Statistics on the midterm grades

June 26, 2013

For this document, I wrote text and code in one document (using LYX and Sweave). What you see in this pdf is actual output from the R code that is printed. Perhaps this helps you in the last strides towards completing your hand-in assignment 1.

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.csv("buec333_summer2013_grades_MIDTERM_WEBCT.csv")
```

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

2   Examining the 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.    NA's
   5.50   12.00   14.00   13.88   16.00   19.00       2

> ## 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
```

---

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

```
>
> ## Sample mean
> mean(df.grades$Midterm.grade,na.rm=TRUE)

[1] 13.88095

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

[1] 2.874864

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

[1] 8.26484

> ## 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 14.0 17.5
```

mean are very close, and are at about 14/20. Nobody managed to get the full 20 points, **even** with the bonus question!
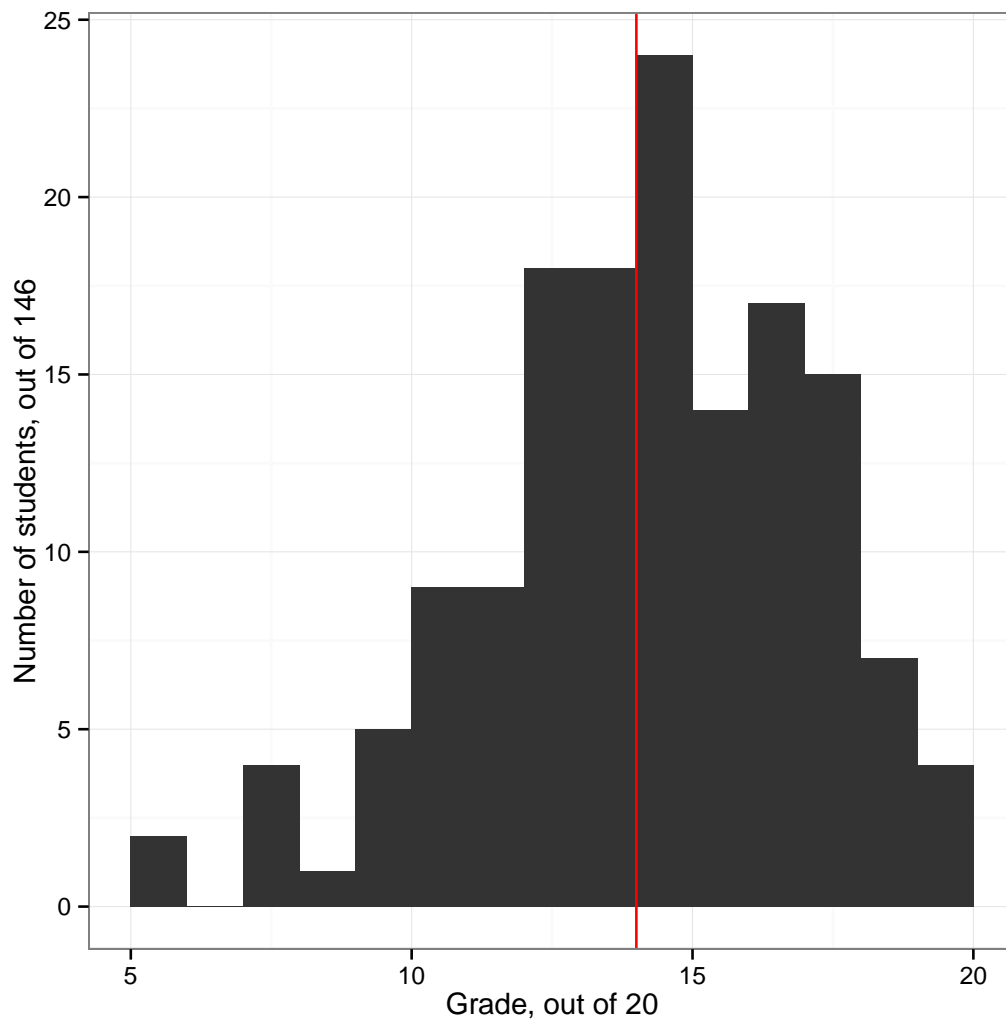
3   Making a picture

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")
```

4 Population / sample

In this exercise, what population is the sample drawn from? How confident can we be that the true mean μ of the midterm grade is in the range, say, $[13, 14]$?