**Q1~Q4**

Q1.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{S_{xx}} = \frac{\sum (x_i y_i - x_i \bar{Y} - \bar{X} y_i + \bar{X}\bar{Y})}{S_{xx}}$$

$$= \frac{\sum (x_i y_i) - \sum (x_i \bar{Y}) - \sum (\bar{X} y_i) + \sum (\bar{X}\bar{Y})}{S_{xx}}$$

$$= \frac{\sum (x_i y_i) - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y}}{S_{xx}} = \frac{\sum (x_i y_i) - n\bar{X}\bar{Y}}{S_{xx}} = \frac{\sum (x_i - \bar{X})}{S_{xx}} y_i$$

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum (x_i - \bar{X}) y_i}{S_{xx}}\right) = \frac{(\sum (x_i - \bar{X}))^2}{(\sum (x_i - \bar{X})^2)^2} Var(y_i) = \frac{1}{S_{xx}} \sigma^2$$

(a). $cov(\hat{\beta}_0, \hat{\beta}_1) = cov(\bar{Y} - \hat{\beta}_1 \bar{X}, \hat{\beta}_1) = cov(\bar{Y}, \hat{\beta}_1) - cov(\hat{\beta}_1 \bar{X}, \hat{\beta}_1)$

$$= 0 - \bar{X} cov(\hat{\beta}_1, \hat{\beta}_1) = -\bar{X} var(\hat{\beta}_1) = \frac{-\bar{X}\sigma^2}{S_{xx}} \ \#.$$

(b) $cov(\bar{y}, \hat{\beta}_1) = cov\left(\frac{\sum y_i}{n}, \frac{\sum (x_i - \bar{X}) y_i}{S_{xx}}\right) = \frac{1}{nS_{xx}} cov\left(\sum y_i, \sum (x_i - \bar{X}) y_i\right)$

$$= \frac{\sum cov(y_i, (x_i - \bar{X}) y_i)}{nS_{xx}} = \frac{\sum (x_i - \bar{X}) cov(y_i, y_i)}{nS_{xx}} = \frac{\overset{0}{(n\bar{X} - n\bar{X})}\sigma^2}{nS_{xx}} = 0 \ \#$$

Q2:

$$SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} \hat{y}_i \bar{y} + \sum_{i=1}^{n} \bar{y}^2$$

$$= \sum_{i=1}^{n} \hat{y}_i^2 - 2\bar{y} \sum_{i=1}^{n} \hat{y}_i + n\bar{y}^2 \ \#$$

Q3: (a)

$$HH = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$$

$$= X\underbrace{(X^TX)^{-1}X^TX}_{=I}(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H \quad \ldots\ldots ①.$$

$$(I-H)^2 = I - 2H + H^2 = I - 2H + H \quad \because H^2 = HH = H$$

$$= I - H \quad \ldots\ldots ②.$$

Based on ① and ②, $H$ is idempotent.

(b).

let $\hat{Y} = X(X^TX)^{-1}X^TY = HY$

$$Var(\hat{Y}) = Var(HY) = HH \cdot Var(Y) = H \cdot Var(Y)$$

$\because Y = X\beta + \varepsilon$, $X\beta$ is a constant

$$\therefore Var(Y) = Var(X\beta + \varepsilon) = Var(\varepsilon) = \sigma^2$$

$$\Rightarrow Var(\hat{Y}) = H \times Var(Y) = H\sigma^2 \quad \#$$

---

Q4:

$$R^2 = \frac{SS_R}{SS_{total}} \Rightarrow \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{=SS_R}, \quad \hat{y}_i = \underline{\hat{\beta}_0 + \hat{\beta}_1 x_i}.$$

$$SS_{Total} = \sum_{i=1}^{n}(y_i - \bar{y})^2, \quad y_i = \underline{\beta_0 + \beta_1 x_i + \varepsilon_i}.$$

We can easily see that $SS_{total}$ must bigger than $SS_R$ which means $R^2$ cannot larger than 1 or smaller than 0 #.

**Q5:**

| | Variables | VIF |
|---|---|---|
| 0 | const | 763.557531 |
| 1 | cylinders | 10.737535 |
| 2 | origin | 1.772386 |
| 3 | acceleration | 2.625806 |
| 4 | model year | 1.244952 |
| 5 | horsepower | 9.943693 |
| 6 | weight | 10.831260 |
| 7 | displacement | 21.836792 |

When constructing a multivariate linear regression model, we need to check whether the model has a collinearity problem, which can lead to enhanced predictive power of certain variables with respect to each other, thus reducing the accuracy of the model.

From the figure, we can see that the VIF values of cylinder, displacement, and weight are all greater than 10, i.e., they are highly correlated, which makes them have better predictive power than other variables. In this condition, we can combine them into one variable.