STAT350 Assignment 3 Solution

Question 1

(i)

We know that the hat matrix $H = X(X'X)^{-1}X'$. To show H is symmetric, we need to prove $H' = H$:

$$\begin{aligned}
H' &= (X(X'X)^{-1}X')' \\
&= (X')'((X'X)^{-1})'X' \text{ (because for two matrices } A \text{ and } B\text{, } (AB)' = B'A'\text{)} \\
&= X(X'X)^{-1}X' \\
&= H
\end{aligned}$$

(ii)

To show H is idempotent, we need to prove $H^2 = H$:

$$\begin{aligned}
H^2 &= X(X'X)^{-1}X'X(X'X)^{-1}X' \\
&= X(X'X)^{-1}X' \text{ (because } (X'X)^{-1}X'X = I\text{)} \\
&= H
\end{aligned}$$

Question 2

(a)

The covariance for the least squares regression estimators $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$.

```
data=data.frame("x1"=c(4.49,3.04,3.94,2.63,4.55,3.88,2.92,2.82,3.17,2.91),
                "x2"=c(2.92,4.33,4.27,1.92,2.47,2.36,3.21,4.22,1.80,2.35),
                "y"=c(-5.32,-9.24,-5.89,1.15,-1.47,1.91,-3.99,-6.82,1.49,-0.89))
model=lm(y~x1+x2,data=data)
sigma_square=2
X=cbind(rep(1,10),data$x1,data$x2)
sigma_square*solve(t(X)%*%X)
```

```
##           [,1]         [,2]         [,3]
## [1,]  7.575272 -1.519895879 -0.721752043
## [2,] -1.519896  0.434863250  0.008757325
## [3,] -0.721752  0.008757325  0.231715454
```

(b)

The estimated $V(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$, where $\hat{\sigma}$ is the residual standard error in the summary table of `lm` function.

```r
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4299 -0.4180 -0.1350  0.3775  2.6050
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.8320     3.1998   2.760 0.028089 *
## x1           -0.2113     0.7667  -0.276 0.790844
## x2           -3.6896     0.5596  -6.593 0.000306 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.644 on 7 degrees of freedom
## Multiple R-squared:  0.8613, Adjusted R-squared:  0.8217
## F-statistic: 21.74 on 2 and 7 DF,  p-value: 0.0009933
```

```r
RSE=1.644
RSE^2*solve(t(X)%*%X)
```

```
##            [,1]        [,2]        [,3]
## [1,] 10.2369804 -2.05393865 -0.97535262
## [2,] -2.0539387  0.58766028  0.01183437
## [3,] -0.9753526  0.01183437  0.31313285
```

(c)

The variance of residuals is $V(e_i) = \sigma^2(1 - h_{ii})$, where h_{ii} is the i^{th} diagonal element of the hat matrix $H = X(X'X)^{-1}X'$.

```r
H=X%*%solve(t(X)%*%X)%*%t(X)
sigma_square*(1-diag(H)[1]) # variance of the 1st residual
```

```
## [1] 1.316208
```

```r
sigma_square*(1-diag(H)[3]) # variance of the 3rd residual
```

```
## [1] 1.295119
```

The variance of the 1st and 3rd residuals are 1.316 and 1.295, respectively.

(d)

The covariance between the i^{th} and j^{th} residual is $Cov(e_i, e_j) = -\sigma^2 h_{ij}$.

```r
-sigma_square*H[1,3]
```

```
## [1] -0.4239149
```

The covariance between the 1st and 3rd residual is -0.424.

Question 3

(a)

```
Dat=read.csv("/Users/dbingham/Desktop/data_computer_experiment.csv")
exp_model=lm(location~., data=Dat)
summary(exp_model)$coef[2,2]^2
```

```
## [1] 0.3999662
```

The estimated variance for the least squares estimator of the regression coefficient for the thickness of the beryllium disk is 0.4.

(b)

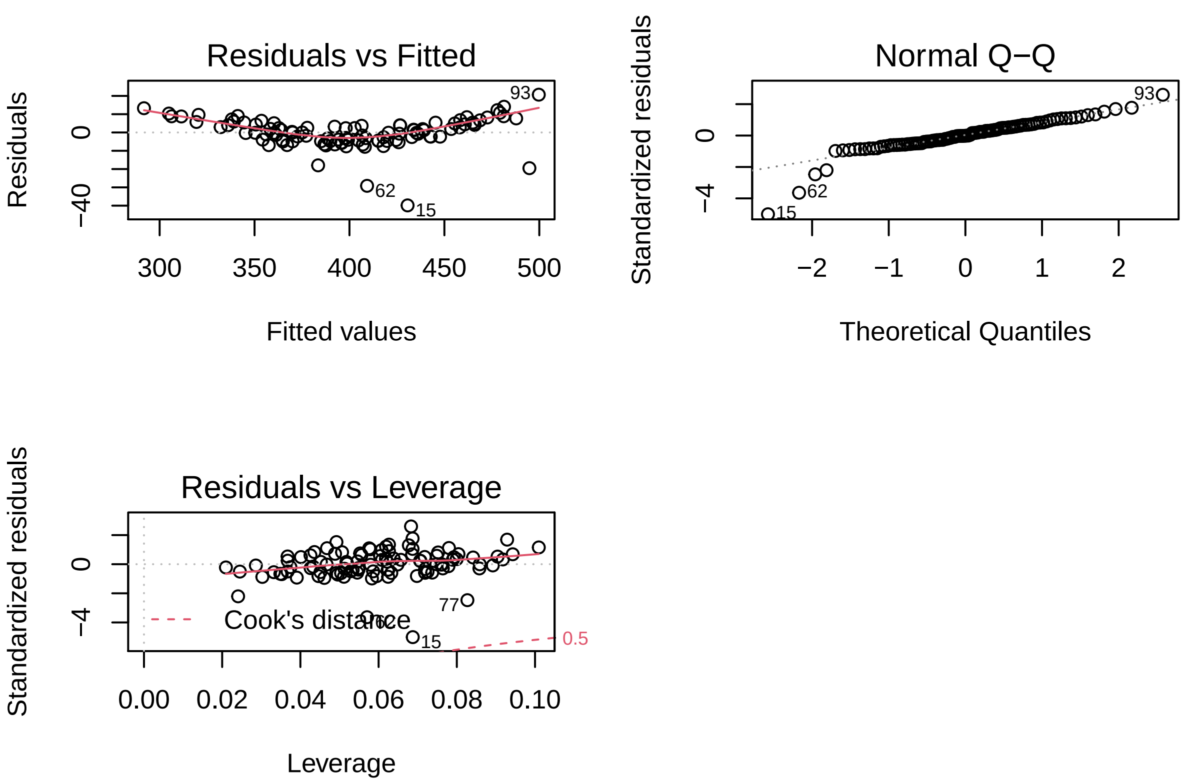
```
vcov(exp_model)[2,6]
```

```
## [1] 3.18406
```

The estimated covariance between the least squares estimators of the regression coefficients for the thickness of the beryllium disk and the wall opacity is 3.184.

(c)

```
par(mfrow=c(2,2))
plot(exp_model,which=1)
plot(exp_model,which=2)
plot(exp_model,which=5)
```



(i)

We use the plot of residuals vs fitted values. The constant variance assumption for the errors appears reasonable.

(ii)

We use the normal Q-Q plot. The normality assumption appears reasonable.

(iii)

We use the plot of residuals vs leverage. There is no large leverage points.

(iv)

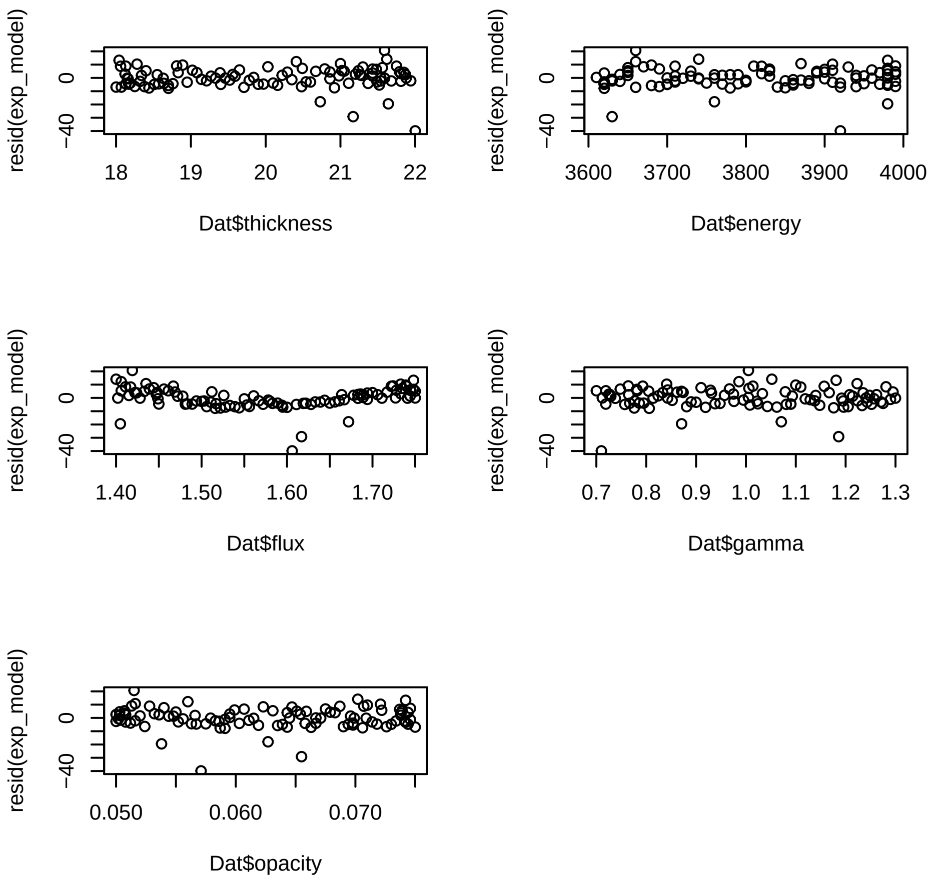
There are some outliners, such as point 15, 62 and 77.

(v)

We use the plot of residuals vs leverage. No points have Cook's distance > 0.5, so there is no influential points.

(vi)

```r
par(mfrow=c(3,2))
plot(Dat$thickness,resid(exp_model))
plot(Dat$energy,resid(exp_model))
plot(Dat$flux,resid(exp_model))
plot(Dat$gamma,resid(exp_model))
plot(Dat$opacity,resid(exp_model))
```



The relationship between the predictors and the response are linear expcept for flux.