

Possible points:95

All data files will be located on the Canvas website. The data files can be relocated to your local working directory and read as follows:

The data here are just text files. Load the data into STATA (note: first line of files contain variable names):

```
. insheet using filename, clear
```

...or load the data into R:

```
> mydata = read.delim("filename")
```

1. [16 pts total] Triglyceride and waist circumference data

Using the regression model output of SPRM Table 2.3 (to answer, you don't need the data, just this table), do the following:

- (a) [4 pts] test $H_0 : \beta_1 = 5$ versus $H_1 : \beta_1 \neq 5$, using significance level $\alpha = 0.1$.
- (b) [4 pts] $H_0 : \beta_0 = -150$ versus $H_1 : \beta_0 \neq -150$, using significance level $\alpha = 0.1$.
- (c) [8 pts] construct the 99% confidence interval for β_1 . How is this possibly more useful than just testing whether β_1 is equal to some specific value?

2. [20 pts] Triglyceride and waist circumference data (cont.)

The data from the SPRM example (Table 2.2) are on the website under `triglyc.txt`. Do the following:

- (a) [8 pts] Create the scatter plot and compute the correlation between the two variables. Run the basic regression model. What is the qualitative relationship between the correlation coefficient and the regression slope?
- (b) [6 pts] Write out what is being tested about the slope (β_1) and state the result of the test. Show how you would test (a) using the computer.
- (c) [6 pts] Use the model to predict (i.e., generate \hat{y}). Calculate the residuals ($y - \hat{y}$) and plot these values against the X variable waist circumference. Does this plot look reasonably like it should? What properties should it have?

3. [10 pts, 5 pts each] Exercise 2.7, (Chatterjee and Hadi text)

The Ascombe example data (shown in Jan 9 notes) are stored in the file `anscombe.txt`. Choose the *first* (y_1 and x_1) and one other of the datasets (data are in X, Y pairs for each of the 4 cases) and complete the following:

- (a) verify that $\hat{\beta}_0$, $\hat{\beta}_1$, $\text{Cor}(Y, X)$, and R^2 are the same between data sets - i.e. run models to confirm -

(b) plot/summarize the residuals for the two datasets chosen. How do these differ?

Note: The data are reported to two decimal places, so you cannot expect accuracy to four decimal places in your results. So, there will be slight differences in the values.

4. [24 pts total] Modified from Chattergee and Hadi text

One may wonder if people of similar heights tend to marry each other. For this purpose, a sample of newly married couples was selected. Let X be the height of the husband (in cm) and Y be the height of the wife (in cm). The data on heights of husbands (mht) and wives (fht) can be found online in the file `heights.txt` (**please note** this ‘old’ data far predates contemporary thinking regarding marriage, and also comes from a time of (more) differential societal roles of men and women. Please do not be offended).

(a) [2 pts] Compute the covariance between the heights of the husbands (males in this data) and wives (females in this data).

(b) [2 pts] What would the covariance be if the heights were measured in inches rather than in centimeters?

(c) [2 pts] Compute the correlation coefficient between the heights of the husbands and wives.

(d) [2 pts] What would the correlation be if the heights were measured in inches rather than in centimeters?

(e) [3 pts] What would the correlation be if every man married a woman exactly 5 centimeters shorter than him?

(f) [2 pts] We wish to fit a regression model relating the heights of husbands and wives. Which one of the two variables would you choose as the response variable? Justify your answer (this is somewhat subjective, depends on what you are studying).

(g) [4 pts] Using the model described in 2.10(f), test the null hypothesis that the slope is zero. If we fail to reject the null hypothesis, what does that imply?

(h) [2 pts] By now, we really should have plotted the data. Plot and comment.

(i) [5 pts] What hypotheses and tests would you choose to test whether people of similar heights tend to marry each other (i.e., tall men marry tall women, and short men marry short women, conditional on the fact that the tall women are, on average, shorter than tall men)? What is your conclusion?

5. [25 pts, 5 pts each] SPRM text, Chapter 1

Please complete the practice exercise at the end of the chapter. You will need to enter the data.