HW1 Solution

Problem 1

(a) We have $RSS(\beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$, and take the partial derivatives:

$$\frac{\partial RSS}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0, \tag{1}$$

$$\frac{\partial RSS}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0. \tag{2}$$

(b) By solving (1), we have:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{3}$$

Plug (3) into (2), we get:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} x_i y_i - \left(\bar{y} - \hat{\beta}_1 \bar{x}\right) \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \\
&= \Big(\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i\Big) - \Big(\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i\Big)\hat{\beta}_1
\end{aligned}$$

Hence we derive that:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4}$$

(c) To have (4) well defined, we assume that the denominator of $\hat{\beta}_1$, i.e. $\sum_{i=1}^{n}(x_i - \bar{x})^2$ is not 0. This is equivalent to say x_i cannot be constant. Otherwise, $\hat{\beta}_1$ does not exist.

(d) Denote $Y = (y_1, \cdots, y_n)'$, $X = (\mathbf{1}_n, \mathbf{x})$ where $\mathbf{x} = (x_1, \cdots, x_n)'$ and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ which we got from (b). Then RSS can be rewritten in matrix form:

$$\begin{aligned}
RSS =& ||Y - X\beta||^2 \\
=& ||Y - X\hat{\beta} + X\hat{\beta} - X\beta||^2 \\
=& ||Y - X\hat{\beta}||^2 + ||X\hat{\beta} - X\beta||^2 + 2\big\langle Y - X\hat{\beta}, X(\hat{\beta} - \beta)\big\rangle \\
=& ||Y - X\hat{\beta}||^2 + ||X\hat{\beta} - X\beta||^2 \\
=& (\beta - \hat{\beta})^T (X^T X)(\beta - \hat{\beta}) + ||Y - X\hat{\beta}||^2
\end{aligned}$$

In the fourth equality, we need to use the normal equations from(b) to have:

$$\big\langle Y - X\hat{\beta}, X(\hat{\beta} - \beta)\big\rangle = (Y^T X - \hat{\beta}^T X^T X)(\hat{\beta} - \beta) = 0$$

Hence we have $b = \hat{\beta}$, $A = X^T X$ and $c = ||Y - X\hat{\beta}||^2$.

(e) We can see that $A = X^TX$ is symmetric matrix. And since $x_i's$ are not constant under the same assumption in (c), we have $rank(A) = rank(X) = 2$, then A is also positive definite.

Since A is positive definite, we have:

$$\begin{aligned} RSS =&(\beta - \hat{\beta})^T(X^TX)(\beta - \hat{\beta}) + ||Y - X\hat{\beta}||^2 \\ \geq& ||Y - X\hat{\beta}||^2 \end{aligned}$$

And the equality achieves only when $\beta = \hat{\beta}$, so the solution in part (b) is the unique minimizer of RSS.

Problem 2

From the normal equations in 1(1), we can see that:

$$\frac{\partial RSS}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = -2\sum_{i=1}^{n}\hat{\epsilon}_i = 0$$

$$\frac{\partial RSS}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = -2\sum_{i=1}^{n}\hat{\epsilon}_i x_i = 0$$

So we can get residual mean and the sample correlation between residuals and predictor as below:

$$\bar{\hat{\epsilon}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i = 0$$

$$cor(\hat{\epsilon}, \mathbf{x}) = \sum_{i=1}^{n}(\hat{\epsilon}_i - \bar{\hat{\epsilon}})(x_i - \bar{x}) = \sum_{i=1}^{n}(\hat{\epsilon}_i - \bar{\hat{\epsilon}})x_i = \sum_{i=1}^{n}\hat{\epsilon}_i x_i = 0$$

Problem 3

Under Gauss-Markov conditions, $E[\epsilon] = 0, var(\epsilon) = \sigma^2 I$.

(a)

$$\begin{aligned} E[\mathbf{y}] &= E[\mathbf{X}\beta_0 + \epsilon] = \mathbf{X}\beta_0 + E[\epsilon] = \mathbf{X}\beta_0 \\ var(\mathbf{y}) &= var(\mathbf{X}\beta_0 + \epsilon) = var(\epsilon) = \sigma^2\mathbf{I}_n \end{aligned}$$

(b) From normal equations, we have known that $\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$.

$$E[\hat{\beta} - \beta_0] = E[(\mathbf{X'X})^{-1}\mathbf{X'y} - \beta_0] = (\mathbf{X'X})^{-1}\mathbf{X'}E[\mathbf{y}] - \beta_0 = (\mathbf{X'X})^{-1}\mathbf{X'X}\beta_0 - \beta_0 = 0$$

$$var(\hat{\beta} - \beta_0) = var((\mathbf{X'X})^{-1}\mathbf{X'y}) = (\mathbf{X'X})^{-1}\mathbf{X'}var(\mathbf{y})\mathbf{X}(\mathbf{X'X})^{-1} = \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1} = \sigma^2(\mathbf{X'X})^{-1}$$

(c) We define the hat matrix $H = X(X'X)^{-1}X'$ first, then $\hat{y} = X\hat{\beta} = Hy$.

$$E[y - \hat{y}] = X\beta_0 - E[X\hat{\beta}] = X\beta_0 - XE[\hat{\beta}] = 0$$

$$var(y - \hat{y}) = var((I - H)y) = (I - H)var(y)(I - H)' = \sigma^2(I - H)^2 = \sigma^2(I - H)$$

Here both H and $(I - H)$ are projection matrices, so they are symmetric and idempotent.

(4) Using the hat matrix H as before, we have:

$$E[\hat{y}] = E[X\hat{\beta}] = XE[\hat{\beta}] = X\beta_0$$

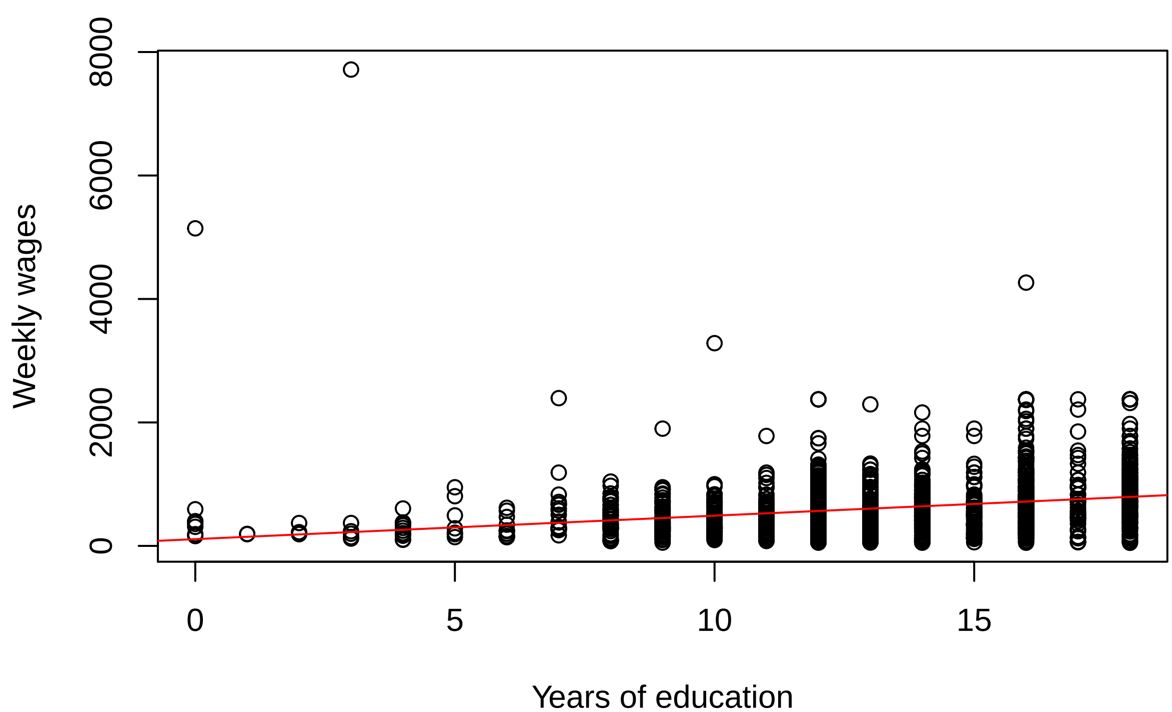
$$var(\hat{y}) = var(Hy) = Hvar(y)H' = \sigma^2 H^2 = \sigma^2 H = \sigma^2 X(X'X)^{-1}X'$$

Or $var(\hat{y}) = var(X\hat{\beta}) = Xvar(\hat{\beta})X' = \sigma^2 X(X'X)^{-1}X'$

Problem 4

(a)

```
library(faraway)
data("uswages")
myfit=lm(wage~educ, data=uswages)
plot(wage~educ, data=uswages, xlab="Years of education", ylab="Weekly wages")
abline(myfit$coefficients[1],myfit$coefficients[2],col="red")
```




```
summary(myfit)
```

```
##
## Call:
## lm(formula = wage ~ educ, data = uswages)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -743.6 -269.5  -67.7   173.0 7492.3
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  109.754     44.616    2.46    0.014 *
## educ          38.011      3.317   11.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 445.5 on 1998 degrees of freedom
## Multiple R-squared:  0.06167,    Adjusted R-squared:  0.0612
## F-statistic: 131.3 on 1 and 1998 DF,  p-value: < 2.2e-16
```

(b)

```r
summary(myfit)$r.sq
```

```
## [1] 0.06167066
```

The coefficient of determination is 0.0617.

(c)

```r
which.max(abs(myfit$residuals))
```

```
## 15387
##  1576
```

The case number is 15387(which is the 1576th observation in the data set).

(d)

```r
mean(myfit$residuals)
```

```
## [1] 5.784567e-15
```

```r
median(myfit$residuals)
```

```
## [1] -67.67192
```

The mean of residuals is 0, which coincides with the normal equations of SLR with intercept that the summation of residuals eqauls 0. The median of residuals is -67.672, which shows that the middle number of ordered residuals is negative. It indicates that the residuals are skewed to the right and we may expect there are some relatively large residuals on the positive part.

(e)

```r
coef(myfit)
```

```
## (Intercept)        educ
##   109.75385    38.01114
```

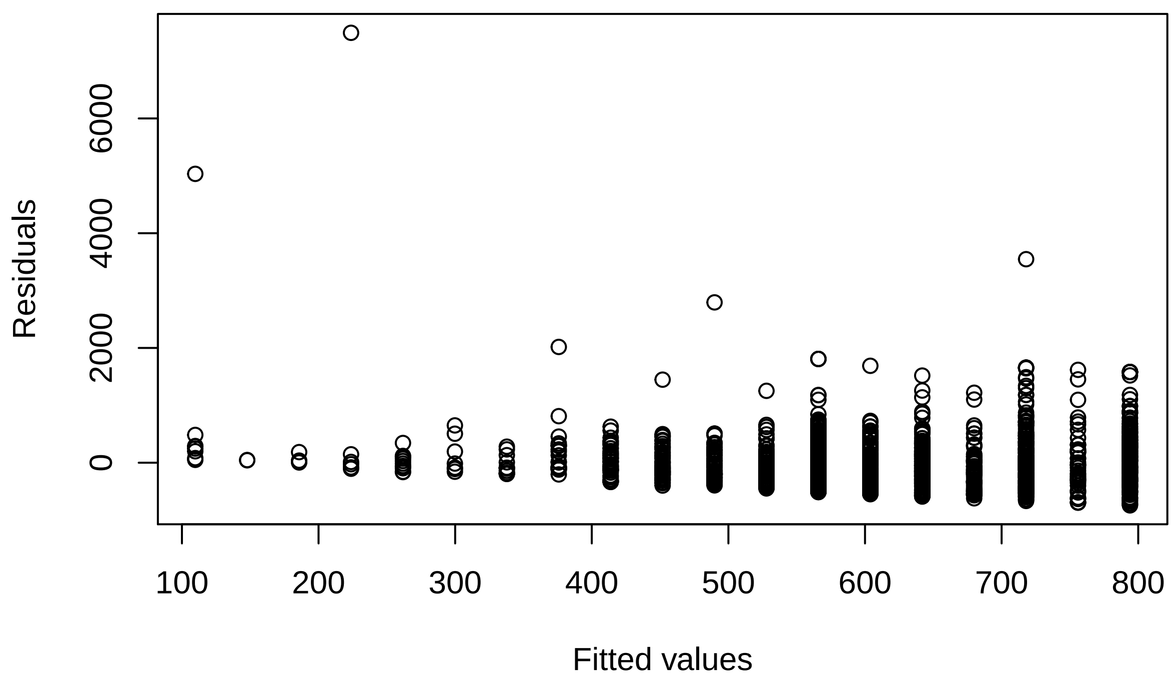
The coefficient of *educ* represents the increment of predicted weekly wage corresponding to a unit increment in *educ* with other predictors(intercept in this single model) remain the same, so the difference of two predicted weekly wages is 38.01114.

(f)

```r
cor(residuals(myfit), fitted(myfit))
```

```
## [1] 1.186985e-16
```

```r
plot(residuals(myfit)~fitted(myfit), xlab="Fitted values", ylab="Residuals")
```



The sample correlation between residuals and fitted values must be zero in linear models. Since the fitted values are the projections of responses onto the space spanned by columns of predictors, while the residuals are the remaining composition (the difference between the observations and the fitted values). Hence, fitted values must be orthogonal to residuals. The zero correlation just reflects this property.