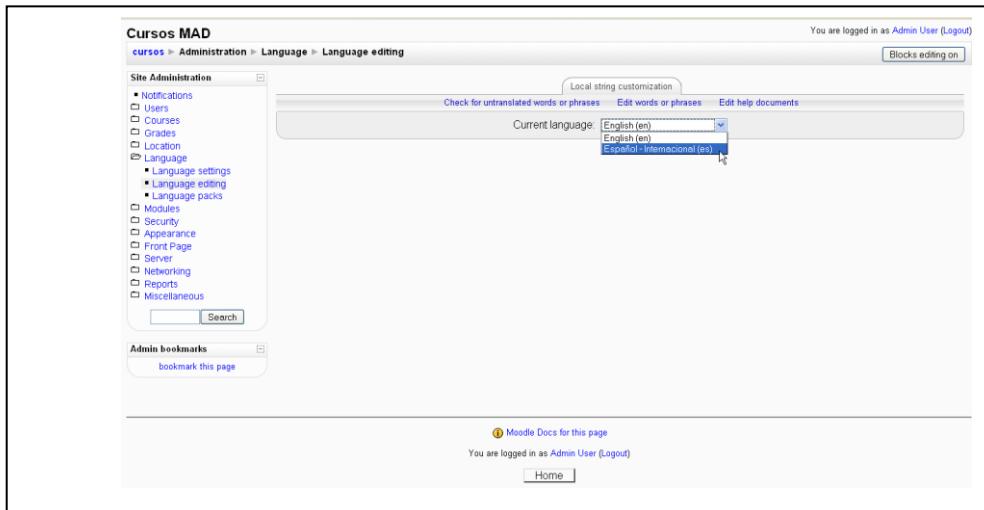


Figura 18.- Idioma actual seleccionado

Una vez seleccionado los apartados se muestran en el idioma elegido:

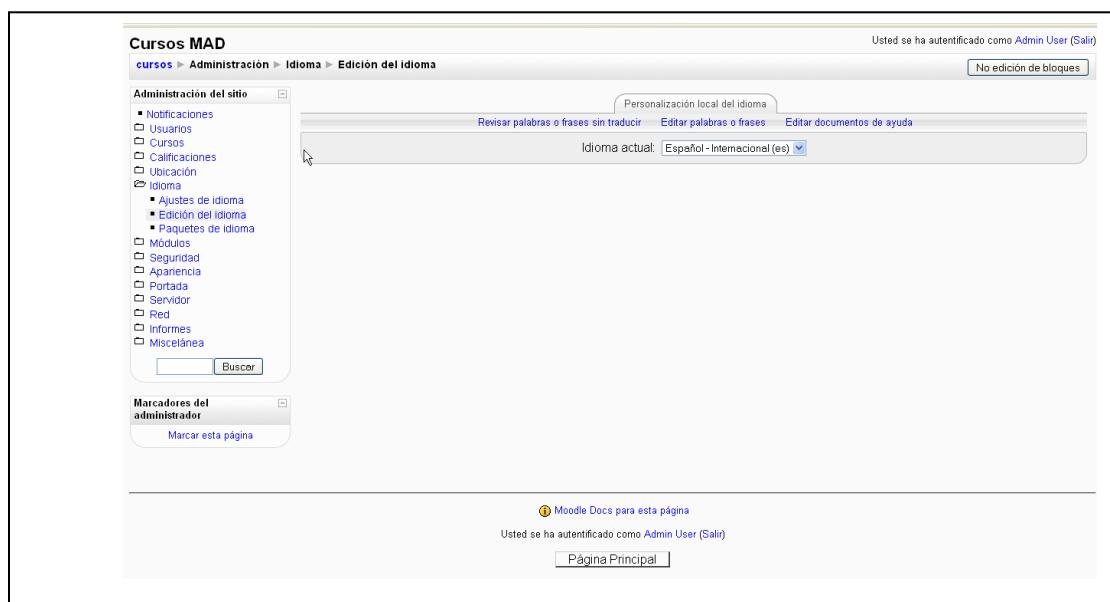


Figura 19.- Idioma actual seleccionado

También en el apartado **Ajustes de idioma** seleccionamos el idioma correspondiente:

The screenshot shows the 'Ajustes de idioma' (Language settings) page in Moodle. On the left, there's a sidebar with 'Administración del sitio' and 'Marcadores del administrador'. The main area is titled 'Ajustes de idioma'. It contains several configuration options with checkboxes and dropdown menus. One dropdown menu, 'Idioma por defecto', has 'English (en)' selected and 'Español - Internacional (es)' as an alternative. Other settings include 'Detectar idioma automáticamente', 'Mostrar menú de idioma', 'Idiomas en el menú de idioma', 'Guardar en caché el menú de idioma', and 'Locale' del sitio.

Figura 20.- Selección del Idioma actual

Nuestra página principal pasa a tener la siguiente forma:

The screenshot shows the Moodle homepage ('Cursos disponibles'). At the top right, it says 'Usted se ha autenticado como Admin User (Salir) Español - Internacional (es)'. Below that is a search bar with 'Activar edición'. The main content area says 'No hay cursos en esta categoría' and has a 'Agregar un nuevo curso' button. On the left is the 'Administración del sitio' sidebar. On the right is a 'Calendario' for December 2010, showing dates from 1 to 31. The moodle logo is at the bottom.

Figura 21.- Vista página principal sitio

6. Ajuste del Calendario.

El siguiente paso es ajustar el Calendario:

Cursos MAD

Usted se ha autenticado como Admin User (Salir)

No edición de bloques

Administración del sitio

- Notificaciones
- Usuarios
- Cursos
- Calificaciones
- Ubicación
- Idioma
- Módulos
- Seguridad
- Apariencia
- Temas
- Calendario
- Editor HTML
- Ajustes HTML
- Moodle Docs
- Mi Moodle
- Gestores de curso
- AJAX y Javascript
- Gestionar marcas
- Portada
- Servidor
- Red
- Informes
- Miscelánea

Buscar

Marcadores del administrador

Marcar esta página

Calendario

Los administradores ven todo

Pueden los administradores ver todos los eventos del calendario o sólo los que les conciernen?

Formato de hora Por defecto Valor por defecto: Por defecto

Puede seleccionar el formato en que se ven las horas (12 o 24) en todo el sitio. Si selecciona la opción por defecto, el formato se elegirá automáticamente de acuerdo con el idioma que utilice en el sitio. Este ajuste puede pasarse por alto mediante las preferencias del usuario.

Comienzo de la semana Lunes Valor por defecto: Domingo

Domingo

Lunes

Martes

Miércoles

Jueves

Viernes

Sábado

¿Cuántos días de la semana se considerarán como 'fin de semana' y se mostrarán con un color diferente?

Días del fin de semana calendario_weekend

Días de previsión 21 Valor por defecto: 21

¿Cuántos días debe considerar el calendario por defecto para eventos próximos?

Eventos de previsión 10 Valor por defecto: 10

¿Cuántos eventos próximos se mostrarán como máximo a los usuarios?

calendar_startwday

calendar_timeformat

calendar_maxevents

calendar_lookahead

Figura 22.- Ajuste Calendario

En especial modificar el día de **comienzo de la semana**.

Cursos MAD

Usted se ha autenticado como Admin User (Salir)

Español - Internacional (es) ▾

Activar edición

Administración del sitio

- Notificaciones
- Usuarios
- Cursos
- Calificaciones
- Ubicación
- Idioma
- Módulos
- Seguridad
- Apariencia
- Portada
- Servidor
- Red
- Informes
- Miscelánea

Buscar

Cursos disponibles

No hay cursos en esta categoría

Agregar un nuevo curso

Cursos MAD

Calendario

diciembre 2010

| Lun | Mar | Mie | Jue | Vie | Sáb | Dom |
|-----|-----|-----|-----|-----|-----|-----|
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | 31 | | |

Usted se ha autenticado como Admin User (Salir)

moodle

Figura 23.- Vista página principal del sitio

Cambiar la apariencia de nuestro Moodle.

El Tema de fondo de Moodle se puede cambiar, modificando la apariencia con la que Moodle nos irá presentando cada una de las páginas, y las distintas secciones dentro de las mismas:

The screenshot shows the Moodle Admin interface under 'Apariencia > Temas'. On the left, a sidebar lists various site administration categories. The main area is titled 'Temas' and displays three theme preview cards: 'chameleon', 'cornflower', and 'custom_corners'. Each card includes a preview image, the theme name, and a list of options: 'Vista previa', 'Información', and an 'Elegir' button. The 'chameleon' theme is currently selected, as indicated by its larger preview image and the highlighted 'Elegir' button.

Figura 24.- Seleccionar un Tema

This screenshot shows the Moodle Admin interface after selecting the 'formal_white' theme. The main area now displays the Moodle homepage with the 'formal_white' theme applied. The theme features a light gray header with the 'moodle' logo, a white navigation bar with blue links, and a white content area. On the right side, there is a sidebar with course links and a search bar. The 'Elegir' button is highlighted with a yellow box and a cursor icon, indicating it has been clicked.

Figura 25.- Tema "formal_white"

Y el aspecto de nuestra página principal pasa a ser el siguiente:

The screenshot shows the Moodle administration interface. At the top, there's a navigation bar with links for 'Notificaciones', 'Usuarios', 'Cursos', 'Calificaciones', 'Ubicación', 'Idioma', 'Módulos', 'Seguridad', 'Apariencia', 'Portada', 'Servidor', 'Red', 'Informes', and 'Miscelánea'. Below this is a search bar with a 'Buscar' button. The main content area is titled 'Cursos disponibles' and displays the message 'No hay cursos en esta categoría'. In the center, there's a button labeled 'Agregar un nuevo curso'. To the right, there's a sidebar titled 'Cursos MAD' and a 'Calendario' section showing the month of December 2010. The calendar grid includes dates from 1 to 31, with some days highlighted in red. At the bottom left, there's a small 'moodle' logo.

Figura 26.- Vista página principal sitio

Y ahora sólo nos queda empezar a agregar cursos, contenidos, alumnos, grupos, actividades, etc ..., pero eso será objeto de otro Taller.

7. Webgrafía

- [1] www.moodle.org Página principal de MOODLE.

Acerca del autor



Miguel Ángel Daza es Técnico Estadístico en el Centro de Investigación Operativa de la Defensa.

Licenciado en Ciencias y Técnicas Estadísticas por la Facultad de Matemáticas de la Universidad de Valencia.

Pertenece al Grupo de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa (GENAEIO) de la Sociedad de Estadística e Investigación Operativa (SEIO.)

Interesado en los entornos virtuales de enseñanza y en especial en el uso de MOODLE. Es, desde 2009, el administrador de la plataforma montada sobre MOODLE para uso del grupo (<http://genaeio.moodlehub.com/>).

La enseñanza de la Estadística en las nuevas titulaciones adaptadas al Espacio Europeo de Educación Superior

Rivera Galicia, Luis F.¹

¹luisf.rivera@uah.es, Departamento de Estadística, E.E. y O.E.I.
Facultad de Ciencias Económicas y Empresariales. Universidad de Alcalá

Resumen

Durante el curso 2009-2010 se han implantado en la Universidad de Alcalá las titulaciones de los nuevos Grados en Ciencias Sociales y Jurídicas, adaptados a las directrices marcadas por el Espacio Europeo de Educación Superior. En este trabajo se presenta cómo ha quedado encuadrada la materia de Estadística en los planes de estudio que se imparten en la Facultad de Ciencias Económicas y Empresariales y en la Escuela de Turismo de la Universidad de Alcalá. Se analizan los resultados académicos y la efectividad de los diversos modelos de evaluación continua empleados frente a la evaluación final tradicional.

Palabras clave: Enseñanza de la Estadística; Espacio Europeo de Educación Superior.

Clasificación AMS: 97C70, 97D40.

1. El Espacio Europeo de Educación Superior.

En los últimos años se ha producido en la Universidad española un profundo proceso de transformación motivado por la decisión de adecuar el mismo a las normativas y directrices marcadas por el denominado Espacio Europeo de Educación Superior (en adelante EEES; en inglés se denomina European Higher Education Area: EHEA). Este proceso, impulsado por la Unión Europea, obliga a que las universidades dejen de tener como marco de referencia el nacional para tener un marco de referencia mucho más amplio, a nivel europeo.

El EEES supone un reto porque obliga a introducir múltiples cambios en cuanto a la concepción de la enseñanza universitaria. En cuanto a su organización, porque persigue establecer una estructura homogénea de titulaciones y de formación continua entendible por todos los Estados miembros, que permita que un estudiante pudiera circular libremente por el sistema universitario europeo. De la misma manera, se ha producido un cambio en los calendarios docentes, ya que se ha trasladado la convocatoria de exámenes extraordinarios de septiembre al mes de julio, lo cual hace que se modifique el modo de hacer la matriculación, los períodos de actividad docente, de exámenes, etcétera.

Desde el punto de vista del docente, es necesario tener en cuenta varios aspectos importantes: En primer lugar, es inevitable que se produzca un cambio importante en las metodologías docentes. Ahora el proceso de enseñanza-aprendizaje no se debe basar sólo en

transmisión de contenidos o conocimientos, sino que es preciso que el estudiante sea capaz de adquirir las competencias y destrezas necesarias para asegurar que se domina una materia. En segundo lugar, se cambia en consecuencia la forma de contabilizar la actividad desarrollada, para que sea posible hacer una justa valoración de la carga de trabajo que debe completar un estudiante para adquirir las competencias que se desarrollan en una materia concreta. Por tanto, el alumno pasa a ser ahora el centro del proceso de enseñanza-aprendizaje, debiendo ser capaz de realizar su propio progreso (Zabalza, 2007, págs. 188-223).

Este hecho motiva la necesidad de organizar el currículum universitario según una nueva norma, que se ha denominado ECTS (European Credit Transfer System). El sistema de estructuración de materias siguiendo los criterios de valoración ECTS es la decisión más importante y determinante para la reorganización de los planes de estudio universitarios (véase, por ejemplo, Goñi Zabala, 2005, pp. 32).

El crédito tradicional, medida de la carga de trabajo en una materia, se establecía en términos de horas de docencia presencial que se impartía en el aula para explicar los contenidos de una materia concreta. Por tanto, medía la carga de trabajo que suponía para el profesor la impartición de los contenidos de una asignatura. En el EEES, en cambio, la medición de la carga de trabajo se establece en función del trabajo desarrollado por el estudiante. Concretamente, se ha establecido un crédito ECTS como el equivalente a 25-30 horas de trabajo del alumno, entre horas de clase presencial teóricas y prácticas, estudio personal, asistencia a seminarios, elaboración de trabajos, proyectos o casos prácticos y la preparación y realización de exámenes o pruebas de evaluación. Para tratar de acotar la dedicación de un estudiante a lo largo del curso escolar, se ha establecido que un curso completo conste de 60 ECTS, lo que supone un total de unas 1500 a 1800 horas de trabajo total de un estudiante. En la Universidad de Alcalá, se ha establecido que la presencialidad de estos estudiantes se sitúe en torno al 30% de su carga de trabajo, por lo que se ha fijado el crédito ECTS equivalente a unas 8 horas de clase presencial.

Basándose, por tanto, en esta contabilización de la carga de trabajo de un estudiante, antes y después de la reforma del EEES, la estructura de los estudios superiores cambia para adaptarse al nuevo esquema, como se muestra en la Figura 1:

| ESTRUCTURA ANTIGUA | | ESTRUCTURA NUEVA | |
|--------------------------------------|---|---|----------------------|
| <u>1 crédito = 10 horas lectivas</u> | | <u>1 crédito ECTS= 25 horas de trabajo del estudiante</u> | |
| 1º ciclo | 2 - 3 años | 1º ciclo | 4 años (240 ECTS)* |
| | Diplomado/a, Arquitecto/a Técnico/a, Ingeniero/a Técnico/a | | Graduado/a |
| | 3 años | | |
| 2º ciclo | 2 años | 2º ciclo | 2 años (60-120 ECTS) |
| | Licenciado/a, Ingeniero/a, Arquitecto/a | | Máster |
| 3º ciclo | 3 -4 años | 3º ciclo | 4 años |
| | Doctor/a | | Doctor/a |

Figura 1: Cambio en la estructura de las Titulaciones universitarias.

Este trabajo se distribuye de la siguiente manera: en primer lugar, se hace un breve recorrido por el proceso seguido por la Universidad de Alcalá (UAH) para implantar los planes de estudio de acuerdo con los conceptos fundamentales del Espacio Europeo de Educación Superior, describiendo la transformación de los planes de estudio de los títulos de la rama de Ciencias Sociales y Jurídicas en los que se imparte la materia de Estadística, en su transición desde las licenciaturas y diplomaturas a los grados. En la sección 3, se presenta cómo se ha implantado la

materia de Estadística, así como los métodos de evaluación empleados en cada una de las titulaciones en las que se imparte. En la sección 4 se analizan los resultados académicos obtenidos por los estudiantes de cada una de las asignaturas impartidas durante el curso 2009-2010. Por último, se incluye un apartado de conclusiones.

2. La implantación de los títulos de Grado en la Universidad de Alcalá.

La Universidad de Alcalá comenzó la impartición de algunos títulos de Grado durante el curso 2008-2009. Concretamente, estos grados fueron los de Ciencias Ambientales, Química, Estudios Ingleses y Derecho. A lo largo de ese curso, la Universidad de Alcalá presentó a la ANECA las memorias de otros muchos grados, entre los que se encontraban los Grados en Turismo, en Administración y Dirección de Empresas, y Economía. Tras su aprobación, se planificó el inicio de impartición de estos estudios para el curso 2009-2010. Otros estudios, como el Grado en Contabilidad y Finanzas o el Grado en Economía y Negocios Internacionales, fueron aprobados a principio del curso 2009-2010, y se fijó su inicio de impartición para el curso 2010-2011.

Durante los últimos años, se ha producido en la UAH una enorme actividad normativa, en la que se ha intentado preparar el paso al modelo de enseñanza del EEEES. Destacan, por su importancia en la organización de los estudios de Grado, los siguientes¹:

- Normativa de Matrícula y Régimen de Permanencia en los Estudios de Grado. Tiene por objeto regular el régimen de matriculación y la permanencia de los estudiantes en los estudios oficiales de Grado de la UAH.
- Normativa de Adaptación al Grado. Esta normativa es de aplicación exclusiva a los estudiantes de la UAH que cursan estudios en un plan de estudio en proceso de extinción como consecuencia de la implantación de un estudio de grado, y aquellos otros que habiendo iniciado estudios en un plan de estudios en proceso de extinción, han optado por adaptar sus estudios al nuevo estudio de grado (es decir, prefieren terminar sus estudios en un Grado equivalente a los estudios de Licenciatura o Diplomatura en los que comenzaron).
- Normativa de Reconocimiento y Transferencia de Créditos. Se entiende por reconocimiento la aceptación por parte de la Universidad de los créditos que, habiendo sido obtenidos en unas enseñanzas oficiales de Grado o en enseñanzas conducentes a títulos antiguos, en la misma o en otra universidad, son computados en otra enseñanza distinta a efectos de la obtención del título oficial de Grado, mientras que la transferencia de créditos consiste en la constancia en el expediente académico del estudiante de la totalidad de los créditos obtenidos en enseñanzas oficiales cursadas con anterioridad, en la misma u otra universidad, que no hayan conducido a la obtención de un título oficial.

¹ Para más información sobre las decisiones normativas de la Universidad de Alcalá, consultese la página http://www.uah.es/acceso_informacion_academica/primero_segundo_ciclo/normativa_academica/inicio.htm. Última consulta el 1 de diciembre de 2010.

- Normativa sobre Evaluación en los Estudios de Grado. El sistema más congruente con la nueva metodología del EEEES es el de la evaluación continua. No obstante, el artículo 144 de los Estatutos de la Universidad de Alcalá reconoce, junto al sistema de evaluación continua como criterio inspirador, el derecho a la realización de pruebas finales. Mientras continúe en vigor esta disposición, es necesario compaginar la evaluación continua con el derecho de los alumnos que lo deseen a la realización de un examen final.

3. La Estadística en los planes de estudio de las titulaciones de Ciencias Sociales y Jurídicas en la UAH.

Hasta el curso 2008-2009, en la UAH había cinco estudios de la rama de Ciencias Sociales y Jurídicas en cuyo currículum se incluía alguna asignatura de Estadística. Concretamente, había tres licenciaturas que se impartían en la Facultad de Ciencias Económicas y Empresariales: la Licenciatura en Administración y Dirección de Empresas, la Licenciatura en Economía, y la Licenciatura en Ciencias Actuariales y Financieras (estudios de segundo ciclo). En el Edificio Multidepartamental que la Universidad de Alcalá posee en Guadalajara se impartían las Diplomaturas en Turismo y en Ciencias Empresariales. Estas últimas titulaciones eran de primer ciclo, y se han adaptado al EEEES transformándose en sendos grados. En la Figura 2 se muestra cómo ha sido la transición entre los planes antiguos y los planes adaptados al EEEES.

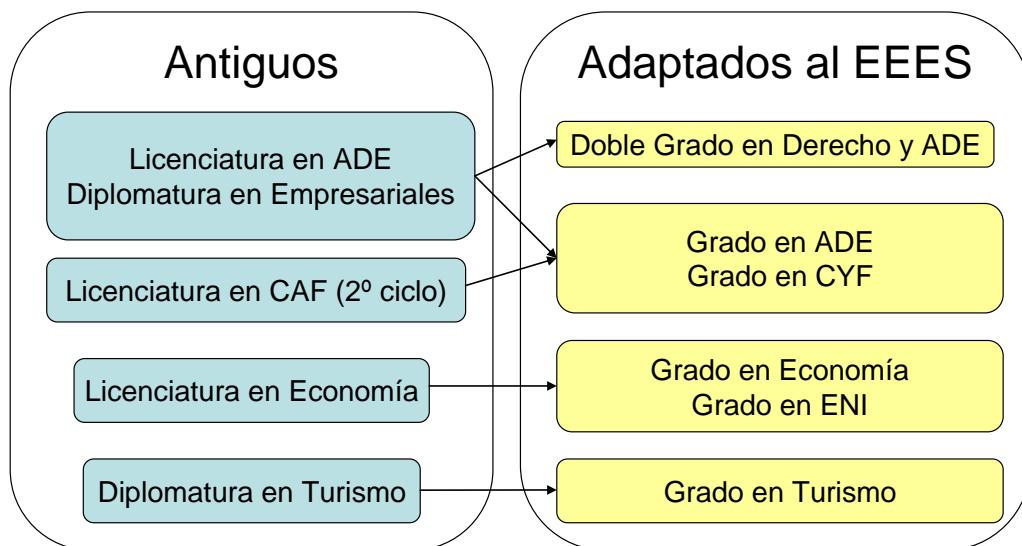


Figura 2: Equivalencia entre Licenciaturas y Grados en el área de Ciencias Sociales y Jurídicas de la UAH.

En cuanto a los contenidos de la Estadística en estos 6 Grados, se han organizado de tres maneras diferentes. En primer lugar, la Estadística del Grado en Turismo se centra en la Estadística Descriptiva y en el estudio de la evolución temporal de magnitudes (a través de series temporales y números índice). En segundo lugar, la Estadística de los Grados de Administración y Dirección de Empresas (ADE) y Contabilidad y Finanzas (CYF), en los que el peso de la Estadística es mayor, llega en su desarrollo hasta los contrastes de hipótesis paramétricas, con lo que se han asignado dos asignaturas de 6 ECTS en cada estudio. De la misma manera, la Estadística del Doble Grado en Derecho y Administración y Dirección de Empresas (DADE) ha quedado estructurada exactamente igual, ya que el plan de estudios del

Doble Grado se ha elaborado yuxtaponiendo las materias específicas de cada uno de los Grados, y respetando las materias comunes. En tercer lugar, en los Grados de Economía (ECO) y Economía y Negocios Internacionales (ENI) se ha configurado una asignatura de Estadística en primer curso, algo más extensa (valorada en 9 créditos ECTS), que incluye toda una visión general de la Estadística (desde la descriptiva hasta la inferencia paramétrica), y en el segundo curso del Grado en Economía se ha incluido, además, una asignatura adicional de Estadística de 6 créditos en la que se pretende profundizar en aspectos concretos como los contrastes no paramétricos y las técnicas de muestreo.

A continuación se presentan las distintas asignaturas de Estadística que hemos impartido durante el curso 2009-2010 en el Departamento de Estadística, Estructura Económica y Organización Económica Internacional de la UAH, especificando la evolución sufrida en el plan de estudios, los contenidos impartidos y la metodología de evaluación utilizada en cada una de ellas.

3.1. La Estadística en los estudios de Turismo.

La Estadística en la Diplomatura en Turismo se impartía en una única asignatura denominada “Estadística Descriptiva”, de 4,5 créditos (3 horas semanales de clase) en el primer cuatrimestre del primer curso. Los contenidos que se desarrollaban se correspondían con la estadística descriptiva unidimensional y bidimensional, el análisis de regresión y correlación, los números índices y las series temporales. Al diseñar el Grado en Turismo, se ha considerado la Estadística como una materia básica, denominada “Estadística Aplicada al Turismo”, con un total de 6 ECTS, y que comprende 3 horas semanales de clase presencial, con los mismos contenidos que tenía la asignatura equivalente en los estudios de Diplomatura. En la Figura 3 se indica cómo se ha transformado la Estadística en los estudios de turismo.

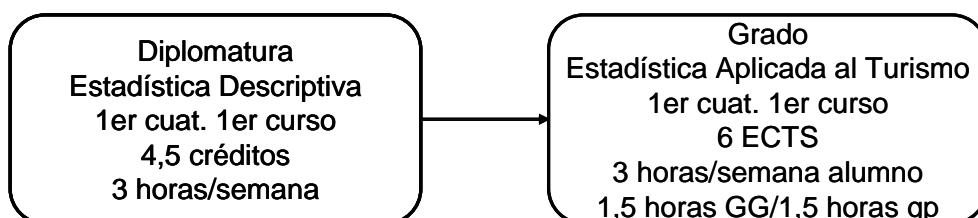


Figura 3: Transformación de la Estadística en los estudios de Turismo.

Esta asignatura ha sido la primera que hemos afrontado los profesores del Departamento de Estadística, Estructura Económica y Organización Económica Internacional en los estudios de Grado. Nos ha servido de banco de pruebas, y hemos podido ensayar los procedimientos de evaluación continua que hemos empleado. En este caso, se planteó un sistema de evaluación continua con el siguiente esquema de valoración:

- 1 punto, por la asistencia del alumno a las clases presenciales. Para evaluar esta presencia, nos hemos visto obligados a pasar lista cada sesión. Para conseguir este punto, se exigía la presencia del estudiante al menos en el 80% de las sesiones.
- 2 puntos, por los ejercicios de clase, que los estudiantes debían entregar a lo largo del curso.
- 7 puntos, por las pruebas de evaluación escritas que se planteaban a lo largo del cuatrimestre (concretamente tres, una correspondiente a la estadística unidi-

mensional, otra a la estadística bidimensional, y la tercera al estudio de la evolución temporal de los fenómenos: series temporales y números índice).

Los estudiantes que no superaban este modelo de evaluación continua tenían la oportunidad de presentarse al examen final, según la normativa de la UAH (según el artículo 144 de los Estatutos de la Universidad, como se ha comentado en el apartado 2). En caso de que el alumno tuviera alguno de los tres parciales suspendidos, le dábamos la posibilidad de que se presentara sólo a esa parte de la materia en el examen final. Este hecho nos creó bastantes dificultades, ya que hubo que preparar cuatro exámenes para el examen final: uno de toda la materia, y tres más que correspondían a cada parte de la asignatura.

3.2. La Estadística en los estudios de Administración y Dirección de Empresas.

La Licenciatura en Administración y Dirección de Empresas (LADE) constaba de 5 cursos. La Estadística se explicaba en una asignatura anual durante el primer curso, denominada “Estadística Empresarial”, con un total de 12 créditos (120 horas de carga docente) que, al ir distribuidos en los dos cuatrimestres, se concretaban en una dedicación de 4 horas de clase semanales. Los contenidos que se desarrollaban se distribuían de la siguiente forma: en el primer cuatrimestre se impartían la estadística descriptiva unidimensional y bidimensional, el análisis de regresión y correlación, los números índices y las series temporales, además del cálculo de probabilidades; en el segundo cuatrimestre, se explicaban las variables aleatorias, los modelos de distribuciones, las distribuciones en el muestreo y la estimación puntual, así como la estimación por intervalos de confianza y los contrastes de hipótesis paramétricos. Al estructurar las materias en el Grado en Administración y Dirección de Empresas, se han desarrollado dos asignaturas, “Estadística Empresarial I” y “Estadística Empresarial II”, situadas en el segundo cuatrimestre del primer curso, y en el primer cuatrimestre del segundo curso, respectivamente, tal y como se muestra en la Figura 4:

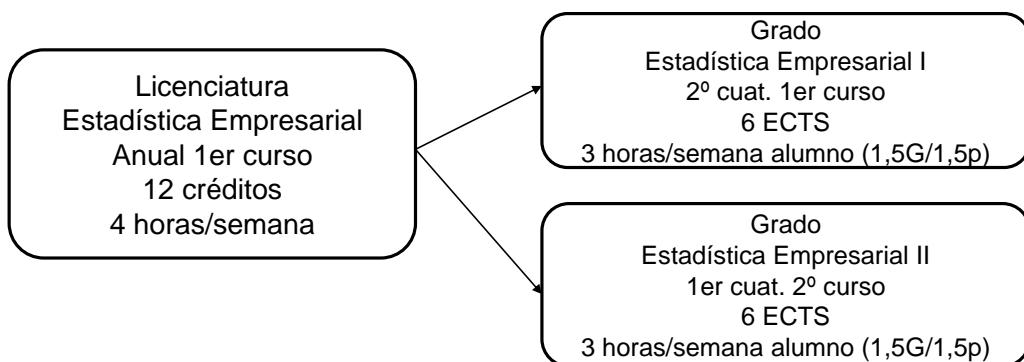


Figura 4: Transformación de la Estadística en los estudios de Administración y Dirección de Empresas.

En la Estadística Empresarial I se imparten los contenidos del primer cuatrimestre de la asignatura Estadística Empresarial de la Licenciatura en ADE, es decir, la estadística unidimensional y bidimensional, el análisis de regresión, los números índice y las series temporales, así como el cálculo de probabilidades. Hemos reservado a la Estadística Empresarial II el estudio de las variables aleatorias, los modelos de distribuciones y toda la parte destinada a inferencia paramétrica (estimación puntual y por intervalos y contraste de hipótesis).

La primera de estas asignaturas la hemos impartido ya a los estudiantes de los primeros cursos del Grado de ADE y del Doble Grado en Derecho y ADE en el segundo cuatrimestre del curso 2009-2010, y hemos comprobado que la reducción de las 4 horas por semana que

teníamos en la Licenciatura a las 3 horas por semana del plan de estudios de Grado nos crea muchas dificultades para conseguir transmitir a los estudiantes todos los contenidos de la materia de manera holgada.

En esta asignatura hemos realizado ya la primera adaptación del modelo de evaluación continua, a partir de la experiencia tenida en el Grado de Turismo. En este caso, se mantuvo el mismo esquema de valoración que en Turismo, pero modificando la forma de superar las pruebas escritas que se hacían a lo largo del cuatrimestre:

- 1 punto, por la asistencia del alumno a las clases presenciales.
- 2 puntos, por los ejercicios de clase, que los estudiantes debían entregar a lo largo del curso.
- 7 puntos, por las pruebas de evaluación escritas que se planteaban a lo largo del cuatrimestre (concretamente tres, una correspondiente a la Estadística Unidimensional, otra a la Estadística Bidimensional, y la tercera, algo más amplia, a las series temporales, los números índice y el cálculo de probabilidades).

Para superar este modelo de evaluación, se exigía que los estudiantes obtuvieran al menos una media de 4 puntos en las pruebas de evaluación escritas, no pudiendo obtener menos de 4 puntos en dos o más de ellas (valoradas sobre 10 puntos). De esta forma, los estudiantes que no superaban este criterio, debían presentarse al examen final de toda la asignatura, sin existir la posibilidad de repetir ningún parcial de los pendientes. De esta forma, se preparó un único examen final para la asignatura, que debían hacer los estudiantes que no hubieran superado la misma siguiendo el modelo de la evaluación continua.

3.3. La Estadística en los estudios de Economía.

La Licenciatura en Economía constaba de 5 cursos. La Estadística se explicaba en dos asignaturas anuales durante los dos primeros cursos, con un total de 22,5 créditos, 10,5 en primer curso y 12 en segundo. Los 10,5 créditos de la asignatura “Estadística Económica I” del primer curso se concretaban en una dedicación de 4 horas de clase semanales durante el primer cuatrimestre y de 3 horas semanales durante el segundo. La asignatura “Estadística Económica II”, que se impartía durante el segundo curso tenía una dedicación de 4 horas a la semana durante todo el año.

La Estadística Económica I se dedicaba al estudio de la estadística descriptiva (incluyendo el análisis descriptivo de las series temporales y los números índice), así como la introducción al cálculo de probabilidades, las variables aleatorias y los modelos de distribuciones de probabilidad. La Estadística Económica II comprendía toda la inferencia paramétrica (estimación puntual y por intervalos y contraste de hipótesis), así como contrastes no paramétricos, el análisis de la varianza y la introducción a las técnicas del muestreo estadístico.

Al estructurar la materia de Estadística en el Grado Economía, se han desarrollado dos asignaturas, denominadas “Estadística Económica I” y “Estadística Económica II”, situadas en el segundo cuatrimestre del primer curso, y en el primer cuatrimestre del segundo curso, respectivamente, tal y como se muestra en la Figura 5:

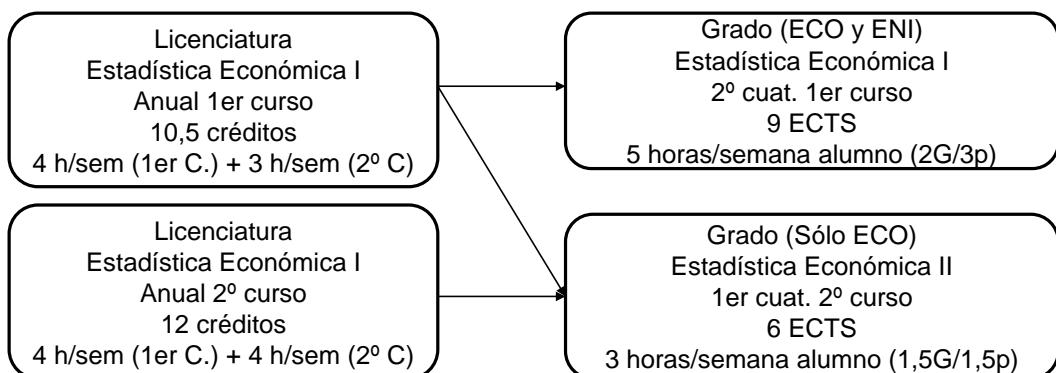


Figura 5: Transformación de la Estadística en los estudios de Economía.

En el caso de los Grados de Economía y de Economía y Negocios Internacionales, se ha añadido un problema en el diseño de estas asignaturas: la Estadística Económica II sólo se imparte para los estudiantes del Grado en Economía, pero la Estadística Económica I se imparte a los estudiantes de ambos grados, con la dificultad añadida de que estos estudiantes reciben su docencia en los mismos grupos, puesto que el primer curso de estos grados es muy similar, con muchas asignaturas en común. Este hecho hace que la Estadística Económica I tenga que ser la misma en ambos Grados, y como en el Grado de ENI no reciben más formación de Estadística, los contenidos de Estadística Económica I deben cubrir hasta el contraste de hipótesis paramétricas. Esto provoca que la amplitud de contenidos haga que no se profundice demasiado en las técnicas estadísticas básicas en este primer curso, algo que luego sí se puede hacer en la asignatura de Estadística Económica II, en la que los estudiantes del Grado en Economía pueden adquirir, además, conocimientos de análisis de la varianza o de las técnicas de muestreo.

Hay que destacar que se han asignado 9 créditos ECTS a la asignatura Estadística Económica I, lo cual ha quedado reflejado en nuestra Facultad en una presencialidad de 5 horas a la semana para cada alumno, una sesión de 2 horas en grupo grande y dos sesiones de hora y media en grupo reducido. En cambio, la Estadística Económica II tiene la misma disposición de las demás asignaturas de 6 ECTS, una hora y media en grupo grande y una hora y media en grupo reducido.

En la asignatura Estadística Económica I del Grado en Economía, impartida ya durante el curso 2009-2010, hemos realizado un proceso diferente de evaluación continua, según el siguiente esquema de valoración:

- 3 puntos, por la asistencia, seguimiento y participación del alumno en las clases presenciales. El seguimiento se ha evaluado en función de los trabajos individuales realizados por los estudiantes.
- 3 puntos, por las pruebas escritas, de carácter teórico, realizadas a lo largo del curso. Concretamente, se han realizado 5 pruebas objetivas (con preguntas tipo test) a lo largo del curso.
- 4 puntos, por la prueba final de evaluación, de tipo práctico, en la que se planteaban a los estudiantes una serie de ejercicios para resolver, bien con calculadora, bien con ordenador.

Para superar este modelo de evaluación, se exigía que los estudiantes obtuvieran al menos una media de 4 puntos sobre 10 en las pruebas de evaluación escritas (de conocimientos teóricos), lo cual hemos comprobado que es muy difícil de alcanzar, debido a que los estudiantes de nuestra Facultad no tienen mucha capacidad para estudiar los contenidos teóricos,

pero sí para su aplicación. Para el curso 2010-2011, se ha cambiado la metodología de la evaluación continua, para incluir en las pruebas escritas de evaluación parcial el desarrollo de ejercicios prácticos, que esperamos que faciliten a nuestros alumnos la superación de la asignatura.

4. Análisis de los resultados académicos.

En este apartado, se realiza el análisis de los resultados académicos de los estudiantes a los que he tenido la oportunidad de impartir clase durante el primer curso de los Grados en Turismo y ADE y en el Doble Grado en Derecho y ADE.

Hay dos hipótesis que se pretenden comprobar en relación a los nuevos métodos de enseñanza y evaluación y los resultados académicos que obtienen los alumnos. Por una parte, parece lógico pensar que el método de evaluación continua facilita a los estudiantes la superación de las materias cursadas. Por otro lado, cabe esperar que los estudiantes se comporten de manera diferente según la titulación elegida, por su distinta procedencia, y por su rendimiento previo, medido en términos de su nota de corte al acceder a la Universidad.

En primer lugar, se ha comprobado cómo el método de evaluación continua ha mejorado sensiblemente los resultados de los estudiantes en relación con la superación de la materia. Por citar algún ejemplo concreto, y tomando como referencia los resultados de los estudiantes del Grado en Turismo, se muestran en la Figura 6 sendos diagramas de caja para explicar la calificación que han obtenido los estudiantes según hayan seguido la evaluación continua o no. Se aprecia que las calificaciones de los estudiantes de la evaluación continua son superiores a los de la evaluación final. De hecho, todos los estudiantes de la evaluación continua han aprobado la asignatura (están por encima de 5 puntos). Hay que recordar, no obstante, que en esta asignatura se permitía a los alumnos de evaluación continua que no superaran este método de evaluación que podrían recurrir a la evaluación final, razón por la cual no suspendió ningún estudiante de evaluación continua.

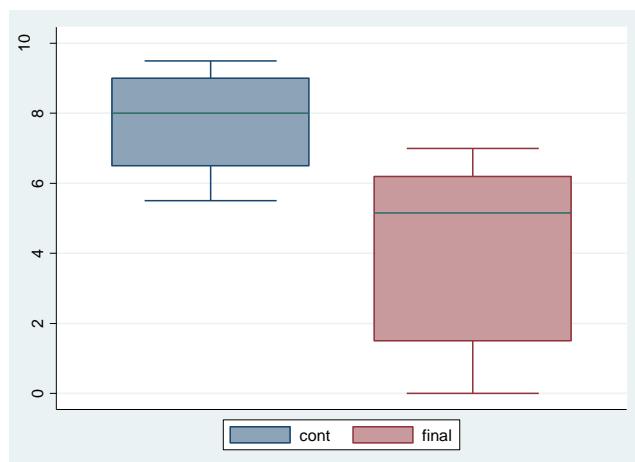


Figura 6: Diferencias entre las calificaciones de los alumnos de evaluación continua y evaluación final en el Grado de Turismo.

En cuanto a la segunda hipótesis planteada, en el Cuadro 1 se muestran las notas de corte obtenidas en las Pruebas de Acceso a la Universidad por los alumnos que fueron admitidos en los Grados en Turismo, en Administración y Dirección de Empresas y en el Doble Grado en

Derecho y ADE. Se puede apreciar que la tipología de estos estudiantes es bastante diferente en cuanto a su perfil de ingreso, ya que la nota de corte que se ha fijado es muy desigual. Mientras que los estudiantes han podido acceder a los estudios del Grado en Turismo incluso en septiembre (aunque no todos los que lo solicitaban, ya que la nota de corte en septiembre es de 5.41), los cupos de admisión se completaron en junio para el Grado en ADE y el Doble Grado en DADE, si bien con casi un punto de diferencia en la nota (5.23 para los estudiantes de ADE y 6.21 para los de DADE). También se muestran los datos del rendimiento académico obtenido en cada titulación, en los que se observa, por un lado, que el porcentaje de alumnos que se presentan es mucho más elevado que en cursos anteriores, en los que el promedio de alumnos presentados rondaba el 50% y, por otro lado, que la proporción de estudiantes que superan el modelo de evaluación elegido es mucho mayor, siendo del 70,3% en Turismo, 74,2% en el Grado de ADE, y del 98,1% en el Doble Grado de DADE. Se aprecia, además, que la tasa de rendimiento, en términos de porcentaje de aprobados, es mayor en el grado en el que los estudiantes entran con una mayor nota de corte, lo que da idea de la mejor selección del estudiantado en estos estudios.

Cuadro 1: Notas de Corte PAU y resultados académicos. Curso 2009-2010.
Grados de Turismo y ADE, y Doble Grado en Derecho y ADE.

| | Turismo | ADE | DADE |
|---------------------|---------------|----------------|---------------|
| Nota de Corte (PAU) | 5.41 (S) | 5.23 (J) | 6.21 (J) |
| Matriculados | 58 | 205 | 56 |
| Presentados | 37 (63.8%) | 178 (86.8%) | 53 (94.6%) |
| Aprobados | 26 (70.3%) | 132 (74.2%) | 52 (98.1%) |
| Suspensos | 11 (29.7%) | 46 (25.8%) | 1 (1.9%) |

Para analizar mejor la diferencia que existe entre los grupos en cuanto a las calificaciones obtenidas por los alumnos en estas materias, se muestran en la Figura 7 los diagramas de caja que representan las puntuaciones obtenidas por los alumnos en los cuatro grupos considerados (grupo de ADE por la mañana, grupo de ADE por la tarde, grupo único de DADE y grupo único de Turismo):

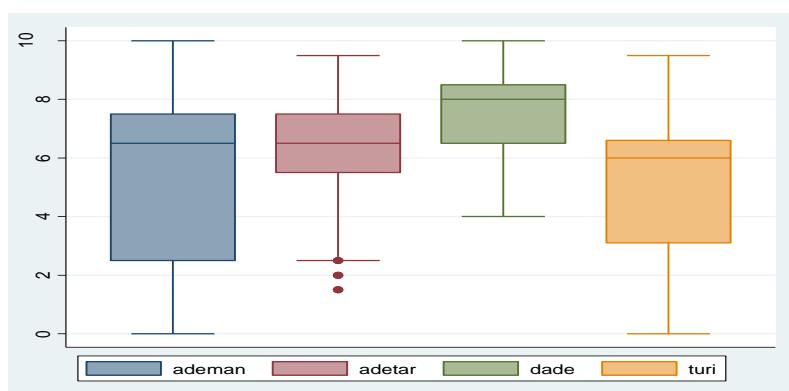


Figura 7: Resultados académicos por titulaciones y grupos.

En la Figura 7 se aprecia, como cabía esperar, que los estudiantes con mejores resultados son los estudiantes del Doble Grado en DADE, seguidos por los de ADE y luego por los de Turismo.

Otra característica importante que se ha detectado en las calificaciones obtenidas por los estudiantes de los Grados de Turismo y ADE es que existe cierta bimodalidad en sus puntuaciones. Si se observa la Figura 8, se aprecian, en las densidades estimadas de las calificaciones de estos grupos, dos modas, una de ellas en torno al 7, que corresponde a los alumnos que han superado holgadamente la asignatura, y otra, en torno al 2 o al 3, que es la calificación que obtienen los estudiantes que no superan la asignatura, que en su mayor parte han elegido el modelo de evaluación basado en una única prueba final.

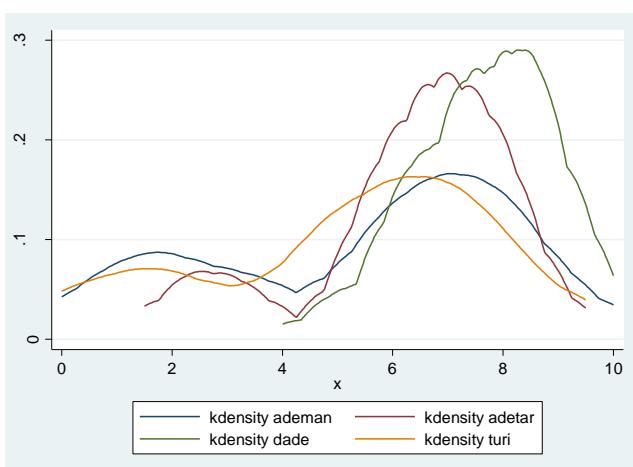


Figura 8: Estimación kernel de la densidad de las calificaciones por titulaciones y grupos.

5. Conclusiones.

El Espacio Europeo de Educación Superior ha provocado una revolución en la Universidad española, en el sentido de que nos ha hecho pensar sobre los métodos de enseñanza y los procedimientos que empleamos en la transmisión del conocimiento. En estos momentos, el estudiante pasa a ser el centro de su propio aprendizaje, y debemos forzar a que tome conciencia de su importancia en el proceso de enseñanza-aprendizaje. Lejos de algunas percepciones negativas, debemos entender el nuevo sistema educativo como una oportunidad para mejorar los mecanismos que empleamos en la Universidad para formar a nuestros estudiantes.

La adaptación de las asignaturas de Estadística en los estudios de Grado en la rama de Ciencias Sociales y Jurídicas en la UAH se ha hecho teniendo en cuenta la configuración de la Estadística en los anteriores estudios de Licenciatura o Diplomatura. La temporalización de los mismos ha cambiado debido a la nueva forma de medir la carga de trabajo. Ahora el ECTS mide la dedicación de los estudiantes, y no la docencia del profesor. En todos los estudios se ha perdido capacidad docente, porque se ha reducido el número de horas presenciales que los estudiantes comparten con el profesor en el aula. Este hecho dificulta la impartición de los contenidos, ya que hay que adecuar lo que se explicaba antes al escaso tiempo del que disponemos ahora.

Si nos fijamos en el protagonismo que deben ejercer ahora los estudiantes, observamos que nuestros alumnos no tienen aún la capacidad de asumir su papel central en el proceso de

enseñanza-aprendizaje, hecho que provoca dificultades para hacerles partícipes del nuevo sistema. Es previsible que esta situación cambie con el tiempo, a medida que la costumbre incite a los estudiantes a asumir su papel principal en la adquisición de conocimientos.

En cuanto a las metodologías de evaluación, se han ensayado varias alternativas dentro del esquema general de la evaluación continua. Se ha comprobado que la valoración de la asistencia y la participación ha sido muy bien acogida por los estudiantes, ya que les permite obtener puntos de una forma sencilla. Desde el punto de vista del profesorado, este aspecto no debería ser tan importante, puesto que, en algunos casos, el alumno acude al aula por el hecho de que se le va a tener en cuenta, sin atender y, en ocasiones, se permite molestar al resto, lo cual perjudica el rendimiento global en el aula. Por otra parte, como se han valorado los ejercicios y trabajos que han hecho los estudiantes a lo largo del curso, se ha comprobado que en muchas ocasiones lo que los alumnos entregaban eran copias de los trabajos de otros compañeros, lo cual hace que sea muy fácil conseguir una valoración positiva sin que las competencias sean convenientemente adquiridas. Todas estas consideraciones nos han llevado, durante el curso 2010-2011 a disminuir la valoración de la asistencia y participación a únicamente 2 puntos.

Con relación a las pruebas de evaluación, también hemos ensayado distintas posibilidades. En primer lugar, se ha desterrado la opción de repetir las pruebas parciales para repescar a los alumnos, y se ha comprobado que es preferible que el alumno realice una única prueba final si no es capaz de superar las pruebas parciales. En cuanto al contenido de estas pruebas parciales, parece recomendable que contengan aspectos tanto teóricos como prácticos, puesto que el perfil de nuestros estudiantes es eminentemente aplicado. Se ha podido comprobar que la evaluación continua ha mejorado enormemente el rendimiento académico de nuestros estudiantes, en términos de alumnos que se presentan a la evaluación y de los que finalmente superan la misma.

6. Bibliografía.

- [1] ANECA (2004). *Libro Blanco del Título de Grado en Turismo*. En http://www.aneca.es/media/359791/libroblanco_turismo_03.pdf. Consultado durante el curso académico 2009-2010.
- [2] ANECA (2005). *Libro Blanco del Título de Grado en Economía y Empresa*. En http://www.aneca.es/media/150292/libroblanco_economia_def.pdf. Consultado durante el curso académico 2009-2010.
- [3] Goñi Zabala, J.M. (2005). *El espacio Europeo de educación superior, un reto para la Universidad: competencias, tareas y evaluación, los ejes del currículum universitario*. Octaedro, Barcelona.
- [4] Zabalza, M.A. (2007). *La enseñanza universitaria. El escenario y sus protagonistas*. Narcea, Madrid.

Acerca del autor



Luis F. Rivera Galicia es Profesor Titular de Escuela Universitaria del área de Métodos Cuantitativos para la Economía y la Empresa. Desarrolla su actividad docente e investigadora en el Departamento de Estadística, Estructura Económica y Organización Económica Internacional. En la actualidad es Vicedecano de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Alcalá.

Licenciado en Ciencias Matemáticas (especialidad Ciencias de la Computación) por la Universidad Complutense y Doctor en Economía por la Universidad de Alcalá, sus principales líneas de investigación son los métodos no supervisados de clasificación estadística y la enseñanza y el aprendizaje de la Estadística en los estudios de la rama de Ciencias Sociales y Jurídicas. Ha publicado varios artículos en revistas especializadas y ha participado en diversos congresos de ámbito nacional e internacional. Ha formado parte del equipo investigador en varios Proyectos de Investigación en convocatorias públicas, y ha dirigido hasta la actualidad cuatro Proyectos de Innovación Docente.

Autores

Daza Árboli, Miguel Ángel. “Nuevas plataformas virtuales para la docencia” pg. 94.

Gámez Mellado, Antonio. “R: Introducción y primeros pasos” ; “R: Estadística Descriptiva Unidimensional y Bidimensional”; “R: Distribuciones e Inferencia” y “Herramientas de autor para la creación de contenidos para la Docencia de la Estadística”. pg. 11, 36, 53 y 77.

Marín Trechera, Luis M. “R: Introducción y primeros pasos” ; “R: Estadística Descriptiva Unidimensional y Bidimensional”; “R: Distribuciones e Inferencia” y “Herramientas de autor para la creación de contenidos para la Docencia de la Estadística”. pg. 11, 36, 53 y 77.

Rivera Galicia, Luis F. “La Enseñanza de la Estadística en las nuevas titulaciones adaptadas al Espacio Europeo de Educación Superior”. pg. 108.

Participantes

Aguilera del Pino, Ana

aaguiler@ugr.es

Universidad de Granada

Artiles Romero, Juan

jartiles@dma.ulpgc.es

Universidad de Las Palmas de Gran Canaria

Daza Árboli, Miguel Ángel

mdazarb@et.mde.es

MINISDEF

Fuentes Jiménez, Miguel

fuentesj@ugr.es

Universidad de Granada

Gámez Mellado, Antonio

antonio.gamez@uca.es

Universidad de Cádiz

García-Ligero Ramírez, María Jesús

mjgarcia@ugr.es

Universidad de Granada

Lara Porras, Ana María

alara@ugr.es

Universidad de Granada

Marmolejo Martín, Juan Antonio

jamarmo@ugr.es

Universidad de Granada

Marín Trechera. Luis Miguel

luis.marin@uca.es

Universidad de Cádiz

Montero Alonso, Miguel Angel

mmontero@ugr.es

Universidad de Granada

Ortiz de Haro, Juan Jesús

jortiz@ugr.es

Universidad de Granada

Quesada Vázquez, Isabel

iquesada@ugr.es

Universidad de Granada

Rivera Galicia, Luis Felipe

luisf.rivera@uah.es

Universidad de Alcalá

Román Román, Patricia

proman@ugr.es

Universidad de Granada

San Martín Pérez, Montserrat

montse.sanmartin@unirioja.es

Universidad de La Rioja

Serrano Romero, Luis

lserrano@ugr.es

Universidad de Granada.

Entidades organizadoras

GENAEIO

**Grupo de Enseñanza y Aprendizaje
de la Estadística e Investigación
Operativa**



**Escuela Universitaria de Ciencias Sociales.
Campus de Melilla.
Universidad De Granada**

Entidades colaboradoras



**Sociedad de
Estadística e Investigación
Operativa**



**Consejería de
Educación de la Ciudad
Autónoma de Melilla**



**Vicerrectorado para
la Garantía de la Calidad de
la Universidad de Granada**