Name _____

Anderson School of Management
UCLA

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

Problem Set #2

This problem set will review material on least squares and the simple linear regression model from Chapters I and II.

Download the latest version of PERregress (ver 1.0-3)!!! You will need it for the firspt

1. The Classic Height-Weight Data

This problem is designed to review the material on the relationship between the least squares regression coefficients and traditional correlation analysis.

The data frame `hgtwgt` in the class dataset area contains information on heights and weights of MM students at Northwestern University and their parents.

a. Many think that mother's height should be a good determinant of child's height. Plot MHGT vs. SHGT (SHGT on vertical axis). Is there a relationship between MHGT and SHGT?

b. Compute correlations between SHGT and each of (MHGT, FHGT and PHGT). Discuss the relative sizes of the correlations you observe. If you have installed version 1.0-3 of PERregress, you can just say "corr(hgtwgt) and it will print out all possible correlations. Otherwise you can use, cor(SHGT,MHGT), cor(SHGT,FHGT), cor(SHGT,PGHT).

c. Compute the intercept and slope least squares estimates for a regression of SHGT on MHGT using only descriptive statistics (means, variances, and covariances). DO NOT use the correlations computed in b.
d. Using the fitted values and least squares residuals, compute SST, SSR and SSE; do not use the anov() command for this!

2. More on the Flat Panel TV dataset

The variable Type in the Flat_Panel_TV dataset is what R calls a "factor."' This is the way R stores a categorical variable. A categorical variable takes on one of a relatively small number of values and these values represent "categories" or a classification of an observation. In this case, we see that the Type variable takes on two values "Plasma" and "LED" for the two types of screen technology. By default, R lists the possible values that a factor takes on in alphabetical order. To see the possible levels that Type takes on, you can say summary(Type) or the command levels(Type).

Let's try to regression Price on Type. This shouldn't work as we shouldn't be able to using character values in an equation!

a. Regress Price on Type

However, R seems to be smart enough to do the regression. The answer to this puzzle is that R must associate a numerical value with each of the categories. In fact, R sets the value for "LED" to 0 and that for "Plasma" to 1.

b. Given the fitted regression in part a), what is the estimated conditional mean of Price given Type="LED"? What is the estimated conditional mean of Price given Type="Plasma"? Interpret the slope coefficient in the regression you ran in part a).

3. Getting to Know the Simple Linear Regression Model

In all parts below, we are referring to the following SLRM:

$$Y_i = 1.0 + .8X_i + \varepsilon_i \qquad \varepsilon_i \sim \text{iid}N(0,4)$$

a. E[Y|X=1] = ?
b. E[Y|X=0] = ?
c. Var(Y|X) = ?
d. Compute 95% prediction interval for Y given X=10
e. If E[X]= 5, what is E[Y] (hint: E[Y]=E[E[Y|X]}?
f. What is probability that Y > 3, given X=1.5?

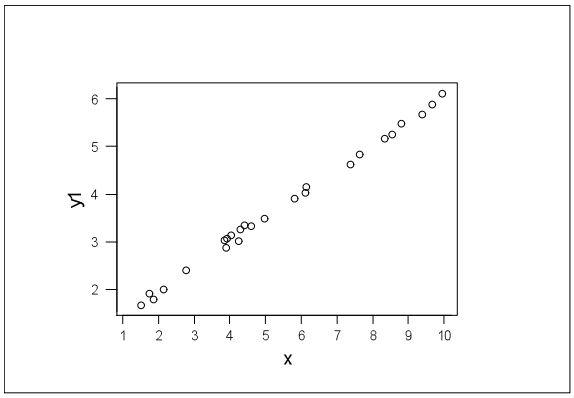
4. Match the plots

There are four plots below. Match the models to the plots

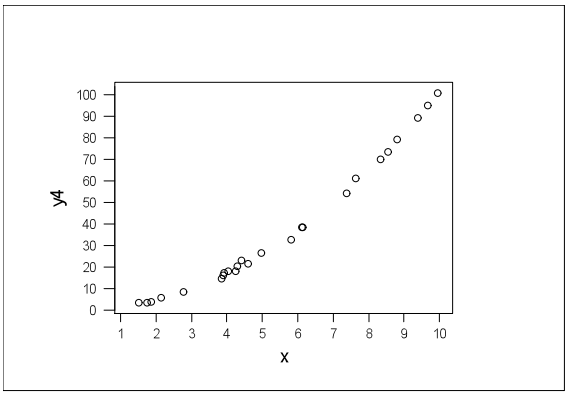
A. E[Y|X] = 1. + .5X σ=1

B. E[Y|X] = 1. - .5X σ=.1
C. E[Y|X] = 1. + .5X σ=.1
D. E[Y|X] = 1. + X^2 σ=.1

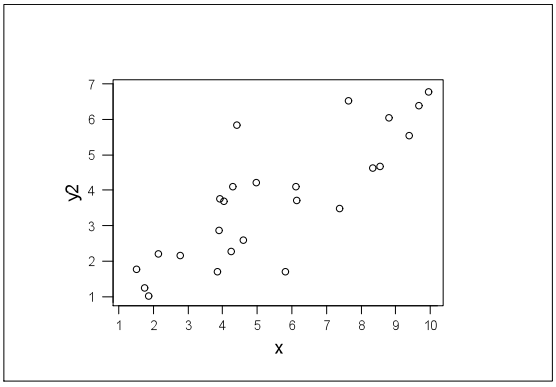
Plot 1



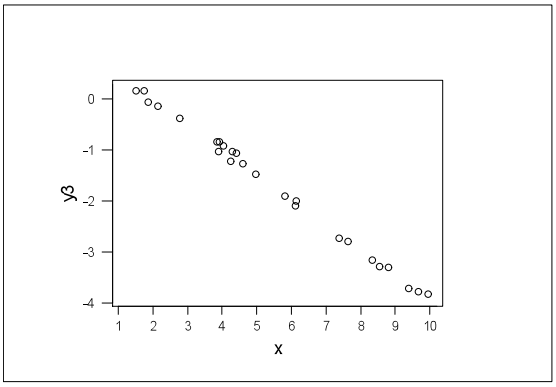
Plot 2



Plot 3



Plot 4



5. <u>More on the Regression Model</u>

Y = 1 + .5X + ε Var(ε) = 2, Var(X) =1

Note: cov(X,ε) = 0

a. compute Var(Y|X=1)

b. compute Var(Y)
c. compute corr(Y,X)

Hints:

1. Var(Y) = Var(1+.5X + ε) = Var(.5X+ε) (why?)

So Y is the weighted sum (weights are .5 for X and for ε) of two variables (X, ε). If we don't know X, then X is random so Var(Y) is the variance of the sum of two r.v.s (see prerequisite material).

2. Cov(X,Y)= E[(X-E[X])Y]=E[(X-E[X])(1 + .5 X + ε)]=
   E[(X-E[X])(.5X))] (why?) = ??