Simple Linear Regression
Homework Problems

Rob McCulloch

1.1. Problem: SLR Model

Suppose we are modeling house price as depending on house size. Price is measured in thousands of dollars and size is measured in thousands of square feet.
Suppose our model is:

$$P = 20 + 50s + \epsilon, \epsilon \sim N(0, 15^2).$$

That is, suppose that somehow we *know* the parameters: $\beta_0 = 20$, $\beta_1 = 50$, and $\sigma = 15$.
(a)
Given you know that a house has size s = 1.6, give a 95% predictive interval for the price of the house.
(b)
Given you know that a house has size s = 2.2, give a 95% predictive interval for the price.

(c)
In our model the slope is 50. What are the units of this number?
(d)
What are the units of the intercept 20?
(e)
What are the units of the standard deviation 15?
(f)
Suppose we change the units of price to dollars and size to square feet. What would the values and units of the intercept, slope, and error standard deviation?
(g)
If we plug s = 1.6 into our model equation (with the original units), P is a constant plus the normal random variable ϵ.

Given s = 1.6, what is the distribution of P ?

3.1. Problem: The Shock Absorber Data

The data comes from a company which supplies a major automobile manufacturer with shock absorbers. An important characteristic is the “force transferred through the shock absorber when the shank is forced out of the cylinder”. If you don’t know what that really means, don’t worry, neither do I.

What we do need to understand is that the manufacturer only considers the shock to be an acceptable part if the force measurement is between 485 and 585.

The shock manufacturer and the auto manufacturer are arguing over the following issue. Before the shock is finally shipped, it is filled with gas. After it is filled with gas, it becomes very difficult to measure the force characteristic we are interested in. The shock manufacturers would like to make the measurement before the shock is filled with gas. The auto maker is concerned that there may be a difference in the force before and after the shock is filled with gas and so would like to make the measurement after it is filled.

The shock maker claims that there is little difference between the before and after measurement so that the before measurement can be used.

To investigate this we have the before (column 1, reboundb) and the after (column 2, rebounda) measurements on 35 shocks (in shock.csv).

Get the shock data (shock.csv) from the webpage.

(a)

Plot reboundb vs. rebounda.

Does this look like the kind of data the simple linear regression model is designed to capture?

```
Excel:
Download the file shock.csv and double click the file icon to get into excel.
Click on a cell in the data.
/Insert/Charts/the one with a picture of a scatter plot.
Right click on each axis  and then choose "Format Axis"
to change the plot range.
Play around with other plot options!!!

R:
sdat = read.csv("http://www.rob-mcculloch.org/data/shock.csv")
plot(sdat)
```

(b)

Run the regression of y=rebounda on x=reboundb.
What is the estimate of the true slope?

```
Excel:
Download the file shock.csv and double click the file icon to get into excel.
(i) /Data/Data Analysis/Regression
(ii) put in y range (e.g. b1:b36) and x range (e.g. a1:a36)
(iii) click labels and then OK.

R:
sdat = read.csv("http://www.rob-mcculloch.org/data/shock.csv")
#put the results of the regression in the data structure sreg (a list)
sreg = lm(rebounda~reboundb,sdat)
#print a summary of sreg
print(summary(sreg))
```

(c)
Given reboundb $= 535$, give the plug-in predictive interval for rebounda.

(d)
Give the 95% confidence interval for β_1.

(e)
Give the 95% confidence interval for β_0.

(f)
Test the null hypothesis (level .05) that $\beta_0 = 0$.

(g)
Test the null hypothesis (level .05) that $\beta_1 = 0$.

(h)
Test the null hypothesis (level .05) that $\beta_1 = 1$.
Why is this an interesting hypothesis to test?

(i)
Is it ok to use the before measurement as a proxy for the after measurement? What does the simple linear regression model tell us about this?