STAT 308 – Homework 6

For the problems in which calculations are needed, please include your `R` code with your answers, otherwise you will not be given full credit. Please upload your assignment by Thursday, October 20, 11:59 pm in a pdf file to Sakai.

1. Use the below ANOVA table from a multiple linear regression with three explanatory variables to answer the following questions.

	df	Sums of Squares	Mean Square	f Value	Pr(>f)
X_1	1	2560.45	2560.45	1852.713	<2.2e-16
$X_2 \mid X_1$	1	173.18	173.18	125.311	<2.2e-16
$X_3 \mid X_1, X_2$	1	0.50	0.5	0.362	0.550
Error	56	77.40	1.382		
Total	59	2811.53	47.653		

a. What are the model sums of squares for the linear model with X_1, X_2, and X_3 included as covariates/predictors?

$SSM_{X_1,X_2,X_3} = SS_{X_1} + SS_{X_2|X_1} + SS_{X_3|X_1,X_2} = 2560.45 + 173.18 + 0.5 = 2734.13$

b. What are the model sums of squares for the linear model with only X_1 and X_2 included as covariates?

$SSM_{X_1,X_2} = SS_{X_1} + SS_{X_2|X_1} = 2560.45 + 173.18 = 2733.63$

c. What are the error sums of squares for the linear model with only X_1 included as a covariate?

$SSE_{X_1} = SST - SS_{X_1} = 2811.53 - 2560.45 = 251.08$

d. What is the mean squared error for the linear model with only X_1 included as a covariate?

$MSE_{X_1} = \frac{SSE_{X_1}}{dfe_{X_1}} = \frac{251.08}{59-1} = 4.33$

2. Reconsider the `economy.csv` dataset from the last homework assignment.

a. Suppose we create a linear model with crude oil in dollars per barrel as the response variable and all of the other variables in the dataset as the explanatory variables. Perform an overall F test (hypothesis test) to determine if this overall linear model is statistically significant. Be sure to include all necessary information to perform a hypothesis test. Assume $\alpha = 0.05$.

Model is $\hat{crude} = \beta_0 + \beta_1 INTEREST + \beta_2 FOREIGN + \beta_3 DJIA + \beta_4 GNP + \beta_5 PURCHASE + \beta_6 CONSUMER$

$$H_0 : \beta_1 = \cdots = \beta_6 = 0$$
$$H_a : \text{At least one } \beta \neq 0$$

```
econ <- read.csv("../Data/economy.csv")
mod <- lm(CRUDE ~ .,econ)
f <- summary(mod)$f
f["value"] # This is the f statistic for testing for an overall model fit
```

```
##    value
## 31.50497
```

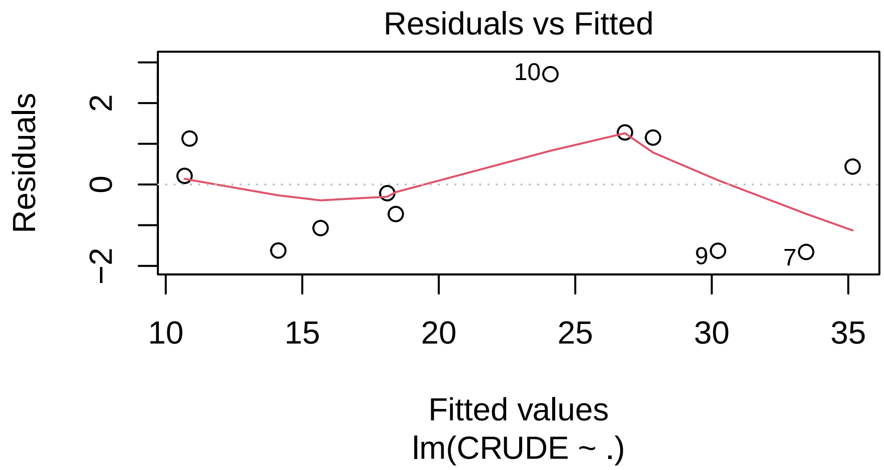
```
pf(f["value"],f["numdf"],f["dendf"],lower.tail=FALSE) # This is the p-value for testing for an overall m
```

```
##        value
## 0.0008088748
```

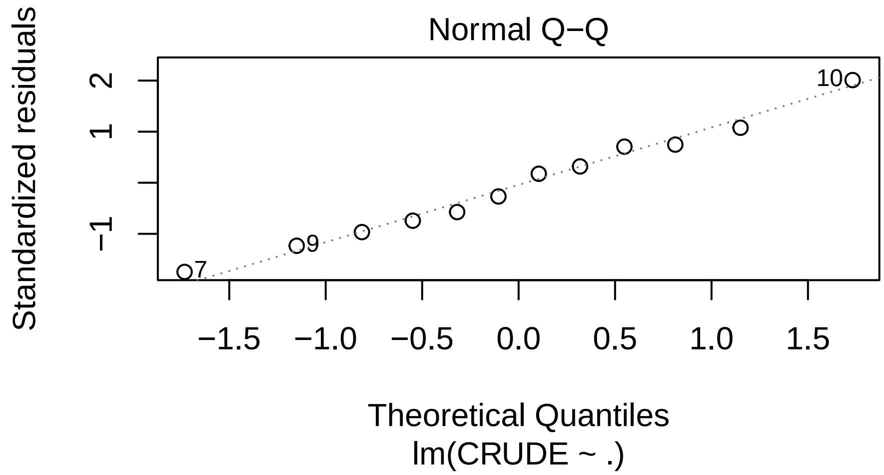
$F = 31.5$, $p - value = 0.00081$. We reject H_0 and can conclude that at least one of interest rate, foreign investments, Dow Jones industrial average, gross national product, purchasing power of U.S. dollar, and consumer debt is linear related to crude oil price.

b. Check if the assumptions of homoscedasticity and normally distributed residuals are violated for the model in (a).

```
plot(mod,1)
```




```
plot(mod,2)
```



The points in the residual plot appear to be randomly and uniformly scattered across all of the fitted values in our dataset, so the assumption of homoscedasticity does not appear to be violated.

The points in the normal QQ plot appear to fall closely along the 45-degree line, so the assumption of normally distributed residuals does not appear to be violated.

c. Suppose we know that dollars per barrel is related linearly with interest rates.

(i). Report and interpret the r^2 value for the linear model for dollars per barrel with interest rate as the only explanatory variable.

```
mod_red <- lm(CRUDE ~ INTEREST,econ)
summary(mod_red)$r.squared
```

```
## [1] 0.9396989
```

93.97% of the variation in crude oil prices can be explained through its linear relationship interest rates.

(ii). Perform a hypothesis test to determine if adding purchasing power of the U.S. dollar to our linear model significantly improves the predictive ability of our model. Be sure to include all necessary information to perform a hypothesis test. Assume $\alpha = 0.05$.

Reduced Model: $CR\hat{U}DE = \beta_0 + \beta_1 INTEREST$ Full Model: $CR\hat{U}DE = \beta_0 + \beta_1 INTEREST + \beta_2 PURCHASE$

$$H_0 : \beta_2 = 0$$

$$H_a : \beta_2 \neq 0$$

```
mod_full <- lm(CRUDE ~ INTEREST + PURCHASE,econ)
anova(mod_red,mod_full)
```

```
## Analysis of Variance Table
##
## Model 1: CRUDE ~ INTEREST
## Model 2: CRUDE ~ INTEREST + PURCHASE
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     10 50.481
## 2      9 26.897  1    23.584 7.8916 0.0204 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$f = 7.892$, $p = 0.0204$. We reject H_0 and can conclude that adding purchasing power to our linear model that already includes interest rates significantly improves the predictions of crude oil prices.

(iii). What is the additional percent variation in crude oil dollars per barrel than can be explained by adding purchasing power of U.S. dollar to our model to our linear model that already includes interest rates (Hint: this is the difference in the r^2s of the two models).

```
summary(mod_full)$r.squared - summary(mod_red)$r.squared
```

```
## [1] 0.02817211
```

2.82%

(iv). Perform a hypothesis test to determine if adding both purchasing power of the U.S. dollar and Dow Jones Industrial Average to our linear model that already includes interest rates significantly improves the predictive ability of our model. Be sure to include all necessary information to perform a hypothesis test. Assume $\alpha = 0.05$.

Reduced Model: $\widehat{CRUDE} = \beta_0 + \beta_1 INTEREST$ Full Model: $\widehat{CRUDE} = \beta_0 + \beta_1 INTEREST + \beta_2 PURCHASE + \beta_3 DJIA$

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_a : \text{At least one } \beta \neq 0$$

```
mod_full <- lm(CRUDE ~ INTEREST + PURCHASE + DJIA,econ)
anova(mod_red,mod_full)
```

```
## Analysis of Variance Table
##
## Model 1: CRUDE ~ INTEREST
## Model 2: CRUDE ~ INTEREST + PURCHASE + DJIA
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     10 50.481
## 2      8 26.895  2    23.585 3.5077 0.08058 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$f = 3.508$, $p = 0.0806$. We fail to reject H_0 and can conclude that adding purchasing power and the Dow Jones industrial average to our linear model that already includes interest rates does not significantly improve the predictions of crude oil prices.