**STATISTICS 110/201, FALL 2017**
**Homework #3**
**Assigned Wed, October 18, Due Wed, October 25**

**Note:** Some students requested homework that isn't so "R intensive." This assignment requires less use of R (although still some), and more use of understanding how things fit together, reading R output, etc.
**For Exercises 1 to 3:** In last week's homework you examined the data on textbook prices shown in exercise 1.26 (***TextPrices***), and identified an outlier, with 400 pages and price of $128.50. For Exercises 1 to 3 below, you will use the original data set, as well as a new data set with the outlier removed.

1. Create a new data set called *NewPrices* by removing the row with the outlier, which is row 4. You can do this in R as follows:
   ```
   > NewPrices<-TextPrices[-4,]
   ```
   a. Use the `summary` command in R to provide summaries of the two variables for the original data set and the new data set. Show the results from R.
   b. What is the mean number of pages for the data without the outlier? If a book could have that number of pages, what would be the predicted price, using the least squares regression line obtained with the NewPrices data? [Hint: You don't need to carry out the regression to answer this.]

2. Find $R^2$ using the TextPrices data and again using the NewPrices data.
   a. Report both values of $R^2$ and interpret the one for NewPrices.
   b. Based on $R^2$, explain which data set does a better job of predicting price from number of pages.

3. Exercise 2.15 on page 83 of the book asks you to use the TextPrices data. Do that exercise with the following modifications:
   - Use the NewPrices data, not the original data, thus conducting the analyses on the 29 books without the outlier.
   - In some editions of the textbook, part (b) has a typo. It should say "Determine a 95% prediction interval…" not a 95% confidence interval (which is what was done in part a).
   - In parts (a) and (b), after you compute the intervals, *interpret* what they represent in the context of this situation (textbook with 450 pages).

4. Do Exercise 2.10 on page 82.
5. Do Exercise 2.17, parts (b) and (c) only. (Exercise is on page 83)
6. Do Exercise 2.44 on pages 93-94. (No computer required!)
7. For this exercise, you will use the Skin Cancer example shown in class on Oct 16, and linked to the course webpage for that day. Use the regression output relating X = latitude and Y = skin cancer mortality to answer the following. (No computer required, except to find t* in part b. But doing this all by hand will make you appreciate the computer!)
   a. Give numerical values for each of the following:
      i. $\hat{\beta}_1$
      ii. Standard error of $\hat{\beta}_1$
      iii. MSE
      iv. SSX (Hint: You can find this using values you just gave above and a formula given in class.)
      v. The predicted skin cancer mortality for Irvine, with latitude = 33.7.
   b. Using relevant formulas (and not the computer; see notes or page 77 of text), find a 95% confidence interval for the mean skin cancer mortality rate for all locations with latitude = 33.7. You may use the computer to find the multiplier *t**. The only additional information you need, not provided in the output on the website, is the sample mean latitude, which is 39.53.