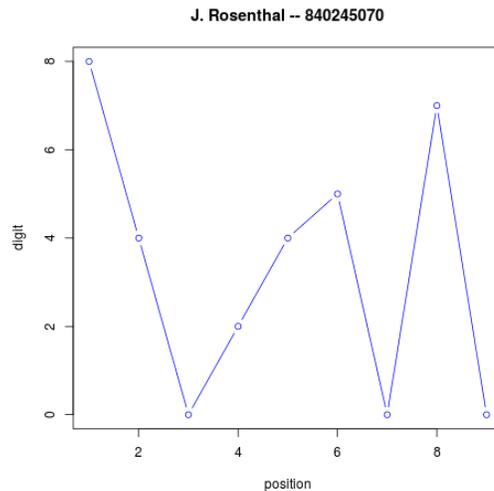


STA130 Homework #1: Studies, R, and Student Survey Data

(Due in tutorial at 2:10 pm sharp on Wednesday, February 1, 2017.)

NOTE: You may discuss this assignment with other students, but you must understand and type and create your solutions entirely by yourself. You can also ask your TA for some assistance, e.g. at the end of tutorial on January 25, or by arranging to meet them separately.

1. Find one interesting “study” which was reported recently in a news article. Then:
 - a) Specify what news article you found it in.
 - b) Explain what issue the study was investigating, and why you find it interesting.
 - c) Explain (briefly) how they conducted the study, and what data/facts/figures they obtained, and what conclusions they formed.
 - d) Discuss whether or not you are convinced by their conclusions, and why.
2. Read the information at www.probability.ca/Rinfo.html about the “R” free statistical software package.
3. Following the instructions on that page, either download R to your own computer, or find a way to use R on campus or online.
4. Use R to add up the digits of your U of T student number. Show the R command used.
5. Use R to create a plot of the digits of your U of T student number, with header stating your name and student number. Show your plot, and the R commands you used. For example, for the instructor’s student number of 840245070 (from when he was an undergraduate student at U of T back in the 1980s), the plot should look something like this:



6. Download and examine the results of our class’s student survey, available at:
<http://probability.ca/sta130/StudentSurvey.csv>
(The questions are listed on lecture slide 17. For Q6, all heights are given in cm.)

CONTINUED ON NEXT PAGE

7. Read the student survey data into R, using a command like:


```
sdata = read.csv("StudentSurvey.csv", as.is=TRUE)
```
8. Then, try each of the following four R commands, and explain in words what each of them is doing: `sdata$Q6` and `mean(sdata$Q6)` and `sdata$Q6[sdata$Q1=="F"]` and `mean(sdata$Q6[sdata$Q1=="F"])`
9. Use R to compute the average height of Female students, and the average height of Male students. How do they compare? Do you think the difference is significant*, or just sampling luck? Why?
10. Try each of the following two R commands, and explain in words what each of them is doing: `table(sdata$Q12)` and `table(sdata$Q12[sdata$Q2=="canada"])`
11. Use R to compute the fraction of students born in Canada who have a romantic partner, and the fraction of students born in China who have a romantic partner. How do they compare? Do you think the difference is significant*, or just sampling luck? Why?
12. Think of two additional interesting comparisons that you can make using the student survey data. For each of them, explain the comparison, and state how they compare in the data, and say whether you think the difference is significant* or just sampling luck, and why.
13. Use R (including the command `pbinom`) to compute a P-value, based on the student survey data, for the hypothesis that “more than 60% of students live off campus”. (So, here the null hypothesis is “60% of students live off campus”, and the alternative hypothesis is “more than 60% of students live off campus”.) Be sure to explain your steps, and what conclusions you reached. (Note that in this question, unlike the previous questions, you do need to compute an actual P-value, not just give your opinion.)

Print your answers to questions 1, 4, 5, 8, 9, 10, 11, 12, and 13 on paper, and hand them in to your TA at the start of tutorial at 2:10 pm on Wednesday February 1. (There will be penalties if it is even one minute late!)

*In this question you do not need to compute a P-value, you just need to give your opinion (with reasons) for why you think it is or is not significant.