SDS 439/5130: Linear Statistical Models

Topic 2. Multiple Linear Regression Model I:

Basic Setting, Matrix Notation, and OLSE

Mengxin (Maxine) Yu [1]

[1] Department of Statistics and Data Science

https://maxineyu.github.io/persona_web/

Washington University in St. Louis

August 26, 2025

Today's Class

1. Multiple Linear Regression Model
   - The Model
   - Ordinary Least Squares Estimators (OLSE)
   - Multiple Linear Regression in R

Multiple Regression Models

- Still have a single response variable Y, but we now have multiple explanatory or predictor variables $X_1, X_2, \ldots, X_k$, which in most applications make more sense.
- Many Examples:
  - Blood Pressure vs. Age, Weight, Diet, Smoking, Fitness Level
  - Traffic Count vs. Time, Location, Population, Month
  - House sale price vs. Taxes, # of baths, living space, # of rooms, etc.
  - Weight vs. Height, Age, and Gender.
- While in this course we always assume that the response is quantitative and continuous, the predictors could be both quantitative or qualitative (or categorical).

Multiple Linear Regression Model

- In a multiple linear regression model, we aim to explain the variation in the response variable Y as a linear function of the explanatory variables $X_1, \ldots, X_k$ up to an individual random error ε specific for each case:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, \ldots, n,$$

  where k is the number of predictors and n is the number of cases.
- The number $p = k + 1$ represents the number of coefficients $\beta_0, \ldots, \beta_k$.
- In the Simple Linear Regression Model, we assume that
  - $\mathbb{E}(\varepsilon_i) = 0$,
  - $\mathrm{Var}(\varepsilon_i) = \sigma^2$, (Homoskedasticity Assumption)
  - $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$, for $i \neq j$.

Comments and Special Cases

The term linear here refers to the parameters, not the predictor variables. So, all the models below are also considered multiple linear regression models:

- Polynomial Regression of order k:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + \varepsilon_i$$

- Regression Model with Interaction of Predictors:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

Interpretation of Coefficients

- No Interactions (i.e., $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$):
  - β_0 is the 'intercept'
  - β_i describes the average change in the value of the response per one unit increase in the variable X_i when the other variables are held constant. β_i measures the 'effect' of the i^{th} variable in the response.
- When interactions are present, the rate of change for one variable can be affected by others. Consider

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

  - The average or mean change in Y per one unit change in X_1 when $X_2 = x_2$ is

$$\mathbb{E}(\Delta Y) = \beta_1 + \beta_3 x_2.$$

  - The average or mean change in Y per one unit change in X_2 when $X_1 = x_1$ is

$$\mathbb{E}(\Delta Y) = \beta_2 + \beta_3 x_2.$$

Matrix Formulation of the Model

The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n,$$

can be expressed in the following matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

or, in short,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

- **y** is an $n \times 1$ vector of observations called response vector
- **X** is an $n \times p$ matrix with the values of the regressors called design matrix
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients called parameter vector
- $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors called error vector

Ordinary Least Squares Estimators (OLSE) I

We can readily generalize the least-squares estimation procedure from the one-predictor case to multiple predictors:

- Given a tentative regression formula $y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_k x_k$, we wish to minimize the Sum of Square Errors (SSE):

$$S(\tilde{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \tilde{e}_i^2 := \sum_{i=1}^{n} \Big(y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \cdots + \tilde{\beta}_k x_{ik}) \Big)^2$$

  over all possible $\tilde{\boldsymbol{\beta}} = [\tilde{\beta}_0, \ldots, \tilde{\beta}_k]'$.[1]

- $\tilde{e}_i := y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \cdots + \tilde{\beta}_k x_{ik})$ represents the fitting error or residual for the i^{th} observation point $(x_{i1}, \ldots, x_{ik}, y_i)$ incurred by the tentative regression formula $y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_k x_k$.

[1] Hereafter, $\boldsymbol{A}'$ or $\boldsymbol{A}^T$ denote the transpose of a matrix or vector $\boldsymbol{A}$.

Ordinary Least Squares Estimators (OLSE) II

- The vector $\hat{\beta}$ where $S(\tilde{\beta})$ attains its minimum (assuming it exists and is unique) is called the (Ordinary) Least Squares Estimator (OLSE) and is denoted by

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \in \mathbb{R}^{p\times 1}.$$

- In other words, $\hat{\boldsymbol{\beta}}$ is such that

$$\mathcal{S}(\hat{\boldsymbol{\beta}}) \leq \mathcal{S}(\tilde{\boldsymbol{\beta}}), \quad \text{for all} \quad \tilde{\boldsymbol{\beta}} \in \mathbb{R}^{p\times 1}.$$

Example 1

To find the OLSE in R, we use the same function `lm` as before. Consider the following data from Table B.4 in Montgomery et al.:

table.b4	*Table B4*

Description

The `table.b4` data frame has 24 observations on property valuation.

Usage

```
data(table.b4)
```

Format

This data frame contains the following columns:

y sale price of the house (in thousands of dollars)

x1 taxes (in thousands of dollars)

x2 number of baths

x3 lot size (in thousands of square feet)

x4 living space (in thousands of square feet)

x5 number of garage stalls

x6 number of rooms

x7 number of bedrooms

x8 age of the home (in years)

x9 number of fireplaces

Example 1. Downloading the data

```
> install.packages('MPV')
trying URL 'http://cran.case.edu/bin/macosx/maveri
Content type 'application/x-gzip' length 176184 by
==================================================
downloaded 172 KB


The downloaded binary packages are in
	/var/folders/sl/qz6p16hd54qbz7z85p1l6ms80000gn/
> library(MPV)
> data(table.b4)
> table.b4
      y     x1  x2     x3    x4  x5 x6 x7 x8 x9
1  29.5 5.0208 1.0 3.5310 1.500 2.0  7  4 62  0
2  27.9 4.5429 1.0 2.2750 1.175 1.0  6  3 40  0
3  25.9 4.5573 1.0 4.0500 1.232 1.0  6  3 54  0
4  29.9 5.0597 1.0 4.4550 1.121 1.0  6  3 42  0
5  29.9 3.8910 1.0 4.4550 0.988 1.0  6  3 56  0
6  30.9 5.8980 1.0 5.8500 1.240 1.0  7  3 51  1
7  28.9 5.6039 1.0 9.5200 1.501 0.0  6  3 32  0
8  35.9 5.8282 1.0 6.4350 1.225 2.0  6  3 32  0
9  31.5 5.3003 1.0 4.9883 1.552 1.0  6  3 30  0
10 31.0 6.2712 1.0 5.5200 0.975 1.0  5  2 30  0
11 30.9 5.9592 1.0 6.6660 1.121 2.0  6  3 32  0
12 30.0 5.0500 1.0 5.0000 1.020 0.0  5  2 46  1
13 36.9 8.2464 1.5 5.1500 1.664 2.0  8  4 50  0
14 41.9 6.6969 1.5 6.9020 1.488 1.5  7  3 22  1
15 40.5 7.7841 1.5 7.1020 1.376 1.0  6  3 17  0
16 43.9 9.0384 1.0 7.8000 1.500 1.5  7  3 23  0
17 37.5 5.9894 1.0 5.5200 1.256 2.0  6  3 40  1
18 37.9 7.5422 1.5 5.0000 1.690 1.0  6  3 22  0
19 44.5 8.7951 1.5 9.8900 1.820 2.0  8  4 50  1
20 37.9 6.0831 1.5 6.7265 1.652 1.0  6  3 44  0
21 38.9 8.3607 1.5 9.1500 1.777 2.0  8  4 48  1
22 36.9 8.1400 1.0 8.0000 1.504 2.0  7  3  3  0
23 45.8 9.1416 1.5 7.3262 1.831 1.5  8  4 31  0
24 25.9 4.9176 1.0 3.4720 0.998 1.0  7  4 42  0
```

Example 1. Running the Multiple Linear Regression

```
> fullmodel<-lm(data=table.b4)
> summary(fullmodel)

Call:
lm(data = table.b4)

Residuals:
   Min     1Q Median     3Q    Max 
-3.720 -1.956 -0.045  1.627  4.253 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 14.92765    5.91285   2.525   0.0243 *
x1           1.92472    1.02990   1.869   0.0827 .
x2           7.00053    4.30037   1.628   0.1258  
x3           0.14918    0.49039   0.304   0.7654  
x4           2.72281    4.35955   0.625   0.5423  
x5           2.00668    1.37351   1.461   0.1661  
x6          -0.41012    2.37854  -0.172   0.8656  
x7          -1.40324    3.39554  -0.413   0.6857  
x8          -0.03715    0.06672  -0.557   0.5865  
x9           1.55945    1.93750   0.805   0.4343  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.949 on 14 degrees of freedom
Multiple R-squared:  0.8531,	Adjusted R-squared:  0.7587 
F-statistic: 9.037 on 9 and 14 DF,  p-value: 0.000185
```

Example 1. Running a Partial Model

```
> partialmodel<-lm(y~x1+x2,data=table.b4)
> summary(partialmodel)

Call:
lm(formula = y ~ x1 + x2, data = table.b4)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7639 -1.9454 -0.1822  1.8068  5.0423

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.0418     2.9585   3.394  0.00273 **
x1            2.7134     0.4849   5.595 1.49e-05 ***
x2            6.1643     3.1864   1.935  0.06663 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.792 on 21 degrees of freedom
Multiple R-squared:  0.8025,	Adjusted R-squared:  0.7837
F-statistic: 42.67 on 2 and 21 DF,  p-value: 4.007e-08
```

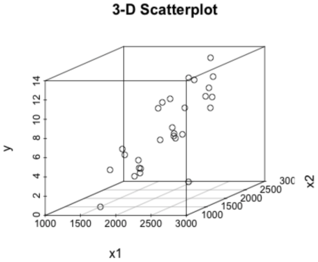
Interpretation And Preliminary Conclusions

1. About $R^2 = 85\%$ of the variability in sale price is accounted for by the linear regression model on $x_1, \ldots, x_9$;
2. However, only x_1 (taxes) and x_2 (number of baths) seem to be significant (see their corresponding p-values to the right are higher than .05; more about this later on)
3. A regression only on these two variables accounts for 80% of the variability in sale prices, a modest reduction from the full model.
4. The coefficients in the full model tell us that, with all other variables remaining constant, the sale price of a house increases on average about 2 thousand for each additional thousand dollars in taxes. Also, with all other variables remaining constant, the sale price increases on average about 7 thousand for each additional bathroom in the house.

Example 2

Using the 1976 NFL data from Table B.1, let's look at some plots first. Here y =number of games won (per 14-game season), x_1 =rushing yards (season) and x_2 =passing yards (season).

```
library(scatterplot3d)
attach(table.b1)
scatterplot3d(x1,x2,y,main="3-D Scatterplot")
```

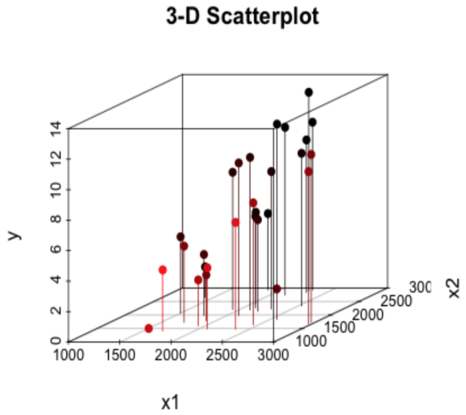


Using the **scatterplot3d** package and function of the same name:

Example 2. Fancier ScatterPlot

3-D Scatterplot with coloring and vertical drop lines:

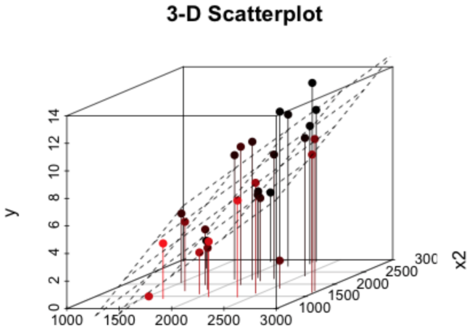
```
scatterplot3d(x1,x2,y, pch=16, highlight.3d=TRUE,
       type="h", main="3-D Scatterplot")
```



Example 2. Regression Plane

3-D Scatterplot with coloring, vertical drop lines and regression plane:

```
s3d <-scatterplot3d(x1,x2,y, pch=16, highlight.3d=TRUE,
type="h", main="3-D Scatterplot")
fit <- lm(y ~ x1+x2)
s3d$plane3d(fit)
```



Example 2. Summary Function

```
> summary(fit)

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

Residuals:
    Min      1Q  Median      3Q     Max 
-6.4613 -1.4362  0.0489  1.6283  3.5213 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.218e+01  3.112e+00  -3.913 0.000619 ***
x1           5.520e-03  1.132e-03   4.875 5.15e-05 ***
x2           3.522e-03  8.749e-04   4.026 0.000464 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.268 on 25 degrees of freedom
Multiple R-squared:  0.6068, Adjusted R-squared:  0.5754 
F-statistic: 19.29 on 2 and 25 DF,  p-value: 8.556e-06
```

Interpretation And Preliminary Conclusions

1. About $R^2 = 60\%$ of the variability in the number of wins in a season is accounted for by the linear regression model on rushing (x_1) and passing (x_2) yards (not a bad fit);
2. The two predictor variables are highly significant.
3. The coefficients tell us that, with all other variables remaining constant, we expect a team to win about 5 more games for each addition 1000 rushing yards per season. Similarly, with all other variables remaining constant, we expect a team to win about 3 more games for each addition 1000 passing yards per season.