Statistics on the grades

July 17, 2013

What you see in this pdf is actual output from the R code that is printed, using your actual grades.

1 Preparing R

The following code and output prepares the R workspace, loads plotting functionality, and loads the data. The data is confidential, so you cannot walk through this code as an exercise. However, you can learn from the code, its output, and the comments.[1]

```
> ## Clear the workspace
> rm(list=ls())
> ## Load a package for making plots
> require(ggplot2)
> ## Load the data
> df.grades <- read.table("333Analyze.txt",header=TRUE)
```

Some of this data is confidential, and we will only be working with the variable "Midterm.grade".

2 Examining the midterm data

Next, we have a look at our data, which should answer most of the questions that students usually ask me about.

```
> ## A fast way to get some information is to issue "summary"
> summary(df.grades$Midterm.grade)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   13.00   14.50   14.36   16.50   20.00

> ## Next, we use specific commands to get some quantities we are particularly interested
> ## Most of this is not necessary, as "summary" has given us most info.
> ## The option "na.rm=TRUE" in the code below asks R to omit the
> ## ... two students who received an NA because they were unable to attend
>
> ## Sample mean
> mean(df.grades$Midterm.grade,na.rm=TRUE)
```

---

[1]There has been an upgrade to the grades since I posted the original document.

```
[1] 14.35897

> ## Standard deviation, variance
> sd(df.grades$Midterm.grade,na.rm=TRUE)

[1] 3.282232

> var(df.grades$Midterm.grade,na.rm=TRUE)

[1] 10.77304

> ## Median, and the first and ninth decile
> quantile(df.grades$Midterm.grade,c(0.1,0.5,0.9),na.rm=TRUE,names=TRUE)

 10%  50%  90%
10.5 14.5 18.0

> ## The maximum?
> max(df.grades$Midterm.grade)

[1] 20
```

What do we learn from this? There are two NA's. The median and the mean are very close, and are at about 14.5. Also, at least one person got the full 20 points! Overall, you did very well. Or: I made the midterm too easy.

Finally, let's make a picture.

```
> qplot(x=Midterm.grade,data=df.grades,binwidth=1) + xlim(5,20) + theme_bw() +
+   xlab("Grade, out of 20") + ylab("Number of students, out of 146") +
+   geom_vline(xintercept=median(df.grades$Midterm.grade,na.rm=TRUE),colour="red")
```
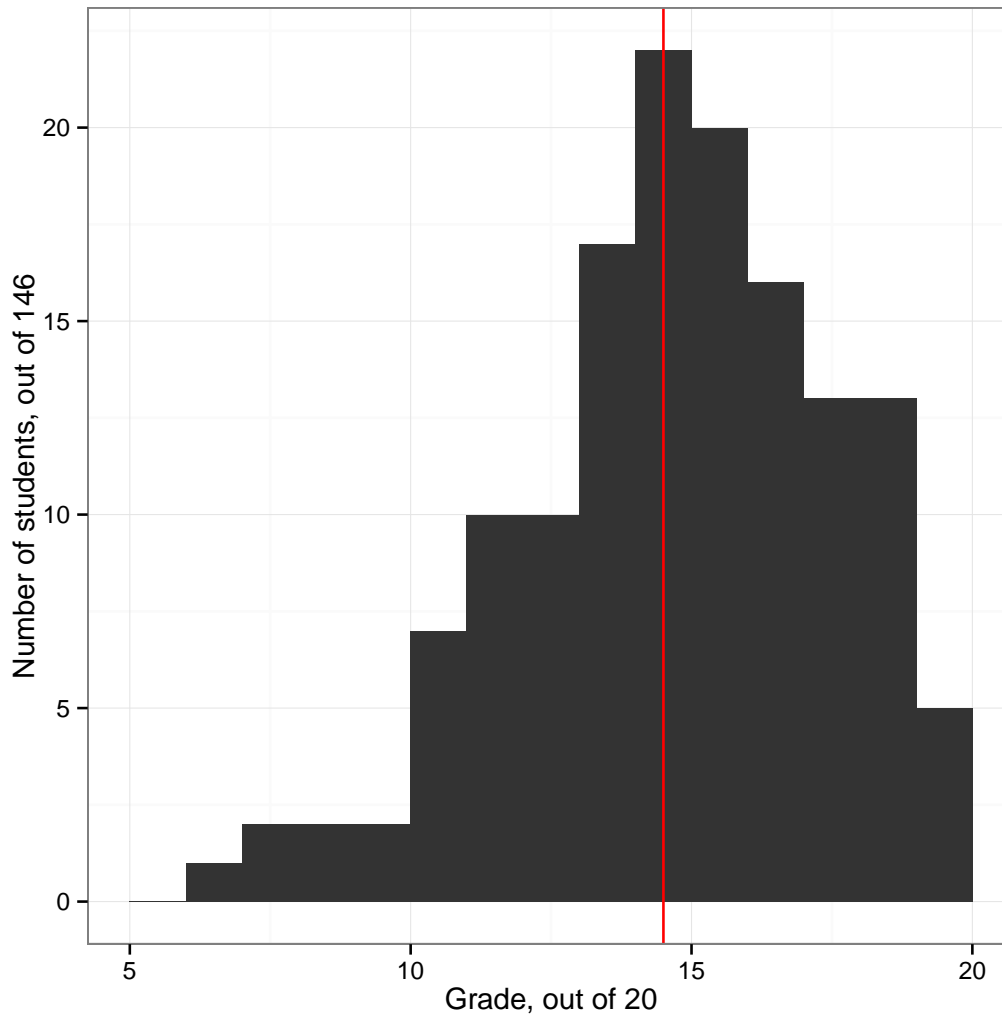
3 Hand-in assignment 1

We do the same analysis as before, but now for the hand-in assignment.

```
> # Make a picture
> qplot(x=HW1,data=df.grades,binwidth=1) + xlim(5,20) + theme_bw() +
+        xlab("Points") + ylab("Number of students, out of 145") +
+        geom_vline(xintercept=median(df.grades$Midterm.grade,na.rm=TRUE),colour="red")
> # Some statistics students usually ask me about
> mean(df.grades$HW1,na.rm=TRUE)

[1] 14.65862

> sd(df.grades$HW1,na.rm=TRUE)

[1] 2.271276

> var(df.grades$HW1,na.rm=TRUE)
```
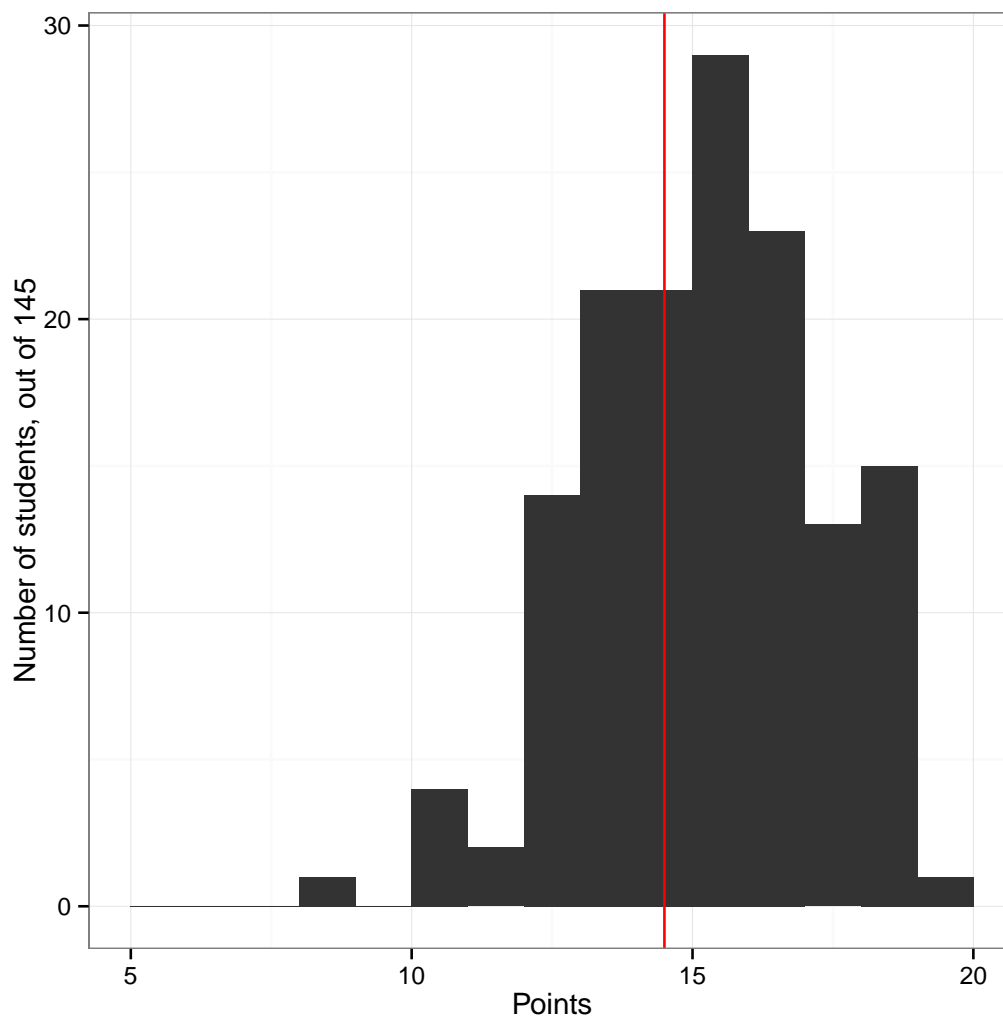
```
[1] 5.158693

> quantile(df.grades$HW1,c(0.1,0.5,0.9),na.rm=TRUE,names=TRUE)

10% 50% 90%
 12   15   18

> max(df.grades$HW1)

[1] 19
```
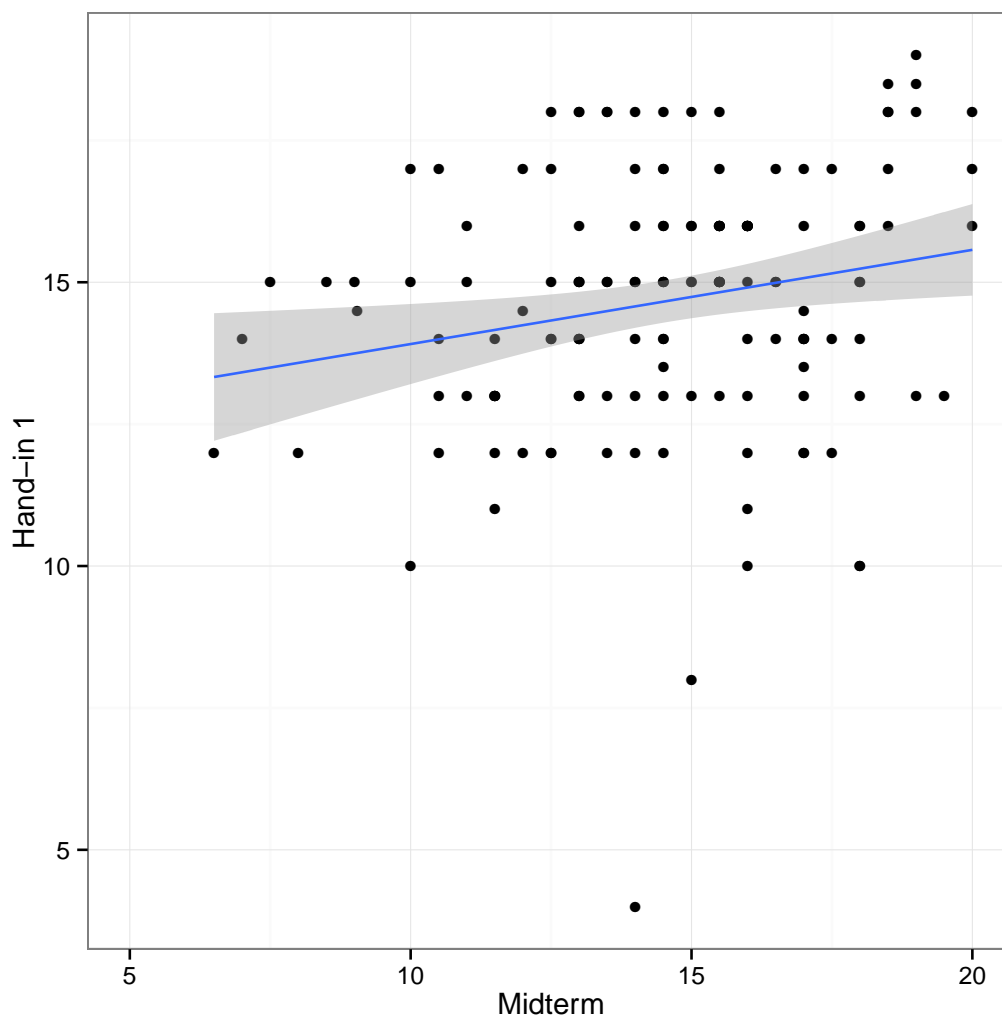

4   A regression exercise

First, let's draw the scatterplot for the points on the two parts of this course. The scatterplot also has the SRF plotted along with it.

```
> qplot(x=Midterm.grade,y=HW1,data=df.grades) + theme_bw() +
+          xlab("Midterm") + ylab("Hand-in 1") + xlim(5,20) +
+          geom_smooth(method="lm")
```

To see the results in the familiar regression framework, consider the following regression output:

```
> summary(lm(HW1~Midterm.grade,data=df.grades))

Call:
lm(formula = HW1 ~ Midterm.grade, data = df.grades)

Residuals:
    Min      1Q  Median      3Q     Max
-10.6107 -1.2767  0.2557  1.2557  3.7214

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.74038    0.83617  15.237   <2e-16 ***
Midterm.grade  0.13359    0.05678   2.353     0.02 *
---
Signif. codes:  0
```