STA302 METHODS OF DATA ANALYSIS 1

MODULE 2: MULTIPLE LINEAR REGRESSION (MLR) MODELS

PROF. KATHERINE DAIGNAULT

MODULE 2 OUTLINE

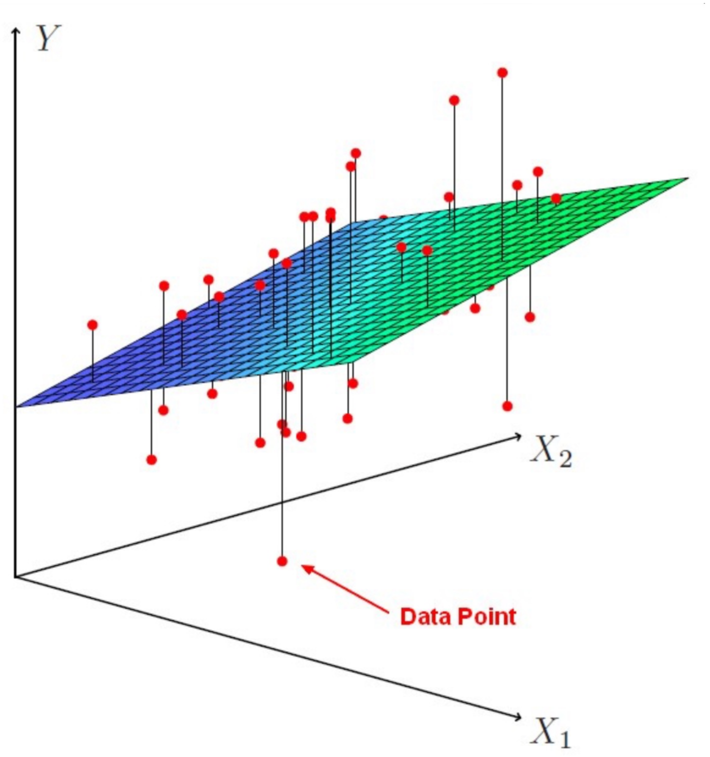
1. Multiple Linear Regression Basics
2. Estimation via Least Squares
3. Application Example
4. Interpretation of Coefficients in Multiple Linear Regression

MULTIPLE PREDICTORS

- Simple linear regression (SLR): $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
  - Functional relationship: $E(Y|X) = \beta_0 + \beta_1 x_i$
  - Statistical error/variation around $E(Y|X)$ given by ε_i
- Multiple linear regression (MLR) involves more than one predictor

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

  - Functional relationship: $E\left(Y \middle| X_1, \ldots, X_p\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$
  - i.e. the mean response is dependent on the value of p predictors
- Sample data: n sets of $\left(y_i, x_{i1}, \ldots, x_{ip}\right)$
- Need to estimate the p-dimensional surface of best fit by estimating $\beta_0, \beta_1, \ldots, \beta_p$



LINEAR REGRESSION IN MATRIX FORM

Simple Linear Regression (SLR)

- Algebraic form: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ for $i = 1, \dots, n$
- Instead create matrices to store these components:

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$n \times 1 \qquad 2 \times 1 \qquad n \times 2 \qquad n \times 1$$

- Each row of $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ is equal to the algebraic form above through matrix multiplication:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Multiple Linear Regression (MLR)

- Algebraic form: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$ for $i = 1, \dots, n$
- Similar matrix components, just augmented with extra predictor information:

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$n \times 1 \qquad (p+1) \times 1 \qquad n \times (p+1) \qquad n \times 1$$

- The matrix expression of the multiple linear regression trend is simply $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

PREDICTOR MATRIX

General Notes

- Each column represents the information of one predictor
- First column ALWAYS just 1's to represent the constant intercept
- Predictors that are continuous or discrete will populate matrix with numbers, e.g.,

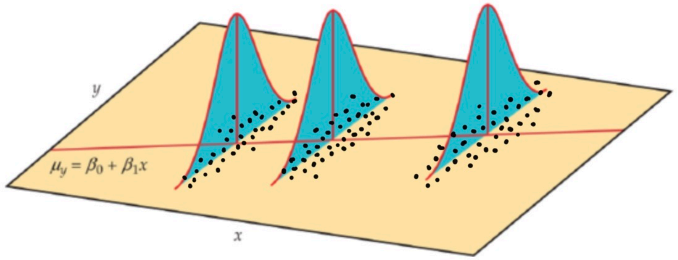
$$\boldsymbol{X} = \begin{pmatrix} 1 & 1.2 & -2 \\ 1 & 5.3 & -1 \\ \vdots & \vdots & \vdots \\ 1 & 3.7 & 0 \\ 1 & 2.4 & -2 \end{pmatrix}$$

Qualitative/Categorical Predictors

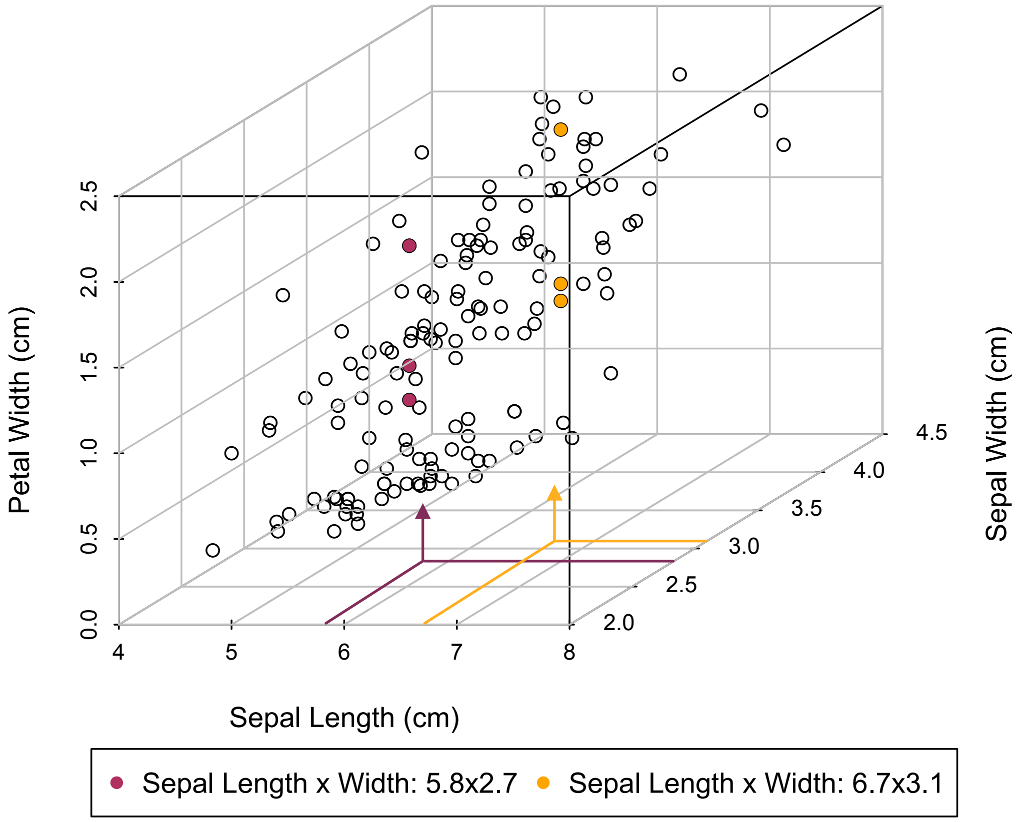
- Suppose one predictor (X_1) is discrete but one predictor is qualitative with values (Yes, No, Maybe)
- Use indicator variables to indicate the value of each entry
  - $X_2 = \begin{cases} 1, & \text{if Yes} \\ 0, & \text{otherwise} \end{cases}$, and $X_3 = \begin{cases} 1, & \text{if No} \\ 0, & \text{otherwise} \end{cases}$
  - If both are 0, then we have Maybe as our entry

$$\boldsymbol{X} = \begin{array}{cl} \begin{pmatrix} 1 & 2 & 1 & 0 \\ 1 & 4 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} & \begin{array}{l} \longrightarrow \text{Yes} \\ \longrightarrow \text{No} \\ \\ \longrightarrow \text{Maybe} \\ \longrightarrow \text{No} \end{array} \end{array}$$

CONDITIONAL MEAN & DISTRIBUTION OF RESPONSES



- Simple linear regression conditions on the value of one predictor
  - At each X value, we have a distribution of Y responses with mean $E(Y|X)$
- Multiple linear regression (MLR) conditions on the values of p predictors
  - E.g. at values (x_1, x_2) of predictors (X_1, X_2), we have a distribution of responses Y with a mean $E(Y|x_1, x_2)$



MODULE 2 OUTLINE

1. Multiple Linear Regression Basics
2. Estimation via Least Squares
3. Application Example
4. Interpretation of Coefficients in Multiple Linear Regression

MINIMIZING RESIDUAL SUMS OF SQUARES… AGAIN

- Still want a "snug" surface through the data, so minimize errors
- In simple linear regression, minimized

$$RSS = \hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}} = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

- Minimize same Residual Sum of Squares, edited to include additional predictors and coefficients:

$$RSS = \hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}} = \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}))^2$$

- Finds estimates $\widehat{\boldsymbol{\beta}}$ that make all residuals $\hat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ as small as possible

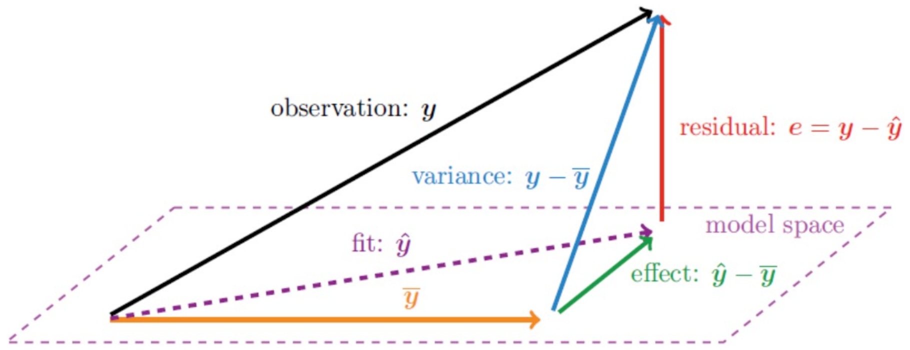


Figure 6.2. Generic linear model.

From Foundations and Application of Statistics by Pruim (2011)

ESTIMATION BY LEAST SQUARES

Least Squares Procedure

** Now require matrix operations, properties, and derivatives

1) Set up the estimating equation for given model with parameters present

2) Take partial derivatives of your estimating equation with respect to each unknown parameter

3) Set each derivative to 0 to obtain score equation

4) Rearrange equations to solve for each unknown parameter.

$$RSS = \hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}} = \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$$

$= \left(\boldsymbol{Y}^T - \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\right)\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$ by properties of transposes

$= \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\boldsymbol{Y} + \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$ by matrix multiplication

$= \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Y}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}$ by properties of transposes and scalars

$$= \boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{Y}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}$$

$$\frac{\partial RSS}{\partial \widehat{\boldsymbol{\beta}}} = -2\boldsymbol{X}^T\boldsymbol{Y} + 2(\boldsymbol{X}^T\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{0}$$

$$\Rightarrow (\boldsymbol{X}^T\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$$

$$\Rightarrow \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

** Assuming the inverse exists*

Theorem 2.14a. Let $u = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$, where $\mathbf{a}' = (a_1, a_2, \ldots, a_p)$ is a vector of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}. \qquad (2.112)$$

Theorem 2.14b. Let $u = \mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}. \qquad (2.113)$$

From Rencher & Schaalje's Linear Models in Statistics, page 56

WORKING WITH MATRIX ESTIMATORS

- Data stored in $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices, but too cumbersome to use for computation
- Instead, component matrices in Least Squares (LS) estimators can be easily calculated:

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} n & \Sigma x_{i1} & \Sigma x_{i2} & \cdots & \Sigma x_{ip} \\ \Sigma x_{i1} & \Sigma x_{i1}^2 & \Sigma x_{i1}x_{i2} & \cdots & \Sigma x_{i1}x_{ip} \\ \Sigma x_{i2} & \Sigma x_{i1}x_{i2} & \Sigma x_{i2}^2 & \cdots & \Sigma x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \Sigma x_{ip} & \Sigma x_{i1}x_{ip} & \Sigma x_{i2}x_{ip} & \cdots & \Sigma x_{ip}^2 \end{pmatrix} \quad , \quad \boldsymbol{X}^T\boldsymbol{Y} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_{i1}y_i \\ \Sigma x_{i2}y_i \\ \vdots \\ \Sigma x_{ip}y_i \end{pmatrix}$$

- Inverting $\boldsymbol{X}^T\boldsymbol{X}$ usually needs the aid of software so is often provided in a question.
- The fitted values $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ are found by multiplying the responses $\boldsymbol{Y}$ by the hat matrix $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$
  - The hat matrix is a projection matrix so has all the same properties (i.e. symmetric, idempotent)

MODULE 2 OUTLINE

1. Multiple Linear Regression Basics
2. Estimation via Least Squares
3. Application Example
4. Interpretation of Coefficients in Multiple Linear Regression

WORKED EXAMPLE (BY HAND)

Consider the sample of 12 observations below for a response variable Y and two predictors X_1 and X_2. Find the estimated multiple linear regression relationship between Y and the two predictors.

TABLE 7.1 Data for Example 7.2

Observation Number	y		x_1	x_2
1	2	1	0	2
2	3	1	2	6
3	2	1	2	7
4	7	1	2	5
5	6	1	4	9
6	8	1	4	8
7	10	1	4	7
8	7	1	6	10
9	8	1	6	11
10	12	1	6	9
11	11	1	8	15
12	14	1	8	13

$\boldsymbol{Y} =$ (the y column) ; (the column of 1s, x_1 and x_2) $= \boldsymbol{X}$

From Rencher & Schaalje's Linear Models in Statistics, page 140

If we were to start from scratch, we'd need our $\boldsymbol{Y}$ and $\boldsymbol{X}$ matrices…

And then we'd need to transpose X and do a lot of matrix multiplication – no fun ☹

Unless we are using software, the following summaries would be provided:

$$\sum_{i=1}^{12} y_i = 90, \quad \sum_{i=1}^{12} x_{i1} = 52, \quad \sum_{i=1}^{12} x_{i2} = 102, \quad \sum_{i=1}^{12} x_{i1}x_{i2} = 536$$

$$\sum_{i=1}^{12} x_{i1}^2 = 296, \quad \sum_{i=1}^{12} x_{i2}^2 = 1004, \quad \sum_{i=1}^{12} y_i x_{i1} = 482, \quad \sum_{i=1}^{12} y_i x_{i2} = 872$$

WORKED EXAMPLE (BY HAND)

Consider the sample of 12 observations below for a response variable Y and two predictors X_1 and X_2. Find the estimated multiple linear regression relationship between Y and the two predictors.

$$\sum_{i=1}^{12} y_i = 90, \quad \sum_{i=1}^{12} x_{i1} = 52, \quad \sum_{i=1}^{12} x_{i2} = 102, \quad \sum_{i=1}^{12} x_{i1}x_{i2} = 536$$

$$\sum_{i=1}^{12} x_{i1}^2 = 296, \quad \sum_{i=1}^{12} x_{i2}^2 = 1004, \quad \sum_{i=1}^{12} y_i x_{i1} = 482, \quad \sum_{i=1}^{12} y_i x_{i2} = 872$$

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} n & \Sigma x_{i1} & \Sigma x_{i2} \\ \Sigma x_{i1} & \Sigma x_{i1}^2 & \Sigma x_{i1}x_{i2} \\ \Sigma x_{i2} & \Sigma x_{i1}x_{i2} & \Sigma x_{i2}^2 \end{pmatrix} = \begin{pmatrix} 12 & 52 & 102 \\ 52 & 296 & 536 \\ 102 & 536 & 1004 \end{pmatrix}$$

** Find inverse using software or provided to you:*

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} \cong \begin{pmatrix} 0.9747634 & 0.2429022 & -0.2287066 \\ 0.2429022 & 0.1620662 & -0.1111987 \\ -0.2287066 & -0.1111987 & 0.08359621 \end{pmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{Y} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_{i1} y_i \\ \Sigma x_{i2} y_i \end{pmatrix} = \begin{pmatrix} 90 \\ 482 \\ 872 \end{pmatrix}$$

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

$$= \begin{pmatrix} \boxed{0.9747634} & \boxed{0.2429022} & \boxed{-0.2287066} \\ 0.2429022 & 0.1620662 & -0.1111987 \\ -0.2287066 & -0.1111987 & 0.08359621 \end{pmatrix} \begin{pmatrix} \boxed{90} \\ \boxed{482} \\ \boxed{872} \end{pmatrix}$$

$$= \begin{pmatrix} 5.375 \\ 3.012 \\ -1.285 \end{pmatrix}$$

$$\boxed{\hat{y}_i = 5.375 + 3.012x_{i1} - 1.285x_{i2}}$$

WORKED EXAMPLE (USING R)

Consider the sample of 12 observations below for a response variable Y and two predictors X_1 and X_2. Find the estimated multiple linear regression relationship between Y and the two predictors.

TABLE 7.1 Data for Example 7.2

Observation Number	y	x_1	x_2
1	2	0	2
2	3	2	6
3	2	2	7
4	7	2	5
5	6	4	9
6	8	4	8
7	10	4	7
8	7	6	10
9	8	6	11
10	12	6	9
11	11	8	15
12	14	8	13

From Rencher & Schaalje's Linear Models in Statistics, page 140

Load the data into R:

```
> y <- c(2,3,2,7,6,8,10,7,8,12,11,14)
> x1 <- c(0,2,2,2,4,4,4,6,6,6,8,8)
> x2 <- c(2,6,7,5,9,8,7,10,11,9,15,13)
> cbind(y, x1, x2)
       y x1 x2
 [1,]  2  0  2
 [2,]  3  2  6
 [3,]  2  2  7
 [4,]  7  2  5
 [5,]  6  4  9
 [6,]  8  4  8
 [7,] 10  4  7
 [8,]  7  6 10
 [9,]  8  6 11
[10,] 12  6  9
[11,] 11  8 15
[12,] 14  8 13
```

Create $\boldsymbol{X}^T\boldsymbol{X}$ matrix:

```
> X <- matrix(cbind(rep(1, 12), x1, x2), ncol=3)
> XtX <- t(X) %*% X
> XtX
     [,1] [,2] [,3]
[1,]   12   52  102
[2,]   52  296  536
[3,]  102  536 1004
```

Invert the matrix:

```
> invXtX <- solve(XtX)
> invXtX
           [,1]       [,2]        [,3]
[1,]  0.9747634  0.2429022 -0.22870662
[2,]  0.2429022  0.1620662 -0.11119874
[3,] -0.2287066 -0.1111987  0.08359621
```

Create $\boldsymbol{X}^T\boldsymbol{Y}$ matrix:

```
> xy <- t(X) %*% as.matrix(y)
> xy
     [,1]
[1,]   90
[2,]  482
[3,]  872
```

Matrix multiplication to get $\widehat{\boldsymbol{\beta}}$ matrix:

```
> invXtX %*% xy
          [,1]
[1,]  5.375394
[2,]  3.011830
[3,] -1.285489
```

Or more simply…

```
> lm(y ~ x1 + x2)

Call:
lm(formula = y ~ x1 + x2)

Coefficients:
(Intercept)           x1           x2
      5.375        3.012       -1.285
```

MAKING PREDICTIONS

Suppose we now want to use this estimated regression relationship $\hat{y}_i = 5.375 + 3.012x_1 - 1.285x_2$ to predict/estimate the mean response when $X_1 = 5$ and $X_2 = 9$.

Predictions by hand

Since $\hat{E}(Y|X_1, X_2) = \hat{y}_i = 5.375 + 3.012x_1 - 1.285x_2$, just evaluate this at $X_1 = 5$ and $X_2 = 9$:

$$\begin{aligned}\hat{y}_i &= 5.375 + 3.012x_1 - 1.285x_2 \\ &= 5.375 + 3.012(5) - 1.285(9) = 8.87\end{aligned}$$

The mean response when $X_1 = 5$ and $X_2 = 9$ is 8.87.

Predictions using R

Two ways to do this in R:

1. Manually:

```
> 5.375 + 3.012*5 - 1.285*9
[1] 8.87
```

2. Using a built-in function with the fitted model:

```
> # save the linear model
> model <- lm(y ~ x1 + x2)
> # use built-in function
> predict(model, newdata = data.frame(x1 = 5, x2 = 9))
       1
8.865142
```

MODULE 2 OUTLINE

1. Multiple Linear Regression Basics
2. Estimation via Least Squares
3. Application Example
4. Interpretation of Coefficients in Multiple Linear Regression

CONDITIONAL NATURE OF MULTIPLE PREDICTOR MODEL

- In previous example, we could fit three models:
  - Simple model with X_1 only: $\hat{y}_i = 1.86 + 1.30x_{i1}$
  - Simple model with X_2 only: $\hat{y}_i = 0.86 + 0.78x_{i2}$
  - Two-predictor model we fit before: $\hat{y}_i = 5.375 + 3.012x_1 - 1.285x_2$
- Why is the relationship between Y and X_2 positive in one model and negative in another?
- Estimation and interpretation of coefficients in multiple predictor models conditions on all other predictors
  - i.e. consider only one fixed value of all other predictors when estimating or interpreting the coefficient of interest.

Figure from Rencher & Schaalje's Linear Models in Statistics, page 140

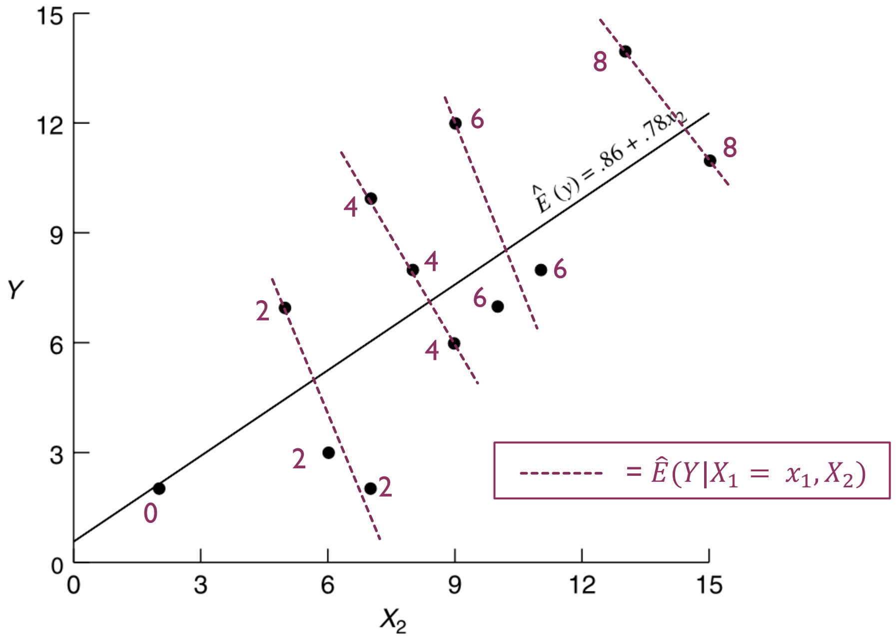
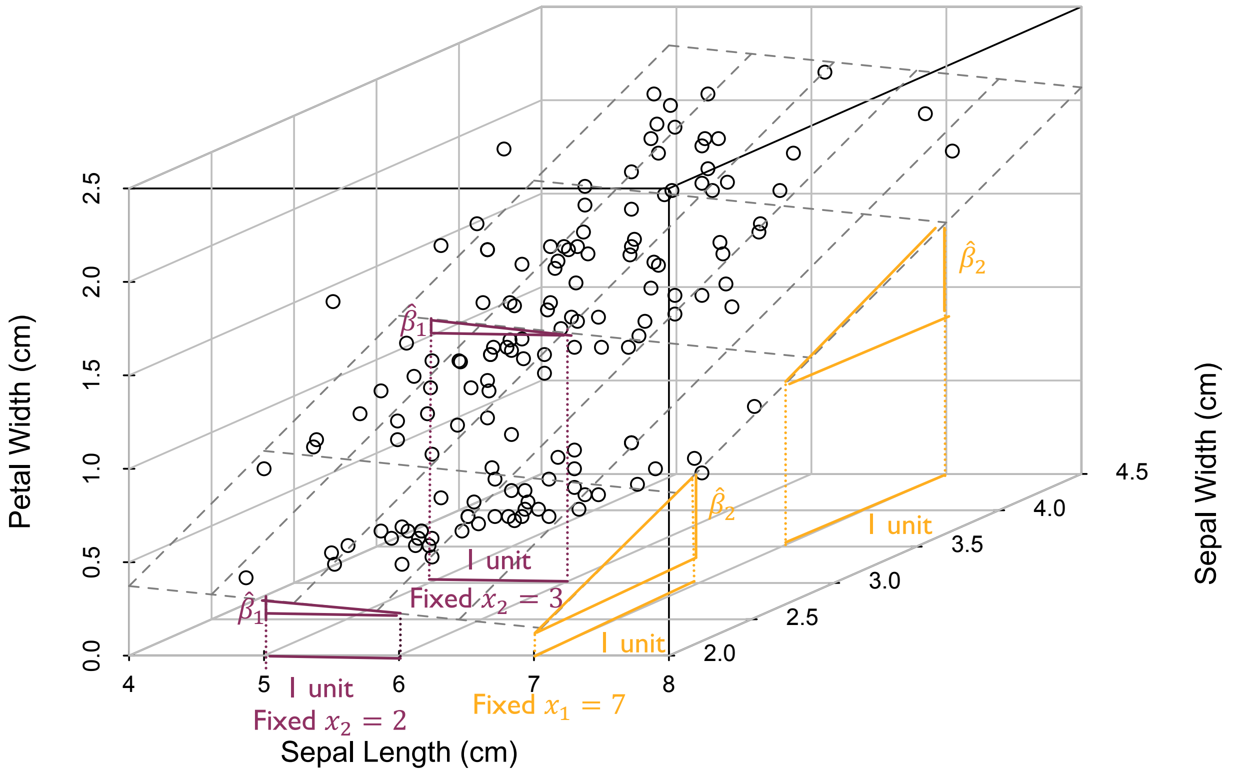


Figure 7.1 Regression of y on x_2 ignoring x_1.

INTERPRETATION OF COEFFICIENTS IN MLR

- In a simple model, the coefficients are interpreted as:
  - $\hat{\beta}_0$ is the mean response when the predictor is zero
  - $\hat{\beta}_1$ is the change mean response for a one-unit increase in the value of the predictor.
- Now each coefficient is interpreted individually so that the change observed in the response is attributed ONLY to that predictor:
  - $\hat{\beta}_0$ is the mean response when ALL predictors have value zero
  - $\hat{\beta}_j$ is the average/mean/expected change in the response for a one-unit increase in X_j when all other predictors are held fixed.



INTERPRETATION WITH INDICATOR VARIABLES

- A third predictor could be added to our 3D scatterplot and multiple linear regression
  - Species is qualitative with 3 levels (Setosa, Virginica, Versicolor)
- Below model gives us 3 regression planes, one for each species:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1\, Sepal\ Length + \hat{\beta}_2\, Sepal\ Width + \hat{\beta}_3 \mathbb{I}(Versicolor) + \hat{\beta}_4 \mathbb{I}(Virginica)$$

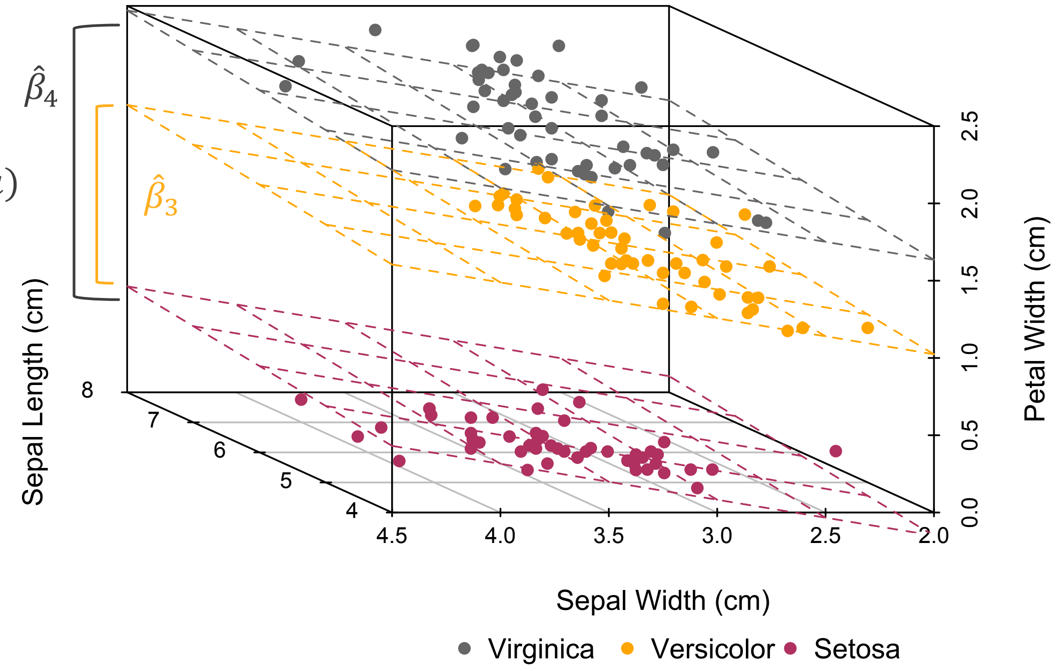
- Gives different intercepts but common slopes (i.e. parallel planes)
- Interpretation must compare each level to the reference/baseline level (i.e. Setosa)

```
> data(iris)
> lm(Petal.Length ~ Sepal.Length + Sepal.Width + Species, data=iris)

Call:
lm(formula = Petal.Length ~ Sepal.Length + Sepal.Width + Species,
    data = iris)

Coefficients:
      (Intercept)       Sepal.Length        Sepal.Width  Speciesversicolor
         -1.63430            0.64631           -0.04058            2.17023
 Speciesvirginica
          3.04911
```

Intercept for Setosa

Difference in Virginica and Setosa intercepts

"Change in mean Petal Width for a versicolor iris compared to a setosa iris with sepal length and width of 0"



INTERPRETATION WITH INTERACTION VARIABLES

- Interactions allow the relationship between response and one predictor to vary according to values of a second predictor
  - i.e., they give different slopes
- Below model interacts Species with Sepal Width to give different slopes of Sepal Width for different species

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \, Sepal\ Length + \hat{\beta}_2 \, Sepal\ Width$$
$$+ \hat{\beta}_3 \, Sepal\ Width * \mathbb{I}(Versicolor) + \hat{\beta}_4 \, Sepal\ Width * \mathbb{I}(Virginica)$$

- Interpret slope relative to reference/baseline level (i.e., setosa here)

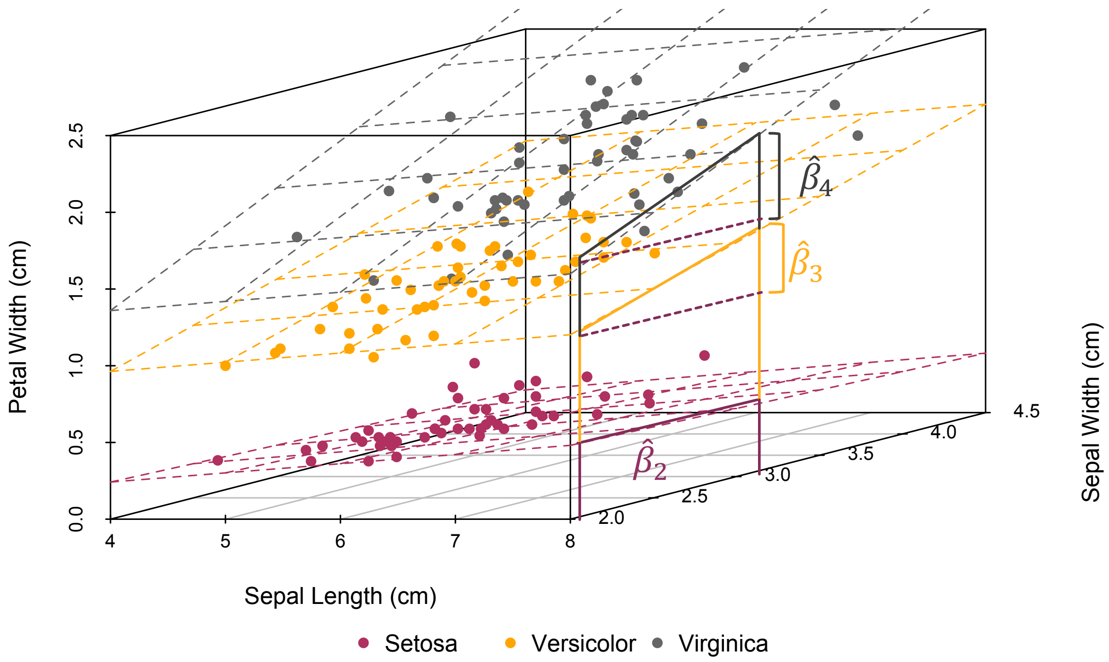
```
> data(iris)
> lm(Petal.Width ~ Sepal.Length + Sepal.Width+ Sepal.Width:Species, data=iris)

Call:
lm(formula = Petal.Width ~ Sepal.Length + Sepal.Width + Sepal.Width:Species,
    data = iris)

Coefficients:
                 (Intercept)               Sepal.Length                Sepal.Width
                     0.07744                    0.05984                   -0.03748
Sepal.Width:Speciesversicolor  Sepal.Width:Speciesvirginica
                     0.36055                    0.55878
```

Interaction term

Slope for setosa

Change in Sepal Width slope between Virginica and Setosa

"Change in mean response for a one-unit increase in Sepal Width for Versicolor compared to Setosa for a fixed Sepal Length"



MODULE TAKE-AWAYS

1. What is similar/different between simple and multiple linear regression? (e.g. estimation, formulae, interpretation, notation, etc.)
2. How do we estimate/compute the coefficients of our model?
3. How do we include qualitative information in our model?
4. What is the correct way to interpret the estimated coefficients?
5. How does the interpretation change when we use indicator variables and interaction terms?