Problem 1
Derive E(SSTO) and E(SSR) under the simple regression model using matrix algebra.

Solution:

- $SSTO = Y'\left[1 - \frac{1}{n}J\right]Y$. Define

$$E[Y] = X\beta \quad \text{and} \quad Cov(Y) = \Sigma = \sigma^2 I$$

Then we see that:

$$\begin{aligned}
E[SSTO] &= E\left[Y'\left[I - \frac{1}{n}J\right]Y\right] \\
&= tr(\left[I - \frac{1}{n}J\right]\Sigma) + \beta' X'\left[I - \frac{1}{n}J\right]X\beta \\
&= \sigma^2 tr(\left[I - \frac{1}{n}J\right]) + \beta' X'\left[I - \frac{1}{n}J\right]X\beta \\
&= (n-1)\sigma^2 + \beta' X'\left[I - \frac{1}{n}J\right]X\beta \\
&= (n-1)\sigma^2 + tr\left(\left[I - \frac{1}{n}J\right]X\beta\beta' X'\right) \\
&= (n-1)\sigma^2 + tr(X\beta\beta' X') - \frac{1}{n}tr(JX\beta\beta' X')
\end{aligned}$$

Then we see that:

$$\begin{aligned}
X\beta\beta' X' &= \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \begin{pmatrix} \beta_0 & \beta_1 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \\
&= \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0^2 & \beta_0\beta_1 \\ \beta_0\beta_1 & \beta_1^2 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \\
&= \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0^2 + \beta_0\beta_1 x_1 & \beta_0^2 + \beta_0\beta_1 x_2 & \dots & \beta_0^2 + \beta_0\beta_1 x_n \\ \beta_0\beta_1 + \beta_1^2 x_1 & \beta_0\beta_1 + \beta_1^2 x_2 & \dots & \beta_0\beta_1 + \beta_1^2 x_n \end{pmatrix}
\end{aligned}$$

$$\implies tr(X\beta\beta' X') = n\beta_0^2 + 2\beta_0\beta_1\sum_{i=1}^n x_i + \beta_1^2\sum_{i=1}^n x_i^2$$

$$JX\beta\beta'X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0^2 + \beta_0\beta_1 x_1 & \beta_0^2 + \beta_0\beta_1 x_2 & \dots & \beta_0^2 + \beta_0\beta_1 x_n \\ \beta_0\beta_1 + \beta_1^2 x_1 & \beta_0\beta_1 + \beta_1^2 x_2 & \dots & \beta_0\beta_1 + \beta_1^2 x_n \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum_{i=1}^n x_i \\ n & \sum_{i=1}^n x_i \\ \vdots & \vdots \\ n & \sum_{i=1}^n x_i \end{pmatrix} \begin{pmatrix} \beta_0^2 + \beta_0\beta_1 x_1 & \beta_0^2 + \beta_0\beta_1 x_2 & \dots & \beta_0^2 + \beta_0\beta_1 x_n \\ \beta_0\beta_1 + \beta_1^2 x_1 & \beta_0\beta_1 + \beta_1^2 x_2 & \dots & \beta_0\beta_1 + \beta_1^2 x_n \end{pmatrix}$$

$$= n \begin{pmatrix} 1 & \overline{X} \\ 1 & \overline{X} \\ \vdots & \vdots \\ 1 & \overline{X} \end{pmatrix} \begin{pmatrix} \beta_0^2 + \beta_0\beta_1 x_1 & \beta_0^2 + \beta_0\beta_1 x_2 & \dots & \beta_0^2 + \beta_0\beta_1 x_n \\ \beta_0\beta_1 + \beta_1^2 x_1 & \beta_0\beta_1 + \beta_1^2 x_2 & \dots & \beta_0\beta_1 + \beta_1^2 x_n \end{pmatrix}$$

$$\implies \frac{1}{n} tr(JX\beta\beta'X') = n\beta_0^2 + 2n\beta_0\beta_1\overline{X} + n\beta_1^2\overline{X}^2$$

Therefore, we see that,

$$\begin{aligned} E[SSTO] &= (n-1)\sigma^2 + n\beta_0^2 + 2n\beta_0\beta_1\overline{X} + \beta_1^2 \sum_{i=1}^n x_i^2 - (n\beta_0^2 + 2n\beta_0\beta_1\overline{X} + n\beta_1^2\overline{X}^2) \\ &= (n-1)\sigma^2 + \beta_1^2 \sum_{i=1}^n x_i^2 - n\beta_1^2\overline{X}^2 \\ &= (n-1)\sigma^2 + \beta_1^2 \sum_{i=1}^n (x_i - \overline{X})^2 \end{aligned}$$

- $SSR = Y'\left[H - \frac{1}{n}J\right]Y$. Observe, we leverage the fact that

$$SSR = SSTO - SSE$$

and the linearity of expectation:

$$E[SSR] = E[SSTO - SSE] = E[SSTO] - E[SSE] = (n-1)\sigma^2 + \beta_1^2 \sum_{i=1}^n (x_i - \overline{X})^2 - (n-1)\sigma^2 = \beta_1^2 \sum_{i=1}^n (x_i - \overline{X})^2$$

Problem 3
For each of the following models, answer whether it can be expressed as a multiple regression model or not. If so, indicate which transformations and/or new variables need to be introduced. (In the following, ϵ_i's denote the error terms and are assumed to be i.i.d. random variables.)

(a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \epsilon_i$

(b) $Y_i = \epsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2)$ with $\epsilon_i > 0$

(c) $Y_i = \beta_0 \exp(\beta_1 X_{i1}) + \epsilon_i$

(d) $Y_i = \frac{1}{1+\exp(\beta_0 + \beta_1 X_{i1} + \epsilon_i)}$

Solution:

(a) Yes. Order and interaction variable sums in X are always expressible as multiple regression models.

(b) Yes. We can rewrite this relationship as:

$$Y_i = \epsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2 + \ln(\epsilon_i))$$

which is well defined since we assume $\epsilon_i > 0$. We can then use the transform:

$$\tilde{Y}_i = \ln(Y_i)$$

So we are left with learning

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2 + \ln(\epsilon_i)$$

which is expressible as a multiple regression model since it's a polynomial.

(c) No. Unless we can assume all of Y_i is positive or negative (or sufficiently close to). Otherwise, there is no way to express this without doing the transformation:

$$\tilde{Y}_i = \ln(Y_i) = \ln(\beta_0) + \beta_1 X_{i1}$$

And then we simply choose β_0 to have sign equivalent to the majority sign of Y_i.

(d) Yes provided $Y_i \in (0, 1)$ for all i. Define the transformation:

$$\tilde{Y}_i = \ln\left(\frac{1}{Y_i} - 1\right)$$

Then we see that

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$$

which is a polynomial in terms of X and is therefore expressible!

Problem 4
Answer the following questions with regard to multiple regression models and provide a brief explanation.

(a) What is the maximum number of X variables that can be included in a multiple regression model (with intercept) that is used to fit a data set with 10 cases?

(b) With 4 predictors, how many X variables are there in the interaction model with all main effects and all interaction terms (2nd order, 3rd order, etc.)?

Solution:

(a) 9. Otherwise you will simply overfit your dataset. Between every 2 points, there is a unique line. Between every 10 points, there is a unique 9-dimensional hyperplane. Anything beyond and you will enter the overparametrized regime.

(b) This is a dynamic programming problem. Observe, you simply add the previous row entry with the previous column entry to get:

order \dependents	x	x, y	x, y, z	x, y, z, w
1	2	3	4	5
2	3	6	10	15
3	4	10	20	35
4	5	15	35	70
5	6	21	56	126

This is equivalent to the number of multi-combinations of choose d dimension elements amongst n variables (including the intercept term), which is

$$\binom{n+d}{d}$$

Therefore, setting $n = 4$, we get:

$$\text{number of X variables of 4 predictors} = \binom{4+d}{d}$$

Problem 5
Tell true or false of the following statements with regard to multiple regression models.

(a) The multiple coefficient of determination R^2 is always larger/not-smaller for models with more X variables.

(b) If all the regression coefficients associated with the X variables are estimated to be zero, then $R^2 = 0$.

(c) The adjusted multiple coefficient of determination R_a^2 may decrease when adding additional X variables into the model.

(d) Models with larger R^2 is always preferred.

(e) If the response vector is a linear combination of the columns of the design matrix X, then the coefficient of multiple determination $R^2 = 1$.

Solution:

(a) True. SSTO remains the same when introducing a new variable.

(b) False. The response $Y = \beta_0 + \epsilon$ will yield R^2 greater than 0.

(c) True. A decrease in SSE may be more than offset by the loss of degrees of freedom in SSE.

(d) False. It may suggest overfit.

(e) True. SSE = 0 when this is true.

Problem 6
Submitted as a Markdown file.

Problem 7
Under the multiple regression model, show that the residuals are uncorrelated with the fitted values and the estimated regression coefficients.

Solution:

- Observe:

$$\begin{aligned}
Cov(e, \hat{Y}) &= Cov\left((I-H)Y, HY\right) \\
&= E\left[(I-H)(Y-\overline{Y})(Y-\overline{Y})^T H^T\right] \\
&= (I-H)E\left[(Y-\overline{Y})(Y-\overline{Y})^T\right] H^T \\
&= (I-H)Cov(Y)H^T \\
&= (I-H)\sigma^2 I H^T \\
&= \sigma^2 (I-H)H \\
&= \sigma^2 (H-H) \\
&= 0
\end{aligned}$$

- $$\begin{aligned}
Cov(e, \hat{\beta}) &= Cov\left((I-H)Y, (X^TX)^{-1}X^TY\right) \\
&= E\left[(I-H)(Y-\overline{Y})(Y-\overline{Y})^TX(X^TX)^{-T}\right] \\
&= (I-H)E\left[(Y-\overline{Y})(Y-\overline{Y})^T\right]X(X^TX)^{-1} \\
&= (I-H)Cov(Y)X(X^TX)^{-1} \\
&= (I-H)\sigma^2IX(X^TX)^{-1} \\
&= \sigma^2(I-(X^TX)^{-1}X^T)X(X^TX)^{-1} \\
&= \sigma^2(X(X^TX)^{-1}-(X^TX)^{-1}X^TX(X^TX)^{-1}) \\
&= \sigma^2(X(X^TX)^{-1}-(X^TX)^{-1}) \\
&= \sigma^2(X-I)(X^TX)^{-1}
\end{aligned}$$

Since we can choose columns in X that yield a basis, we see that, $col(X) \perp (X-I)(X^TX)^{-1}$. Therefore,

$$(X-I)(X^TX)^{-1} = 0$$

which leaves us with

$$Cov(e, \hat{\beta}) = 0$$