

Scrivere report statistici con e Sweave

Michele Scandola^{a,1} Nadia Baltieri^{b,2}

^aIRCCS 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

^bDipartimento di Filologia, Letteratura e Linguistica
Università degli Studi di Verona

GulT meeting 2014

¹michele.scandola@univr.it


²nadia.baltieri@univr.it


Cos'è un report statistico



Un report statistico è un documento che illustra sinteticamente i fenomeni oggetto di indagine al fine di comunicare i risultati che provengono dalla ricerca sperimentale, dalla raccolta dei dati di interviste e questionari o da altre fonti relative a una particolare area o progetto.


Elementi fondamentali di un report statistico

- ▶ Introduzione
- ▶ Review dello stato dell'arte
- ▶ Formulazione di ipotesi
- ▶ Descrizione del metodo
- ▶ Dati

 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,  è un dialetto di S, ambiente e linguaggio di programmazione statistico sviluppato nei laboratori Bell da John Chambers e colleghi.

 è disponibile gratuitamente sotto licenza GNU e conta attualmente più di 5200 pacchetti statistici, scaricabili gratuitamente tramite il Comprehensive  Archive Network: <http://cran.r-project.org/mirrors.html>

Sweave è un ambiente che permette di integrare  e \LaTeX in modo da avere l'elaborazione statistica ed i vantaggi del *literate programming*. Tramite il suo utilizzo si possono creare report statistici dinamici, aggiornabili ad ogni compilazione.

Codice R (1)

```
## librerie R di utilizzo
library(ggplot2)## libreria per i grafici
library(afex)## librerie per l'ANOVA di terzo tipo

## creazione dei dati
dati = data.frame(
  y = c(10,33,44,77,52,34,76,21,87,32,15,31,51,16,
        49,52,56,68,26,72,16,13,16,16,56,47,38,19,
        46,35,59,15,74,58,35,62,12,51,45,78,56,27,
        15,30,23,65,46,51,55,48,12,23,59,52,23,90),
  group = gl(2,28,length=56,labels=c("A","B")),
  condition=gl(2,14,length=56,labels=c("I","II")),
  subject = c(rep(1:14,2),rep(15:28,2)))
```

Codice R (2)

```
## visualizzazione dei dati
dati

## controllo se le condizioni sono bilanciate
table(dati$group,dati$condition)

## grafico boxplot
ggplot(dati)+
  geom_boxplot(aes(x=group,y=y))+
  facet_grid(.~condition)+xlab("")
```

Codice R (3)

```
## informazioni riassuntive della
## distribuzione: media e dev. standard
sommario=aggregate(y~group*condition,data=dati,
  FUN=mean)
sommario$SD=aggregate(y~group*condition,data=dati,
  FUN=sd)$y

## visualizzazione delle informazioni riassuntive
sommario
```


Codice R (4)

```
## grafico delle medie e deviazioni standard
ggplot(sommario)+
  geom_bar(aes(x=group,y=y),stat="identity")+
  geom_errorbar(aes(x=group,ymin=y-SD,ymax=y+SD),
  width=.2)+
  facet_grid(.~condition)+xlab("")

## analisi statistica
aov.car(y~group*condition+Error(subject/condition),
data=dati)
```

Codice Sweave (1)

```
<<loading,echo=TRUE>>=
library(ggplot2)## libreria per i grafici
library(afex)## librerie per l'ANOVA di terzo tipo
library(xtable)## per tabelle formattate TeX

## creazione dei dati
dati = data.frame(
  y = c(10,33,44,77,52,34,76,21,87,32,15,31,51,16,
        49,52,56,68,26,72,16,13,16,16,56,47,38,19,
        46,35,59,15,74,58,35,62,12,51,45,78,56,27,
        15,30,23,65,46,51,55,48,12,23,59,52,23,90),
  group = gl(2,28,length=56,labels=c("A","B")),
  condition=gl(2,14,length=56,labels=c("I","II")),
  subject = c(rep(1:14,2),rep(15:28,2)))
```

@

Codice Sweave (2)

```
<<datatablevisualization,echo=TRUE,results=tex>>=  
## visualizzazione dei dati  
xtable(dati,caption="Tabella dei dati per esteso")  
  
## controllo se le condizioni sono bilanciate  
xtable(table(dati$group,dati$condition),  
        caption="Tabella bilanciamento condizioni")  
@
```

Codice Sweave (3)

```
<<boxplot,echo=TRUE,fig=TRUE>>=  
## grafico boxplot  
ggplot(dati)+  
  geom_boxplot(aes(x=group,y=y))+  
  facet_grid(.~condition)+xlab("")  
@
```

Codice Sweave (4)

```
<<medie,echo=TRUE,results=tex>>=
## informazioni riassuntive della
## distribuzione: media e dev. standard
sommario=aggregate(y~group*condition,data=dati,
  FUN=mean)
sommario$SD=aggregate(y~group*condition,data=dati,
  FUN=sd)$y

## visualizzazione delle informazioni riassuntive
xtable(sommario,caption="Medie e dev. st.")
@
```


Codice Sweave (5)

```
<<medievisualizzazione,echo=TRUE,fig=TRUE>>=  
## grafico delle medie e deviazioni standard  
ggplot(sommario)+  
  geom_bar(aes(x=group,y=y),stat="identity")+  
  geom_errorbar(aes(x=group,ymin=y-SD,ymax=y+SD),  
    width=.2)+  
  facet_grid(.~condition)+xlab("")
```

@

Codice Sweave (5)

```
<<anova,echo=TRUE,results=tex>>=
## analisi statistica
xtable(aov.car(y~group*condition+
  Error(subject/condition),
  data=dati),caption="Analisi della varianza")
@
```

Abbiamo visto come con Sweave sia possibile integrare  e L^AT_EX.

Questo permette di creare report statistici dinamici, trasparenti e che facilitano la riproducibilità della ricerca.

Grazie dell'attenzione

Per ulteriori informazioni scrivere a:

`michele.scandola@univr.it`

o consultare:

<http://www.stat.uni-muenchen.de/~leisch/Sweave/>
e l'articolo apparso su ArsTEXnica (Crivellaro, 2011).