

BUEC-333, Summer 2013

Hand-in assignment #1

Grade distribution and tiny solutions

Questions

First, download the data set “testscores_california_1999.csv”, from the course website. There, you also find a description of the data. Second, reread the slides and notes of Lecture 1.

1. Install, then start RStudio.¹ Set your working directory to the folder where you saved your data. Now, load and inspect the data. You could, for example, cut and paste the following commands

```
#Read the data file and save the results
csdata <- read.csv("testscores_california_1999.csv")
#Have a quick look at the data
#Just to check that the data import went well
#First 6 lines
head(csdata)
#Last 6 lines
tail(csdata)
#Display a summary of the data
summary(csdata)
```

From the output from that last command, you can read the maximum value for “comp_stu”. What is that value? What does it mean? Are you surprised? **1 point. You’d expect them to write something like: this is long ago, so there were not a lot of computers (per student) yet. Any reasonable answer goes.**

2. Give the sample mean, minimum, maximum, and sample standard deviation of read_scr.² **1 point.**
3. Is the sample standard deviation you computed under (2) an estimand, an estimator, or an estimate? **1 point. Estimate. No excuses!**

¹More information about R and RStudio at this section of the course website: <http://www.sfu.ca/~cmuris/2013-Summer-333/index.html#R>

²Remember: Always provide a short, written, answer; write/type the R code you used to obtain the answer; and attach a hard copy of the R output that supports your written answer. Exception: purely theoretical questions.

4. Is the number you got a random variable? **1 point. No! Once you have the number, it is not random anymore. Basically the same as 3.**
5. Construct a 95% confidence interval for the mean of “read_scr”. Also, construct a 99% confidence interval for the mean of “avg_inc”. **2 points.**
6. For the previous question, what is the population you chose? **1 point.**
7. From question (5): What is the interpretation of the 99% confidence interval that you computed? **2 points. A very wrong answer (0 points): “There is a 99% probability that the true parameter / μ lies in the interval. Perfect answer: “If you repeat this procedure with 100 different samples, then the true parameter will lie in 99 of those intervals, on average.” Or something that uses the word “coverage probability”. 1 point: something in between, far enough away from the 0-point answer.**
8. Estimate the coefficients in the linear regression of “read_scr” on “str”. Include a constant. Report the coefficient estimates and standard errors, and interpret the coefficient estimates. **2 points. 1 point for the estimate and standard error. 1 point for the interpretation. 0 points if that interpretation is not precise.**
9. Construct a 95% confidence interval for the regression coefficient of “str”. **1 point.**
10. You expect the coefficient of str to be negative. Formulate an appropriate null and alternative hypothesis; formulate a decision rule (use significance level of 5%); use R to compute the necessary values; draw the conclusion. **2 points. 1 for setting up the hypotheses, and drawing the conclusion. 1 point for all the mechanics in between.**
11. What is the interpretation of the conclusion of the test in (10)? What do you conclude about the effect of class size on student test performance? **1 point for a reasonable answer**
12. There other variables that could have an effect on student test scores. List two or three such variables in this data set, and explain why they could have an effect on test scores, Do you expect the effects to be positive or negative? **2 points for an answer that is very good. 1 point for somebody who mentions reasonable variables but does not give an explanation why they would matter.**
13. Estimate the coefficients in the linear regression of “read_scr” on “str” and the variables you came up with under (12). Include a constant. Report the coefficient estimate and standard error for “str”. **1 point.**
14. For bonus points, explain (or speculate about) the difference between your answer under (13) versus (8). **1 bonus point for “omitted variables” or something that smells like it. (I only started talking about OV this week. That’s why it is a bonus question.)**
15. If you could gather additional data to answer this question in a better way, what data would you gather? **1 point max.**