

Homework 1 - Solutions (Template)

Your Name

2018-02-15

Contents

Problem 1 Tables	1
Problem 2 A Regression Example	1
Problem 3 Plots	2

Problem 1 Tables

For this problem I create an artificial data frame:

```
myDataFrame <- data.frame(Names = LETTERS[1:3], Values = runif(3))
myDataFrame
```

```
##   Names Values
## 1     A 0.88189
## 2     B 0.08093
## 3     C 0.75094
```

A better way to present this data

```
myDataFrame <- data.frame(Names = LETTERS[1:3], Values = runif(3))
knitr::kable(myDataFrame)
```

Names	Values
A	0.0611
B	0.6065
C	0.6909

Problem 2 A Regression Example

For this problem I study the relationship between `weight` and `height` using `women` dataset in the base package.

The model is

$$\text{weight}_i = \beta_0 + \beta_1 \text{height}_i + u_i$$

and we would like to obtain the OLS estimates of β_1 and β_2 .

Let's run the regression and save the output as an object called `myRegression`:

```
myRegression <- lm(weight ~ height, data = women)
myRegression
```

```
##
## Call:
```

```
## lm(formula = weight ~ height, data = women)
##
## Coefficients:
## (Intercept)      height
##      -87.52         3.45
```

Therefore, we obtain the following estimates: $\hat{\beta}_1 = -87.5167$ and $\hat{\beta}_2 = 3.45$.

and hence the sample regression line is:

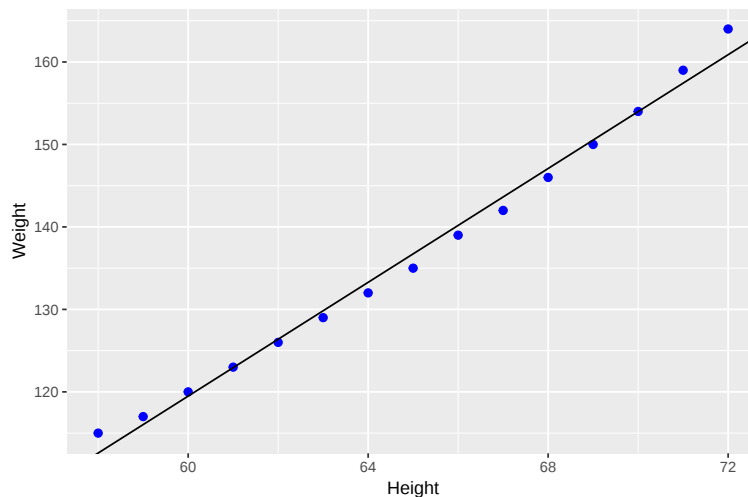
$$\widehat{\text{weight}}_i = -87.5167 + 3.45 \widehat{\text{height}}_i$$

Please look at R markdown file carefully. I didn't *manually* type the numerical values of the estimates!!!

Also, look at how the following sentence is written in the .Rmd file: `women` dataset includes 15 observations.

Finally, let's depict the fitted line

```
myReg <- lm(weight ~ height, data = women)
b_0 = coef(myReg)[1]
b_1 = coef(myReg)[2]
ggplot(data=women, aes(x=height, y=weight)) +
  geom_point(color="blue", size=2) +
  geom_abline(intercept = b_0, slope = b_1) +
  labs(title="", y="Weight", x="Height")
```

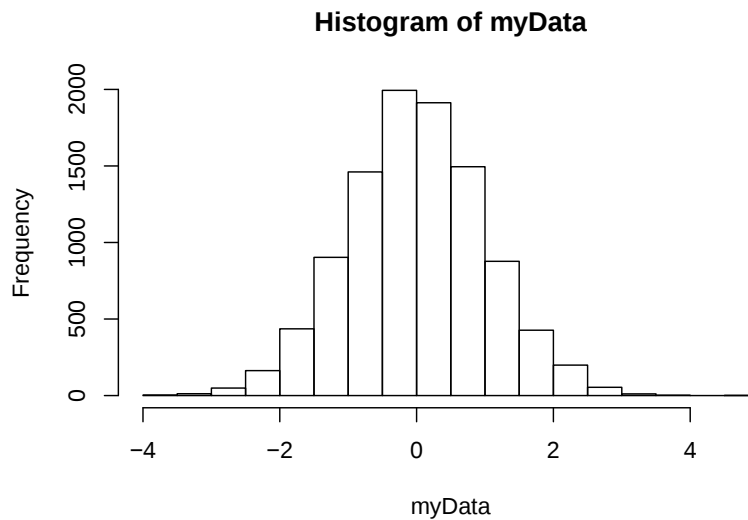


Problem 3 Plots

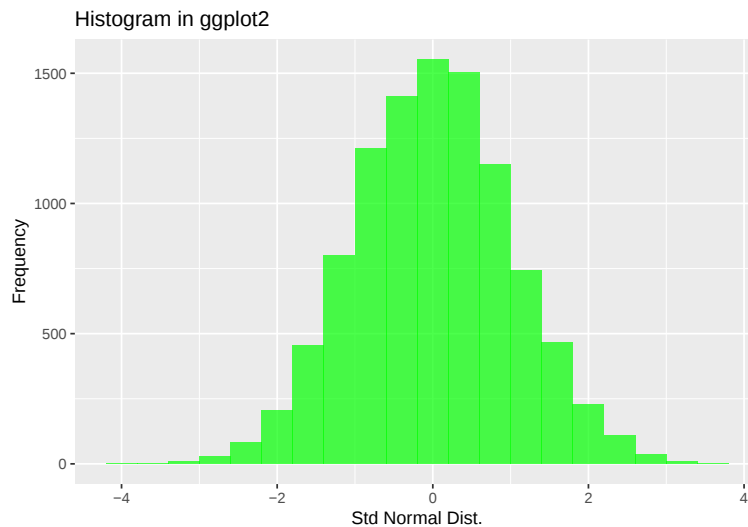
For this problem I make a few simple plots.

3.a. A histogram

```
# note that we are overwriting fig.asp option
myData <- rnorm(10000)
hist(myData)
```



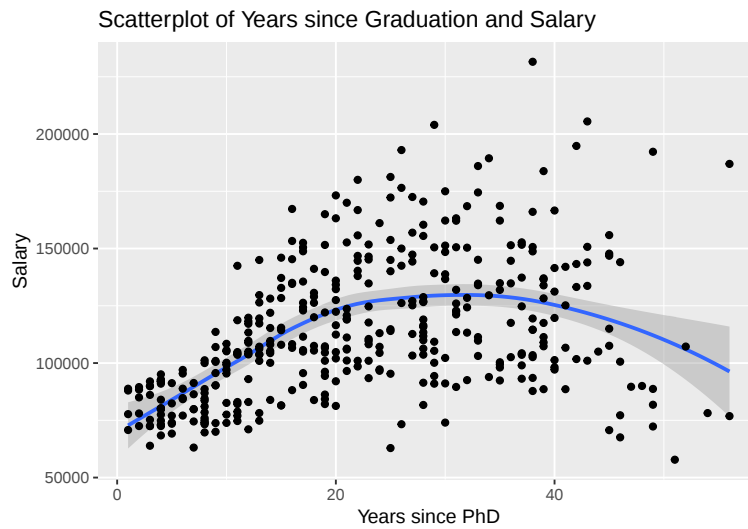
```
myData <- data.frame(myData=rnorm(10000))
ggplot(myData, aes(x=myData)) +
  geom_histogram(binwidth = 0.4, fill="green", alpha=0.7) + # produces a histogram
  labs(title="Histogram in ggplot2",
       x="Std Normal Dist.", y="Frequency")
```



3.b. A scatterplot

A scatterplot using ggplot2 and car dataset:

```
library(ggplot2)
library(car)
data(Salaries, package="carData")
ggplot(data=Salaries, aes(x=yrs.since.phd, y=salary)) +
  geom_smooth() +
  geom_point() +
  labs(title="Scatterplot of Years since Graduation and Salary",
       x="Years since PhD", y="Salary")
```



3.c. An enhanced scatterplot

Now, using `mtcars` dataset I plot an enhanced scatterplot. Note that code is not shown in the final report. Look at the source code to see how to do this by setting `echo=FALSE`.

