STAT_757_Assignment1

DUE 2/4/2018 11:59PM

AG Schissler, modified from P. Hurtado

1/24/2018

i. DataCamp: Introduction to R [30 points]

Please complete the course an Introduction to R. You should have received an email with an invitation link. Please email me if you did not. If you already know R, please talk to me in class and follow up with an email to opt out.

ii. Instructions for the rest of this assignment

The purpose of this portion of the assignment is to get a little experience making R Markdown documents as a way of nicely formatting output from R code while exploring the datasets from Sheather Ch.1 and learning to generate realizations of random variables (aka "fake data"). Modify this RMarkdown file (STAT_757_Assignment1.Rmd) and compile your document as a PDF (or Word document if you're having LaTeX issuse) and naming it according to the format SURNAME-FIRSTNAME-Assignment1.pdf, and emailing that PDF to the instructor by the due date listed above.

2. Reproduce the plots from Sheather Ch.1 [40 points]

Modify this file so that it reproduces all the output from the R script located at http://www.stat.tamu.edu/~sheather/book/docs/rcode/Chapter1.R. I've done the plots for the first dataset for you below. Remember that you will need to download each of the four data sets from http://www.stat.tamu.edu/~sheather/book/data_sets.php, and set your working directory (under the "Session" menu in Rstudio) appropriately. (And yes, this really is as easy as copying the blocks of R code for each dataset into this document into the appropriate places!) Need help? First, see http://rmarkdown.rstudio.com. Especially the resources under Learning More (http://rmarkdown.rstudio.com/#learning-more).

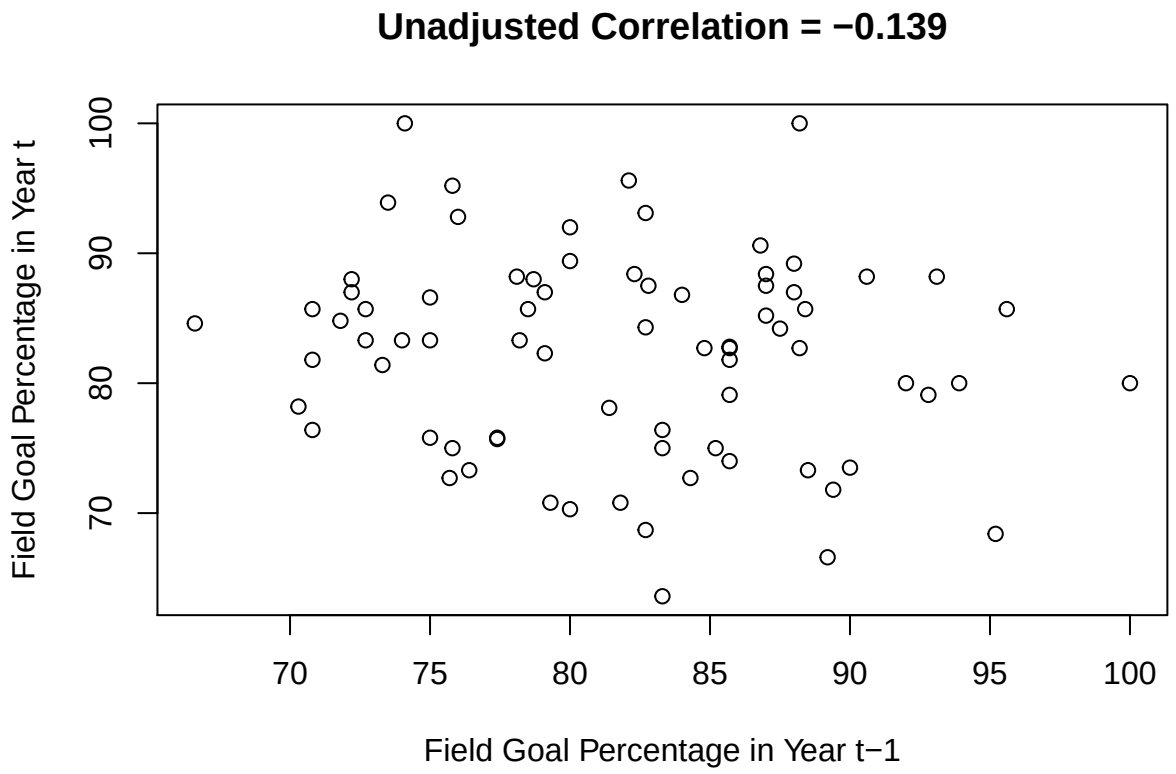
Below are the plots that appear in Chapter 1 of the textbook. They were created from the R script http://www.stat.tamu.edu/~sheather/book/docs/rcode/Chapter1.R and the data files at http://www.stat.tamu.edu/~sheather/book/data_sets.php.

NFL Kicker Data

```
kicker <- read.csv("~/OneDrive - University of Nevada, Reno/Teaching/STAT_757/Sheather_data/Data/FieldG
## Sorry this line is too long, the data are labeled 'FieldGoals2003to2006.csv'

attach(kicker) ## THIS IS NOT USUALLY RECOMMENDED, ASK ME IN CLASS WHY NOT.

#Figure 1.1 on page 2
plot(kicker$FGtM1,kicker$FGt,
main="Unadjusted Correlation = -0.139",
xlab="Field Goal Percentage in Year t-1",ylab="Field Goal Percentage in Year t")
```




```
#p-values on page 3
fit.1 <- lm(FGt~FGtM1 +Name +FGtM1:Name,data=kicker)
anova(fit.1)
```

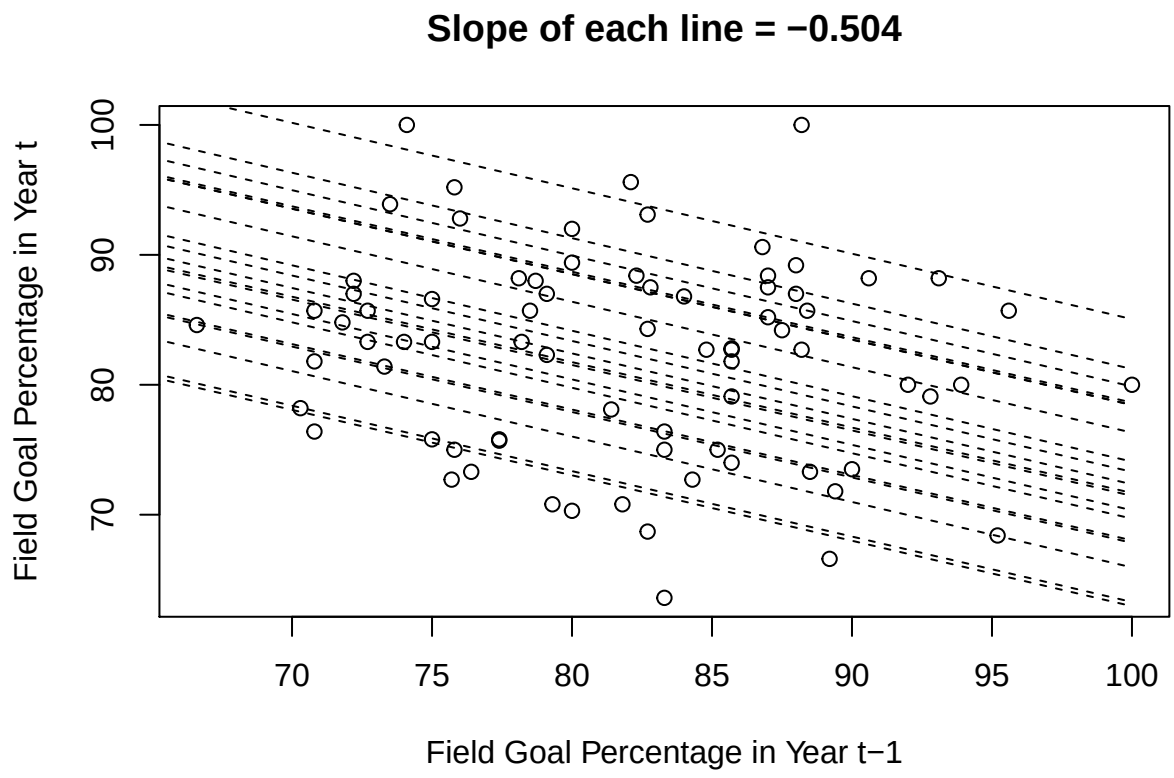
```
## Analysis of Variance Table
##
## Response: FGt
##             Df Sum Sq Mean Sq F value Pr(>F)
## FGtM1        1     87    87.2    1.90 0.1760
## Name        18   2252   125.1    2.73 0.0046
## FGtM1:Name 18    418    23.2    0.51 0.9386
## Residuals  38   1743    45.9
```

```
#slope and interecepts of lines in Figure 1.2 on page 3
fit.2 <- lm(FGt ~ Name + FGtM1,data=kicker)
fit.2
```

```
##
## Call:
## lm(formula = FGt ~ Name + FGtM1, data = kicker)
##
## Coefficients:
##               (Intercept)           NameDavid Akers
##                   126.687                   -4.646
##           NameJason Elam          NameJason Hanson
##                    -3.017                    2.117
##            NameJay Feely             NameJeff Reed
##                   -10.374                   -8.296
##         NameJeff Wilkins           NameJohn Carney
##                     2.310                   -5.977
##            NameJohn Hall            NameKris Brown
```

```
##                   -8.486                  -13.360
##          NameMatt Stover       NameMike Vanderjagt
##                   8.736                    4.896
##         NameNeil Rackers           NameOlindo Mare
##                   -6.620                  -13.036
##          NamePhil Dawson          NameRian Lindell
##                   3.552                   -4.867
##        NameRyan Longwell  NameSebastian Janikowski
##                   -2.231                  -3.976
##        NameShayne Graham                    FGtM1
##                   2.135                   -0.504
```

```
#Figure 1.2 on page 3
plot(kicker$FGtM1,kicker$FGt,
main="Slope of each line = -0.504",
xlab="Field Goal Percentage in Year t-1",
ylab="Field Goal Percentage in Year t")
tt <- seq(60,100,length=1001)
slope.piece <- summary(fit.2)$coef[20]*tt
lines(tt,summary(fit.2)$coef[1]+slope.piece,lty=2)
for (i in 2:19)
{lines(tt,summary(fit.2)$coef[1]+summary(fit.2)$coef[i]+slope.piece,lty=2)}
```




```
detach(kicker)
```

Newspaper circulation

and so on...

3. Generating fake data [30 points]

3.1 Generate 100 random variates from a normal distribution with mean 0 and standard deviation of 100. Summarize and plot the data. (Set a seed to make it reproducible).

3.2 Generate 1000 random variates from a beta distribution with the parameters α and β both equal to 2. Summarize and plot the data. (Set a seed to make it reproducible).

3.3 Generate 10000 random variates from a binomial distribution with the parameters $n = 10$ and $p = 0.2$. Summarize and plot the data. (Set a seed to make it reproducible).