Using knitr to produce reports from R scripts

David A. Stephens

September 26, 2016

Abstract

This document shows you how to use the knitr package in R or RStudio to create pdf documents based on R scripts.

The file that created this pdf document is knit-01-lm.Rnw, or the equivalent file RStudio-knit-01-lm.Rnw for use in RStudio.

- Using R: when issuing the commands

  ```
  library(knitr)
  knit('knit-01-lm.Rnw')
  ```

  whilst running R in a folder containing the file knit-01-lm.Rnw, the code will produce a Latex file knit-01-lm.tex which can then be compiled using pdflatex.

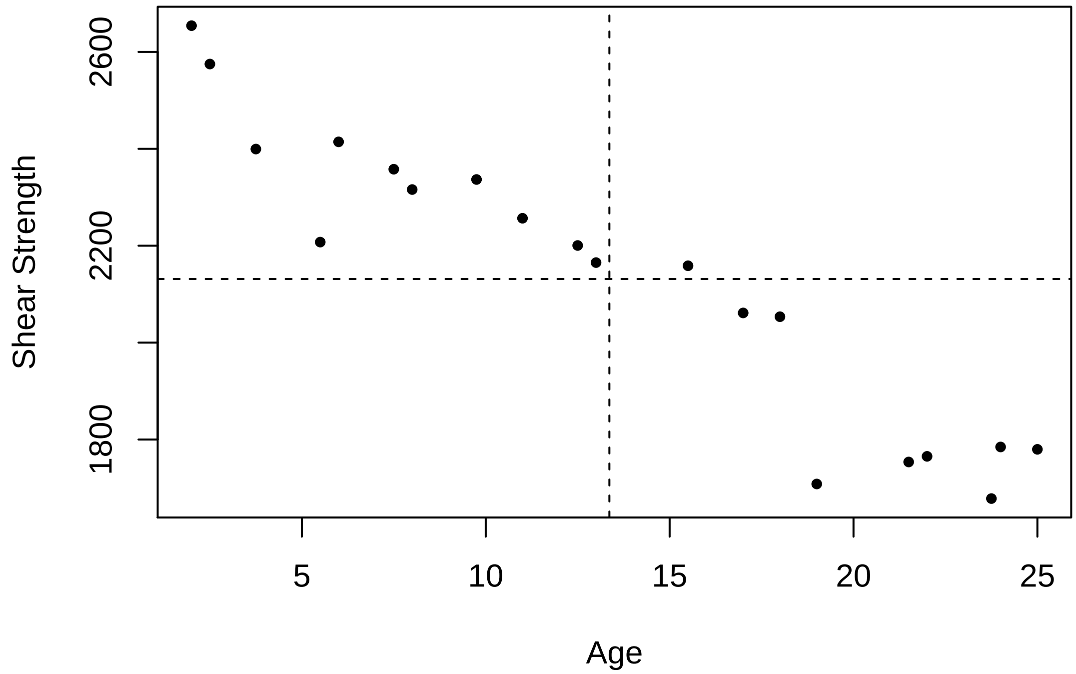
- Using RStudio: you can open the RStudio-knit-01-lm.Rnw file directly in RStudio, and compile it using the 'Compile pdf' button at the top of the R/Sweave script window that opens automatically when you click on the .Rnw file. Note that you should set the file to be compiled correctly by going to the

  Tools ⟶ *Global Options*

  pulldown menu, and hitting the Sweave symbol in the left-hand sidebar. Then from the top pulldown menu *Weave Rnw files using* select the option 'knitr'.

Here is a plot of the rocket propulsion data from the textbook, as studied in class:

```
data.source<-"http://www.math.mcgill.ca/dstephens/Regression/Data/2-1-RocketProp.csv"
RocketProp<-read.csv(file=data.source)
names(RocketProp)<-c('i','Strength','Age')
x<-RocketProp$Age
y<-RocketProp$Strength
plot(x,y,pch=19,cex=0.6,xlab='Age',ylab='Shear Strength')
xmean<-mean(x)
ymean<-mean(y)
abline(v=xmean,h=ymean,lty=2)
```



The mean of the x values is 13.3625, and the mean of the y values is 2131.3575.

```
############################################
#Fit the simple linear regression using lm

fit.RP<-lm(y~x)
summary(fit.RP)

+
+ Call:
+ lm(formula = y ~ x)
+
+ Residuals:
+     Min      1Q  Median      3Q     Max
+ -215.98  -50.68   28.74   66.61  106.76
+
+ Coefficients:
+              Estimate Std. Error t value Pr(>|t|)
+ (Intercept) 2627.822     44.184   59.48  < 2e-16 ***
+ x            -37.154      2.889  -12.86 1.64e-10 ***
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
+
+ Residual standard error: 96.11 on 18 degrees of freedom
+ Multiple R-squared:  0.9018,Adjusted R-squared:  0.8964
+ F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10

coef(fit.RP)

+ (Intercept)           x
+  2627.82236   -37.15359
```

From this, we can deduce that the estimate of the intercept is $\widehat{\beta}_0 = 2627.822359$, and the estimate of the slope is $\widehat{\beta}_1 = -37.1535909$, that is, the line of best fit is

$$y = 2627.822359 + -37.1535909x$$

However, this is reporting the estimates to too many decimal places: we can reduce that by using the `round` function: for example `round(coef(fit.RP)[1],3)` gives the result to 3 decimal places. We have

$$y = 2627.822 + -37.154x.$$

The line of best fit is good (see plot below).

Here is an alternative way to use the `lm` function in `R`, by using the `data=RocketProp` argument; this allows you to call the variables by name in the command

`fit.RP<-lm(Strength Age,data=RocketProp)`

```
fit.RP<-lm(Strength~Age,data=RocketProp)
plot(x,y,pch=19,cex=0.6,xlab='Age',ylab='Shear Strength')
coef(fit.RP)

+ (Intercept)          Age
+  2627.82236   -37.15359

abline(coef(fit.RP),col='red')
title('Line of best fit for Rocket Propulsion Data')
```