

# Generación automática de reportes con R y L<sup>A</sup>T<sub>E</sub>X

Automatic report generation with R and L<sup>A</sup>T<sub>E</sub>X

MARIO ALFONSO MORALES RIVERA \*

---

## Resumen

R es un lenguaje y entorno para calculo estadístico y gráficos. Es un proyecto GNU similar al lenguaje y entorno S que fue desarrollado en los laboratorios Bell por John Chambers y colegas. R puede considerarse como otra implementación del lenguaje S.

R proporciona una amplia variedad de técnicas estadísticas y gráficas y es altamente extensible. A menudo el lenguaje S es el vinculo escogido por investigadores en metodología estadística, y R proporciona una ruta de código abierto para la participación en esa actividad (R Development Core Team 2006).

L<sup>A</sup>T<sub>E</sub>X es un sistema de preparación de documentos para composición de texto de alta calidad. Este es usado a menudo para escribir medianos a grandes documentos científicos y técnicos, pero puede usarse para casi cualquier tipo de publicación.

L<sup>A</sup>T<sub>E</sub>X fue desarrollado en 1985 por Leslie Lamport, y ahora lo mantiene y desarrolla el proyecto L<sup>A</sup>T<sub>E</sub>X3. Al igual que R, L<sup>A</sup>T<sub>E</sub>X es software libre y en este momento es el formato de documento más empleado entre matemáticos y estadísticos (The LaTeX3 project 2006).

La función Sweave() de R proporciona un marco flexible para mezclar texto y código R para la generación automática de documentos. Un archivo fuente simple contiene el texto de documentación y el código R, los cuales son *entrelazados* dentro de un documento final que contiene el texto de documentación junto con el código R y/o la salida del código (texto, gráficos). El código R del análisis completo es incrustado dentro de un documento L<sup>A</sup>T<sub>E</sub>X usando la sintaxis noweb (Ramsey, 1998). Por tanto, todo el poder de L<sup>A</sup>T<sub>E</sub>X (para composición de texto de alta calidad) y de R (para análisis de datos) puede usarse simultáneamente (Leisch 2005).

Este tutorial pretende dar ejemplos concretos acerca de la forma como trabaja Sweave, incluyendo la configuración del sistema L<sup>A</sup>T<sub>E</sub>X.

**Palabras Claves:** R, L<sup>A</sup>T<sub>E</sub>X, Sweave, noweb

## Abstract

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories by John Chambers and colleagues. R can be considered as a different implementation of S.

R provides a wide variety of statistical and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity (R Development Core Team 2006).

L<sup>A</sup>T<sub>E</sub>X is a document preparation system for high-quality typesetting. It is most often used for medium-to-large technical or scientific documents, but it can be used for almost any form of publishing.

It was first developed in 1985 by Leslie Lamport, and is now being maintained and developed by the LaTeX3 project. Like R, L<sup>A</sup>T<sub>E</sub>X is free software and it is the most commonly used document format among mathematicians and statisticians at the present time (The LaTeX3 project 2006).

The function Sweave() from R provides a flexible framework for mixing text and R code for automatic document generation. A single source file contains both documentation text and R code, which are then woven into a final document containing, the documentation text together with the R code and/or the output of the code (text, graphs). The S code of the complete analysis is embedded

---

\*Profesor asistente, Departamento de Matemáticas y Estadística, Universidad de Córdoba  
mmorales@sinu.unicordoba.edu.co

into a  $\LaTeX$  document using the noweb syntax (Ramsey, 1998). Hence, the full power of  $\LaTeX$  (for high-quality typesetting) and  $\mathbf{R}$  (for data analysis) can be used simultaneously (Leisch 2005).

This tutorial wants to give concrete examples about of how **Sweave** work, including the  $\LaTeX$  system configuration.

**Keywords:**  $\mathbf{R}$ ,  $\LaTeX$ , Sweave, noweb.

## 1. Introducción

El análisis estadístico tiene dos etapas bien diferenciadas. Por una parte está el análisis de los datos, que involucra la utilización de un paquete (software) estadístico para producir números, tablas y gráficos y por otra está la escritura del informe final (reporte) que requiere de la utilización de un procesador de texto.

La mayoría de los paquetes estadísticos tienen la opción de guardar sus salidas (resultados) en archivos con formato `.doc`, `.rtf` o `.txt`, los cuales permiten usar un procesador de texto como Microsoft Word u Open Office para abrirlo, así que *copiando y pegando*, se construye el reporte final con los comentarios del análisis.

Si deseamos usar  $\LaTeX$ <sup>1</sup> para la composición del texto del reporte o artículo, encontramos que la *metodología* de copiar y pegar resulta poco eficiente, por varias razones:

- $\LaTeX$  tiene formato muy particular para la inclusión de tablas.
- $\LaTeX$  tiene un comando especial para la inclusión de gráficos, con opciones que varían según el formato del archivo a importar.
- “ $\LaTeX$  ignora –en el archivo fuente– cualquier espacio que siga a otro espacio; es decir un espacio es lo mismo que 10 espacios o 100 espacios” (Rodrigo De Castro Korgi 2003)
- Las salidas de los paquetes usan caracteres como %, # y & entre otros, que son reservados de  $\LaTeX$  y si no se emplean adecuadamente generan errores al momento de compilar

Si insistimos en usar  $\LaTeX$  y el *método copiar–pegar* para la edición del reporte, debemos estar preparados para perder mucho tiempo dándole formato a las tablas y salidas del paquete estadístico para obtener un documento de alta calidad en cuanto a su presentación y estética. La filosofía de  $\LaTeX$  es: *usted se dedica a la producción intelectual, y no se preocupa por la presentación del documento, de eso se encarga  $\LaTeX$* . La pérdida de tiempo dándole formato al texto y tablas de un análisis estadístico va totalmente en contra de este concepto.

$\mathbf{R}$  y **S-PLUS** son dos herramientas<sup>2</sup> ampliamente usadas por la comunidad estadística para el análisis de datos y producción de gráficos de alta calidad. Una característica importante de estos dos entornos es su alta compatibilidad con  $\LaTeX$ . Los creadores del lenguaje **S**, el equipo de desarrollo de  $\mathbf{R}$  y los usuarios que escriben extensiones (librerías), se preocupan por facilitar la tarea de inclusión de los resultados del análisis de los datos y los gráficos en documentos  $\LaTeX$ , es así como los usuarios de  $\mathbf{R}$  contamos con el entorno gráfico `picTeX` que permite generar directamente, desde  $\mathbf{R}$ , gráficos en el código de  $\LaTeX$ ; la librería `xtable` que permite generar el código  $\LaTeX$  para tablas, marcos de datos y matrices de tal forma que funcione *copiar–pegar*, la función `latex()` de la librería `Hmisc` que convierte objetos de  $\mathbf{R}$  a código  $\LaTeX$  y la función `Sweave()`, objeto de este tutorial, que toma un archivo que contiene el texto con los comentarios del análisis, formulas, el código de  $\mathbf{R}$  usado para realizar los cálculos y generar los gráficos, y entrega un archivo plano con extensión `.tex` en el que deja sin modificar todo lo que tiene que ver con fórmulas y texto y adiciona: las salidas de  $\mathbf{R}$ , el código usado (opcional) y el comando para la inserción de

<sup>1</sup>El uso de  $\LaTeX$  para la composición de texto se hace indispensable (por no decir obligatorio) cuando, por ejemplo, en el ejercicio de la docencia en estadística se desean escribir notas de clases o tutoriales que incluyen texto, fórmulas, salidas de un programa de análisis de datos, como tablas y gráficos, el código usado para obtener dichas salidas y mucha referencia cruzada.

<sup>2</sup> $\mathbf{R}$  se puede descargar gratuitamente de <http://www.r-project.org/>, **S-PLUS** es comercial y su distribuidor autorizado para América Latina es MS MIAMI, <http://www.msiami.com/>

los gráficos. El archivo `.tex` se compila con  $\text{\LaTeX}$  y se obtiene un documento de alta calidad *sin pérdida de tiempo*. **Sweave** es compatible con la filosofía de  $\text{\LaTeX}$  en el sentido que nos permite dedicar nuestro escaso y valioso tiempo a planear el análisis, escribir el código (*script*) para obtener cálculos y gráficos y a comentar los resultados. De la presentación del documento se encarga **Sweave** junto con  $\text{\LaTeX}$ .

Se enfatiza la frase *sin pérdida de tiempo*, porque en realidad si la hay. Los que inician el aprendizaje de  $\text{\LaTeX}$  y de R tienen que invertir una cantidad considerable de tiempo para empezar a tener resultados aceptables, **Sweave** no es la excepción. Con este documento se ofrece un tutorial para que el lector reduzca considerablemente el tiempo de aprendizaje de **Sweave**, de tal forma que el interesado no se frustre como ocurre con muchos que intentan aprender a usar  $\text{\LaTeX}$  y R sin ninguna asistencia adicional.

## 2. Cómo funciona Sweave?

**Sweave** es un paquete de R escrito por Friedrich Leisch<sup>3</sup> que permite al investigador *entrelazar* código R en un documento  $\text{\LaTeX}$  usando las sintaxis *noweb*. Noweb es una herramienta de programación *literal* que permite combinar código fuente de programas y la documentación correspondiente dentro de un archivo simple. Un archivo noweb contiene una secuencia de segmentos de código y documentación, los cuales reciben el nombre de *chunks* (Leisch 2005).

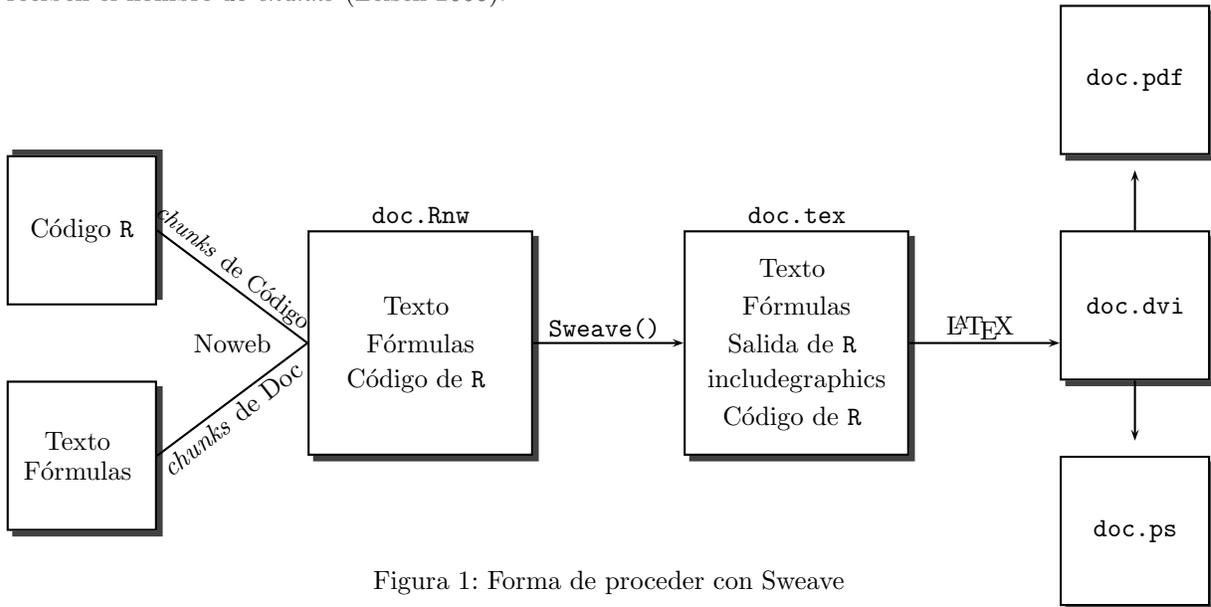


Figura 1: Forma de proceder con Sweave

Para la generación automática de un informe o reporte, se procede como se muestra en la figura 1. El código de R se mezcla con el texto usando *chunks* de código y *chunks* de documentación de acuerdo con la sintaxis noweb, este archivo se guarda como `doc.Rnw`<sup>4</sup>. Desde la línea de comandos de R se ejecuta la orden

```
> Sweave(doc.Rnw)
```

Si no ocurren errores de sintaxis, **Sweave** genera el archivo `doc.tex` que contiene la documentación, la salida del código R que estaba en los *chunks* de código y/o el código mismo, esto lo hace dentro de los entornos  $\text{\LaTeX}$  `Sinput` y `Soutput`, respectivamente. Si el código genera gráficos y se tiene la opción `fig=true` (sección 3.2), **Sweave** coloca en el archivo `doc.tex` el comando `\includegraphics{}` apropiado. El archivo `doc.tex` se compila con  $\text{\LaTeX}$  para obtener, finalmente el archivo `doc.dvi` y opcionalmente, archivos `doc.pdf` y/o `doc.ps`.

<sup>3</sup><http://www.ci.tuwien.ac.at/~leisch/>

<sup>4</sup>Otras extensiones que se pueden usar son: `.nw`, `.rnw`, `.snw` y `.Snw`

Hay que tener en cuenta que **Sweave** deja sin modificar todo lo que esté fuera de los *chunks*, así que el resto del documento `.Rnw` debe respetar todo lo que tiene que ver con la sintaxis de  $\text{\LaTeX}$ .

Esta forma de proceder tiene varias ventajas:

- Si se hace algún cambio en los datos, solo tenemos que modificar el archivo `doc.Rnw` y volver a correr el código, ejecutando nuevamente la orden `Sweave(doc.Rnw)` y el reporte queda actualizado automáticamente.
- **Sweave** puede usarse con otras librerías de R como `xtable` que produce resultados en código  $\text{\LaTeX}$ .
- **Sweave** produce gráficos en formato PostScript y/o PDF y agrega el comando `\includegraphics{}` apropiado ahorrándonos la tarea de hacerlo manualmente.
- Se reduce de la posibilidad de error porque no hay necesidad de *copiar y pagar* manualmente la salida de R al documento `.tex`.

### 3. Ejemplos

En esta sección se discuten las principales opciones disponibles para los *chunks* de código a través de ejemplos. Para los ejemplos se usa el marco de datos `PlantGrowth` que pertenece a los datos de ejemplos de R, y contiene los resultados de un experimento agrícola para comparar crecimientos de plantas (medido por el peso húmedo) obtenidos bajo un control y dos condiciones diferentes de tratamientos.

En los ejemplos no se comenta el código de R ni el de  $\text{\LaTeX}$  porque se asume que el lector está familiarizado con ellos<sup>5</sup>. Solo se comenta lo referente a los *chunks* y sus diferentes opciones.

#### 3.1. Inclusión de salida y código de R

El recuadro de la izquierda, en la figura 2, muestra la sintaxis más simple de un *chunk*. La línea 1 es el inicio del *chunk* de código, la línea 4 indica su final y al mismo tiempo el inicio de un *chunk* de documentación. Cualquier texto que quede después del signo `=` (en la misma línea 1) es ignorado por R al momento de ejecutar el código, es decir, se asume como un comentario.

Cuando se corre `Sweave()` sobre el archivo `.Rnw` que contiene este código, en el archivo `.tex` que se genera, se reemplaza el *chunk* por el texto que se muestra en el centro de la figura 2. Los entornos `Schunk`, `Sinput` y `Soutput`, están definidos en el paquete `Sweave.sty` de  $\text{\LaTeX}$ , cuya instalación se discute en la sección 4. A la derecha, en la figura 2 se muestra lo que se obtiene definitivamente en el documento `.dvi`.

Por defecto **Sweave** coloca el código y la salida, pero eso no siempre es deseable ya que si se está escribiendo un informe, es casi seguro que el lector no estará interesado en conocer el código de programación usado para hacer cálculos, obtener tablas y los gráficos; pero si escribimos unas notas de clase o un tutorial, quizás estemos interesados en mostrar a nuestros estudiantes el código utilizado para una tarea específica sin mostrar los resultados. **Sweave** prevé todas esas posibilidades y da al usuario total control a través de las opciones que se escriben dentro de `< < > >=`, las principales se explican a continuación:

```

<<echo=false>>=
  data(PlantGrowth)
  summary(PlantGrowth)
@

```

Le indica a **Sweave** que no queremos el código en el documento final, de esta forma solo coloca en el documento `.tex` las filas 1 y 6 a 15 (figura 2, centro). `echo=false` se puede cambiar por `echo=FALSE` o por `echo=F` con los mismos resultados.

---

<sup>5</sup>Si no es así, un buen comienzo es (W N Venables and D M Smith and R Development Core 2004) para R y (Rodrigo De Castro Korgi 2003) para  $\text{\LaTeX}$ .

<pre>doc.Rnw &lt;&lt;&gt;&gt;=   data(PlantGrowth)   summary(PlantGrowth) @</pre>	<pre>doc.tex 1 \begin{Schunk} 2 \begin{Sinput} 3 &gt; data(PlantGrowth) 4 &gt; summary(PlantGrowth) 5 \end{Sinput} 6 \begin{Soutput} 7   weight      group 8   Min.       :3.590   ctrl:10 9   1st Qu.    :4.550   trt1:10 10  Median     :5.155   trt2:10 11  Mean       :5.073 12  3rd Qu.    :5.530 13  Max.       :6.310 14 \end{Soutput} 15 \end{Schunk}</pre>	<pre>doc.dvi &gt; data(PlantGrowth) &gt; summary(PlantGrowth)    weight      group Min.       :3.590   ctrl:10 1st Qu.    :4.550   trt1:10 Median     :5.155   trt2:10 Mean       :5.073 3rd Qu.    :5.530 Max.       :6.310</pre>
---	---	--

Figura 2: Un *chunk* de código (izquierda), el resultado después de ejecutar Sweave(`doc.Rnw`) (centro) y como queda en el documento final (derecha)

<pre>&lt;&lt;results=hide&gt;&gt;=   data(PlantGrowth)   summary(PlantGrowth) @</pre>	<p>Le indica a Sweave que en el documento final no queremos la salida, sólo el código, de esa forma aparecen en el documento <code>.tex</code> las filas 1 a 5 y la 15 (figura 2, centro). Otros dos posibles valores son <code>results=tex</code> que se discutirá en la sección 3.3 y <code>results=verbatim</code> que es la opción por defecto.</p>
---	---

<pre>&lt;&lt;results=hide,echo=false&gt;&gt;=   data(PlantGrowth)   summary(PlantGrowth) @</pre>	<p>Omite tanto la salida como el código. En el documento <code>.tex</code> no aparece nada en esa posición, pero R carga los datos y realiza el cálculo de las estadísticas de resumen. Esa opción es útil cuando hacemos lectura de datos, definimos funciones o se inicializan variables que serán usadas en los cálculos que siguen.</p>
--	---

### 3.2. Inclusión de gráficos

Para la inclusión automática de gráficos se usa la opción `fig=true` en el *chunk* que crea el gráfico, veamos un ejemplo:

<pre>doc.Rnw \begin{figure}[ht] \centering &lt;&lt;echo=false,fig=true&gt;&gt;=   boxplot(weight~group,ylab="Peso") @ \caption{Boxplot}\label{boxplot} \end{figure}</pre>	<pre>doc.tex 1 \begin{figure}[ht] 2 \centering 3 \includegraphics{doc-003} 4 \caption{Boxplot}\label{boxplot} 5 \end{figure}</pre>
---	--

En las filas 3 a 5 del recuadro de la izquierda se encuentra el *chunk* que contiene el código R para hacer el gráfico. En el recuadro de la derecha, el *chunk* se reemplaza por el código  $\text{\LaTeX}$  `\includegraphics{doc-003}`

(fila 3), que inserta, en el documento final, el archivo `doc-003.ps`. El nombre del archivo gráfico `doc-003.ps` lo toma del nombre del archivo fuente `doc.Rnw` y el número corresponde al del *chunk* que produjo el gráfico (en este caso el tercero). El gráfico creado no se inserta en este documento para ahorrar espacio.

Es posible controlar el alto y ancho del gráfico creado por medio de las opciones `width` y `height`, por defecto estos dos parámetros están fijados en 6 pulgadas.<sup>6</sup>

```
\begin{figure}[ht]
\centering
<<echo=false,fig=true,width=7,height=8>>=
boxplot(weight~group,ylab="Peso")
@
\caption{Boxplot}\label{boxplot}
\end{figure}
```

Crea el Boxplot en un dispositivo gráfico con 7 pulgadas de ancho y 8 de alto.

Sweave guarda, por defecto, el mismo gráfico en formato PDF y PostScript. Si se compila con  $\LaTeX$  se inserta el archivo PostScript y si se compila con PDF $\LaTeX$  se inserta el archivo PDF, si el usuario no quiere crear el gráfico en formato PDF (porque siempre compila con  $\LaTeX$ ), debe insertar la opción `pdf=FALSE`, de igual forma, si no quiere crear el archivo en formato PostScript debe insertar la opción `eps=FALSE`.

### 3.3. Inclusión de tablas en formato $\LaTeX$

Podemos combinar la librería `xtable` de R junto con la opción `results=tex` para insertar tablas en formato  $\LaTeX$  directamente, veamos un ejemplo:

```
<<echo=F,results=tex>>=
library(xtable)
g<-lm(weight~group,data=PlantGrowth)
xtable(anova(g),"Análisis de varianza")
@
```

En el archivo `.tex` queda el código  $\LaTeX$  que genera la tabla 1.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	3.77	1.88	4.85	0.0159
Residuals	27	10.49	0.39		

Tabla 1: Análisis de varianza

### 3.4. Inserción de números en el texto

Es posible insertar automáticamente números en el texto usando la función `\Sexpr`, de la siguiente forma:

```
<<echo=F>>=
sw<-shapiro.test(residuals(g))
```

<sup>6</sup>Esas dimensiones son para el dispositivo gráfico de R, ellas no afectan el tamaño de las figuras en el archivo `.dvi`. El paquete `Sweave.sty` fija el ancho de las figuras que  $\LaTeX$  inserta en un 80% del parámetro `\textwidth`, por medio del comando `\setkeys{Gin}{width=0.8\textwidth}`, así por ejemplo, en este documento, que tiene la línea `\setlength{\textwidth}{16cm}` en el preámbulo, se garantiza que los gráficos insertados no tendrán un ancho mayor que 12.8 cm.

```
@
El valor del estadístico de Shapiro--Wilks es  $W = \text{round}(sw\$statistic,3)$ 
con  $p = \text{round}(sw\$p.value,3)$  por tanto no rechazamos la hipótesis de
normalidad de los errores.
```

Sweave coloca en el archivo `doc.tex` lo siguiente:

```
doc.tex
El valor del estadístico de Shapiro--Wilks es  $W=0.966$  con  $p=0.438$  por tanto
no rechazamos la hipótesis de normalidad de los errores.
```

y en el archivo `doc.dvi` queda así:

```
El valor del estadístico de Shapiro--Wilks es  $W = 0.966$  con  $p = 0.438$  por tanto no rechazamos la
hipótesis de normalidad de los errores.
```

### 3.5. Reutilización de código

Un *chunk* puede reutilizarse en otro que se defina posteriormente, para eso es necesario usar la opción `label=nombre`, para asignar un nombre al código que servirá para el llamado posterior. Se ilustra esto en el siguiente ejemplo.

```
Chunk a
<<a,echo=false>>=
sigma2<-summary(g)$sigma^2
XtX<-crossprod(model.matrix(g))
b<-matrix(coef(g))
@
```

```
Chunk b
<<b,echo=false>>=
est<-t(q)%*%b
var<-t(q)%*%solve(XtX)%*%q*sigma2
@
```

```
Chunk c
<<c,echo=false>>=
q<-matrix(c(0,0,1))
<<b>>
@
```

```
Chunk d
<<d,echo=false>>=
q<-matrix(c(0,-1,-1))
<<b>>
@
```

El *Chunk b* se está utilizando en el *Chunk c* para obtener la estimación de la función paramétrica  $\alpha_3 - \alpha_1$ <sup>7</sup> y su varianza. Lo mismo se hace en el *Chunk d* para la función  $2\alpha_1 - \alpha_2 - \alpha_3$ .

En estos ejemplos el nombre de los *chunks* es lo primero que aparece dentro de `<< >>`, esto siempre debe ser así a menos que el nombre se asigne mediante la opción `label=nombre`, en cuyo caso puede ir en cualquier posición.

La asignación de nombre a los *chunks*, además de la reutilización, permite definir código en un lugar del documento y obtener su resultado en otro, el siguiente ejemplo ilustra esa situación.

```
La creación
<<label=figura2,echo=false,fig=false>>=
plot(g,which=2)
@
```

```
El llamado
\begin{figure}[ht]
<<echo=false,results=hide,fig=true>>=
<<figura2>>
@
\caption{Normal Q--Q plot } \label{qqplot}
\end{figure}
```

1  
2  
3  
4  
5  
6

El *chunk* de código de la izquierda contiene el código de R que crea el *Normal Q-Q plot*, a éste se le asigna el nombre `figura2` por medio de la opción `label=figura2`. En la línea 3 del código de la derecha se hace

<sup>7</sup>Para obtener una solución de las ecuaciones normales, R hace una reparametrización, para lo cual impone la restricción  $\alpha_1^0 = 0$

el llamado, el cual puede estar en cualquier lugar del documento `.Rnw`, después de la creación, en ese sitio aparecerá el comando `\includegraphics{}` de  $\text{\LaTeX}$  que inserta el gráfico.

Los nombres de los *chunks* también se utilizan cuando se recupera el código `R` de un archivo `.Rnw`, con la función `Stangle`, esto se estudiará en la sección 3.6.

### 3.6. Recuperación de código

Si desde la consola de `R` ejecutamos la orden

```
> Stangle(doc.Rnw)
```

`R` recupera del archivo `doc.Rnw` el código que está dentro de los *chunk* y los guarda en un archivo de extensión `.r`, de la siguiente forma: Para cada *chunk* que tengan la opción `split=true`, crea un archivo `doc-nombre.r`, si el *chunk* tiene nombre, de lo contrario `nombre` se reemplaza por su número. Si hay varios *chunks* con el mismo nombre, todos se ubican en el mismo archivo. Por otra parte los *chunk* que tengan `split=false` (opción por defecto) se ubican en el archivo `doc.r`.

### 3.7. Asignación global de opciones.

Otra forma de especificar las opciones para controlar la manera como se transfieren los *chunks* y sus salidas (texto y figuras) del archivo `.Rnw` al archivo `.tex`, es mediante el comando `\SweaveOpts{}`, el cual modifica las opciones por defecto a partir del lugar donde aparece. Por ejemplo, si queremos fijar para todo el documento la opción `echo=false`, agregamos la línea `\SweaveOpts{echo=false}` al inicio, de esa forma no es necesario hacerlo en cada uno de los *chunks*. En cualquier lugar del documento podemos volver a activar `echo=true` mediante `\SweaveOpts{echo=true}`. Se puede fijar más de una opción separándolas por comas, por ejemplo `\SweaveOpts{echo=true,pdf=false,fig=true}`.

## 4. Configuración de $\text{\LaTeX}$

En principio `R` y `Sweave` están configurados para que el usuario reciba el archivo `.tex` y sin hacer nada más, compile con  $\text{\LaTeX}$  para obtener el documento `.dvi`, sin embargo eso es cierto sólo para usuarios de sistemas Unix (como Linux, por ejemplo), los usuarios de sistemas Windows, tienen que realizar una tarea adicional de configuración que se explica a continuación.

### 4.1. Instalación del paquete `Sweave.sty`

Cuando `Sweave` toma el archivo `.Rnw` y no encuentra en el preámbulo la línea `\usepackage{Sweave}`, agrega automáticamente en el documento `.tex` la línea <sup>8</sup>

```
\usepackage{C:/ARCHIV~1/R/R-22~1.0/share/texmf/Sweave}
```

que debería indicarle a  $\text{\LaTeX}$  dónde encontrar el paquete `Sweave.sty`, que es donde se definen los entornos `Schunk`, `Sinput` y `Soutput` necesarios para la presentación de las salidas y el código de `R`.

Al intentar compilar un documento con esa línea en el preámbulo, se genera un error de compilación, eso se debe a que  $\text{\LaTeX}$  no soporta espacios en blanco en la definición de las rutas de los archivos. Hay tres soluciones posibles para el inconveniente:<sup>9</sup>

1. (**Recomendada**) Instalar manualmente el paquete `Sweave.sty`. Para eso se debe crear una carpeta con el nombre `Sweave` en la ruta `C:\texmf\tex\latex` y copiar en ella el archivo `Sweave.sty` que se encuentra en la carpeta `...R\R-2.2.0\share\texmf`. Luego se debe actualizar la base de datos de los nombres de archivos instalados, en el caso se  $\text{\MiKTeX}$  se hace así:

<sup>8</sup>Puede variar de acuerdo con la configuración del sistema.

<sup>9</sup>Se asume que se está usando  $\text{\MiKTeX}$  con la instalación por defecto.

- Inicio > Todos los programas.
- MiKTeX > MiKTeX options.
- En la pestaña **General** en el recuadro **file name database**, click en Refresh Now.
- Espere unos pocos segundos y click en aceptar.

Al hacerlo de esa forma se tiene la ventaja que una vez realizado el procedimiento, no tenemos que preocuparnos más por ese asunto.

2. (**Poco práctica**) Copiar el archivo `Sweave.sty` en la misma carpeta donde tiene el archivo `.tex`. Esta opción tiene la desventaja que si se está trabajando en varios proyectos, en carpetas diferentes, se requiere una copia del archivo por cada uno, y cada vez que inicia un nuevo proyecto tendría que repetir el proceso de copia.
3. Instale R en una ruta sin espacios en blanco, por ejemplo en `C:\R\R-2.2.0`

Si se eligió la solución 1 o 2 dadas en la sección anterior, es necesario agregar manualmente, en el preámbulo del documento `.Rnw` la línea `\usepackage{Sweave}`, que invoca al paquete `Sweave.sty`.

## 5. Un ejemplo completo.

El siguiente es un ejemplo de cómo debe quedar definitivamente un archivo `.Rnw`. El documento `.tex` y el `.dvi` no se muestran por falta de espacio.

```
\documentclass[12pt]{report}
\usepackage{amsmath,amsthm,amsfonts,amssymb}
\usepackage[latin1]{inputenc}
\usepackage[spanish,activeacute]{babel}
\usepackage{Sweave}
\SweaveOpts{pdf=false}
\begin{document}
En el análisis exploratorio de los datos calculamos las estadísticas
de resumen para cada grupo, las cuales se muestran en la siguiente
salida de \texttt{R}.

<<echo=false>>=
  attach(PlantGrowth)
  by(weight,group,summary)
@

En la figura \ref{boxplot} se muestra el Boxplot de peso para cada
grupo, el cual muestra que no parece haber problemas con el supuesto
de homogeneidad de la varianza de los grupos.
\begin{figure}[ht]
\centering

  <<echo=false,fig=true,width=8,height=8>>=
  boxplot(weight~group,ylab="Peso")
  @

\caption{Boxplot}\label{boxplot}
\end{figure}
A continuación ajustamos el modelo de una vía de clasificación. En
la tabla \ref{anova} se muestra el análisis de de varianza.
```

```

<<echo=F,results=tex>>=
  g<-lm(weight~group,data=PlantGrowth)
  xtable(anova(g),"Análisis de varianza","anova")
@

El p valor de la prueba ( $p=\text{Sexpr}\{\text{round}(\text{anova}(g)\$`Pr(>F)`[1],4)\}$ )
indica que los efectos de los tratamientos difieren.

<<echo=true>>=
  sw<-shapiro.test(residuals(g))
@

El valor del estadístico de Shapiro--Wilks es
 $W=\text{Sexpr}\{\text{round}(sw\$statistic,3)\}$  con
 $p=\text{Sexpr}\{\text{round}(sw\$p.value,3)\}$  por tanto se verifica el supuesto de
normalidad.

\end{document}

```

## 6. Conclusión

En este documento se ha mostrado un *panorama* general del funcionamiento de **Sweave** y su versatilidad, para mayores detalles debe consultarse (Leisch 2005). Con este tutorial, el lector que posea experiencia con  $\text{\LaTeX}$  y **R** puede empezar a producir documentos con cierto nivel de calidad en poco tiempo. Se espera que los lectores no familiarizados con estos dos paquetes se hayan percatado de lo práctico que resulta esta forma de escribir informes y entiendan que cualquier cantidad de tiempo que se gaste en su aprendizaje, debe tomarse como *una inversión* que será recuperada con creces posteriormente. Ténganse en cuenta que  $\text{\LaTeX}$  y **R** son dos herramientas poderosas y útiles distribuidas bajo **Licencia Pública General** y por tanto se convierten en una buena opción para la enseñanza, dado que los estudiantes, las universidades y centros de investigación pueden adquirirlos sin ningún costo.

## Referencias

- Leisch, F. (2005), *Sweave User Manual*.
- Neely, W. W. (n.d.), *Using Sweave For Integrating R and \LaTeX*.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rodrigo De Castro Korgi (2003), *El Universo \LaTeX*, Universidad Nacional de Colombia.
- Rudloff, P. (2005), *An Introduction to Sweave*.
- The  $\text{\LaTeX}$ 3 project (2006), *The \LaTeX3 project*, <http://www.latex-project.org/>.
- W N Venables and D M Smith and R Development Core (2004), *An Introduction to R*, Network Theory Limited.