Name ________________________________

Anderson School of Managment
UCLA

MGMT 264B — Mr. Rossi
Regression with Applications to Marketing and Finance

Problem Set #1

This problem set is designed to review material on summation symbols, correlation/covariance and least squares.

1. Summation Notation: Part A

i	Y_i
1	2.0
2	-2.0
3	3.0
4	-3.0

a. compute $\sum_{i=1}^{4} Y_i$

b. compute $\sum_{i=1}^{4} \left(Y_i - \bar{Y}\right)^2$

Moral of this story: just because sum of items = 0 doesn't mean sum of items squared is 0.

2. Summation Notation: Part B

suppose we have a dataset, $Y_1, Y_2, \ldots, Y_N$. These are a GENERAL set of numbers. You need to show a) and b) below for the general case NOT a specific set of numbers. To do so, you will have to review the definition of the sample mean and sample variance using summation notation.

a. let $Z_i = cY_i$, verify (using summation notation) that

$$\bar{Z} = c\bar{Y}$$

and

$$s_Z^2 = c^2 s_Y^2$$

b. Let $Z_i = Y_i + k$. Verify that

$$\bar{Z} = \bar{Y} + k$$

and

$$s_Z^2 = s_Y^2$$

3. <u>Normal Random Variables and Covariance</u>

X ~ N(1,4) (this means X is normally distributed with mean 1 and variance 4).

X ~ N(2,2), y ~ N(2,2), cov(X,Y)=1, compute Var(.2X + .8Y)

If cov(X,Y) = 0, explain why the variance of the weighted sum will be lower.

4. <u>Some Simple Least Squares Calculations</u>

This problem will review the material on least squares calculations contained in Chapter I.

A company sets different prices for a particular stereo system in eight different regions of the country. The table below shows the numbers of units sold (in 1000s of units) and the corresponding prices (in hundreds of dollars).

Sales	420	380	350	400	440	380	450	420
Price	5.5	6.0	6.5	6.0	5.0	6.5	4.5	5.0

a. Using a hand calculator or Excel, regress sales on price and obtain the intercept and slope estimates

b. enter the data into R by creating the variable Sales and Price.

e.g. `Sales=c(420, 380, …, 420)`

c. plot the data using the `plot` command

d. Fit the regression using the `lm()` command as in the class notes.

e. Construct the residuals manually. First, construct the fitted values and then subtract the fitted values from the actual y values to create the residuals. DO NOT use the automatically created fitted.values and residuals.

f. Verify that the residuals calculated in e are uncorrelated with the X variable (Price).

g. Show that the predictor, $\hat{Y} = \bar{Y}$ (i.e. $b_0 = \bar{Y};\ b_1 = 0$), is not the least squares predictor by constructing the residuals for this and plotting them against X. Describe intuitively why this predictor is an inefficient predictor (doesn't use the sample information well).

h. If price were expressed in cents (currently given in 100$), how would b_0 and b_1 change? Note: do not rerun the regression, use the least squares formulas to relate slope and intercept from the original regression to the slope and intercept in the new regression. Hint: if you are using a calculator and a computer, you are doing this part incorrectly.

5. <u>The Relationship between b and r</u>

The `housepr` dataset contains data on house prices and size. WITHOUT running the regression of Price (dep var) on Size (indep var), compute the least squares slope and intercept, b_0 and b_1 .

You will need the descriptive stats. Note: you will not actually have to retrieve that dataset to compute the least squares coefficients.

```
> cor(Size,Price)
[1] 0.90921
> descStat(housepr)
         Mean Median     SD  IQR SE Mean 95% CI-L 95% CI-U NMissing
Size    1.853    1.6  0.841  1.2   0.217    1.428    2.279        0
Price 104.467   95.0 32.724 35.0   8.449   87.907  121.027        0
Number of Observations =  15
```