Homework Assignment 3

Problem 1

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

(i) In Excel, regress sales on price and obtain the intercept and slope estimates.

(ii) Present a plot with the data and the regression line

(iii) Based on this analysis, briefly describe your understanding of the relationship between sales and prices.

Problem 2: Match the Plots

Below (Figure 1) are 4 different scatter plots of an outcome variable y versus predictor x followed by 4 four regression output summaries labeled A, B, C and D. Match the outputs with the plots.
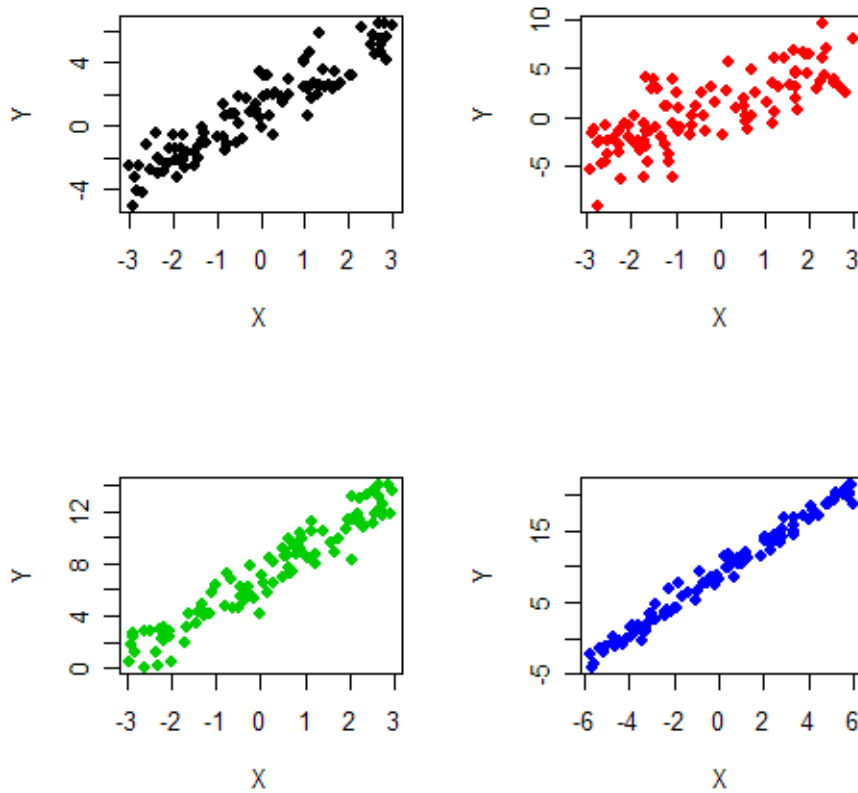

Figure 1: Scatter Plots

Regression A:

```
Coefficients:
                Estimate    Std. Error
(Intercept)      7.03747      0.12302
(Slope)          2.18658      0.07801
---
Residual standard error: 1.226
R-Squared: 0.8891
```

Regression B:

```
Coefficients:
                Estimate    Std. Error
(Intercept)       1.1491      0.1013
(Slope)           1.4896      0.0583

Residual standard error: 1.012
R-Squared: 0.8695
```

Regression C:

```
Coefficients:
                Estimate    Std. Error
(Intercept)       1.2486      0.2053
(Slope)           1.5659      0.1119

Residual standard error: 2.052
R-Squared: 0.6666
```

Regression D:

```
Coefficients:
                Estimate    Std. Error
(Intercept)       9.0225      0.0904
(Slope)           2.0718      0.0270
---
Residual standard error: 0.902
R-Squared: 0.9835
```

Problem 3

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:

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

(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 the error 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, P is a constant plus the normal random variables ϵ. Given $s = 1.6$, what is the distribution of P?

4

Problem 4: 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. The data for this problem is in **shock.xls** available in the class website.

(a) Plot the before measurement vs. the after measurement. Does this look like the kind of data we can use the simple linear regression model to think about? Why does it make sense to choose the after measurement as "Y" and the before measurement as "X"?

(b) What are 95% confidence intervals for both the slope and the intercept?

(c) Test the null hypothesis $H_o : \beta_0 = 0$.

(d) From the shock maker's point of view, what hypotheses would be of interest to test for the slope and intercept. That is, what would the shock maker like the true intercept and slope to be?

Test whether the intercept is equal to the value proposed by the shock maker.
Test whether the slope is equal to the value proposed by the shock maker.

(f) Suppose the before measurement is 550.
What is the plug-in predictive interval given x-before=550.

What does this interval suggest about the acceptability of the shock absorber?

Problem 5

The data file for this question is available in the course website.

Consider the regression model

$$Apple_t = \alpha + \beta SP500_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

where $Apple_t$ represents the return on Apple Computers in month t and $SP500_t$ represents the return on the S&P 500 in month t.

(a) What is the interpretation of β in terms of a measure of risk of the stock?

(b) What is the interpretation of α?

(b) Plot *Apple* against $SP500$. What graphical evidence is there of a relationship between *Apple* and $SP500$? Does the relationship appear to be linear? Why or why not?

(c) Estimate β. What does this estimate tell you about the risk of Apple?

(d) Is the estimate of β obtained in part (c) the actual value of β? Why or why not?

(e) Now consider the regression models

$$Intel_t = \alpha + \beta SP500_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

where $Intel_t$ represents the return on Intel in month t, and

$$Safeway_t = \alpha + \beta SP500_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

where $Safeway_t$ represents the return on Safeway in month t. How does the beta risk of the three companies compare? What do their α's tell you?

(f) On an idiosyncratic basis, ie, not considering co-movements with the market, which of the 3 stocks has the highest risk?

**Problem 6: Presidential Election**
(Imagine this is October 2012!)

Five Thirty Eight

Nate Silver's Political Calculus

This question is based on an analysis presented in the New York Times blog
FiveThirtyEight by Nate Silver...

According to the blog post, the most accurate economic indicator to predict the results of the
election for an incumbent president is the election year "equivalent non-farm payroll growth"
(in number of jobs from Jan to Oct). In the blog they provided a picture summarizing the
relationship between this economic indicator and the election results. The analysis was
based on the last 16 elections where a sitting president was seeking re-election. The picture
is displayed below:
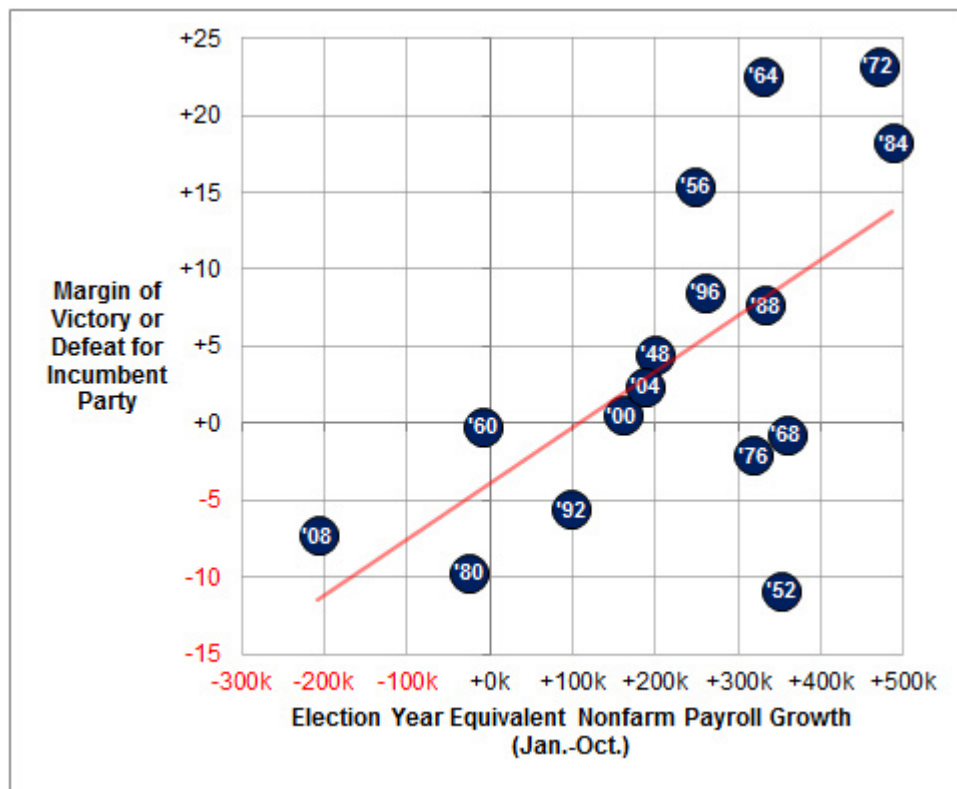
This week, The Department of Labor released the most recent employment numbers which
says that the economy has added, on average, 100,000 non-farm jobs since January 2012.

Now, I am a gambling person and I need your help... Currently, on intrade.com, I can buy or sell a future contract on Obama's re-election for $64.7. This contract pays $100 if Obama wins (see below).

1. Based on the model presented in the New York Times blog and the numbers released by the Department of Labor how should I bet on the Intrade website (i.e., should I buy or sell the future contract)? Why?

2. Do you think the answer in the question above is a good idea? Why or why not?