Problem 1
Tell true or false of the following statements and briefly explain your answer.

(a) If the response variable is uncorrelated with all X variables in the model, then the least-squares estimated regression coefficients of the X variables are all zero.

(b) Even when the X variables are perfectly correlated, we might still get a good fit of the data.

(c) Taking transformations of the X variables as in the standardized regression model (referred to as correlation transformation) will not change coefficients of multiple determination.

(d) In a regression model, it is possible that none of the X variables is statistically significant when being tested individually, while there is a significant regression relation between the response variable and the set of X variables as a whole.

(e) In a regression model, it is possible that some of the X variables are statistically significant when being tested individually, while there is no significant regression relation between the response variable and the set of X variables as a whole.

(f) If an X variable is uncorrelated with the rest of the X variables, then in the standardized regression model, the variance of its least-squares estimated regression coefficient equals to the error variance.

(g) If an X variable is uncorrelated with the response variable and also is uncorrelated with the rest of the X variables, then its least-squares estimated regression coefficient must be zero.

(h) If the coefficient of multiple determination of regressing an X variable to the rest of the X variables is large, then its least-squares estimated regression coefficient tends to have large sampling variability.

Solution:

(a) True. $\beta = \sqrt{n-1}S_Y r_{XX}^{-1} r_{XY} = 0$

(b) True. This is the idea of multicollinearity.

(c) True. You have not changes the overall relationship between the explanatory and response variables. You have only scaled them.

(d) True. A simple example is $Y = X_1 + X_2$. Only regression on one variable at a time, it's easy to get a coefficient of zero.

(e) Mostly false. You shouldn't have a test where too many explanatory variables stop the significance of a subset.

(f) False. It will equal $\sqrt{n-1}S_Y$.

(g) True. This is because the coefficient should be $\sqrt{n-1}S_Y r_{XX}^{-1} r_{XY} e_k$ where k is that variable. Since $r_{XX}^{-1} r_{XY} e_k = 0$, it should drop out entirely.

(h) False. The two concepts are independent.

Problem 2
When $X_1, \ldots, X_{p-1}$ are uncorrelated, show the following results. *Hint: Show these results under the standardized regression model and then transform them back to the original model.*

(a) The fitted regression coefficients of regressing Y on $(X_1, \ldots, X_{p-1})$ equal to the fitted regression coefficients of regressing Y on each individual X_j $(j = 1, \ldots, p-1)$ alone.

(b) Let $X(-j) := \{X_k : 1 \leq k \leq p-1, k \neq j\}$. Show that $SSR(X_j|X(-j)) = SSR(X_j)$, where $SSR(X_j)$ denotes the regression sum of squares when regressing Y on X_j alone.

Solution:

(a) When the variables $X_1, \dots, X_{p-1}$ are uncorrelated, then

$$r_{XX} = I_{p-1}$$

Then we see that

$$\hat{\beta} = \left(\frac{\overline{Y}}{\sqrt{n-1} S_Y r_{XX}^{-1} r_{XY}} \right) = \left(\frac{\overline{Y}}{\sqrt{n-1} S_Y r_{XY}} \right)$$

where each r_{XY} is equal to the coefficient between each explanatory variable with the response variable.

(b) Observe,

$$SSR(X_j|X_{(-j)}) = SSR(X_1, \dots, X_{p-1}) - SSR(X_{(-j)}) = SSR(X_j)$$

I'm a little tired ... I feel like we didn't get a lot of time to finish this and do this quiz this week.

Problem 3

Variance Inflation Factor for models with $2X$ variables. Show that for a model with two X variables, X_1 and X_2, the variance inflation factors are

$$VIF_1 = VIF_2 = \frac{1}{1-R_1^2} = \frac{1}{1-R_2^2}$$

(Hint: Note $R_1^2 = R_2^2 = r_{12}^2$, where r_{12} is the sample correlation coefficient between X_1 and X_2.)

Solution: Observe, we see that $X_1 \sim X_2$ and $X_2 \sim X_1$ has the same coefficient of determination, specifically,

$$R_1^2 = R_2^2 = \left(\frac{Cov(X_1, X_2)}{\sqrt{Var(X_1, X_2)}} \right)^2 = r_{12}^2$$

Now, for the correlation matrix, we know that

$$r_{XX} = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}$$

Now, we see that,

$$r_{XX}^{-1} = \frac{1}{det(r_{XX})} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix} = \frac{1}{1-r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{pmatrix}$$

Since VIF_k are the diagonal elements, we see that,

$$VIF_1 = VIF_2 = \frac{1}{1-r_{12}^2} = \frac{1}{1-R_1^2} = \frac{1}{1-R_2^2}$$