2019

Question 1

a)

Already done in 2014 - i) is done as a question itself, ii) is done as part of one of the questions.

b)

$$\begin{aligned}
\text{Cov}(AY, BZ) &= \text{E}((AY - E(AY))(BZ - E(BZ))^T) \\
&= \text{E}((AY - AE(Y))(Z^T B^T - E(Z)B^T)) \\
&= \text{E}(A(Y - E(Y))(Z - E(Z))^T B^T) \\
&= A\text{Cov}(Y, Z)B^T \\
&= AVB^T
\end{aligned}$$

c)

1. Just use the formula you just derived.

$$\begin{aligned}
\text{Cov}((I - H)y, (X^T X)^{-1} X^T y) &= (I - H)\text{Cov}(y, y)X(X^T X)^{-1} \\
&= (X(X^T X)^{-1} - X(X^T X)^{-1} X^T X(X^T X)^{-1})\text{Cov}(y, y) \\
&= (X(X^T X)^{-1} - X(X^T X)^{-1})\text{Cov}(y, y) \\
&= \mathbf{0}
\end{aligned}$$

2. Use the expectation of a quadratic form, given to you in the formula sheet.

3.

$$\begin{aligned}
E(\hat{\sigma}^2) &= \frac{E(y^T(I - H)y)}{n - p} \\
&= \frac{tr((I - H)\sigma^2 I) + B^T X^T (I - X(X^T X)^{-1} X^T)XB}{n - p} \\
&= \frac{tr((I - H)\sigma^2 I) + (B^T X^T - B^T X^T)XB}{n - p} \\
&= \frac{tr((I_n - X(X^T X)^{-1} X^T))\sigma^2 I}{n - p} \\
&= \frac{\sigma^2 tr(I_n) - \sigma^2 tr(X(X^T X)^{-1} X^T)}{n - p} \\
&= \frac{n\sigma^2 - \sigma^2 tr((X^T X)^{-1} X^T X)}{n - p} \\
&= \frac{n\sigma^2 - p\sigma^2}{n - p} \\
&= \sigma^2
\end{aligned}$$

Thereby, unbiased.

Question 2

a)

1. (a) Errors, and hence responses, are uncorrelated.
   (b) Responses are a linear combination of predictors.
   (c) Error variance is constant.
   (d) Errors are normally distributed $\sim N(0, \sigma^2)$, and thus responses are also normally distributed.

2. First diagnostic plot has a fan shape in the residuals vs fitted plot, which indicates non-constant variance. It has skew in the QQ plot, which violates the normality assumption. It also has some shape to the to the residuals vs fitted plot, which may mean the responses are not a linear combination of predictors.

   Second diagnostic has less/no fan shape - but still has a minor shape to the graph. Normality looks a lot better with less skew. The second data set is of course much better.

3. You can transform the data to stabilise variance, or you can used weighted least squares regression.

b)

1. 77.79% (Not *too* sure whether to use adjusted or not here... but purely by definition), R^2 should be used here?

2. The p-value is $< 2 \times 10^{-16}$. This indicates that with a significance level of 5%, we can reject the null hypothesis that $\beta_{weight} = 0$, and that is has statistical significance in ouir model.

3. $t_{0.975,394} = 1.966$. Therefore, we have:

$$(-2.424 \times 10^{-4}) \pm 1.966 \times (2.652 \times 10^{-5})$$

4. We have the statistic:

$$\frac{R(\text{displacement}|\text{weight, cylinders}) + R(\text{weight}|\text{cylinders})/2}{SS_{res}/394} = 85.129$$

   This is much larger tha $F_{0.95,2,394}$, so we can reject the null hypothesis.

5. This is just a t-test testing whether $\beta_{displacement} = 0$. The p-value is 0.0317. There is enough evidence to reject the null hypothesis.

Question 3

This is already done in 2014.

Question 4

a)

We didn't learn logistic regression.

b)

1. Consider that $H_{ii} = x_i^T (X^T X)^{-1} x_i$.

$$
\begin{aligned}
\frac{\sum_{i=1} \mathrm{Var}(\hat{y}_i)}{\sigma^2} &= \frac{\sum_{i=1}^n \mathrm{Var}(x_i^T b)}{\sigma^2} \\
&= \frac{\sum_{i=1}^n x_i^T \mathrm{Var}(y_i) x_i}{\sigma^2} \\
&= \frac{\sum_{i=1}^n \sigma^2 x_i^T (X^T X)^{-1} x_i}{\sigma^2} \\
&= \frac{\sum_{i=1}^n \sigma^2 H_{ii}}{\sigma^2} \\
&= \sum_{i=1}^n H_{ii} \\
&= p
\end{aligned}
$$

2. R^2 considers the raw SS_{reg}/SS_{res}. Adjusted adjusts for the fact that for GLM, then more params you add, the higher R^2 becomes - so it weights the SS_{reg} by the number of parameters, and the SS_{res} by the number of parameters. Mallow's C_p considers a subset model's bias-variance tradeoff; considering whether the usage of p parameters effectively reduces the variance of the model compared to the full model's variance.