Homework 1

RABE 2.2

Explain why you would or would not agree with each of the following statements:

a. $\mathrm{Cov}(Y, X)$ and $\mathrm{Cov}(X, Y)$ can take values between $-\infty$ and $+\infty$
b. If $\mathrm{Cov}(Y, X) = 0$ or $\mathrm{Cov}(X, Y) = 0$ one can conclude that there is not relationship between Y and X
c. The least squares line fitted to the points in the scatter plot made by the points $(Y_i, \hat{Y}_i)$ has a zero intercept and a unit slope

RABE 2.3

Using the regression output listed below ("Computer Repair Data Regression Table"), test the following hypotheses using $\alpha = 0.1$:

a. $H_0 : \beta_1 = 15$ vs $H_A : \beta_1 \neq 15$
b. $H_0 : \beta_1 = 15$ vs $H_A : \beta_1 > 15$
c. $H_0 : \beta_1 = 0$ vs $H_A : \beta_1 \neq 0$
d. $H_0 : \beta_1 = 5$ vs $H_A : \beta_1 \neq 5$

RABE 2.4

Using the regression output listed below ("Computer Repair Data Regression Table"), construct the 99% confidence interval for β_0

Computer Repair Data Regression Table (RABE Table 2.9)

```
# data available here: http://www1.aucegypt.edu/faculty/hadi/RABE5/Data5/P031.txt
computer.repair = read.csv('/home/peter/Desktop/Teaching/STAT_4400_5400/data/computer_repair.tsv',sep='
summary(lm(Minutes ~ Units, data=computer.repair))
```

```
##
## Call:
## lm(formula = Minutes ~ Units, data = computer.repair)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.2318 -3.3415 -0.7143  4.7769  7.8033
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.162      3.355    1.24    0.239
## Units         15.509      0.505   30.71 8.92e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.392 on 12 degrees of freedom
## Multiple R-squared:  0.9874, Adjusted R-squared:  0.9864
## F-statistic: 943.2 on 1 and 12 DF,  p-value: 8.916e-13
```

RABE 2.12

In order to investigate the feasibility of starting a Sunday edition for a large metropolitan newspaper, information was obtained from a sample of 34 news-papers concerning their daily and Sunday circulations (in thousands). The data are online here: http://www1.aucegypt.edu/faculty/hadi/RABE5/Data5/P054.txt.

a. Construct a scatter plot of Sunday circulation versus daily circulation. Does the plot suggest a linear relationship between Daily and Sunday circulation? Do you think this is a plausible relationship?
b. Fit a regression line predicting Sunday circulation from Daily circulation
c. Obtain the 95% confidence itnervals for β_0 and β_1
d. Is there a significant relationship between Sunday circulation and Daily circulation? Justify your answer by a statistical test. Indicate what hypothesis you are testing, and your conclusion.
e. What proportion of the variability in Sunday circulation is accounted for by Daily circulation?
f. Provide an interval estimate (based on 95% level) for the true average Sunday circulation of newspapers with Daily circulation of 500,000
g. The particular newspaper that is considering a Sunday a edition has a Daily circulation of 500,000. Provide an interval estiamte (based on the the 95% level) for the predicted Sunday circulation of this paper. How does this interval differ from that given in (f)

(h) Another newspaper being considered as a candidate for a Sunday edition has a daily circulation of 2,000,000. Provide an interval estimate for the predicted Sunday circulation for this paper. How does this interval compare with the one given in (g)? Do you think it is likely to be accurate?

RABE 2.13

Let $y_1, y_2, ..., y_n$ be a sample drawn from a normal population with unknown mean μ and unknown variance σ^2 ($y_i \sim N(\mu, \sigma^2)$). One way to estimate μ is to fit the linear model:

$$y_i = \mu + \epsilon; i = 1, 2, ...n$$

And use the least squares method estimator (LSE), that is to chose μ such that it minimizes the sum of squares $\sum\limits_{i=1}^{n}(y_i - \mu)^2$. Another way is to use least absolute value estimator (LAVE), that is, to minimize the sum of the absolute values: $\sum\limits_{i=1}^{n}|y_i - \mu|^2$

a. Show that the LSE of μ is the sample mean $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$
b. Show that the LAV of μ is the sample median
c. State one advantage and one disadvantage each of the LSE and the LAVE