

Scrivere report statistici con e Sweave

Michele Scandola, Nadia Baltieri

Sommario

In questo articolo si cercherà di fornire alcune informazioni introduttive sull'uso di **Sweave**, uno strumento che combina codice $\text{\LaTeX}$ con codice R, un linguaggio di programmazione per il calcolo statistico che negli ultimi anni sta guadagnando sempre più consensi, al punto tale da diventare il linguaggio di riferimento della comunità statistica.

Abstract

The purpose of this brief article is to give an introduction to **Sweave**, a tool that allows to combine $\text{\LaTeX}$ code and R code. R is a statistical programming language that is gaining popularity in the last years, thus becoming the referential programming language for the statistics community.

1 Introduzione

Ai lettori di $\text{\LaTeX}$ probabilmente non serve una descrizione di $\text{\LaTeX}$, ma forse è opportuno spendere qualche parola introduttiva sull'ambiente di sviluppo R.

R è un ambiente e un linguaggio di programmazione per il calcolo statistico e la rappresentazione grafica dei dati. Apparso per la prima volta nel 1993, R è un dialetto di S, ambiente e linguaggio di programmazione statistico sviluppato nei laboratori Bell da John Chambers e colleghi (HORNIK, 2013).

R è disponibile gratuitamente sotto licenza GNU GPL e conta attualmente più di 5200 pacchetti statistici, scaricabili gratuitamente tramite il Comprehensive R Archive Network (CRAN <http://cran.r-project.org/mirrors.html>). Tali pacchetti coprono la maggior parte delle metodologie statistiche e in taluni casi, come nei *mixed effect models* (PINHEIRO e BATES, 2000; BATES, 2010) e nella statistica bayesiana (KRUSCHKE, 2011), rappresentano il cosiddetto “gold standard”. Nei pacchetti spesso si trovano le più moderne tecniche statistiche: in molti casi essi sono scritti direttamente dagli ideatori delle ultime novità nel campo della statistica.

Fra i numerosi pacchetti che si trovano nel CRAN ve ne sono anche alcuni (più di 40) che permettono delle efficaci ed eleganti rappresentazioni grafiche dei dati usando concetti innovativi e sperimentali, come, ad esempio, la “grammatica dei grafici” (WILKINSON, 2005), pienamente sviluppata in ggplot2 (WICKHAM, 2009).

Altri pacchetti e funzionalità di R riguardano gli strumenti per la gestione e/o il trattamento dei dati, come ad esempio tuneR (LIGGES *et al.*, 2013), utile per il processamento dei dati audio, o il pacchetto RODBC (RIPLEY e LAPSLEY, 2010), necessario per le connessioni ODBC¹.

Oggigiorno R non è ancora così diffuso nella comunità scientifica, questo probabilmente anche a causa della grande diffusione di programmi di più lungo corso e dotati di una comoda interfaccia grafica, come SPSS (SPSS ITALIA, 2014) e SAS (SAS INSTITUTE, 2014), che permettono anche a chi ha una conoscenza minima del tema di fare analisi statistiche. Ciò nonostante R risulta essere il più usato fra i “software minori” in ambito statistico, mostrando una tendenza in rapida crescita (MUENCHEN, 2014). Per gli statistici di professione R ormai è il software di riferimento.

2 I report statistici

Un report statistico viene redatto con lo scopo di informare i lettori sui risultati che possono provenire dalla ricerca sperimentale, dai dati di interviste e questionari o da altre fonti relative a una particolare area o progetto (vedi Figura 1). Qualsiasi ricerca che implichi la raccolta di dati quantitativi (e a volte anche qualitativi) richiede un'analisi statistica, o per lo meno l'utilizzo di statistiche descrittive. Né l'analisi statistica né la ricerca in sé è utile finché non viene redatta in un formato leggibile e comprensibile. La preparazione del report statistico fa parte di questo processo comunicativo.

Se a un profano può sembrare che realizzare report statistici sia un compito banale, ossia un semplice copia-incolla dei risultati, motivazioni che risiedono nelle fondamenta della sperimentazione scientifica mostrano come ciò non basti. In ogni esperimento a carattere scientifico la metodologia utilizzata deve essere chiara e trasparente, per permettere la riproducibilità dell'esperimento e la conseguente verifica dei risultati.

Anche se un report statistico non è un articolo scientifico, elementi quali *introduzione*, *review dello stato dell'arte della letteratura*, *descrizione delle ipotesi e metodo* sono caldamente consigliati, se non strettamente necessari. Questo permette di inquadrare la ricerca in un contesto più ampio

1. le connessioni ODBC sono delle interfacce native che permettono l'accesso al *database management system* utilizzabile da quei linguaggi di programmazione in grado di chiamare funzioni di librerie native (da Wikipedia, <http://it.wikipedia.org/wiki/ODBC>)

anche dai “non addetti ai lavori”. L’*introduzione* dovrebbe spiegare la motivazione e la natura della ricerca in atto, nella *review dello stato dell’arte della letteratura* si dovrebbe cercare di indicare i principali lavori scientifici che stanno alla base del proprio lavoro, in modo da giustificare la formulazione delle *ipotesi*. Nel *metodo* va invece descritto come sono stati raccolti i dati, la natura degli esperimenti o dei questionari, la tipologia di partecipanti e qualsiasi aggiustamento che è stato apportato durante lo svolgimento del progetto.

La descrizione delle analisi statistiche utilizzate è una parte molto importante del *metodo*, anche se spesso negli articoli scientifici viene trascurata e si tende, invece, ad includerla nei *risultati*. Nel report statistico questa parte è fondamentale; si indica quale software è stato utilizzato per l’analisi dei dati, la sua versione e se sono stati utilizzati pacchetti aggiuntivi. Inoltre, si dichiara esattamente che tipo di analisi è stata effettuata e, se pertinente, con quali correzioni ai test. Infine, sono importanti le motivazioni sulla scelta dell’uso di alcune analisi e non di altre.

Tutte queste informazioni sono fondamentali per la riproducibilità della ricerca scientifica e per la valutazione della correttezza o dell’opportunità delle tecniche statistiche utilizzate.

Una sezione che dovrebbe essere particolarmente accurata ed estesa, nei report statistici, è la descrizione dei dati. Va descritto accuratamente il campione da cui sono stati presi i dati, utilizzando statistiche descrittive riguardanti l’età (media, deviazione standard, *range*), la distribuzione di genere (numero di maschi e femmine), se pertinente, la manualità (destro o mancino), il grado di istruzione (frequenza per titolo di studio, media e deviazione standard degli anni di studio), eccetera. La stessa cosa va fatta per i dati di interesse della ricerca, magari aiutandosi con grafici, come i *boxplot* o gli istogrammi.

Un altro metodo che rende ancor più trasparente l’analisi, con conseguente giovamento dell’applicazione del metodo scientifico, è rendere disponibili tutti i dati e, se possibile, il codice delle analisi effettuate. I dati possono essere salvati come foglio Excel o in altro formato, mentre il codice delle analisi viene salvato nel formato del software utilizzato.

Nella pratica comune, però, spesso queste indicazioni non vengono seguite e il codice delle analisi statistiche e i dati analizzati vengono tenuti separati, sebbene siano entrambi una parte fondamentale per la verifica dei risultati. Quando si analizzano i dati, è necessario testarli in più modalità per capirne le caratteristiche e comprendere quali siano i modelli statistici a loro più adatti. Riprendere i dati e le analisi più e più volte può portare ad avere numerose versioni di dati, analisi e report e può essere difficile comprendere a quali analisi e

a quali dati corrisponda ciascun report (LEISCH, 2003).

Una soluzione a tutto questo è l’integrazione di dati, delle analisi e del report statistico, tramite il *literate statistical practice* (ROSSINI *et al.*, 2003). In questo modo le analisi sono incorporate all’interno dello stesso report. Questa metodologia non è applicabile a tutti i software statistici ma è facilmente effettuabile con R, nei limiti della dimestichezza che l’utente ha con R e più in generale con il mezzo informatico. Nel prossimo paragrafo parleremo di come questo tema sia stato affrontato in R.

3 La Reproducible Research

Il CRAN offre l’opportunità di vedere i pacchetti al suo interno organizzati in tematiche (attualmente più di 30) che vanno dalla statistica bayesiana, passando per l’econometria, la genetica, il *machine learning*, finendo in strumenti statistici utili per il Web.

La tematica che ci interessa è quella che si occupa della riproducibilità della ricerca (KUHN, 2014), ossia di tutti quei pacchetti che permettono di preparare report statistici ben impaginati e comprensibili anche a chi non usa R e non ha familiarità con le sue funzioni, faticando quindi a capirne l’output.

L’idea alla base di tutti questi pacchetti è quella di scrivere un unico script contenente codice R di analisi e codice per la creazione del report. In questo modo, se si vuole fare un aggiornamento dei dati, non è necessario riscrivere tutto il testo ma basta eseguire nuovamente lo script.

In breve, R offre pacchetti che permettono di esportare i report in svariati formati, fra cui HTML, markdown, Rich Text Format e formati Word proprietari.

La possibilità più supportata al momento è L^AT_EX, tramite Sweave (LEISCH, 2002) e pacchetti opzionali come xtable (DAHL, 2013), Hmisc (HARRELL JR *et al.*, 2014), animation (XIE, 2013) e tikzDevice (SHARPSTEEN e BRACKEN, 2013).

4 Sweave

Sweave fornisce un framework flessibile che permette di mettere insieme, come già detto, la documentazione delle analisi, il codice R e il suo output in un unico file. I file Sweave sono semplici file testuali, con l’estensione .nw, .rnw, .Rnw, .snw o .Snw.²

All’interno di un file Sweave troviamo *code chunks* e *documentation chunks*. All’interno dei *code chunks* troviamo il codice R, mentre all’interno dei *documentation chunks* si scrive il codice L^AT_EX.

2. Per approfondimenti sull’utilizzo di Sweave, visita la pagina <http://www.stat.uni-muenchen.de/~leisch/Sweave/> e consultare l’articolo apparso su ArSTeXnica (CRIVELLARO, 2011).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	Length	delayStart	delayStop	RTStart	RTStop	VAS ok	VAS fail	VAS down	Answer	fileAudio					
2	Sub02	2	295715	296324	297357	304907	0	0	0	Riesce	sub2_voice_1_20415136353536000000000000.wav					
3	Sub02	3	305013	324179	327419	330749	100	100	100	0 Riesce	sub2_voice_2_20415136353536000000000000.wav					
4	Sub02	8	305013	324179	342688	330749	100	100	100	0 Riesce	sub2_voice_2_20415136353536000000000000.wav					
5	Sub02	4	330780	382063	383857	385146	100	100	100	0 Riesce	sub2_voice_3_20415136353536000000000000.wav					
6	Sub02	10	385176	396664	399997	402468	100	100	100	0 Riesce	sub2_voice_4_20415136353536000000000000.wav					
7	Sub02	2	402500	408295	409335	414900	50	50	50	0 Riesce	sub2_voice_5_20415136353536000000000000.wav					
8	Sub02	6	414930	422539	425046	426830	100	100	100	0 Si blocca	sub2_voice_6_20415136353536000000000000.wav					
9	Sub02	8	426862	435826	439084	442399	50	50	50	0 Riesce	sub2_voice_7_20415136353536000000000000.wav					
10	Sub02	2	442431	449508	450529	459210	0	0	0	0 Si blocca	sub2_voice_8_20415136353536000000000000.wav					
11	Sub02	6	459242	468494	469494	468595	100	100	100	0 Riesce	sub2_voice_9_20415136353536000000000000.wav					
12	Sub02	2	468624	473567	474568	477098	0	0	0	0 Si blocca	sub2_voice_10_20415136353536000000000000.wav					
13	Sub02	4	477130	482446	484177	485707	100	100	100	0 Si blocca	sub2_voice_11_20415136353536000000000000.wav					
14	Sub02	4	485739	492612	494388	496747	50	50	50	0 Riesce	sub2_voice_12_20415136353536000000000000.wav					
15	Sub02	6	496780	503432	505912	507387	100	100	100	0 Riesce	sub2_voice_13_20415136353536000000000000.wav					
16	Sub02	8	507560	513336	516825	518081	100	100	100	0 Si blocca	sub2_voice_14_20415136353536000000000000.wav					
17	Sub02	8	518115	525666	528836	530075	50	50	50	0 Riesce	sub2_voice_15_20415136353536000000000000.wav					
18	Sub02	4	530068	535370	537038	540562	50	50	50	0 Cade	sub2_voice_16_20415136353536000000000000.wav					
19	Sub02	2	540578	545078	546078	550966	50	50	50	0 Riesce	sub2_voice_17_20415136353536000000000000.wav					
20	Sub02	2	550985	556159	557183	559674	0	0	0	0 Si blocca	sub2_voice_18_20415136353536000000000000.wav					
21	Sub02	8	559905	569462	572247	573898	50	50	50	0 Cade	sub2_voice_19_20415136353536000000000000.wav					
22	Sub02	2	573935	581089	582108	582925	100	100	100	0 Si blocca	sub2_voice_20_20415136353536000000000000.wav					
23	Sub02	2	588324	594007	595037	598407	0	0	0	0 Riesce	sub2_voice_21_20415136353536000000000000.wav					
24	Sub02	6	598439	603439	605988	609188	100	100	100	0 Si blocca	sub2_voice_22_20415136353536000000000000.wav					
25	Sub02	10	609219	615157	619094	620566	100	100	100	100 Cade	sub2_voice_23_20415136353536000000000000.wav					
26	Sub02	10	620600	627665	631598	633478	100	100	100	0 Si blocca	sub2_voice_24_20415136353536000000000000.wav					
27	Sub02	2	633502	642558	643593	647893	100	100	100	0 Si blocca	sub2_voice_25_20415136353536000000000000.wav					
28	Sub02	2	642538	647746	648769	652159	0	0	0	0 Riesce	sub2_voice_26_20415136353536000000000000.wav					
29	Sub02	8	652188	657469	660675	661389	100	100	100	0 Riesce	sub2_voice_27_20415136353536000000000000.wav					
30	Sub02	2	661475	665918	666910	669113	50	50	50	0 Riesce	sub2_voice_28_20415136353536000000000000.wav					
31	Sub02	6	669146	674454	676943	685461	50	50	50	0 Riesce	sub2_voice_29_20415136353536000000000000.wav					
32	Sub02	4	685493	690668	692404	693917	50	50	50	0 Si blocca	sub2_voice_30_20415136353536000000000000.wav					
33	Sub02	4	693939	700481	701986	702947	100	100	100	0 Riesce	sub2_voice_31_20415136353536000000000000.wav					
34	Sub02	8	702980	707804	710534	712087	100	100	100	0 Riesce	sub2_voice_32_20415136353536000000000000.wav					
35	Sub02	6	712116	716741	717145	720496	0	0	0	0 Si blocca	sub2_voice_33_20415136353536000000000000.wav					
36	Sub02	6	720222	725878	728356	731575	0	0	0	0 Si blocca	sub2_voice_34_20415136353536000000000000.wav					
37	Sub02	2	731607	736683	737579	739422	0	0	0	0 Riesce	sub2_voice_35_20415136353536000000000000.wav					
38	Sub02	2	739455	744534	745545	749053	0	0	0	0 Riesce	sub2_voice_36_20415136353536000000000000.wav					
39	Sub02	2	749083	754245	755263	757038	0	0	0	0 Si blocca	sub2_voice_37_20415136353536000000000000.wav					
40	Sub02	4	757065	762579	764095	766165	50	50	50	0 Riesce	sub2_voice_38_20415136353536000000000000.wav					
41	Sub02	10	766197	773986	777919	778901	100	100	100	100 Cade	sub2_voice_39_20415136353536000000000000.wav					
42	Sub02	6	778929	784069	786554	789597	100	100	100	0 Si blocca	sub2_voice_40_20415136353536000000000000.wav					
43	Sub02	4	789628	794665	796139	798884	50	50	50	0 Riesce	sub2_voice_41_20415136353536000000000000.wav					

```
y=c(rnorm(20,mean=10,sd=3),
    rnorm(20,mean=5,sd=2)),
# all'interno del data frame
# creo una variabile
# y composta da due serie
# di 20 numeri casuali
```

`\SweaveOpts{op1=val1, op2=val2, ...}`

che permette di modificare le opzioni generiche del report. Il comando può essere scritto in qualsiasi punto di un *documentation chunk*.

Le opzioni principali sono:

label: con questa opzione si può indicare il nome del *code chunk*. È sempre utile e consigliato impiegarlo con label univoci per ogni *chunk*, sia per un possibile *debugging* del codice sia per un eventuale riutilizzo del *chunk*. Inoltre, se nello stesso *chunk* vengono create immagini, l'opzione *label* viene utilizzata per creare immagini automaticamente salvate col nome del *chunk* o, se il documento lo prevede, per creare file multipli di output. Il *label* può essere dichiarato come prima opzione del *chunk*, senza specificare che è un *label* (es. `<<postHoc-tests>>=`). Quando è dichiarato in qualsiasi altro ordine è necessario specificare che si tratta di un *label* (es. `<<opt1=value1,label=postHoc-tests>>=`).

echo: questa opzione, che può essere settata a TRUE o FALSE, viene utilizzata per indicare se si vuole che i comandi R usati nel *code chunk* siano poi riportati nel report statistico oppure no. Per una migliore trasparenza nella ricerca scientifica è bene che quest'opzione sia sempre settata a TRUE.

results: può prendere come valori *hide*, *verbatim* o *tex*. In questo modo è possibile indicare se e come si vuole che i risultati del *code chunk* vengano riportati, vale a dire se si vuole che vengano inseriti come codice LATEX (opzione *tex*), che vengano inseriti all'interno dell'ambiente *verbatim*, oppure se si preferisce che non siano riportati affatto (opzione *hide*).

fig: con questa opzione si indica che come output dal *code chunk* si vuole produrre un'immagine da inserire nel documento. Se impostata a TRUE, Sweave creerà, infatti, un'immagine in formato *pdf*, che sarà nominata con lo stesso nome del label del *code chunk*. In caso di più immagini, non è possibile usare sempre lo stesso label perché ogni nuova immagine verrebbe memorizzata sostituendo quella precedentemente salvata.

Un semplice esempio di come scrivere una pagina in Sweave è mostrato nel codice seguente (per l'output, vedere la Figura 2).

```
\documentclass{article}

\title{Un semplice esempio di come
scrivere una pagina in
\emph{Sweave}}
\author{Michele Scandola}

\begin{document}
```

```
\SweaveOpts{concordance=TRUE,
width=12, height=8}

\maketitle

\section{Introduzione}

In questa pagina \emph{Sweave} si
mostra un semplice esempio di come
integrare codice \LaTeX a codice
\emph{R}.

<<loading,echo=TRUE,results=hide>>=
... # vedere esempio di codice
# riportato in precedenza
@

\section{Plot dei dati}

<<boxplot,echo=TRUE,fig=TRUE>>=
boxplot(y~group,data=dati)
# con questo comando si esegue
# un boxplot dei dati creati in
# precedenza divisi per gruppo
@

\section{T-test sui dati}

<<ttest,echo=TRUE,
results=verbatim>>=
t.test(subset(dati$y,
dati$group=="A"),
subset(dati$y,
dati$group=="B"))
# con questo comando si esegue
# un test statistico t sulla
# differenza media fra i dati
# provenienti dal gruppo "A"
# ed il gruppo "B"
@

\end{document}
```

Come si può vedere da questo esempio con Sweave è possibile riportare non solo i risultati ma anche la analisi effettuate, favorendo così la riproducibilità della ricerca. Ciò nonostante, bisogna tenere presente che l'utilizzo di Sweave non è di per sé garanzia di trasparenza e riproducibilità: in ogni caso il documento deve essere redatto in modo trasparente (ossia i comandi utilizzati devono essere resi visibili, tramite l'opzione *echo=TRUE*) e i dati devono essere accessibili.

È possibile rendere accessibili i dati sia inserendoli nel documento, preferibilmente in forma estesa (non riportando, ad esempio, solo le medie e deviazioni standard), sia rappresentandoli in maniera aggregata (tramite *boxplot* o altre tipologie di grafici). L'accessibilità dei dati è stata recentemente valutata anche per la realizzazione di report nell'ambito di articoli scientifici (SIMONSOHN, 2013). La proposta è stata facilmente resa effettuabile negli ultimi anni, per quanto si tratti di una oppor-

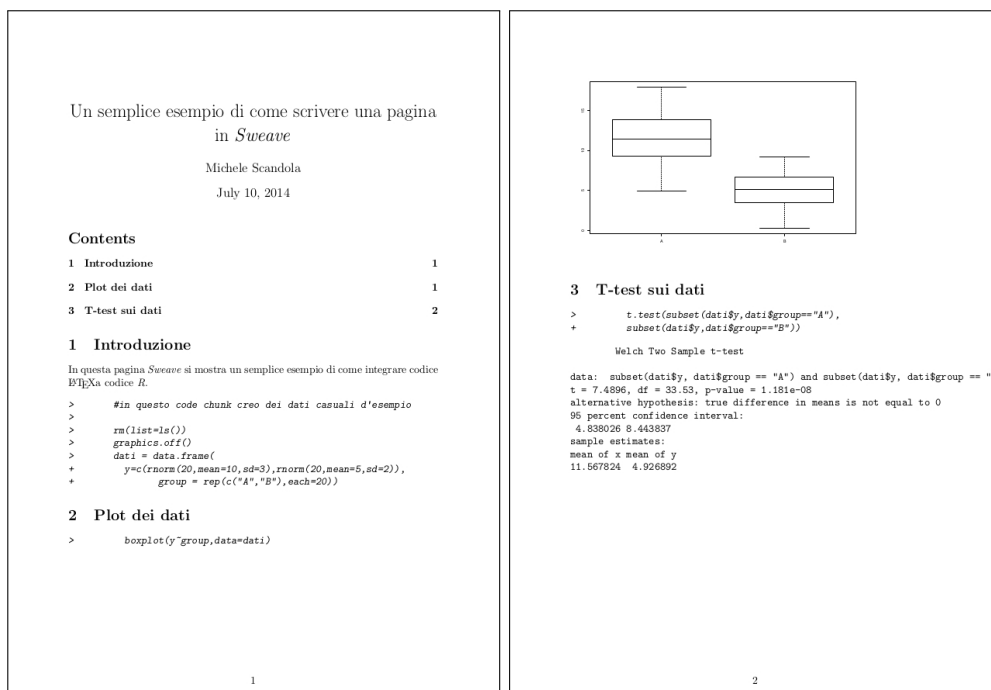


FIGURA 2: Output della pagina d'esempio in Sweave.

tunità non ancora completamente sfruttata, grazie alla possibilità di caricare materiale supplementare online in importanti riviste scientifiche internazionali, come nelle riviste della serie di *Frontiers*, su *Cortex*, ecc. Il procedimento per fare ciò è molto semplice ed è possibile eseguirlo nella fase di caricamento dell'articolo nei siti di riferimento delle varie riviste citate e non solo.

Al fine di realizzare report statistici completamente trasparenti, appare però opportuno rendere accessibili non solo i dati ma anche il codice *Sweave* e, conseguentemente, il codice delle analisi statistiche, caricando anch'esso online come materiale supplementare.

4.1 Come usare Sweave su progetti che includono molteplici files

Nella scrittura di un documento molto ampio, come può essere una tesi di dottorato o un libro, è normale, pratico e utile non limitarsi alla scrittura di un unico file ma suddividerlo in un progetto composto da più files.

Un ampio progetto fatto in *Sweave* può essere costituito da un file principale *.Rnw* contenente diverse chiamate a `\input` per collegarsi ad altri files. Tali files possono essere:

- files *.tex*
- files *.Rnw* o simili, che necessitano di essere convertiti in files *.tex*
- alcune figure statiche
- eccetera

Procedere in tal modo significa che bisogna eseguire il comando *Sweave* su ogni file *.Rnw* in

modo da convertirli in files *.tex* e, solo in seguito, lanciare il compilatore *TeX* che si vuole usare sul file principale.

Questo procedimento può risultare piuttosto macchinoso, soprattutto se si vuole controllare che effetto hanno avuto alcuni cambiamenti appena fatti in un file secondario sul documento complessivo. Un modo per semplificare questo processo è usare il comando *SweaveAll* nel pacchetto *patchDVI* (MURDOCH, 2013a) per R. Questa funzione necessita che due variabili vengano impostate per funzionare in modo ottimale:

.TexRoot una stringa al cui interno si trova il nome del file principale del progetto

.SweaveFiles un array di stringhe al cui interno si trovano i nomi di tutti i files che si vuole vengano compilati col comando *Sweave*

In particolare, il file *.Rnw* principale deve contenere un code chunk simile al seguente codice:

```
<<echo=FALSE>>=
.SweaveFiles = c("main.Rnw",
                  "Chapter1.Rnw",
                  "Chapter2.Rnw",
                  "Chapter3.Rnw")

@
```

In questo modo, i files *main.Rnw*, *Chapter1.Rnw*, *Chapter2.Rnw* e *Chapter3.Rnw* verranno compilati con *Sweave*. Nell'esempio precedente abbiamo inserito anche il file principale *main.Rnw*, in tal modo anche il codice R presente in esso verrà ricompilato con gli altri files. Nel caso non vi fosse codice R nel file principale, lo si può omettere dalla stringa *.SweaveFiles*.

In tutti i *files* subordinati (che sono inclusi nel *file* principale tramite il comando `\input`) è invece necessario l'assegnamento della variabile `.TexRoot`. Un esempio è riportato nel codice seguente:

```
<<echo=FALSE>>
.TexRoot = "main.tex"
@
```

In questo modo il comando *SweaveAll* è informato che, dopo aver eseguito **Sweave** su tutti i *files* contenuti in *.SweaveFiles*, deve eseguire il compilatore $\text{\TeX}$ sul *file* contenuto in *.TexRoot*.³

5 Conclusione

Questo interesse del mondo scientifico a rendere disponibili a tutti i dati e i metodi di analisi utilizzati dimostra come si senta la necessità di una maggior trasparenza e controllo. In questo modo chiunque può verificare i calcoli eseguiti e le conclusioni da essi dedotte, tutelando così la ricerca scientifica da ogni possibile obiezione che potrebbe essere sollevata in ambito accademico in relazione alla trasparenza della metodologia di analisi impiegata. In questo sia R che **Sweave** possono giocare un ruolo importante.

Riferimenti bibliografici

- BATES, D. M. (2010). *lme4: Mixed-effects modeling with R*. URL <http://lme4.r-forge.r-project.org/book>.
- CRIVELLARO, M. (2011). «Un dialogo tra GNU-R e LaTeX: xtable e Sweave». *ArsTeXnica*, **11**, pp. 31–36. URL <http://www.guitex.org/home/images/ArsTeXnica/AT011/AT11-crivellaro.pdf>.
- DAHL, D. B. (2013). *xtable: Export tables to LaTeX or HTML*. URL <http://cran.r-project.org/package=xtable>. R package version 1.7-1.
- HARRELL JR, F. E. *et al.* (2014). *Hmisc: Harrell Miscellaneous*. URL <http://cran.r-project.org/package=Hmisc>. R package version 3.14-0.
- HORNIK, K. (2013). «The R-project website». URL <http://www.r-project.org>.
- KRUSCHKE, J. K. (2011). *Doing Bayesian data analysis : a tutorial with R and BUGS*. Academic Press, Burlington, MA. URL http://www.worldcat.org/search?qt=worldcat_org_all&q=0123814855.
- KUHN, M. (2014). «CRAN Task View: Reproducible Research». URL <http://cran.r-project.org/web/views/ReproducibleResearch.html>.
- LEISCH, F. (2002). «Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis». In *Compstat 2002 — Proceedings in Computational Statistics*, a cura di W. HÄRDLE e B. RÖNZ. Physica Verlag, Heidelberg, pp. 575–580. URL <http://www.stat.uni-muenchen.de/~leisch/Sweave>. ISBN 3-7908-1517-9.
- (2003). «Sweave and Beyond: Computations on Text Documents». In *Proceedings of 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, a cura di K. HORNIK, F. LEISCH e A. ZEILEIS. Vienna, Austria, Dsc. URL <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/Leisch.pdf>.
- LIGGES, U., KREY, S., MERSMANN, O. e SCHNACKENBERG, S. (2013). *tuneR: Analysis of music*. URL <http://r-forge.r-project.org/projects/tuner>.
- MUENCHEN, R. A. (2014). «The Popularity of Data Analysis Software». URL <http://r4stats.com/articles/popularity>.
- MURDOCH, D. (2013a). *patchDVI: Package to patch .dvi files*. URL <http://cran.r-project.org/package=patchDVI>. R package version 1.9.1601.
- (2013b). «The patchDVI package». URL <http://cran.r-project.org/web/packages/patchDVI/vignettes/patchDVI.pdf>.
- PINHEIRO, J. C. e BATES, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer. ISBN 0-387-98957-0.
- RIPLEY, B. e LAPSLEY, M. (2010). *RODBC: ODBC Database Access*. URL <http://cran.r-project.org/package=RODBC>. R package version 1.3-2. Michael Lapsley ha collaborato dal 1999 fino all'ottobre 2002.
- ROSSINI, A., HEIBERGER, R. M., SPARAPANI, R. A., MÄCHLER, M. e HORNIK, K. (2003). «Emacs Speaks Statistics (ESS): A multi-platform, multi-package intelligent environment for statistical analysis».
- SAS INSTITUTE (2014). «The SAS website». URL <http://www.sas.com/offices/europe/italy>.
- SHARPSTEEN, C. e BRACKEN, C. (2013). *tikzDevice: R Graphics Output in LaTeX Format*. URL <http://cran.r-project.org/package=tikzDevice>. R package version 0.7.0.
- SIMONSOHN, U. (2013). «Just post it: the lesson from two cases of fabricated data detected by statistics alone». *Psychological Science*, **24** (10), pp. 1875–1888.

3. Per ulteriori approfondimenti consultare (MURDOCH, 2013b).

- SPSS ITALIA (2014). «The SPSS website». URL <http://www.spss.it>.
- WICKHAM, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. URL <http://had.co.nz/ggplot2/book>.
- WILKINSON, L. (2005). *The Grammar of Graphics (Statistics and Computing)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- XIE, Y. (2013). «animation: An R Package for Creating Animations and Demonstrating Statistical Methods». *Journal of Statistical Software*, **53** (1), pp. 1–27. URL <http://www.jstatsoft.org/v53/i01>.

- ▷ Michele Scandola
IRCCS Fondazione S. Lucia, Rome, Italy
SCNLab, Dipartimento di Psicologia, Università degli Studi di Roma “La Sapienza”
NPSY-Lab.VR, Dipartimento di Filosofia, Pedagogia e Psicologia, Università degli Studi di Verona
`michele dot scandola at univr dot it`
- ▷ Nadia Baltieri
Dipartimento di Filologia, Letteratura e Linguistica, Università degli Studi di Verona
`nadia dot baltieri at univr dot it`