Chapter 5

STAT 3240

Michael McIsaac

UPEI

5: Matrix Approach to Simple Linear Regression Analysis

Learning Objectives for Sections 5.8-5.10

After Sections 5.8-5.10, you should be able to

- Write simple linear regression in matrix terms
- Write simple least squares estimation in matrix terms

5.9: Simple Linear Regression Model in Matrix Terms

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1, \ldots, n$$

This means

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned}$$

In matrix terms,

$$\underset{n\times 1}{\mathbb{Y}} = \underset{n\times 2}{\mathbb{X}} \underset{2\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Our usual regression assumptions:

$$E\{\underset{n\times 1}{\varepsilon}\} = E\left\{\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}\right\} = \begin{bmatrix}E\{\varepsilon_1\}\\ E\{\varepsilon_2\}\\ \vdots\\ E\{\varepsilon_n\}\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix} = \underset{n\times 1}{0}$$

$$\underset{n\times n}{\sigma^2\{\varepsilon\}} = \begin{bmatrix}\sigma^2 & 0 & \cdots & 0\\ 0 & \sigma^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma^2\end{bmatrix} = \sigma^2 \underset{n\times n}{\mathbb{I}}$$

Note that the Variance-Covariance Matrix of Random Vector is

$$\underset{n\times n}{\sigma^2\{\mathbb{Y}\}} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1,Y_2\} & \cdots & \sigma\{Y_1,Y_n\} \\ \sigma\{Y_2,Y_1\} & \sigma^2\{Y_2\} & \cdots & \sigma\{Y_2,Y_n\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{Y_n,Y_1\} & \sigma\{Y_n,Y_2\} & \cdots & \sigma^2\{Y_n\} \end{bmatrix} = E\{[\mathbb{Y}-E[\mathbb{Y}]][\mathbb{Y}-E[\mathbb{Y}]]'\}$$

For our illustration, we have:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \mathbf{E}\left\{\begin{bmatrix} Y_1 - E\{Y_1\} \\ Y_2 - E\{Y_2\} \\ Y_3 - E\{Y_3\} \end{bmatrix} [Y_1 - E\{Y_1\} \quad Y_2 - E\{Y_2\} \quad Y_3 - E\{Y_3\}]\right\}$$

Multiplying the two matrices and then taking expectations, we obtain:

Location in Product	Term	Expected Value
Row 1, column 1	$(Y_1 - E\{Y_1\})^2$	$\sigma^2\{Y_1\}$
Row 1, column 2	$(Y_1 - E\{Y_1\})(Y_2 - E\{Y_2\})$	$\sigma\{Y_1, Y_2\}$
Row 1, column 3	$(Y_1 - E\{Y_1\})(Y_3 - E\{Y_3\})$	$\sigma\{Y_1, Y_3\}$
Row 2, column 1	$(Y_2 - E\{Y_2\})(Y_1 - E\{Y_1\})$	$\sigma\{Y_2, Y_1\}$
etc.	etc.	etc.

$$
\begin{aligned}
E\{\underset{n\times 1}{\mathbb{Y}}\} &= E\{\underset{n\times 2}{\mathbb{X}}\underset{2\times 1}{\beta} + \underset{n\times 1}{\varepsilon}\} \\
&= E\left\{\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}\right\} \\
&= E\left\{\begin{bmatrix} \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 X_n + \varepsilon_n \end{bmatrix}\right\} \\
&= \begin{bmatrix} E\{\beta_0 + \beta_1 X_1 + \varepsilon_1\} \\ E\{\beta_0 + \beta_1 X_2 + \varepsilon_2\} \\ \vdots \\ E\{\beta_0 + \beta_1 X_n + \varepsilon_n\} \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} \\
&= \underset{n\times 2}{\mathbb{X}}\underset{2\times 1}{\beta}
\end{aligned}
$$

5.10: Least Squares Estimation of Regression Parameters

To derive the normal equations by the method of least squares, we minimize the quantity

$$\begin{aligned} Q &= \sum \left[Y_i - (\beta_0 + \beta_1 X_i)\right]^2 \\ &= (\mathbb{Y} - \mathbb{X}\beta)'(\mathbb{Y} - \mathbb{X}\beta) \\ &= \mathbb{Y}'\mathbb{Y} - \beta'\mathbb{X}'\mathbb{Y} - \mathbb{Y}'\mathbb{X}\beta + \beta'\mathbb{X}'\mathbb{X}\beta \\ &= \mathbb{Y}'\mathbb{Y} - 2\beta'\mathbb{X}'\mathbb{Y} + \beta'\mathbb{X}'\mathbb{X}\beta \end{aligned}$$

So, we find the values $\boldsymbol{b}$ such that

$$0 = \left.\frac{\partial}{\partial\beta} Q\right|_{\beta=b} = \left. -2\mathbb{X}'\mathbb{Y} + 2\mathbb{X}'\mathbb{X}\beta\right|_{\beta=b}$$

That is,

$$\mathrm{b} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}$$

Note that

$$\mathbb{X}'\mathbb{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

and

$$\mathbb{X}'\mathbb{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

The Canadian Survey of Household Spending (http://dli-idd-nesstar.statcan.gc.ca.proxy.library.upei.ca/webview/) is carried out annually across Canada. The main purpose of the survey is to obtain detailed information about household spending. Information is also collected about dwelling characteristics as well as household equipment.

```
#A subset of the latest Survey of Household Spending data are displa
spending_subset %>% datatable()
```

Show 200 entries Search:

	province	type_of_dwelling	income	marital_status	age_gr
1	NL	single_detached	68000	never_married	30-34
2	NL	single_detached	48000	never_married	25-29
3	NL	single_detached	30000	married	35-39
4	NL	row_house	30000	never_married	30-34
5	NL	single_detached	35000	married	25-29

Showing 1 to 30 of 30 entries Previous 1 Next

```
clothing_model = lm(clothing_expenditure~income, data=spending_subse
msummary(clothing_model)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.064e+03  7.832e+02   1.358   0.1853
## income      3.050e-02  1.651e-02   1.847   0.0753 .
##
## Residual standard error: 1663 on 28 degrees of freedom
## Multiple R-squared:  0.1086,	Adjusted R-squared:  0.07681
## F-statistic: 3.413 on 1 and 28 DF,  p-value: 0.07528
```

For the Canadian Survey of Household Spending example, the matrix representation of the simple linear regression involves the following vectors and matrices:

```
X = cbind(1, spending_subset$income)
Y = cbind(spending_subset$clothing_expenditure)
```

```
X %>%  datatable()
```

Show 200 entries
Search:

V1	V2
1	68000
1	48000
1	30000
1	30000
1	35000

Showing 1 to 30 of 30 entries

Previous 1 Next

```
Y%>%  datatable()
```

Show 200 entries
Search:

	V1
	4200
	1930
	2340
	2900
	1300

Showing 1 to 30 of 30 entries

Previous 1 Next

```r
# X'X
t(X) %*% X
```

```
##         [,1]       [,2]
## [1,]      30 1.3120e+06
## [2,] 1312000 6.7524e+10
```

```r
# (X'X)^(-1)
solve(t(X) %*% X)
```

```
##               [,1]          [,2]
## [1,]  2.218440e-01 -4.310458e-06
## [2,] -4.310458e-06  9.856230e-11
```

```r
# X'Y
t(X)%*% Y
```

```
##            [,1]
## [1,]      71922
## [2,] 3454802000
```

```r
# (X'X)^(-1) X'Y
solve(t(X) %*% X) %*% t(X)%*% Y
```

```
##              [,1]
## [1,] 1.063687e+03
## [2,] 3.049648e-02
```

- Interpret $(X'X)^{-1}X'Y$ in your own words.

- This is a way to find the slope and intercept of a regression model.

- This is using to find the estimated regression coefficients by the matrix method.

- In your own words, describe the value of this matrix approach to simple linear regression analysis?

- the matrix approach using matrix to contain the values of error, predictor variables, response variable and other value, and do the operations base on the matrix algebra rules. It should be an efficient way for computer to arrange, store and do the calculations.

- It permits extensive systems of equations and large arrays of data to be denoted compactly and operated upon efficiently.

- A matrix approach is not needed for simple linear regression, it is easier to use the method we have used earlier. But since we already know simple linear regression it is a good way of introducing matrix approach which will be very useful when we are dealing with multiple regression.

Recap: Sections 5.8-5.10

After Sections 5.8-5.10, you should be able to

- Write simple linear regression in matrix terms
- Write simple least squares estimation in matrix terms

Learning Objectives for Sections 5.11-5.13

After Sections 5.11-5.13, you should be able to

- Write fitted values and residuals in matrix terms
- Write ANOVA and regression inferences in matrix terms

5.11 Fitted Values and Residuals

In matrix terms,

$$\underset{n\times 1}{\hat{\mathbb{Y}}} = \underset{n\times 2}{\mathbb{X}}\ \underset{2\times 1}{\mathrm{b}}$$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix}$$

Remember that

$$\mathrm{b} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}$$

So,

$$\hat{\mathbb{Y}} = \mathbb{X}\ (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}$$

We call $\underset{n\times n}{\mathbb{H}} = \mathbb{X}\ (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$ the Hat Matrix because it puts a hat on Y (i.e., $\hat{\mathbb{Y}} = \mathbb{H}\mathbb{Y}$)

Note that the fitted values $\hat{\mathbb{Y}}$ are linear combinations of the observed values $\mathbb{Y}$ with weights given by the predictor variables through $\mathbb{H} = \mathbb{X}\,(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$.

Looking at the hat matrix can tell us how much influence each observation of X has on the overall fit; this is useful in diagnosing influential observations (Chapter 10).

An important property of $\mathbb{H}$ is that it is symmetric $(\mathbb{H}' = \mathbb{H})$ and idempotent $(\mathbb{H}\mathbb{H} = \mathbb{H})$.

Both of these properties should be obvious if you remember that

$$\mathbb{H} = \mathbb{X}\,(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'.$$

Residuals

In matrix notation,

$$\begin{aligned}\underset{n\times 1}{\mathbf{e}} &= \underset{n\times 1}{\mathbb{Y}} - \underset{n\times 1}{\hat{\mathbb{Y}}} \\ &= \underset{n\times 1}{\mathbb{Y}} - \underset{n\times 2}{\mathbb{X}}\ \underset{2\times 1}{\mathbf{b}} \\ &= \underset{n\times 1}{\mathbb{Y}} - \underset{n\times n}{\mathbb{H}}\ \underset{n\times 1}{\mathbb{Y}} \\ &= \left(\underset{n\times n}{\mathbb{I}} - \underset{n\times n}{\mathbb{H}}\right)\underset{n\times 1}{\mathbb{Y}}\end{aligned}$$

Remember that $\sigma^2\{aX\} = a^2\sigma^2\{X\}$.

Similarly,

$$\sigma^2\{\mathbb{A}\mathbb{Y}\} = \mathbb{A}\ \sigma^2\{\mathbb{Y}\}\ \mathbb{A}'.$$

Remember also that $\sigma^2\{\mathbb{Y}\} = \sigma^2\mathbb{I}$.

So, the variance-covariance matrix of the residuals can be expressed as

$$\begin{aligned}\underset{n\times n}{\sigma^2\{\mathrm{e}\}} &= (\mathbb{I}-\mathbb{H})\,\sigma^2\{\mathbb{Y}\}(\mathbb{I}-\mathbb{H})' \\ &= (\mathbb{I}-\mathbb{H})\left(\sigma^2\mathbb{I}\right)(\mathbb{I}-\mathbb{H})' \\ &= \sigma^2\,(\mathbb{I}-\mathbb{H})\,(\mathbb{I}-\mathbb{H}) \\ &= \sigma^2\,(\mathbb{I}-\mathbb{H})\end{aligned}$$

is estimated by

$$\underset{n\times n}{\mathrm{s}^2\{\mathrm{e}\}} = MSE\left(\underset{n\times n}{\mathbb{I}} - \underset{n\times n}{\mathbb{H}}\right)$$

Note that, like $\underset{n\times n}{\mathbb{H}}$, the matrix $\underset{n\times n}{\mathbb{I}} - \underset{n\times n}{\mathbb{H}}$ is symmetric and idempotent.

5.12 Analysis of Variance Results

$$
\begin{aligned}
SSE = \mathrm{e}'\mathrm{e} &= (\mathbb{Y} - \mathbb{X}\mathrm{b})'(\mathbb{Y} - \mathbb{X}\mathrm{b}) \\
&= \mathbb{Y}'\mathbb{Y} - 2\mathrm{b}'\mathbb{X}'\mathbb{Y} + \mathrm{b}'\mathbb{X}'\mathbb{X}\mathrm{b} \\
&= \mathbb{Y}'\mathbb{Y} - 2\mathrm{b}'\mathbb{X}'\mathbb{Y} + \mathrm{b}'\mathbb{X}'\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y} \\
&= \mathbb{Y}'\mathbb{Y} - 2\mathrm{b}'\mathbb{X}'\mathbb{Y} + \mathrm{b}'\mathbb{X}'\mathbb{Y} \\
&= \mathbb{Y}'\mathbb{Y} - \mathrm{b}'\mathbb{X}'\mathbb{Y}
\end{aligned}
$$

$$
SSTO = \sum (Y_i - \bar{Y})^2 = \mathbb{Y}'\mathbb{Y} - \frac{1}{n}\mathbb{Y}'\mathbb{J}\mathbb{Y}\,,
$$

where

$$
\mathbb{J} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}
$$

$$
SSR = SSTO - SSE = \mathrm{b}'\mathbb{X}'\mathbb{Y} - \frac{1}{n}\mathbb{Y}'\mathbb{J}\mathbb{Y}\,.
$$

Note that we can write these in *quadratic form* (i.e., in the form $\mathbb{Y}'\mathbb{A}\mathbb{Y}$, where $\mathbb{A}$ is symmetric):

$$SSTO = \mathbb{Y}'\mathbb{Y} - \frac{1}{n}\mathbb{Y}'\mathbb{J}\mathbb{Y} = \mathbb{Y}'(\mathbb{I} - \frac{1}{n}\mathbb{J})\mathbb{Y}$$

$$SSE = \mathbb{Y}'\mathbb{Y} - \mathrm{b}'\mathbb{X}'\mathbb{Y} = \mathbb{Y}'\mathbb{Y} - \hat{\mathbb{Y}}'\mathbb{Y} = \mathbb{Y}'\mathbb{Y} - \mathbb{Y}'\mathbb{H}'\mathbb{Y} = \mathbb{Y}'(\mathbb{I} - \mathbb{H})\mathbb{Y}$$

$$SSR = \mathrm{b}'\mathbb{X}'\mathbb{Y} - \frac{1}{n}\mathbb{Y}'\mathbb{J}\mathbb{Y} = \hat{\mathbb{Y}}'\mathbb{Y} - \frac{1}{n}\mathbb{Y}'\mathbb{J}\mathbb{Y} = \mathbb{Y}'\mathbb{H}\mathbb{Y} - \frac{1}{n}\mathbb{Y}'\mathbb{J}\mathbb{Y} = \mathbb{Y}'(\mathbb{H} - \frac{1}{n}\mathbb{J})\mathbb{Y}$$

5.13 Inferences in Regression Analysis

The variance-covariance matrix of $\mathbf{b}$, the estimator of β,

$$\underset{2\times 2}{\sigma^2\{\mathrm{b}\}} = \begin{bmatrix} \sigma^2\{\mathrm{b}_0\} & \sigma\{\mathrm{b}_0, \mathrm{b}_1\} \\ \sigma\{\mathrm{b}_1, \mathrm{b}_0\} & \sigma^2\{\mathrm{b}_1\} \end{bmatrix}$$

is the variance

$$\begin{aligned} \sigma^2\{\mathrm{b}\} &= \sigma^2\{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}\} \\ &= (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\sigma^2\{\mathbb{Y}\}\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1} \\ &= (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'(\sigma^2\mathbb{I})\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1} \\ &= \sigma^2(\mathbb{X}'\mathbb{X})^{-1}(\mathbb{X}'\mathbb{X})(\mathbb{X}'\mathbb{X})^{-1} \\ &= \sigma^2(\mathbb{X}'\mathbb{X})^{-1} \end{aligned}$$

So,

$$\underset{2\times 2}{\sigma^2\{\mathrm{b}\}} = \sigma^2(\mathbb{X}'\mathbb{X})^{-1} = \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum(X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum(X_i - \bar{X})^2} & \frac{1}{\sum(X_i - \bar{X})^2} \end{bmatrix}$$

The estimated variance-covariance matrix of $\mathbf{b}$ is

$$\underset{2\times 2}{\sigma^2\{\mathrm{b}\}} = MSE(\mathbb{X}'\mathbb{X})^{-1}$$

Mean Response

To estimate the mean response at X_h, we define the vector

$$\underset{2\times1}{\mathbb{X}_\mathrm{h}} = \begin{bmatrix} 1 \\ X_h \end{bmatrix},$$

so we can write

$$\hat{Y}_h = \mathbb{X}_\mathrm{h}'\mathrm{b} = \begin{bmatrix} 1 & X_h \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1 X_h].$$

This has variance

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\{\mathbb{X}_\mathrm{h}'\mathrm{b}\} = \mathbb{X}_\mathrm{h}'\sigma^2\{\mathrm{b}\}\mathbb{X}_\mathrm{h} = \mathbb{X}_\mathrm{h}'\sigma^2(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}_\mathrm{h} = \sigma^2\mathbb{X}_\mathrm{h}'(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}_\mathrm{h}$$

Note that this reduces to the familiar expression

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right],$$

where we can see explicitly that the variance expression contains contributions from $\sigma^2\{b_0\}, \sigma^2\{b_1\}$, and $\sigma\{b_0, b_1\}$ which it must since

Prediction of New Observation

$$s^2\{pred\} = MSE(1 + \mathbb{X}_{\mathrm{h}}'(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}_{\mathrm{h}})$$

The Canadian Survey of Household Spending (http://dli-idd-nesstar.statcan.gc.ca.proxy.library.upei.ca/webview/) is carried out annually across Canada. The main purpose of the survey is to obtain detailed information about household spending. Information is also collected about dwelling characteristics as well as household equipment.

```
#A subset of the latest Survey of Household Spending data are displa
spending_subset %>% datatable()
```

Show 200 entries Search:

	province	type_of_dwelling	income	marital_status	age_gr
1	NL	single_detached	68000	never_married	30-34
2	NL	single_detached	48000	never_married	25-29
3	NL	single_detached	30000	married	35-39
4	NL	row_house	30000	never_married	30-34
5	NL	single_detached	35000	married	25-29

Showing 1 to 30 of 30 entries Previous 1 Next

27 / 32

```
clothing_model = lm(clothing_expenditure~income, data=spending_subse
msummary(clothing_model)
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.064e+03  7.832e+02   1.358   0.1853
## income      3.050e-02  1.651e-02   1.847   0.0753 .
##
## Residual standard error: 1663 on 28 degrees of freedom
## Multiple R-squared:  0.1086,    Adjusted R-squared:  0.07681
## F-statistic: 3.413 on 1 and 28 DF,  p-value: 0.07528
```

For the Canadian Survey of Household Spending example, the matrix representation of the simple linear regression involves the following vectors and matrices:

```
X = cbind(1, spending_subset$income)
Y = cbind(spending_subset$clothing_expenditure)
```

```
X %>%  datatable()
```

Show 200 entries
Search:

V1	V2
1	68000
1	48000
1	30000
1	30000
1	35000

Showing 1 to 30 of 30 entries

Previous 1 Next

```
Y%>%  datatable()
```

Show 200 entries
Search:

V1
4200
1930
2340
2900
1300

Showing 1 to 30 of 30 entries

Previous 1 Next

29 / 32

```r
# X'X
t(X) %*% X
```

```
##          [,1]       [,2]
## [1,]       30 1.3120e+06
## [2,] 1312000 6.7524e+10
```

```r
# (X'X)^(-1)
solve(t(X) %*% X)
```

```
##               [,1]          [,2]
## [1,]  2.218440e-01 -4.310458e-06
## [2,] -4.310458e-06  9.856230e-11
```

```r
# X'Y
t(X)%*% Y
```

```
##             [,1]
## [1,]       71922
## [2,] 3454802000
```

```r
# (X'X)^(-1) X'Y
solve(t(X) %*% X) %*% t(X)%*% Y
```

```
##              [,1]
## [1,] 1.063687e+03
## [2,] 3.049648e-02
```

- What are the dimensions of the hat matrix here?
  - (30×30)
- Find $s^2\{pred\}$ for someone with an income equal to $60,000:
  - `(msummary(clothing_model)$sigma)^2 * (1+cbind(1, 60000)%*%solve(t(X) %*% X)%*% t(cbind(1, 60000)))` $= 2929101$

```r
# Y
Y%>%  datatable()
```

Show 200 entries
Search:

V1
4200
1930
2340
2900
1300

Showing 1 to 30 of 30 entries

Previous 1 Next

```r
#  X (X'X)^(-1) X'Y
X %*% solve(t(X) %*% X) %*% t(X)
```

Show 200 entries
Search:

V1
3137.44786711173
2527.51830630536
1978.58170157963
1978.58170157963
2131.06409178122

Showing 1 to 30 of 30 entries

Previous 1 Next

- What are the residuals corresponding to the first 3 individuals?
  $(1062.552, -597.5183, 361.4183)$

Recap: Sections 5.11-5.13

After Sections 5.11-5.13, you should be able to

- Write fitted values and residuals in matrix terms
- Write ANOVA and regression inferences in matrix terms