1. Consider the following data set on Y and two independent variables x_1 and x_2:

Y	x_1	x_2
5	1	1
5	1	-1
6	-1	1
8	-1	-1

I want you to do the following parts **by hand**, and show all of your work. You can use R to check your work.

(a) Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

for $i = 1, 2, 3, 4$. Write this model as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

(b) Calculate the least squares estimate $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. *Hint:* $\mathbf{X}'\mathbf{X}$ should be a diagonal matrix so inverting it is easy.

(c) Calculate the hat matrix $\mathbf{H}$. Calculate the fitted value vector $\widehat{\mathbf{Y}}$ and the residual vector $\mathbf{e}$. Show $\widehat{\mathbf{Y}}$ and $\mathbf{e}$ are orthogonal vectors in $\mathbb{R}^4$. Also, show the residuals sum to zero.

(d) Find the (estimated) variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$. What is the estimated standard error of $\widehat{\beta}_1$?

2. The brake horsepower (`HORSE`, Y) developed by an automobile engine is thought to be a function of the engine speed in revolutions per minute (`RPM`, x_1), the road octane number of the fuel (`OCT`, x_2), and the engine compression (`COM`, x_3). An experiment is run in a laboratory at twelve different times; on each run, the temperature (`TEMP`, x_4) is also recorded. The data from the experiment are below.

Y	x_1	x_2	x_3	x_4
225	2000	90	100	71.2
212	1800	94	95	70.3
229	2400	88	110	72.3
222	1900	91	96	69.9
219	1600	86	100	73.2
278	2500	96	110	70.0
246	3000	94	98	70.7
237	3200	90	100	70.8
233	2800	88	105	72.1
224	3400	86	97	71.8
223	1800	90	100	71.1
230	2500	89	104	70.6

Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i,$$

for $i = 1, 2, ..., 12$, or, in matrix notation,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

(a) Use R to calculate $\mathbf{X}'\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{X}'\mathbf{Y}$, and finally $\widehat{\boldsymbol{\beta}}$ as we did in the notes for the waste data in Example 12.2.
(b) Use R to calculate the hat matrix $\mathbf{H}$.

3. Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2(3x_i^2 - 2) + \epsilon_i,$$

for $i = 1, 2, 3$, where $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$.
(a) Put this model into $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ form.
(b) Find the least squares estimates of β_0, β_1, and β_2. *Hint:* $\mathbf{X}'\mathbf{X}$ is diagonal, so inverting this matrix is easy.
(c) Show that the least squares estimates of β_0 and β_1 are unchanged if $\beta_2 = 0$. Why do you think this happens? *Hint:* What do you note about the 3 column vectors of $\mathbf{X}$ in part (a)?

4. *Hat matrix fun.* Consider the multiple linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{X}$ is $n \times p$. Let $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ denote the hat matrix. Let $\mathbf{I}$ denote the identity matrix that has the same dimensions as $\mathbf{H}$.
(a) What are the dimensions of $\mathbf{H}$?
(b) Show that $\mathbf{H}$ is symmetric and idempotent.
(c) Show that $\mathbf{I} - \mathbf{H}$ is symmetric and idempotent.
(d) Show that $\mathbf{H}\mathbf{X} = \mathbf{X}$ and $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$.
(e) Show that $\mathbf{H}\mathbf{Y} = \mathbf{X}\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is the least-squares estimator of $\boldsymbol{\beta}$. Note that $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ is the vector of least-squares fitted values.
(f) Show that $(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{e}$, where $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$ is the vector of least-squares residuals.
(g) Show that $(\mathbf{H}\mathbf{Y})'(\mathbf{I} - \mathbf{H})\mathbf{Y} = 0$. What does this imply about the residuals and fitted values from a least-squares fit?
(h) Show that $(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$, the error (residual) sum of squares.