

(Les rogamos mantengan las indicaciones de esta plantilla)

Creación de material docente dinámico con LaTeX

Autoría: Virgilio Gómez Rubio

Temática: Estadística, LaTeX

Palabras clave: Estadística, LaTeX, R, C
Resumen

LaTeX es problemente el procesador de texto más utilizado en Matemáticas para la elaboración de todo tipo de documentos debido a la facilidad para incluir ecuaciones y fórmulas complejas. Aunque menos conocido que LaTeX, Donald Kunth desarrollo un sistema que permitía utilizar LaTeX para producir documentación de alto nivel para programas escritos en C. De esta manera, la documentación se incluía directamente en el código fuente y podía ser extraída utilizando distintos programas.

Los desarrolladores del programa R, muy utilizado en estadística, han creado un sistema similar que permite integrar R y LaTeX. De esta manera, es posible crear documentos en LaTeX que incorporan trozos de código R. Este documento intermedio puede ser procesado de manera que el código en R se ejecuta y es sustituido por su correspondiente salida, ya sea texto, gráfico e incluso tablas de resultados.

En mi charla presentaré este sistema y cómo puede ser utilizado para el desarrollo de material docente. También haré una demostración de cómo funciona este sistema con ejemplos sencillos.

1. INTRODUCCIÓN

LaTeX (Lanport, 1994) es probablemente el editor de texto preferido por los matemáticos debido a la facilidad con la que se pueden incluir ecuaciones y gráficas de alta calidad en el texto. Además, LaTeX permite organizar automáticamente el diseño del texto, e incluye numerosos paquetes que permiten, entre otras cosas, determinar cómo las palabras se deben separar al final de la línea en el texto, incluir bibliografía y citas de manera automática en el texto, etc.

A partir de LaTeX, Donald Kunth también ideó un sistema llamado Cweb (Knuth y Levy, 1993) para poder documentar programas en C utilizando LaTeX. De esta manera, en el mismo fichero de código fuente se incluye el código del programa en C más el texto explicativo en LaTeX. De esta manera, además del código de la función en C, también encontramos una descripción de la misma, de qué hace y qué argumentos necesita.

Esto tiene la ventaja de que tenemos código y documentación en un mismo fichero, de manera que si necesitamos cambiar el código también podemos actualizar la documentación de manera sencilla. La documentación se incluye en el programa como si fueran comentarios dentro del código. Así, podemos compilar el código C sin problemas. Para generar la documentación del programa necesitamos extraerlo del código C y generar ficheros LaTeX.

Aunque este sistema fue desarrollado para C y LaTeX, otros sistemas parecidos han aparecido posteriormente. Uno de ellos se ha desarrollado para el programa R, y es en el que nos vamos a centrar.

2. EL LENGUAJE R

El programa R (R Core Team, 2010) es una implementación libre y gratuita del lenguaje S, que fue desarrollado por John Chambers y sus colegas en los Laboratorios Bell para proporcionar un lenguaje que permitiera realizar análisis estadístico de datos de una manera más flexible. Según Chambers, el objetivo de este lenguaje es “convertir ideas en software de una manera rápida y fiable”. En la actualidad, R es uno de los programas estadísticos más utilizados en todo el mundo, tanto en universidades como compañías privadas (los usan desde Google hasta grandes compañías farmacéuticas).

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

Este documento puede ser procesado de dos formas distintas:

- Ejecutar el código R y sustituirlo por su salida para crear un documento LaTeX, que ya podemos compilar para generar un documento final.
- Extraer el código R del documento Rnw. Esto permite ejecutar todo el código de una para obtener los resultados. Por tanto, todo lo que no sea código R desaparece.

Estas dos acciones se realizan con las funciones Sweave() y Stangle() de R, respectivamente, que toman como argumento el nombre del fichero .Rnw. La Figura 1 muestra un diagrama en el que se muestra el proceso para pasar del documento Rnw (que incluye LaTeX y R) hasta el documento final en PDF.



Figura 1: Cómo pasar de un fichero Rnw al documento final en PDF.

3. EJEMPLO: ANÁLISIS DE DATOS CON R Y LATEX

A modo de ejemplo vamos a ver cómo se realizaría un análisis de datos sencillo de dos variables. Como ya hemos dicho, la estructura del documento es la de un documento LaTeX en la que se incluye código en R para ser ejecutado y reemplazado por su resultado. En el Apéndice 1 aparece el fichero Rnw completo que utilizaremos en este ejemplo. En el texto iremos reproduciendo pequeñas porciones del mismo y las iremos comentando.

En primer lugar, puede verse como todos los trozos de código R introducidos en el texto se encuentran insertados dentro de un entorno definido así:

```
<<>=
```

```
...
```

```
@
```

Ésto nos permite definir claramente dónde incluir el código en R. Además, es posible modificar el comportamiento del código mediante diversas opciones (que se incluyen en el encabezado) para, por ejemplo, ocultar el código, mostrar gráficas y tablas de datos, o simplemente mostrar el código pero no ejecutarlo.

En este ejemplo vamos a analizar dos variables, que se corresponden con datos de calor producido por 13 tipos de cemento dependiendo de la concertación de componentes activos en cada uno de ellos. Estos datos forman parte de un banco de datos más grande que se encuentra disponible en el paquete MASS de R bajo el nombre 'cement'.

Así, para definir las dos variables que vamos a analizar en nuestro ejemplo escribiríamos el siguiente código en R:

```
<<>=
```

```
#Calor
```

```
y<-c(78.5, 74.3, 104.3, 87.6, 95.9, 109.2, 102.7, 72.5, 93.1, 115.9,  
83.8, 113.3, 109.4)
```

```
##% de componentes activos
```

```
x<-c(60, 52, 20, 47, 33, 22, 6, 44, 22, 26, 34, 12,  
12)
```

```
@
```

A continuación nos interesa realizar una regresión lineal y mostrar los resultados, para lo cual escribiríamos el siguiente texto:

```
<<>=  
rl<-lm(y~x)  
summary(rl)  
  
@
```

Como comentábamos antes, cuando compilemos el fichero Rnw estas líneas de códigos serán reemplazadas por su correspondiente salida (ver apéndice 2).

Finalmente, también podemos insertar gráficos de manera automática en el texto. Esto se hace insertando la palabra clave 'fig=TRUE' en la cabecera del código en R:

```
<<fig=TRUE>>=  
  
plot(x,y, main="Calor producido por distintos tipos de cemento",  
      xlab="% componentes activos",  
      ylab="Calor (cal/g)")  
  
abline(rl)#Recta de regresión ajustada  
  
@
```

Cuando compilemos en documento, R creará una figura a partir de esos comandos y sustituirá el código por un comando para incluir la gráfica en el documento. Por defecto, todas las gráficas se crean en formato EPS y PDF. La gráfica producida por el código anterior puede verse en la Figura 2.

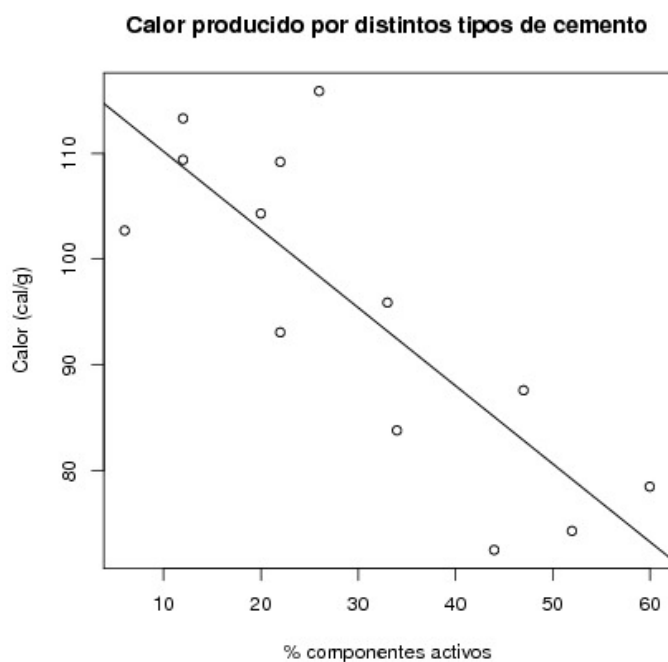


Figura 2: Recta de regresión ajustada a los datos.

Es posible incluir otras opciones para controlar el comportamiento de los trozos de código R que hay en el texto. Así, es posible hacer que el código R no aparezca en el texto final o que se muestre pero no se ejecute, entre otras opciones.

4. POSIBILIDADES PARA LA DOCENCIA

Desde el punto de vista docente, combinar R y LaTeX puede ser muy útil. Algunas posibilidades que encontramos interesantes son:

- Desarrollo de apuntes, en los que se incluyen numerosos ejemplos generados automáticamente. Así, cada vez que cambiamos datos o figuras éstas se pueden actualizar automáticamente simplemente recompilando los ficheros Rnw y LaTeX.
- Desarrollo de listas de ejercicios personalizadas para los alumnos. A partir de un código (el DNI, por ejemplo) es posible generar una lista de valores al azar para cada

alumno que permiten asignarles tareas completamente al azar. Luego estas tareas se pueden incluir en un fichero LaTeX de manera que se puede personalizar la hoja de ejercicios para cada alumno.

- Desarrollo de prácticas de laboratorio con R. De esta manera los alumnos pueden tener, por un lado, la descripción de los métodos estadísticos que van a utilizar y, por otro, el código en R necesario para su ejecución.
- Desarrollo de informes de análisis de datos. De esta manera, se pueden ir analizando los datos de manera interactiva de manera que los resultados se actualizan automáticamente. Si con posterioridad se detectase un error en los datos, bastaría con corregirlo y recompilar el documento para obtener resultados actualizados.

REFERENCIAS

Knuth, D. y Levy, S. (1993). The CWEB System of Structured Documentation. Reading, Massachusetts: Addison-Wesley.

Lamport, L. (1994). LaTeX: A Document Preparation System. Reading, Massachusetts: Addison-Wesley, 2nd. ed.

Leisch, F. (2002) Sweave, part I: Mixing R and Latex. R News, 2(3):28-31.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Apéndice 1. Fichero ejemplo.Rnw

```
\documentclass[a4paper,12pt]{article}
```

```
\begin{document}
```

```
\section{Introducción}
```

Éste es un fichero de ejemplo para ver cómo funcionan LaTeX y R juntos para producir documentos dinámicos. Vamos a realizar un análisis de datos muy sencillo.

```
\section{Datos}
```

Los datos que tenemos representan el calor producido (en cal/g) por trece tipos de cemento y de su concentración de componentes activos.

Los datos los podemos cargar así:

```
<<>>=
```

```
#Calor
```

```
y<-c(78.5, 74.3, 104.3, 87.6, 95.9, 109.2, 102.7, 72.5, 93.1, 115.9,
```

```
83.8, 113.3, 109.4)
```

```
##% de componentes activos
```

```
x<-c(60, 52, 20, 47, 33, 22, 6, 44, 22, 26, 34, 12,12)
```

```
@
```

```
\section{Resultados}
```

Si hacemos una regresión lineal con estos datos, tenemos:

```
<<>>=
```

```
rl<-lm(y~x)
```

```
summary(rl)
```

```
@
```

Además, podemos hacer una gráfica con los datos y la recta ajustada:

```
<<fig=TRUE>>=
```

```
plot(x,y, main="Calor producido por distintos tipos de cemento",
```

```
      xlab="% componentes activos", ylab="Calor (cal/g)")
```

```
abline(r1)
```

```
@
```

```
\end{document}
```

Apéndice 2. Fichero ejemplo.tex

```
\documentclass[a4paper,12pt]{article}
```

```
\usepackage{Sweave}
```

```
\begin{document}
```

```
\section{Introducción}
```

Éste es un fichero de ejemplo para ver cómo funcionan LaTeX y R juntos para producir documentos dinámicos. Vamos a realizar un análisis de datos muy sencillo.


```
\section{Datos}
```

Los datos que tenemos representan el calor producido (en cal/g) por trece tipos de cemento y de su concentración de componentes activos.

Los datos los podemos cargar así:

```
\begin{Schunk}
```

```
\begin{Sinput}
```

```
> y <- c(78.5, 74.3, 104.3, 87.6, 95.9, 109.2, 102.7, 72.5, 93.1,
```

```
+ 115.9, 83.8, 113.3, 109.4)
```

```
> x <- c(60, 52, 20, 47, 33, 22, 6, 44, 22, 26, 34, 12, 12)
```

```
\end{Sinput}
```

```
\end{Schunk}
```

```
\section{Resultados}
```

Si hacemos una regresión lineal con estos datos, tenemos:

```
\begin{Schunk}
```

```
\begin{Sinput}
```

```
> rl <- lm(y ~ x)
```

```
> summary(rl)
```

```
\end{Sinput}
```

```
\begin{Soutput}
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-12.589	-8.228	1.495	4.726	17.524

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	117.5679	5.2622	22.342	1.62e-10 ***
x	-0.7382	0.1546	-4.775	0.000576 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.964 on 11 degrees of freedom

Multiple R-squared: 0.6745, Adjusted R-squared: 0.645

F-statistic: 22.8 on 1 and 11 DF, p-value: 0.0005762

```
\end{Soutput}
```

```
\end{Schunk}
```

Además, podemos hacer una gráfica con los datos y la recta ajustada:

```
\begin{Schunk}
```

```
\begin{Sinput}
```

```
> plot(x, y, main = "Calor producido por distintos tipos de cemento",
```

```
+   xlab = "% componentes activos", ylab = "Calor (cal/g)")
```

```
> abline(rl)
```

```
\end{Sinput}
```

```
\end{Schunk}
```

```
\includegraphics{ejemplo-003}
```

```
\end{document}
```

- **Autoría**

Nombre completo (y función que desempeña en el centro)

Datos de contacto: Virgilio Gómez Rubio, Profesor Contratado Doctor

CENTRO: Escuela de Ingenieros Industriales - Albacete, Universidad de Castilla-La Mancha

TFNO: 967 59 92 00 ext. 2443 / 647 75 08 59

CORREO: Virgilio.Gomez@uclm.es

PÁGINA WEB: <http://www.uclm.es/profesorado/vgomez>

copyright© La revista digital "Programas" se concibe como un servicio al profesorado para facilitar la difusión de sus experiencias y reflexiones. Los textos publicados aquí pueden ser copiados y distribuidos públicamente, siempre que cite la autoría y "Programas". No los utilice para fines comerciales y no haga con ellos obra derivada.