Module 5 - MLR models

STAT 401

Section 1: SLR models in matrix notation

The mathematics of SLR models can involve substantial algebra. Many calculations are made easier using matrix algebra. We define

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Similarly, we define

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

We can then write the SLR model as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Our model assumptions take the form of

$$E(\mathbf{e}) = E\left(\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}\right) = \begin{bmatrix} E(e_1) \\ E(e_2) \\ \vdots \\ E(e_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

and

$$V(\mathbf{e}) = \begin{bmatrix} V(e_1) & Cov(e_1, e_2) & \dots & Cov(e_1, e_n) \\ Cov(e_2, e_1) & V(e_2) & \dots & Cov(e_2, e_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(e_n, e_1) & Cov(e_n, e_2) & \dots & V(e_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

and

$$\mathbf{e} \sim N_n(\mathbf{0},\ \sigma^2\mathbf{I})$$

The OLS estimators are found by minimizing the function

$$RSS(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\mathbf{b})$$

which, after differentiating, is done by solving for $\mathbf{b}$ in:

$$-2\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{0} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

We can verify that this is the same answer we got with algebra:

$$\mathrm{X}^T\mathrm{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$(\mathrm{X}^T\mathrm{X})^{-1} = \frac{1}{SXX} \begin{bmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \frac{1}{SXX}\begin{bmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}\begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$= \begin{bmatrix} \bar{y} - \hat{\beta}_1\bar{x} \\ \frac{SXY}{SXX} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

```r
X <- cbind(1, Heights$mheight)
y <- matrix(Heights$dheight, 1375, 1)
solve(t(X) %*% X, t(X) %*% y)

##             [,1]
## [1,] 29.917437
## [2,]  0.541747

lm(dheight ~ mheight, data = Heights)

##
## Call:
## lm(formula = dheight ~ mheight, data = Heights)
##
## Coefficients:
## (Intercept)      mheight
##     29.9174       0.5417
```

Section 2: MLR models

Multiple linear regression models extend S.L.R. models by including more than one predictor. The general form of the model including p predictors is

$$E(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \quad \Leftrightarrow$$
$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + e_i,$$

or simply,

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \Leftrightarrow \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

The elements in $\boldsymbol{\beta}$ are frequently referred to as *regression coefficients*. The M.L.R. model assumptions are

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \Leftrightarrow \quad E(\mathbf{e}) = \mathbf{0}$$

$$V(\mathbf{y}) = \sigma^2\mathbf{I} \quad \Leftrightarrow \quad V(\mathbf{e}) = \sigma^2\mathbf{I}$$

$$\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{I}) \quad \Leftrightarrow \quad \mathbf{e} \sim N_n(\mathbf{0},\ \sigma^2\mathbf{I})$$

Section 3: Fitting MLR models

The derivation of $\hat{\boldsymbol{\beta}}$ for MLR models is identical to that for SLR models as presented in Section 1. Namely,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Like $\hat{\beta}_0$ and $\hat{\beta}_1$ in the SLR model, $\hat{\boldsymbol{\beta}}$ is a random variable with an expectation and variance.

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\mathbf{y}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

$$
\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= V((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T V(\mathbf{y})\mathbf{X}^T(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}
\end{aligned}
$$

An estimator for σ^2 follows similar logic as in the SLR case. We have assumed that $V(\mathbf{e}) = \sigma^2 \mathbf{I}$. See that,

$$V(\mathbf{e}) = E(\mathbf{e}\mathbf{e}^T) - E(\mathbf{e})E(\mathbf{e})^T = E(\mathbf{e}\mathbf{e}^T)$$

since $E(\mathbf{e}) = \mathbf{0}$. Therefore $E(\mathbf{e}\mathbf{e}^T) = \sigma^2 \mathbf{I}$. Since only the diagonal is non-zero, the sample version of the parameter $E(\mathbf{e}\mathbf{e}^T)$ would be

$$\frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2.$$

The above estimator can be re-written as

$$\frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{RSS}{n}$$

However, there are not n independent pieces of information in $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ since $p+1$ parameters $(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ must be estimated to compute $\mathbf{X}\hat{\boldsymbol{\beta}}$.

Thus, the estimator becomes

$$\hat{\sigma}^2 = \frac{RSS}{n - (p+1)}.$$

This estimator is unbiased for σ^2.

In the case that the errors are normally distributed, we can additionally say:

$$\hat{\boldsymbol{\beta}} \sim N_{p+1}(\boldsymbol{\beta},\ \sigma^2(\mathsf{X}^\top \mathsf{X})^{-1})$$

and

$$\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

with $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ being statistically independent.

Section 4: Interpretation of regression coefficients

The elements of $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$ can be interpreted as follows:

We interpret β_0 as the mean response when $x_1, x_2, \ldots, x_p$ are set to zero.

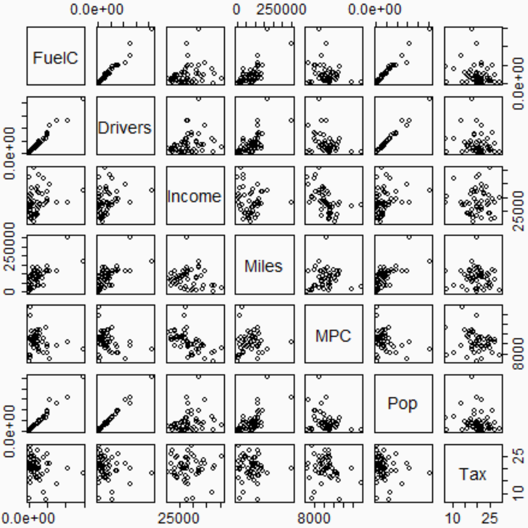
We interpret β_i for $i = 1, \ldots, p$ as the change in the mean response when x_i increases by 1 unit *and all other xs are held constant*.

For example, consider the `fuel2001` data, which reports the fuel consumption in thousands of gallons (`FuelC`) in 2001 for all 50 states and Washington DC. It also contains six other quantitative variables:

1. `Drivers`, the number of licensed drivers in the state
2. `Income`, annual personal income (per capita)
3. `Miles`, miles of federal-aid highway in the state
4. `MPC`, estimated miles driven per capita
5. `Pop`, population over age 15
6. `Tax`, gasoline state tax rate, cents per gallon

```r
head(fuel2001)
```

```
##      Drivers    FuelC Income  Miles      MPC      Pop  Tax
## AL   3559897  2382507  23471  94440 12737.00  3451586 18.0
## AK    472211   235400  30064  13628  7639.16   457728  8.0
## AZ   3550367  2428430  25578  55245  9411.55  3907526 18.0
## AR   1961883  1358174  22257  98132 11268.40  2072622 21.7
## CA  21623793 14691753  32275 168771  8923.89 25599275 18.0
## CO   3287922  2048664  32949  85854  9722.73  3322455 22.0
```


FuelC
Drivers
Income
Miles
MPC
Pop
Tax
0.0e+00
0 250000
0.0e+00
0.0e+00
0.0e+00
25000
0 250000
8000
0.0e+00
0.0e+00
25000
8000
10 25
10 25

```r
summary(lm(FuelC ~ ., data = fuel2001))
```

```
##
## Call:
## lm(formula = FuelC ~ ., data = fuel2001)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1480910  -158802    19267   174208  1090089
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.902e+05  8.199e+05  -0.598 0.552983
## Drivers      6.368e-01  1.452e-01   4.386 7.09e-05 ***
## Income       7.690e+00  1.632e+01   0.471 0.639793
## Miles        5.850e+00  1.621e+00   3.608 0.000784 ***
## MPC          4.562e+01  3.565e+01   1.280 0.207337
## Pop         -1.945e-02  1.245e-01  -0.156 0.876586
## Tax         -2.087e+04  1.324e+04  -1.576 0.122235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 398400 on 44 degrees of freedom
## Multiple R-squared:  0.9808,^^IAdjusted R-squared:  0.9782
## F-statistic: 374.6 on 6 and 44 DF,  p-value: < 2.2e-16
```

To interpret these:

The mean fuel consumption in the population, when drivers, income, miles of highway, miles driven, population, and tax rates are all 0, is estimated to be -490,200 thousands of gallons.

When personal per capita income increases by $1.00 and drivers, miles of highway, miles driven, population, and tax rates are held constant, the mean fuel consumption increases by 7.69 thousands of gallons.

The values of $\hat{\beta}$ span multiple orders of magnitude so that it is difficult to compare them. To remedy this, we center and scale the predictors and refit the model.

```r
fuel2001.stn <- as.data.frame(cbind(fuel2001$FuelC, scale(fuel2001[, -2])))
colnames(fuel2001.stn)[1] <- c("FuelC")
summary(Model <- lm(FuelC ~ ., data = fuel2001.stn))
```

```
##
## Call:
## lm(formula = FuelC ~ ., data = fuel2001.stn)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1480910  -158802    19267   174208  1090089
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2542786      55789  45.578  < 2e-16 ***
## Drivers      2541616     579451   4.386 7.09e-05 ***
## Income         34234      72646   0.471 0.639793
## Miles         309972      85910   3.608 0.000784 ***
## MPC            93044      72704   1.280 0.207337
## Pop           -91609     586467  -0.156 0.876586
## Tax           -94831      60179  -1.576 0.122235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 398400 on 44 degrees of freedom
## Multiple R-squared:  0.9808,^^IAdjusted R-squared:  0.9782
## F-statistic: 374.6 on 6 and 44 DF,  p-value: < 2.2e-16
```

To interpret these:

The mean fuel consumption in the population, when drivers, income, miles of highway, miles driven, population, and tax rates are all at their means, is estimated to be 2,542,786 thousands of gallons.

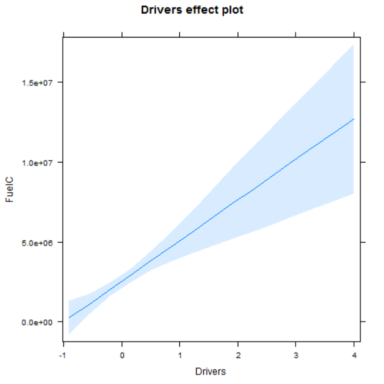
When personal per capita income increases by one standard deviation and drivers, miles of highway, miles driven, population, and tax rates are held constant, the mean fuel consumption increases by 34,234 thousands of gallons.

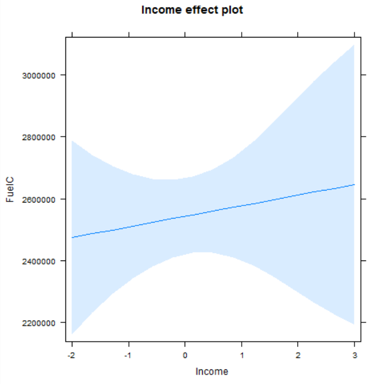
Notice that standardizing did not affect $\hat{\sigma}^2$ and its degrees of freedom.

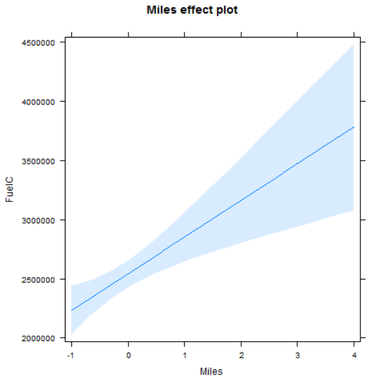
Section 5: Effect plots

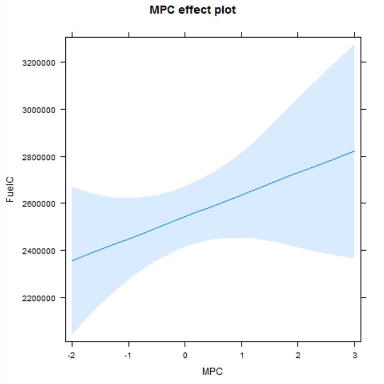
Effects plots for the `fuel2001` data are given illustrate the interpretations discussed in Section 4. Recall that the fitted model is

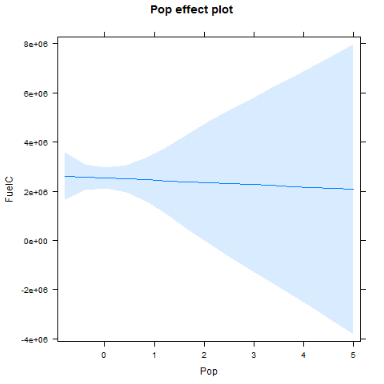
$$
\begin{aligned}
E(Y) = & 2,542,786 + 2,541,616 * x_1 + 34,234 * x_2 + \\
& 309,972 * x_3 + 93,044 * x_4 - 91,609 * x_5 - \\
& 94,831 * x_6
\end{aligned}
$$

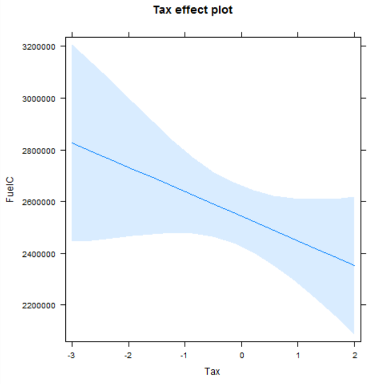

Drivers effect plot
1.5e+07
1.0e+07
5.0e+06
0.0e+00
FuelC
-1
0
1
2
3
4
Drivers


Income effect plot
3000000
2800000
2600000
2400000
2200000
FuelC
-2
-1
0
1
2
3
Income


Miles effect plot
FuelC
4500000
4000000
3500000
3000000
2500000
2000000
-1
0
1
2
3
4
Miles


MPC effect plot
FuelC
3200000
3000000
2800000
2600000
2400000
2200000
-2
-1
0
1
2
3
MPC


Pop effect plot
8e+06
6e+06
4e+06
2e+06
0e+00
-2e+06
-4e+06
FuelC
0
1
2
3
4
5
Pop


Tax effect plot
3200000
3000000
2800000
2600000
2400000
2200000
FuelC
-3
-2
-1
0
1
2
Tax

```r
fuel2001.stn <- as.data.frame(cbind(fuel2001$FuelC,
    scale(fuel2001[, -2])))
colnames(fuel2001.stn)[1] <- c("FuelC")
Model <- lm(FuelC ~ ., data = fuel2001.stn)
Effect("Drivers", Model)
```

```
##
##  Drivers effect
## Drivers
##       -0.9        0.5          2          3          4
##   255331.6  3813594.1  7626018.2 10167634.2 12709250.3
```

```r
2542786 + 2541616 * 2 + 34234 * mean(fuel2001.stn$Income) +
    309972 * mean(fuel2001.stn$Miles) +
    93044 * mean(fuel2001.stn$MPC) -
    91609 * mean(fuel2001.stn$Pop) -
    94831 * mean(fuel2001.stn$Tax)
```

```
## [1] 7626018
```