R实战第二版笔记-R语言创建动态报告

2020-07-01

Table of Contents

# 加载经常用的R包

# R语言报告

最后一项统计分析或者绘图的完成并不意味着研究过程的完成。你总要与他人交流研究结果。这意味着把分析整理到某种报告里面。

有三种常见的创建报告情景。

* 第一种：创建一个包含代码和结果的报告，便于记住六个月前做过的事情。如果要重做之前的事情，从单个完整的文档做起比从多个相关的文档做起要更加容易。
* 第二种：为老师、主管、客户、政府代表、网络观众或者杂志编辑创建一份报告。你需要注 意清晰性和吸引性，而且这份报告可能只需要创建一次。
* 第三种：为日常需求创建一份特定类型的报告。这有可能是关于产品或者资源使用量的每月 报告，可能是关于金融的每周分析，也可能是关于网络流量的每小时更新一次的报告。每一种情况中，数据会有所变化，但是分析过程和报告结构保持不变。

把R的输出合并到报告的一种方法是：进行分析，复制和粘贴每一个图表到一个字处理文档 中，接着重新整理结果格式。这个方法一般来说非常耗时、低效，让人心烦意乱。尽管R创建的图片很现代，但它的文字输出却很复古——由等宽字体组成并用空格实现列对齐的表格。如果数据有所变化的话，你不得不重复整个过程！

举个例子，假设你在使用回归分析来研究一份女性样本中体重和身高的关系。 R允许你提取 lm()函数的等宽输出：

library(pacman)  
p\_load(tidyverse)  
women %>%   
 lm(weight ~ height,data = .) %>%   
 summary()

##   
## Call:  
## lm(formula = weight ~ height, data = .)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.7333 -1.1333 -0.3833 0.7417 3.1167   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -87.51667 5.93694 -14.74 1.71e-09 \*\*\*  
## height 3.45000 0.09114 37.85 1.09e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.525 on 13 degrees of freedom  
## Multiple R-squared: 0.991, Adjusted R-squared: 0.9903   
## F-statistic: 1433 on 1 and 13 DF, p-value: 1.091e-14

women %>%   
 lm(weight ~ height,data = .) %>%   
 stargazer::stargazer(type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## weight   
## -----------------------------------------------  
## height 3.450\*\*\*   
## (0.091)   
##   
## Constant -87.517\*\*\*   
## (5.937)   
##   
## -----------------------------------------------  
## Observations 15   
## R2 0.991   
## Adjusted R2 0.990   
## Residual Std. Error 1.525 (df = 13)   
## F Statistic 1,433.024\*\*\* (df = 1; 13)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

women %>%   
 lm(weight ~ height,data = .) %>%   
 xtable::xtable(label = "HTML")

## Warning: 'coef'部分匹配到'coefficients'

## # A tibble: 2 x 4  
## Estimate `Std. Error` `t value` `Pr(>|t|)`  
## <dbl> <dbl> <dbl> <dbl>  
## 1 -87.5 5.94 -14.7 1.71e- 9  
## 2 3.45 0.0911 37.9 1.09e-14

# 用模版生成报告

根据你起步的模版文件和用来处理模版的函数，可以创建出不同的报告格式（HTML网页文件、 Microsoft Word文档、 OpenOffice Writer文档、PDF报告、文章和图书）。 它们被称为**动态报告**，动态之处在于改变数据和重新处理模版文件的话会生成一份新的报告。

# 用 R 和 Markdown 创建动态报告

n <- nrow(women)  
fit <- lm(weight ~ height, data=women)  
sfit <- summary(fit)  
b <- coefficients(fit)

Linear regression was used to model the relationship between weights and height in a sample of 15 women. The equation **weight = -87.5166667 + 3.45 height** accounted for 0.99% of the variance in weights. The ANOVA table is given below.

library(xtable)  
options(xtable.comment=FALSE)  
print(xtable(sfit), type="html", html.table.attributes="border=0")

## Warning: 'coef'部分匹配到'coefficients'

Estimate

Std. Error

t value

Pr(>|t|)

(Intercept)

-87.5167

5.9369

-14.74

0.0000

height

3.4500

0.0911

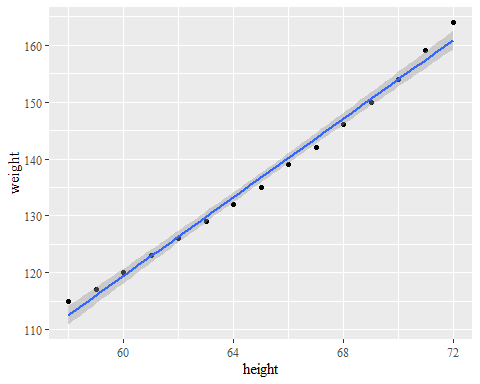
37.86

0.0000

The regression is plotted in the following figure.

library(ggplot2)  
ggplot(data = women, aes(x = height, y = weight)) +  
 geom\_point() + geom\_smooth(method = "lm") +  
 theme(text = element\_text(family = enfont))

## `geom\_smooth()` using formula 'y ~ x'



Markdown语法可以方便地快速创建简单的文件。你可以访问其主页<http://daringfireball.net/projects/markdown/>和[rmarkdown文档页面](http://rmarkdown.rstudio.com/)来学习Markdown。如果想创建复杂的文档，比如说**出版水平的文章和图书**，你可能需要使用LaTeX作为你的标记语言。在下一节，你会使用LaTeX和knitr包来创建高质量排版的文档。

# 用 R 和 LaTeX 创建动态报告

# 用 R 和 Open Document 创建动态报告

# 用 R 和 Microsoft Word 创建动态报告