1.1. **Question 1 (Octave).** Download and install the GNU Octave linear algebra software using the instructions posted in the *Notes and Handouts* section of the MTH396 web page.

Download the data files `computational_assignment3_octave_x.csv` and `computational_assignment3_octave_y.csv` from the links in the assignments section of the MTH396 web page.

Start Octave and load the Y vector using the command:

```
Y=dlmread("computational_assignment3_octave_y.csv",',')
```

This represents miles per gallon data from the 2009 EPA data base.

Now load the X matrix with the command:

```
X=dlmread("computational_assignment3_octave_x.csv",',')
```

The first column of the X matrix is all ones and the second column represents the engine horsepower.

The purpose of this exercise is to use Octave to carry out the matrix operations used to perform a simple linear regression.

The estimates of the parameters are given by $(X'X)^{-1}X'Y$, which in Octave is coded as:

```
beta=inverse(X'*X)*X'*Y
```

The vector of errors or residuals is obtained as $(Y - X\beta)'(Y - X\beta)$,

```
e=Y-X*beta
```

The dataset has 781 elements, so Y, $X\beta$, and e are in $\mathbb{R}^{781}$. X has two linearly independent columns, so the dimension of the column space has dimension 2. The nullspace has dimension $781 - 2 = 779$.

An estimate of σ_e^2, the variance of each error term, is given by $e'e$ divided by the dimension of the column space. Calculate this estimate.

1.2. **Question 2 (R).** We will perform a simple linear regression, that is, a model of the form $Y = X\beta + e$ where (in the row view),

$$Y_i = \beta_1 + \beta_2 \cdot X_i + e_i$$

R is set up to automatically generate the appropriate X matrix from a specification of the model, so all we have to do is get the data into R and run the linear model procedure.

Because R can read data from a URL, the easiest way to get the data into R is just to read the `.csv` file from the link on a web page. The syntax for this command is:

```
epa<-read.table("URL",sep=",",fill=TRUE,header=TRUE)
```

To read the EPA data into R, copy and paste this into the R command line, then replace `URL` with:

```
http://www.sandgquinn.org/stonehill/MA225/notes/09tstcar.csv
```

This will read the entire 2009 EPA test data matrix into a data frame in R. To display the strucure of the result, type:

`str(epa)` The columns we are interested in are `etw`, which contains the vehicle weight, and `mpg`, which contains the mileage rating. However, first we need to restrict the data to only cars and only highway mileage. To do this, we will create a new dataframe that is a subset of our original one using the command:

```
epaCH<-subset(epa, C.H=="H" & car.truck=="C")
```

Now we make the column names of the data frame `epaCH` visible using

```
attach(epaCH)
```

You can verify that the subset and attach worked by typing

```
mpg
```

There should be 781 values in the `mpg` vector.

Now we have R perform the regression. Since R is designed to do this sort of thing, all we have to do is type:

```
summary(lm(mpg ~ etw))
```

In the "Coefficients:" section of the summary, the estimated values of beta are in the "Estimate" column. How do these compare with the values we obtained from Octave using $(X'X)^{-1}X'Y$?

Notice the "Residual standard error:" entry. How does the value compare to the square root of the value of $e'e/779$ from Octave?

How does the degrees of freedom value relate to the geometric interpretation of the linear model?