

# An Introduction to Multiple Linear Regression

Dr. Austin Brown

Kennesaw State University

```
knitr::opts_chunk$set(echo = T, include=T, echo=FALSE, message=FALSE)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages -----
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2     3.4.4      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.0
## v purrr       1.0.2
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org)
```

# Introduction

- ▶ Some of the materials in today's lecture were adapted from those created by Dr. Taasobshirazi as well as my former professor, Dr. Khalil Shafie (Thank you Drs. S & T!).

# Table of Contents

- ▶ Multiple Linear Regression Models
- ▶ Estimation of Model Parameters
- ▶ Hypothesis Testing in Multiple Linear Regression Models
- ▶ Confidence Intervals in Multiple Regression

# Multiple Linear Regression Models

- ▶ In our prior two classes, we've discussed simple linear regression models where we have a single predictor variable explaining a single response variable.
- ▶ In some cases, this may be appropriate. However, I would say I've pretty much never seen a single variable model in an academic paper and certainly not in industry application.
- ▶ Most of the time, based on our knowledge/experience, we can identify several predictor variables. In the baseball example, I would say that there are certainly other variables (team batting average, for example) which could aid in explaining the variability observed in MLB team regular season wins and in turn, improve predictive capability.
- ▶ Fortunately, we can do this using multiple linear regression.

# Multiple Linear Regression Models

- ▶ The multiple linear regression model is simply an extension of the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

- ▶ In general, we can have  $k$  predictors. All of the assumptions and interpretations we learned about for SLR still hold for MLR (e.g., assumptions on the errors and interpretation of the  $\beta_j$ 's)

## Multiple Linear Regression Models

- There are however, a few differences. For starters, with a multiple linear regression model, we are no longer fitting a line, but rather, an object.

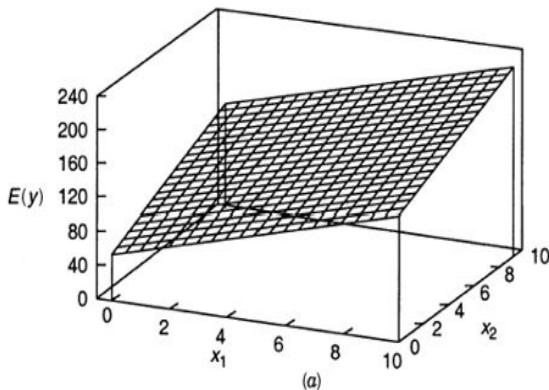


Figure 1: Figure 3.2 from Text

# Multiple Linear Regression Models

- ▶ We have an additional assumption of independence among our predictor variables, which we'll talk more at length about later on.
- ▶ In short, this is an important assumption because correlation between your predictors can cause the precision of your estimates to decrease (another way of saying this is that the variability of the estimates increases).
- ▶ This assumption can't be 100% met in practice, but we need to be able to mitigate its effect to the greatest degree possible.



## Estimation of Model Parameters

- ▶ Now that we have an understanding that MLR is a very logical extension of SLR, we can get into the nuts and bolts of how it works beginning with the estimation of the model parameters (e.g., the  $\hat{\beta}_j$ 's,  $\hat{\sigma}^2$ , etc.).
- ▶ But first, one of the main differences between MLR and SLR is some of the notation we use. Specifically, a MLR model is typically written in matrix notation (SLR can also be written in matrix notation).
- ▶ While it can be a bit daunting at first, this is actually a quite clever way of working with the models as it makes estimation much easier/cleaner.
- ▶ Let's first do a brief overview of matrices and vectors and some of their properties to make sure we're all on the same page.

# Review of Matrices and Vectors

- ▶ A vector (or column vector) is a list of  $m \times 1$  values:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$

- ▶ For example:

$$\mathbf{v} = \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}$$

# Review of Matrices and Vectors

- ▶ A matrix is an  $m \times n$  array of values:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- ▶ The subscript denotes the row and column.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 0 & -1 \\ -2 & 5 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix}$$

# Review of Matrices and Vectors

- In regression, we can use matrices and vectors to store our and conceptualize data:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

# Review of Matrices and Vectors

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

## Review of Matrices and Vectors

- To add or subtract two matrices or vectors, they have to be of the same dimension (same number of rows and columns).

$$\mathbf{A} = \begin{bmatrix} 5 & 3 & 2 \\ -2 & 1 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 2 & 0 & 1 \\ 3 & 7 & 2 \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 7 & 3 & 3 \\ 1 & 8 & 2 \end{bmatrix}$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 3 & 3 & 1 \\ -5 & -6 & -2 \end{bmatrix}$$

# Review of Matrices and Vectors

- ▶ We can also multiply matrices and vectors together, but the rules are a little different.
- ▶ We can multiply a matrix or vector and a constant number referred to as a **scalar**. Here, each element of the matrix or vector is multiplied by that constant.

$$\mathbf{v} = 3 \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 \\ -6 \\ 3 \end{bmatrix}$$

## Review of Matrices and Vectors

$$\mathbf{B} = 2 \begin{bmatrix} 5 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 4 & 2 \\ -4 & 2 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$



# Review of Matrices and Vectors

- ▶ To multiply two matrices or vectors together, they have to conform, meaning that the number of columns of the left matrix is equal to the number of rows of the right matrix.
- ▶ For example, a matrix with three rows and two columns conform with a matrix with two rows and three columns. But if the second matrix had four rows, they couldn't be multiplied together.

## Review of Matrices and Vectors

- For example, suppose we have the below two matrices:

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 0 & -1 \\ -2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} 2 \times 3 + 3 \times 0 & 2 \times (-2) + 3 \times 1 \\ 0 \times 3 + (-1) \times 0 & 0 \times (-2) + (-1) \times 1 \\ -2 \times 3 + 5 \times 0 & -2 \times (-2) + 5 \times 1 \end{bmatrix}$$

# Review of Matrices and Vectors

- ▶ For us, the reason why we write the  $X$  matrix the way we do is so the result we obtain follows the general regression equation we wrote earlier.

$$Y = X\beta + \varepsilon$$

- ▶ Our fitted equation is denoted:

$$\hat{Y} = X\hat{\beta}$$

# Review of Matrices and Vectors

- ▶ There are two other concepts I want to briefly review. The first is the concept of the transpose of a vector or matrix.
- ▶ Basically, the transpose of a matrix or vector is just where we flip the rows and columns. If we want to transpose a matrix, say  $\mathbf{A}$ , I typically denote its transpose as  $\mathbf{A}^T$ .

# Review of Matrices and Vectors

```
A <- matrix(c(2,3,  
              0,-1,  
              -2,5),byrow=T,ncol=2,nrow=3)
```

A

```
##      [,1] [,2]  
## [1,]    2    3  
## [2,]    0   -1  
## [3,]   -2    5
```

```
t(A)
```

```
##      [,1] [,2] [,3]  
## [1,]    2    0   -2  
## [2,]    3   -1    5
```

# Review of Matrices and Vectors

- ▶ The other concept is that of the inverse of a square matrix (square meaning it has the same number of rows and columns).
- ▶ The true inverse of a square matrix is sort of like division. If the true inverse of a square matrix exists, it has the following property:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

- ▶ where  $\mathbf{I}$  denotes the identity matrix, which is a square matrix where all the elements are zero besides those on the main diagonal.

# Review of Matrices and Vectors

```
A <- matrix(c(2,15,  
              9,-8),byrow=T,ncol=2)  
## Inverse ##  
solve(A)
```

```
##           [,1]      [,2]  
## [1,] 0.05298013 0.09933775  
## [2,] 0.05960265 -0.01324503
```

```
## Identity Matrix ##  
round(solve(A)%*%A)
```

```
##           [,1] [,2]  
## [1,]      1   0  
## [2,]      0   1
```

# Estimation of Model Parameters

- ▶ Okay, so how do we estimate all of our  $\hat{\beta}_j$ 's?
- ▶ Well, we still use least-squares estimation as that concept is still sound and logical even as we expand into a multivariate space.
- ▶ To do this, we need to minimize:

$$(Y - X\hat{\beta})^T(Y - X\hat{\beta})$$



# Estimation of Model Parameters

- ▶ If the columns of  $X$  are linearly independent (i.e., one isn't a linear combination of the others), then our estimated  $\hat{\beta}$  vector is:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

- ▶ Notice the dimension of  $X^T X$  makes it a square matrix and thus an inverse may exist for it.
- ▶ Also, we could estimate the parameters of simple linear regression using this exact same function and yield the exact same results as we did before.

# Estimation of Model Parameters

- ▶ Okay, let's do an example using the `sales` dataframe where we are looking to predict the sales our company makes using advertising expenditures in the areas of TV, newspaper, and radio.

# Hypothesis Testing in Multiple Linear Regression

- ▶ Fitting the model using R isn't too rough once you understand how to use the functions.
- ▶ How do we now test the significance of the regression and the significance of the individual predictors?
- ▶ Conceptually, there's no difference between what we did last week with SLR and how it works for MLR.

# Hypothesis Testing in Multiple Linear Regression

- ▶ To test the significance of the regression, we're testing:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \quad \text{for at least one } \beta_j$$

# Hypothesis Testing in Multiple Linear Regression

- ▶ Then, recall from last week that total variability,  $SS_{Total}$  can be partitioned into  $SS_{Regression} + SS_{Error}$ .
- ▶ The omnibus test statistic used to test the overall significance of the model as designated by the null and alternative hypotheses on the prior slide, we use:

$$F_0 = \frac{SSR/k}{SSE/(n-k-1)} \sim F(k, n - (k + 1))$$

- ▶ If  $F_0 > F_{\alpha, k, n-k-1}$ , then this is significant evidence to reject  $H_0$ . In other words, one of our  $\beta_j$ 's was found to be potentially significantly different from 0.

# Hypothesis Testing in Multiple Linear Regression

- ▶ Note, especially with MLR, it is best practice to perform the omnibus test prior to going and checking any individual coefficient's test.
- ▶ However, we typically want to know which predictors are significantly contributing to the model. If we have a significant omnibus test, we can go ahead and test each individual predictor much like we did with SLR.

# Hypothesis Testing in Multiple Linear Regression

- ▶ For an individual regression coefficient, say  $\beta_j$ , just like with SLR, we are testing:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

- ▶ Our test statistic is:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t(n - (k + 1))$$

- ▶ where  $C_{jj}$  is the  $j$ th diagonal element of the  $X^T X$  square matrix. Notice, the degrees of freedom of the null distribution are the same degrees of freedom for  $SSE$ .

# Hypothesis Testing in Multiple Linear Regression

- ▶ Also of note, in MLR, the individual coefficient tests are really *marginal* tests, meaning that they consider all of the other variables entered in the model. Which is to say, these tests are really examining the added contribution of the particular variable under consideration considering the other variables are all ready entered in the model.
- ▶ Let's go ahead and perform the individual coefficient tests for our Sales & Penguin Data.



## Confidence Intervals in Multiple Regression

- ▶ Just as we did with the SLR coefficients, we can also create confidence intervals for the estimated coefficients in MLR models as well.
- ▶ In general, a confidence interval for  $\beta_j$  has the form:

$$\hat{\beta}_j \pm t_{\alpha/2, n-k-1} \sqrt{\hat{\sigma}^2 C_{jj}}$$

- ▶ Given that this output is automatically provided by `broom::tidy`, we don't have to do much of the heavy lifting ourselves.

## Confidence Intervals in Multiple Regression

- ▶ As before, we can also create interval estimates for our fitted values,  $\hat{y}_i$ 's (and remember, this is really fitting an interval estimate to a mean). To do this, we need to use the vector of explanatory variable observations to aid in our estimation of the standard error.
- ▶ For observation  $y_i$ , it has an associated row vector of observations from the  $X$  matrix, that your text calls,  $x_0$ .

$$x_0 = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix}$$

# Confidence Intervals in Multiple Regression

- ▶ The form of the CI for the mean response (or fitted value) is:

$$\hat{y}_i \pm t_{\alpha/2, n-k-1} \sqrt{\hat{\sigma}^2 \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0}$$

- ▶ Let's see how to do this a little more easily using the functionality of R.