Name _____

Anderson School of Managment
UCLA

Mgt 264b Mr. Rossi
Regression with Applications to Marketing and Finance

Problem Set #3

This problem set is designed to reinforce the material on the sampling distribution of least squares estimators in Chapter III and provoke further thought on the meaning of the regression model.

1. More on the Least Squares and Intuition from the Regession Model

   Consider the following data set:

X	Y
0	0
0	1
1	1
1	2

   a. Plot the data
   b. Compute the regression line by using the intuition that the regression line approximates the conditional mean of Y given X. What is the mean of Y given X=0,1? DON'T run a regression in R!
   c. Verify your intuition by running least squares
   d. Show that the line passing thru (0,0) and (1,2) has a higher sum of squared errors.

2. More on Height/Weight Dataset

   a. Is a linear regression model appropriate for the relationship between SHGT and MHGT? What is important here (remember the three key characteristics of a regression model – see slides in Chapter II)?

   b. "Regression to the Mean"

   The British statistician, Galton, studied the relationship between the heights of sons and their fathers and invented the term "regression." In this particular study, he showed that the height of sons of tall fathers

would tend to "regress" or move closer to the mean height of all sons. Similarly, sons of short fathers seemed to be closer to average height than their fathers. This phenomenon has come to be known as "regression to the mean." This question is designed to explain to you why this might be the case

Let X be the height of father, Y that of son. One might expect that s_X should be very close to s_Y.

i. If $s_x = s_y$, explain why $b_1 = r$
ii. Given the result in i, explain why, in this case, $0 < b_1 < 1$.
iii. Show that (write down the equation for the least squares fitted line and substitute in for b_0)
$$\hat{Y}_i - \bar{Y} = b_1 \left(X_i - \bar{X} \right)$$
iv. Explain how the result in iii) relates to the concept of regression to the mean.

c. Rerun the regression of SHGT on MHGT and confirm that regression to the mean still happens today!!

d. Compute Var(SHGT) and Var(SHGT | MHGT).

3. Some Experimental Design

Suppose you were designing a test marketing study to determine the best price for Dove bars. This means that we want to determine the profit-maximizing price. To compute profit, we need the sales (in units) of Dove bars for any given price. To predict sales, the company wants to run a regression of sales on price.

In this situation, we are going to experimentally alter the sales price of a Dove bar to help learn about the relationship between sales and price. The experimental design consists of which prices we will try in the market place.

Different prices will be tested only by using multiple test markets (only one price per market). This means it could be expensive to undertake a pricing test.

If you are only able to test two prices, what does the formula for the variance of the slope coefficient (price sensitivity) suggest you should do in terms of setting price?

Suppose you are able to test three different prices. What would the value of this third price point be in terms of learning about the relationship between sales and price.

Note: I am looking for a discussion of the implication of the formulas not specific numbers!

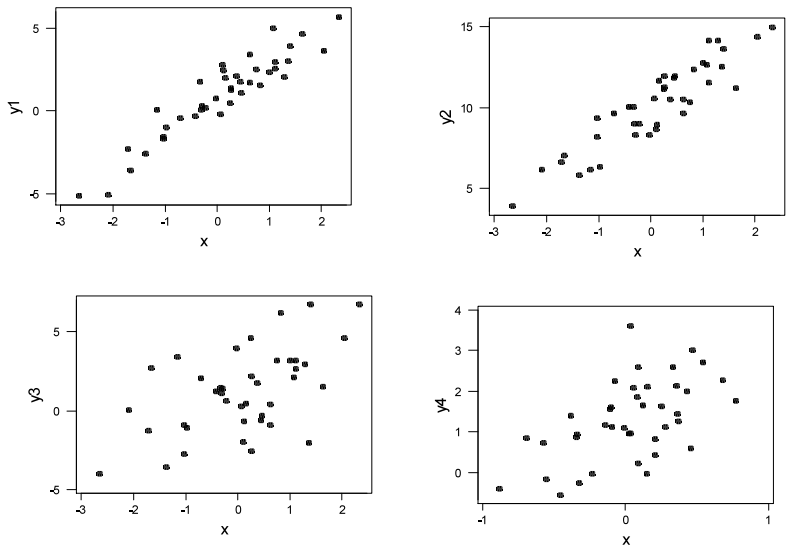
4. Match the Plots

Below are plots of y1,y2,y3, and y4 versus an x variable. Below the plots are summaries of the fitted regressions. Match the plots to the regression output.

Provide an explanation of your choices!!!

Hints: remember s!!!

Recall the variance of (b_1) given in chapter III. What does that say about the printout (when will it be small or large?)



Output

Reg A:
y = 1.25 + 1.47x s=.82 std err(b1)=.35

Reg B:
y = 1.06 + 1.30x s=2.25 std err(b1) =.32

Reg C:
y = 9.99 +2.15x s=1.08 std err(b1)= .15

Reg D:
y =.798 + 2.09x s=.89 std err(b1)= .13