Math 3339
Homework 12 (Chapter 8)

Name: James Richardson PeopleSoft ID: 1555520

Instructions:

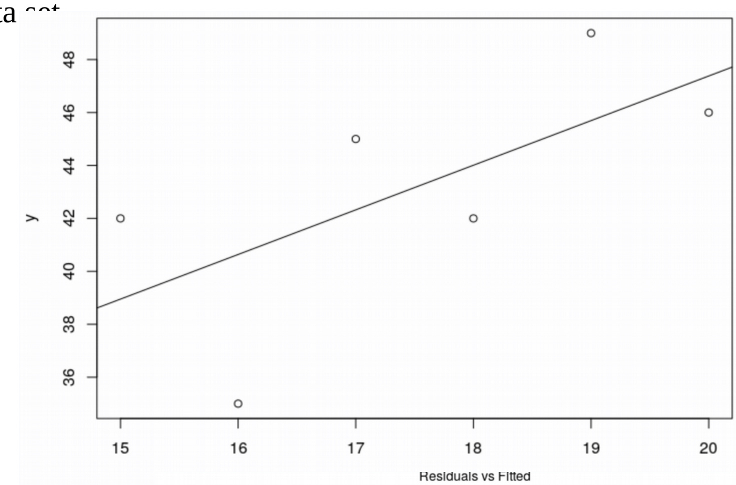
- Homework will NOT be accepted through email or in person. Homework must be submitted through CourseWare BEFORE the deadline.
- Print out this file and complete the problems.
- Use blue or black ink or a dark pencil.
- Write your solutions in the space provided. You must show all work for full credit.
- Submit this assignment at http://www.casa.uh.edu under "Assignments" and choose **HW12**
- Total points: **15.**

1. *The following data is looking at how long it takes to get to work. Let x = commuting distance (miles) and y = commuting time (minutes)

x	15	16	17	18	19	20
y	42	35	45	42	49	46

a. Give a scatterplot of this data and comment on the direction, form and strength of this relationship.

b. Determine the least-squares estimate equation for this data set.

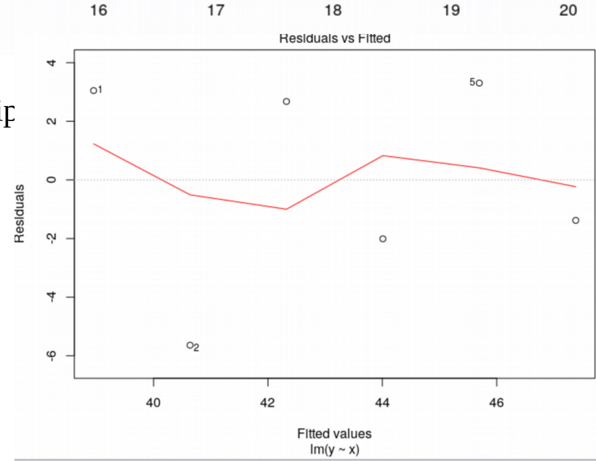


There is a strong positive linear relationship in the data

c. Give the r^2, comment on what that means.

R^2=0.65806
This means that 65% of the model can be explained by explained by the data

d. Give the residual plot based on the least-squares estimate equation.

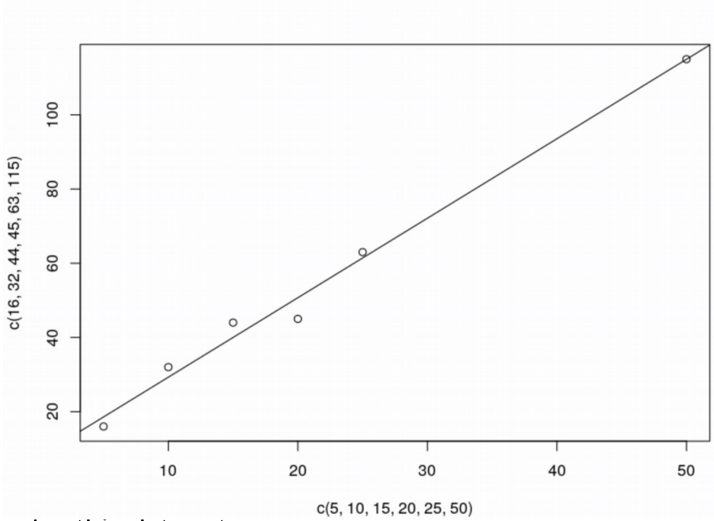


e. Test if this least-squares estimate equation specify a useful relationship between commuting distance and commuting time.

2. *The following another set of data that looking at how long it takes to get to work. Let x = commuting distance (miles) and y = commuting time (minutes)

x	5	10	15	20	25	50
y	16	32	44	45	63	115

a. Give a scatterplot of this data and comment on the direction, form and strength of this relationship.



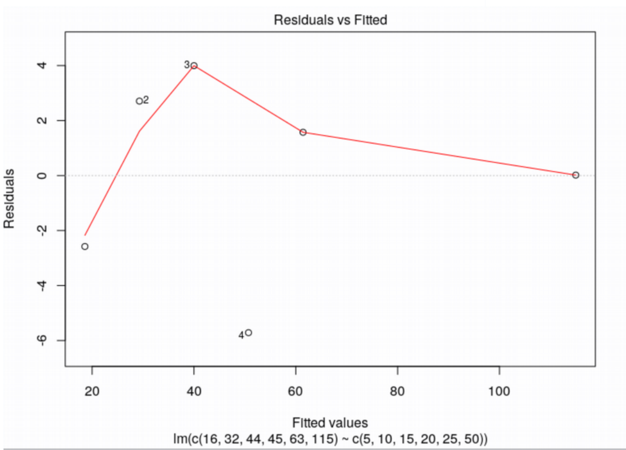
b. Determine the least-squares estimate equation for this data set.

2.142x+7.869

c. Give the r^2. comment on what that means.

```
> cor(c(5,10,15,20,25,50),c(16,32,44,45,63,115))
[1] 0.9944653
```

d. Give the residual plot based on the least-squares estimate equation.



e. Test if this least-squares estimate equation specify a useful relationship between commuting distance and commuting time.

f. Compare this least-square estimate equation to the previous least-squares estimate equation in problem 1. In which situation would the least-squares equation be least effective? Justify your answer.

6. The R datasets package has a data frame called "airquality", which lists ozone concentration, solar radiation, wind speed and temperature in New York for 154 days in 1973. Some of the data values are missing, but R will automatically omit those cases with missing data. Fit a linear model with Ozone as the response and Wind as the X variable. Find 90% confidence intervals for the expected ozone concentration when wind speed is 0 and for the expected increase in ozone concentration for a unit increase in wind speed.

Slope interval =
b +/- t* stderr

= -5.55 + c(-1, 1)*qt(.95, 114)
= (-7.20494 -3.89506)

E[0] = -5.5509(0) + 96.873
= 96.873

```
> mylm = lm(airquality$Ozone~airquality$Wind)
> summary(mylm)

Call:
lm(formula = airquality$Ozone ~ airquality$Wind)

Residuals:
    Min      1Q  Median      3Q     Max 
-51.572 -18.854  -4.868  15.234  90.000 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      96.8729     7.2387   13.38  < 2e-16 ***
airquality$Wind  -5.5509     0.6904   -8.04 9.27e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26.47 on 114 degrees of freedom
  (37 observations deleted due to missingness)
Multiple R-squared:  0.3619,    Adjusted R-squared:  0.3563 
F-statistic: 64.64 on 1 and 114 DF,  p-value: 9.272e-13
```

```
> -5.55 + c(-1, 1)*qt(.95, 114)
[1] -7.20833 -3.89167
> mylm

Call:
lm(formula = airquality$Ozone ~ airquality$Wind)

Coefficients:
    (Intercept)  airquality$Wind  
         96.873           -5.551  
```

∗ Problems came from Devore, Jay and Berk, Kenneth, *Modern Mathematical Statistics with Applications*, Thomson Brooks/Cole, 2007.

7. With the airquality data, test the null hypothesis $H_0 : \beta_1 = -5$ against the one-sided alternative $\beta_1 > -5$. The output of lm does not give you the answer directly, but it does give you the estimated value of β_1 and its standard error. You know that the test statistic has the student-t distribution with $n - 2$ degrees of freedom. Give a p-value. Warning: Because of missing data, $n = 116$ not 154.

```
> mylm = lm(airquality$Ozone~airquality$Wind)
> summary(mylm)

Call:
lm(formula = airquality$Ozone ~ airquality$Wind)

Residuals:
    Min      1Q  Median      3Q     Max
-51.572 -18.854  -4.868  15.234  90.000

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      96.8729     7.2387   13.38  < 2e-16 ***
airquality$Wind  -5.5509     0.6904   -8.04 9.27e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 26.47 on 114 degrees of freedom
  (37 observations deleted due to missingness)
Multiple R-squared:  0.3619,    Adjusted R-squared:  0.3563
F-statistic: 64.64 on 1 and 114 DF,  p-value: 9.272e-13
```

Stderr coeficient = .6904
Stderr = 26.47

Test stat = b1 / stderr
= -5.55/26.47
= -0.2096

> pvalue = pt(-0.2096, 114)
[1] 0.417177

.417 > a = .10

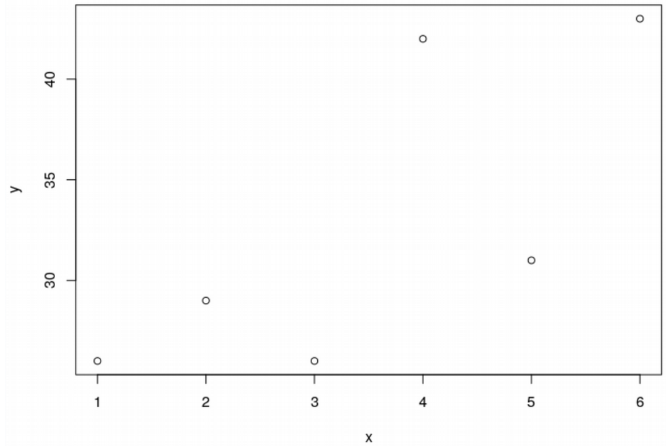
Fail to reject the null hypothesis

∗ Problems came from Devore, Jay and Berk, Kenneth, *Modern Mathematical Statistics with Applications*, Thomson Brooks/Cole, 2007.

5. The cost of a home depends on the number of bedrooms in the house. Suppose the following data is recorded for homes in a given town

price (in thousands)	300	250	400	550	317	389	425	289	389	559
No. bedrooms	3	3	4	5	4	3	6	3	4	5

a) Make a scatterplot



b) Fit the data with a least squares regression line.

3.057x+22.133

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   22.133      5.414   4.088   0.0150 *
x              3.057      1.390   2.199   0.0927 .
---
```

c) Give a 95% confidence interval for the slope.

3.057(+/-)qt(.975, 8)*1.390
[-0.1483, 6.26]

d) If one house has one more number of rooms than another house, how much additional cost would we expect for the price?

In this case it would be the slope. 3.057*1000
So about $3057

e) Test the hypothesis that an extra bedroom costs $60,000 against the alternative that it cost s more.