Homework 1

DS502/MA543

So that everyone in the class has the maximum access to background material Homework 1 will focus on problems from the book. Your answers will be graded both on your **demonstrated understanding** of the concepts **from the book** and **from the class,** as well as the **clarity of your explanations**.

For example, given a question such as:

Describe the null hypothesis to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of **sales***,* **TV***,* **radio***, and* **newspaper***, rather than in terms of the coefficients of the model.*

An **unacceptable**, though perhaps strictly true, answer would be:

It appears the TV and Radio matter more than newspaper.

An **acceptable** answer would be:

The small p-values in Table 3.4 for TV and radio indicate that, holding the other predictors constant, each of these predictors will, with high likelihood, change the amount of sales. For example, fixing the amount of TV and newspaper advertising, while increasing the amount of radio advertising, will very likely lead to an increase in sales because the p-value in Table 3.4 is very small for radio and radio's coefficient is large and positive. On the other hand, since the p-value for newspaper is quite large, the data indicates that newspaper advertising in unlikely to have any effect on sales when TV and radio are held fixed. Note, Table 3.3 does show a small p-value for a single linear regression across newspaper and sales, but this is likely because newspaper advertising is predictive of TV and radio advertising, not necessarily because newspaper advertising directly influences sales.

Please bring the written portion of your answers in **hard copy** to class on Thursday, February 12. Please make sure that both team members names appear on the submission and that every submission is **stand-alone (i.e., does make the grader read or run your code)**! In addition, for any problem on which you write code, plus submit your code **by email** to the TA Jiani Yin (jianiyin@wpi.edu) **before the start of class** on Tuesday September 23. You will **not be graded your programming style**, but having access to your code will allow us to more easily give **partial credit**. You are also welcome to use any of the R scripts in the book and/or those I provide in class to help you with the homework assignment. To keep things organized, please send your R files to Jiani using the following naming convention:

<Last name person 1>_<First name person 1>_<Last name person 2>_<First name person 2>_HW1_<question number>.R

For example, if Jiani and Randy were a team, the file containing the code for problem 4 would be named:

Yin_Jiani_Paffenroth_Randy_HW1_4.R

Homework questions

1. (10 points) Section 2.4, page 52, question 1

2. (10 points) Section 2.4, page 52-53, question 3

3. (5 points) Section 2.4, page 53, question 6

4. (15 points) Section 2.4, page 54-55, question 8 (The R scripts I provide in class should help you get started with this question.)

5. (10 points) Section 2.4, page 56, question 9

6. (5 points) Section 3.7, page 120, question 1 (Yes, this is the question I provided an answer to above. Feel free to use my answer as a starting point, and perhaps add a bit more detail about the effect of TV on sales.)

7. (10 points) Section 3.7, page 121, question 5

8. (5 points) Section 3.7, page 121, question 6

9. (15 points) Section 3.7, page 121-122, question 8

10. (15 points) Section 3.7, page 122, question 9 (Note, this question is the most open ended of the R programming questions, in particular parts (e) and (f). The book will provide insights into what needs to be done here. See how far you can get.)