MPCS 53110 Foundations of Computational Data Analysis — Winter 2017

Homework 1. This problem set is due Tuesday January 10 at 5:30 pm.

Reading: Probability & Statistics with R = P&S with R, chapter 1, sections 1.1– 1.5; chapter 2; chapter 3, sections 3.1–3.2.

Written assignment

- Solve the following "DO" exercises and homework problems "HW".
- "DO" exercises are strongly recommended to check your understanding of the concepts. **Do not submit them**.
- **Problems labeled "HW" must be submitted**.
- You are responsible for the material covered in **both** "DO" exercises and HW problems.

"Do" Exercises (*not* to be submitted):

This week's "Do" exercises are from the textbook: Probability & Statistics with R = P&S with R

1. **"DO"** Exercises 1.5.4, 1.5.13, and 1.5.15, on pages 20–21.

2. **"DO"** Exercise 1.8.10, on page 48.

3. **"DO"** Exercise 2.2.9, on page 60.

4. **"DO"** Exercises 2.3.2 and 2.3.5 on page 71.

5. **"DO"** Exercises 2.4.3 and 2.4.11, on pages 77–79.

6. **"DO"** Exercises 2.5.1 and 2.5.6, on pages 88–89.

7. **"DO"** Exercise 2.6.6, on page 96.

8. **"DO"** Exercises 3.2.1, 3.2.7, and 3.2.9, on pages 109–110.

Homework Problems (to be submitted Tuesday January 10 at 5:30 pm):

- **Collaboration policy: If you work with others, indicate their names as part of your submission. You must answer each question by yourself without assistance. It is a violation of this policy to submit a solution that you cannot explain orally to the instructor/TAs.**
- **Looking for solutions to problems on the internet, even when acknowledged, is STRONGLY DISCOURAGED.**
- **Write out your work for every "theory" problem. If you just write your answer without showing your work, you will not receive credit.**

1. **HW** Consider a triangle and a point chosen at random within the triangle (with uniform probability). Let X be the distance from the point to the base of the triangle. Given the height of the triangle, find the CDF and the PDF of X. Show your work and explain your answer. (4 points)

2. **HW** P&S with R exercises:
   - 3.2.3 (4 points)
   - 3.2.5b (4 points)

- 3.2.6 (3 points)
- 3.2.8 (6 points)
- 3.2.11 (6 points)

on pages 109–110.

3. **HW** Write an R function that, given two numeric vectors of the same length, computes their dot product ...
   - ... using a for-loop to keep track of a running total.
   - ... in one line, using component-wise vector multiplication (*).
   - ... in one line, using matrix multiplication (%*%). (Make sure you do not return a 1×1 matrix.)

   You may assume all inputs are of the correct format.
   (2 points for each part)

4. **HW** This problem uses the "mtcars" dataset built into R. Each observation (row) corresponds to a car, and each variable (column) corresponds to attributes of the cars. For each of the following, give a single command that does the specified task. For details on what the variables mean, enter "?mtcars".
   - Find the mean mpg of all cars with 4 cylinders.
   - Draw a bar graph indicating the frequency of each number of cylinders among the dataset.
   - Draw a scatterplot of mpg versus weight.
   - Find the proportion of cars with manual transmission.

   (2 points for each part)

5. **HW** Discover some interesting statistical difference in the "mtcars" dataset between cars with automatic and manual transmission. For example, we see a large difference in average fuel efficiency between automatic and manual cars:

```
> mean(mtcars[mtcars$am == 0,]$mpg)
[1] 17.14737
> mean(mtcars[mtcars$am == 1,]$mpg)
[1] 24.39231
```

   Give the commands you used to find it. (6 points)

Gerry Brady, David Ramsey, Liwen Zhang, and Colin Hudler
Wednesday January 4 12:02:23 CST 2017