IEOR165

University of California, Berkeley

Midterm Exam I, 2011

Name:

Student ID:

[IEOR165 Midterm Exam]

[IEOR165 Midterm Exam]

1. [**5+10+10+5**] The stem and leaf plot in Figure 1 shows the populations of the 50 states and Washington DC, in millions of people, as reported in the 2000 census.
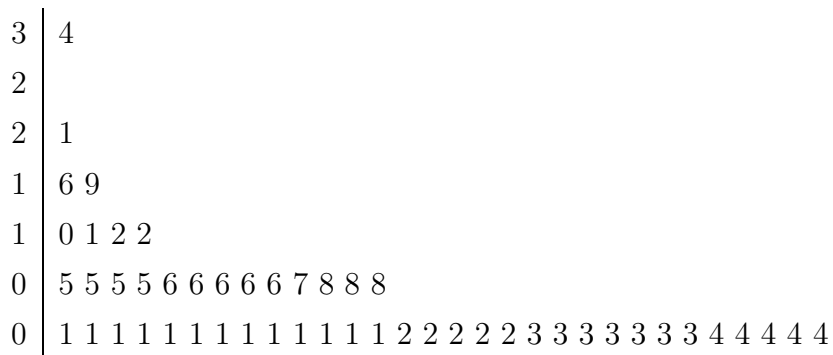
```
3 | 4
2 |
2 | 1
1 | 6 9
1 | 0 1 2 2
0 | 5 5 5 5 6 6 6 6 6 7 8 8 8
0 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 3 4 4 4 4 4
```

Figure 1: Population of the 50 states and Washington DC (1|2 means 12 million)

(a) Which measures of center and spread are most appropriate?

(b) From the stem and leaf plot, find the median and IQR

(c) Draw a box plot.

(d) Briefly describe the distribution (modality, shape, are there outliers)

2. [**5+5+5**] One measure of the size of a car engine is its "displacement", which is the total volume (in liters or cubic inches) if its cylinders. Summary statistics for several models of new cars are shown in Figure 2. Thee displacements were measured in cubic inches.

(a) Why might the mean be much larger than the median?

Count	38
Mean	177.29
Median	148.5
Standard Deviation	88.88
25th percentile	105
75th percentile	231

Figure 2: Summary Statistics

(b) your neighbor is bragging about the 227 cubic inch engine he bought in his new car. Is that engine unusually large? Explain.

(c) We can convert all the data from cubic inches to cubic centimeters (cc) by multiplying by 16.4. For example, a 200 cubic inch engine has a displacement of 3280 cc ($= 200 \times 16.4$). How would the conversion affect each of the summary statistics?

3. [**5+5+5+5+5**] The *Housing Cost Index* is a measure of housing costs in the United States that is managed by the Office of Federal Housing Enterprize Oversight. Figure 3 is a scatter plot of the *Housing Cost Index* (HCI) versus the *Median Family Income* (MFI) for each of the 50 states. The correlation is 0.65. The mean HCI is 338.2 with a standard deviation is 116.55. The mean MFI is $46,234$ with a standard deviation of 7072.47.
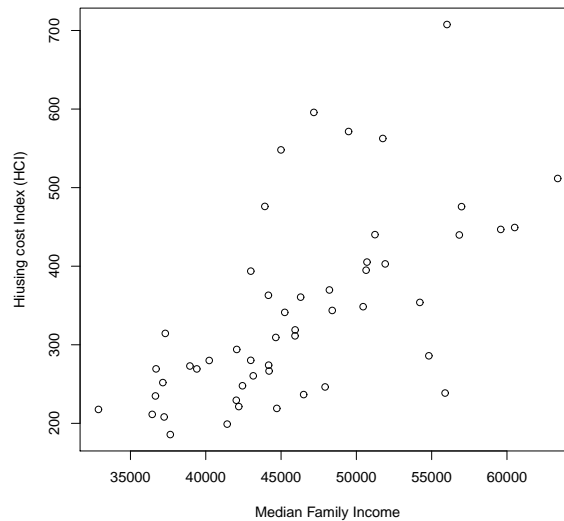

Figure 3: *Housing Cost Index* v's *Median Family Income* for each of the 50 states.

(a) Is a regression analysis appropriate? Explain.

(b) What is the equation that predicts Housing Cost Index from Median Family Income?

(c) For a state with MFI $=\$44,993$, what would be the predicted HCI?

(d) If you standardize both variables, what would be the regression equation that predicts standardized HCI from standardized MFI?

(e) What is the regression equation that predicts MFI from the HCI?

4. [**10+10**] Figure 4 shows the scatter plot of response y versus explanatory variable x. R output for the regression of y against x is shown on the last page of the exam.

   (a) What is the R^2, the correlation between x and y, the standard deviation of the residuals, and the regression equation? If a linear model is appropriate, what is the interpretation of the R^2?

   (b) Using the plots in Figure 4 and the regression output, comment on the appropriateness of a linear model to describe the relationship between x and y. Explain, if necessary, how you might go about improving the model. (Precise "numbers" are not needed; just a description of what you would do and the reasons is sufficient).
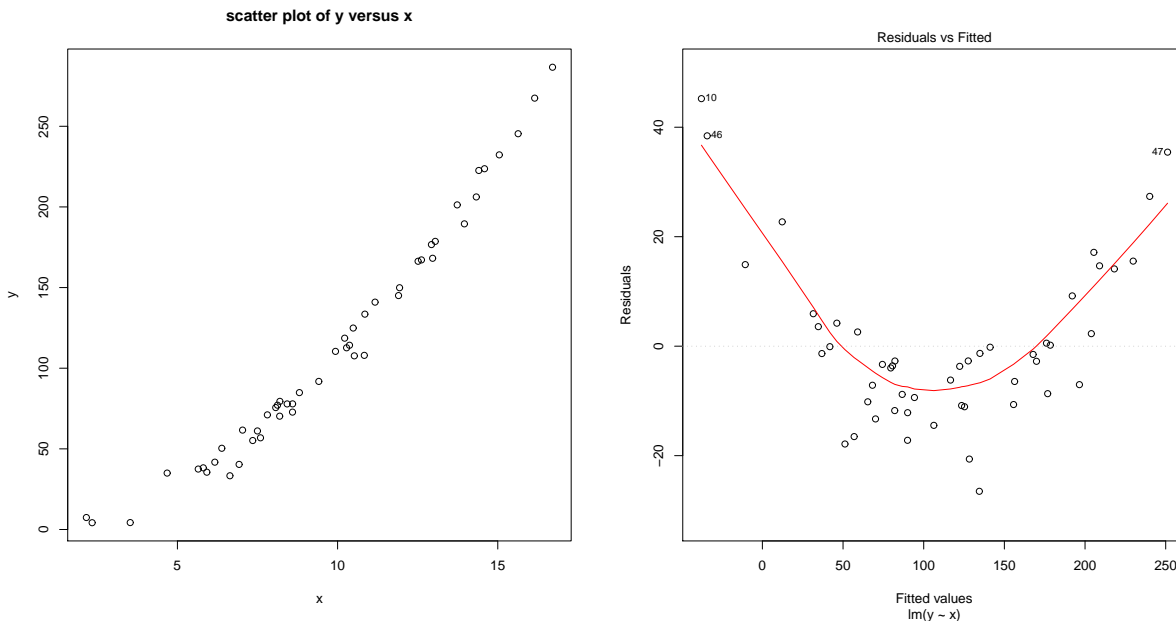

Figure 4: R output for regression of y on x.

5. What score do you think you will get for this exam? (You will receive an additional 5 points bonus if you are within 5 points of your actual score).

```
> summary(lm(y~x))

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q  Median       3Q      Max
 -26.512   -9.967  -2.736    4.045   45.220

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -80.6442     6.2906  -12.82   <2e-16 ***
x            19.8573     0.6072   32.70   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.16 on 48 degrees of freedom
Multiple R-squared: 0.9571,     Adjusted R-squared: 0.9562
F-statistic:  1070 on 1 and 48 DF,  p-value: < 2.2e-16

>
```