Applied Regression

Multiple Linear Regression Model (MLR Model)

Introduction

Module 4 Lecture - 4-1

Multiple Linear Regression

Example-4.1 - Consider the National Football League data with the following variables (Data file is in folder - data_table_B1.xlsx)

y: Games won (per 14-game season)

x_1: Rushing yards (season)

x_2**: Passing yards (season)**

x_3: Punting average (yards/punt)

x_4: Field goal percentage (FGs made/FGs attempted 2season)

x_5: Turnover differential (turnovers acquiredturnovers lost)

x_6: Penalty yards (season)

x_7**: Percent rushing (rushing plays/total plays)**

x_8**: Opponents rushing yards (season)**

x_9: Opponents passing yards (season)

Here is the part of the raw data:

y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
10	2113	1985	38.9	64.7	4	868	59.7	2205	1917
11	2003	2855	38.8	61.3	3	615	55	2096	1575
11	2957	1737	40.1	60	14	914	65.6	1847	2175
13	2285	2905	41.6	45.3	-4	957	61.4	1903	2476
10	2971	1666	39.2	53.8	15	836	66.1	1457	1866
11	2309	2927	39.7	74.1	8	786	61	1848	2339
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
4	2080	1492	35.5	68.8	-8	722	57.8	2053	2550
10	2301	2835	35.3	74.1	2	683	59.7	1979	2110
6	2040	2416	38.7	50	0	576	54.9	2048	2628
8	2447	1638	39.9	57.1	-8	848	65.3	1786	1776
2	1416	2649	37.4	56.3	-22	684	43.8	2876	2524
0	1503	1503	39.3	47	-9	875	53.5	2560	2241

First we will consider building a MLR model with 3 independent ($x_2, x_7,$ & x_8) and the dependent variables - Y.

(Reasoning for that will be answered in Module-6).

The following questions will be answered:

a. Fit a multiple linear regression model relating the number of games won (y) to the team's passing yardage (x_2), the percentage of rushing plays (x_7), and the opponents yards rushing (x_8).

b. Construct the analysis-of-variance table and test for significance of regression.

c. Calculate t statistics for testing the hypotheses
$H_0 : \beta_2 = 0, H_0 : \beta_7 = 0,$ *and* $H_0 : \beta_8 = 0.$

What conclusions can you draw about the roles the variables $x_2, x_7,$ and , x_8 play in the model?

d. Calculate R^2 and for this model.

e. Using the partial F test, determine the contribution of x_7 to the model.

How is partial F statistic related to the t test for β_7 calculated in part(c)?

Each line (only with $y, x_2, x_7 \,\&\, x_8$) in the previous data set is one data point. Hence each data point in MLR model is a vector with one dependent variable (y) and k independent variables x's. Hence the MLR model with k many independent variable is written as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + e$$

In general the ith data data point in MLR (with k-many independent variables) is written as $(Y_i, X_{i1}, X_{i2}, X_{i3}, \ldots, X_{ik})$. Thinking in terms of ith data point the above relationship looks like

$$y_i = \beta_0 + \beta_1.X_{i1} + \beta_2.X_{i2} + \beta_3.X_{i3} + \ldots + \beta_k.X_{ik} + e_i \quad \text{for i = 1,2 ,3, ...,n}$$

So for our model, the first data point (line 1) satisfies the equation:

$$10 = \beta_0 + 1985 \times \beta_2 + 59.7 \times \beta_7 + 2205 \times \beta_8 + e_1$$

and other points satisfy the similar equations. As a result, there are n many equations for n data points which can be expressed in the matrix form. So in general, the data set is a (n x k) matrix. To understand the MLR models and the results we need to review little matrix and vector algebra.

Introduction to matrix algebra

Definition of a matrix:

A (r x c) (pronounced as r by c) matrix is a rectangular array of symbols or numbers arranged in r rows and c columns. A matrix is almost always denoted by a single capital letter (may be in boldface type).

Here are three examples of simple matrices. The matrix A is a 3 x 3 square matrix, B is a 2 x 3 matrix and T is a 3 x 2 matrix.

$$A = \begin{pmatrix} 1 & 7 & 8 \\ 4 & 11 & 34 \\ 5 & 3 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 4 & 8 \\ 11.2 & 2.9 & 3 \end{pmatrix} \qquad T = \begin{pmatrix} 2 & 4 \\ 2 & 8 \\ 2.9 & 3 \end{pmatrix}$$

An element in a matrix is denoted by it its position in the matrix. For example, number 34 in A is in the intersection of 2nd row and 3rd column, hence it is denoted by A_{23}. So $A_{23} = 34$. Similarly $B_{12} = 4$ and $T_{31} = 2.9$

Definition of a vector and a scalar

A column vector is an r x 1 matrix, that is, a matrix with only one column and many rows. A row vector is an 1 x c matrix, that is, a matrix with only one row and many columns.

The following vector S is a 3 x 1 column vector and V is a 1 x 3 row vector :

$$S = \begin{pmatrix} 1 \\ 34 \\ 5 \end{pmatrix} \qquad V = \begin{pmatrix} 2 & 4 & 8 \end{pmatrix}$$

Any 1 x 1 "matrix" is called a **scalar**, but it's just an ordinary number.

Matrix addition: Two matrices cannot not be added or subtracted unless they are of the same dimensions (i.e same number of rows and columns). Then addition of matrix just adds the two numbers in the same position.

Add the entry in ijth position of the first matrix with ijth position of the 2nd matrix, and that becomes the ijth entry of the combined matrix.

Below is an example of addition of two matrices.

$$\begin{pmatrix} 1 & 7 & 8 \\ 4 & 1 & 4 \\ 5 & -3 & 4 \end{pmatrix} \quad + \quad \begin{pmatrix} 2 & 4 & 8 \\ 2 & 9 & -3 \\ 2 & 0 & 6 \end{pmatrix} \quad = \quad \begin{pmatrix} 3 & 11 & 16 \\ 6 & 10 & 1 \\ 7 & -3 & 10 \end{pmatrix} = L$$

Definition of the transpose of a matrix

The transpose of a matrix A is a matrix, denoted A' or A^T, whose rows are the columns of A and whose columns are the rows of A all in the same order. As a result ijth element of A becomes the jith element of A^T. If order of A is r x c, then A^T is c x r matrix. For example, the transpose of the 2 x 3 matrix A:

$$A = \begin{pmatrix} 1 & 7 & 8 \\ 4 & 11 & 34 \end{pmatrix} \quad \Rightarrow \quad A^T = \begin{pmatrix} 1 & 4 \\ 7 & 11 \\ 8 & 34 \end{pmatrix}$$

Observe that $A_{21} = 4$ as a result $A^T_{12} = 4$.

Matrix Multiplication:

Two matrices can be multiplied together only if the number of columns of the first matrix equals the number of rows of the second matrix. Then, when you multiply the two matrices:

The number of rows of the resulting matrix equals the number of rows of the first matrix, and the number of columns of the resulting matrix equals the number of columns of the second matrix.

For example, if A is a 2 x 3 matrix and B is a 3 x 4 matrix, then the matrix multiplication AB is possible. The resulting matrix C = AB has 2 rows and 4 columns. That is, C is a 2 by 4 matrix. Note that the matrix multiplication BA is not possible.

Now to get the elements of the new matrix C, we do the following,

The entry in the ith row and jth column of C is the inner product that is, element-by-element products added together of the ith row of A with the jth column of B.

So $C_{ij} = \sum_{k=1}^{3} A_{ik} \times B_{kj}$

$$A = \begin{pmatrix} 1 & 9 & 7 \\ 8 & 1 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 & 2 & 1 & 5 \\ 5 & 4 & 7 & 3 \\ 6 & 9 & 6 & 8 \end{pmatrix} \quad \Rightarrow \quad C = \begin{pmatrix} 90 & 101 & 106 & 88 \\ 41 & 38 & 27 & 59 \end{pmatrix}$$

Observe that $C_{11} = 1 \times (3) + 9 \times (5) + 7 \times (6) = 90$

$C_{23} = 8 \times (1) + 1 \times (7) + 2 \times (6) = 27$

Remark: If matrix C is the product of A and B, (i.e $C = AB$) then $C^T = (AB)^T = B^T A^T$.

There is no direct division concept but indirectly there is. To understand that we need to first define what is called identity matrix.

Identity matrix of order n is a square matrix with n rows and n columns with all the diagonal elements equals to 1 and all non diagonal elements equal to 0.

The identity matrix plays the same role as the number 1 in ordinary arithmetic. For example, if you multiply any number by 1, it remains the same. Similarly, if you multiply by any identity matrix, the resulting matrix is the same.

$$\text{So identity matrix of order } 3 = I_{3 \text{ x } 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Definition of the inverse of a matrix(Only possible for Square Matrix).

The inverse of a square matrix A, denoted by A^{-1} is the unique matrix such that $A \times A^{-1} = A^{-1} \times A = I$

It is the same concept as reciprocal. The reciprocal of 3 is 1/3 and if you multiply 3 and 1/3 you get 1.

With all these concepts now we can write the MLR model as follows:

If we plug in all the data points (n) in the model and create the n-equations, it looks like

$$\begin{aligned}
Y_1 &= \beta_0 \ + \ \beta_1 \times X_{11} \ + \ \beta_2 \times X_{12} \ + \ \ldots \ + \ \beta_k \times X_{1k} \ + \ e_1 \\
Y_2 &= \beta_0 \ + \ \beta_1 \times X_{21} \ + \ \beta_2 \times X_{22} \ + \ \ldots \ + \ \beta_k \times X_{2k} \ + \ e_2 \\
.... &= .... \\
Y_i &= \beta_0 \ + \ \beta_1 \times X_{i1} \ + \ \beta_2 \times X_{i2} \ + \ \ldots \ + \ \beta_k \times X_{ik} \ + \ e_i \\
.... &= .... \\
Y_n &= \beta_0 \ + \ \beta_1 \times X_{n1} \ + \ \beta_2 \times X_{n2} \ + \ \ldots \ + \ \beta_k \times X_{nk} \ + \ e_n
\end{aligned}$$

All the above equations can be written in the matrix form as follows:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ ... \\ Y_i \\ ... \\ Y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} 1 & X_{11} & X_{12} & .... & X_{1k} \\ 1 & X_{21} & X_{22} & .... & X_{2k} \\ .. & ... & ... & .... & ... \\ 1 & X_{i1} & X_{i2} & .... & X_{ik} \\ .. & ... & ... & .... & ... \\ 1 & X_{n1} & X_{n2} & .... & X_{nk} \end{pmatrix}_{n\times(k+1)} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ ..... \\ \beta_k \end{pmatrix}_{(k+1)\times 1} + \begin{pmatrix} e_1 \\ e_2 \\ ... \\ e_i \\ ... \\ e_n \end{pmatrix}_{n\times 1}$$

As a result, MLR model (in matrix form) is written as

$$Y_{n\times 1} = X_{n\times(k+1)} \times \beta_{(k+1)\times 1} \ + \ e_{n\times 1}$$

Now given the n data points which are all 1×(k+1) row vector $((Y_i X_{i1} X_{i2} \ldots X_{ik}))$, we should be able to estimate all the β_i's. It is done using the same concept of least square. We can write the sum of squared error as

$$SSE = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1.x_{i1} - \ldots - \beta_k.x_{ik})^2$$

and then take partial derivatives w.r.t each β_i's and create (k+1) equations. Finally solving those equations we get the estimated value of the β-vector as

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

Observe that $(X'X)^{-1}X'$ is a matrix of order (k + 1) by n and each β_i is a linear combination of Y_i s.

Now if we need to predict for a new values of independent variables $(X_{01}, X_{02}, \ldots, X_{0k})$, we use the equation,

$$\hat{Y} = X_0'.\hat{\beta} \implies \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \times X_{01} + \hat{\beta}_2 \times X_{02} + \ldots + \hat{\beta}_k \times X_{0k}$$

If we plug in all the X's in the data, we can get the n fitted values as

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = \mathbf{H}Y \quad \text{where } \mathbf{H} = X(X'X)^{-1}X'$$

Note that in the above equation, both Y and $\hat{Y}$ are $(n \times 1)$ column vector where $\hat{y}_i$ is the corresponding fitted value of the observed y_i. So basically the $(n \times n)$ H matrix put the hat on Y and that's why sometimes it is called "hat-matrix".

This n x n matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$ maps the vector of observed values into a vector of fitted values. The hat matrix and its properties play a central role in regression analysis.

The difference between the observed value y_i and the corresponding fitted value is the residual $e_i = y_i - \hat{y}_i$.

The n residuals may be conveniently written in matrix notation as

$$\begin{aligned} e_{n\times 1} &= y_{n\times 1} - \hat{y}_{n\times 1} = y_{n\times 1} - X\hat{\beta} \\ &= y_{n\times 1} - H_{n\times n}\ y_{n\times 1} = (I - H)_{n\times n}\ y_{n\times 1} \end{aligned}$$

where I is the n x n identity matrix.

<u>Properties of Least Squares Estimators:</u>

It is easy to show that if the model is correct then the estimators are unbiased meaning $E(\hat{\beta}) = \beta$ (β as a vector).

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1}X'Y] = (X'X)^{-1}X'E[Y] \\ &= (X'X)^{-1}X'E[(X\beta + e)] \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'E[e] \\ &= \beta + (X'X)^{-1}X'E[e] = \beta \end{aligned}$$

as $(X'X)^{-1}(X'X) = I$ and E[e] $= 0$

As the random variable $\hat{\beta}$ is a vector, we need to look at Variance-Covariance matrix instead of just variance as in SLR model. Variance-Covariance matrix is a matrix where ijth element provides the covariance between ith and jth random variables when $i \neq j$ and the variance $i = j$.

Result: $T_{m\times 1}$ is a random vector then $Cov(MT) = MCov(T)M'$ where M is a known matrix.

Using the above results, we an see that

$$\begin{aligned}
Cov(\hat{\beta}) &= Cov((X'X)^{-1}X'Y) \\
&= (X'X)^{-1}X' \; Cov(Y) \; X(X'X)^{-1} \\
&= (X'X)^{-1}X' \; \sigma^2 I \; X(X'X)^{-1} \\
&= \; \sigma^2 (X'X)^{-1}X' \; X(X'X)^{-1} = \sigma^2 \, (X'X)^{-1}
\end{aligned}$$

If we let $C = (X'X)^{-1}$ and C_{ij} represent the ijth element of C then

$$C_{ij} = \begin{cases} Cov(\hat{\beta}_i, \hat{\beta}_j) & \text{for } i \neq j \\ Var(\hat{\beta}_i) & \text{for } i = j \end{cases}$$

These C_{ij}'s help to develop the inference for all β's and their combinations.

Estimation of σ^2:

As in SLR model, we built the estimation of σ^2 based on error sum of squares (i.e SSE).

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = e'e$$

As $e = y - X\hat{\beta}$ we substitute the value of e in above line and get

$$SSE = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - 2\hat{\beta}X'y + \hat{\beta}X'X\hat{\beta} = y'y - \hat{\beta}X'y$$

We get the last equality because $X'X\hat{\beta} = X'y$

Then $\hat{\sigma}^2$ = Mean Square Error = $\frac{SSE}{n-(k+1)}$

Hence the standard error of estimate $\hat{\sigma} = \sqrt{MSE} = \sqrt{\frac{SSE}{n-(k+1)}}$

Example-4.1: (Revisit) Going back to our problem, it can be easily seen that here we have 3 (=k) independent variables and 28(=n) observations.

The multiple regression model is $Y = \beta_0 + \beta_2 X_2 + \beta_7 X_7 + \beta_8 X_8 + e$

In matrix form with 28 observations it translates to,

$$\underline{\mathbf{Y}} \quad = \quad \underline{\mathbf{X}} \quad \times \quad \underline{\beta} \quad + \quad \underline{\mathbf{e}}$$

$$\begin{pmatrix} 10 \\ 11 \\ 11 \\ \cdots \\ \cdots \\ 0 \end{pmatrix}_{28\times 1} = \begin{pmatrix} 1 & 1985 & 59.7 & 2205 \\ 1 & 2855 & 55 & 2096 \\ 1 & 1737 & 65.6 & 1847 \\ .. & \cdots & \cdots & \cdots \\ .. & \cdots & \cdots & \cdots \\ 1 & 1503 & 53.5 & 2560 \end{pmatrix}_{28\times(3+1)} \times \begin{pmatrix} \beta_0 \\ \beta_2 \\ \beta_7 \\ \beta_8 \end{pmatrix}_{(3+1)\times 1} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \cdots \\ \cdots \\ e_{28} \end{pmatrix}_{28\times 1}$$

Though it is possible to calculate each part of each formula using matrix calculation, instead we will rely on the softwares and the results.

Though we will calculate and recognize all the numbers in the out of R but for now we should be able to pick up the estimated coefficients $(\hat{\beta})$ and SSE from the output and the calculate the estimated standard error $(\hat{\sigma})$.

Observe that estimated coefficients are (from the output at the end)

$$\hat{\beta}_0 = -1.8084,\ \ \hat{\beta}_2 = 0.0036,\ \ \hat{\beta}_7 = 0.1940 \text{ and } \ \hat{\beta}_8 = -0.0048$$

The estimated equation is :$Y = -1.8084 + 0.0036\,X_2 + 0.1940\,X_7 - 0.0048\,X_8$

Interpretation:

$\hat{\beta}_2$ - If the passing yards can be increased by 100 yards then number of games own will increase by 0.36 (keeping every other variable same).

$\hat{\beta}_7$ - If rushing percent goes up by 10% then number of game won will go up by 1.9 (almost 2).

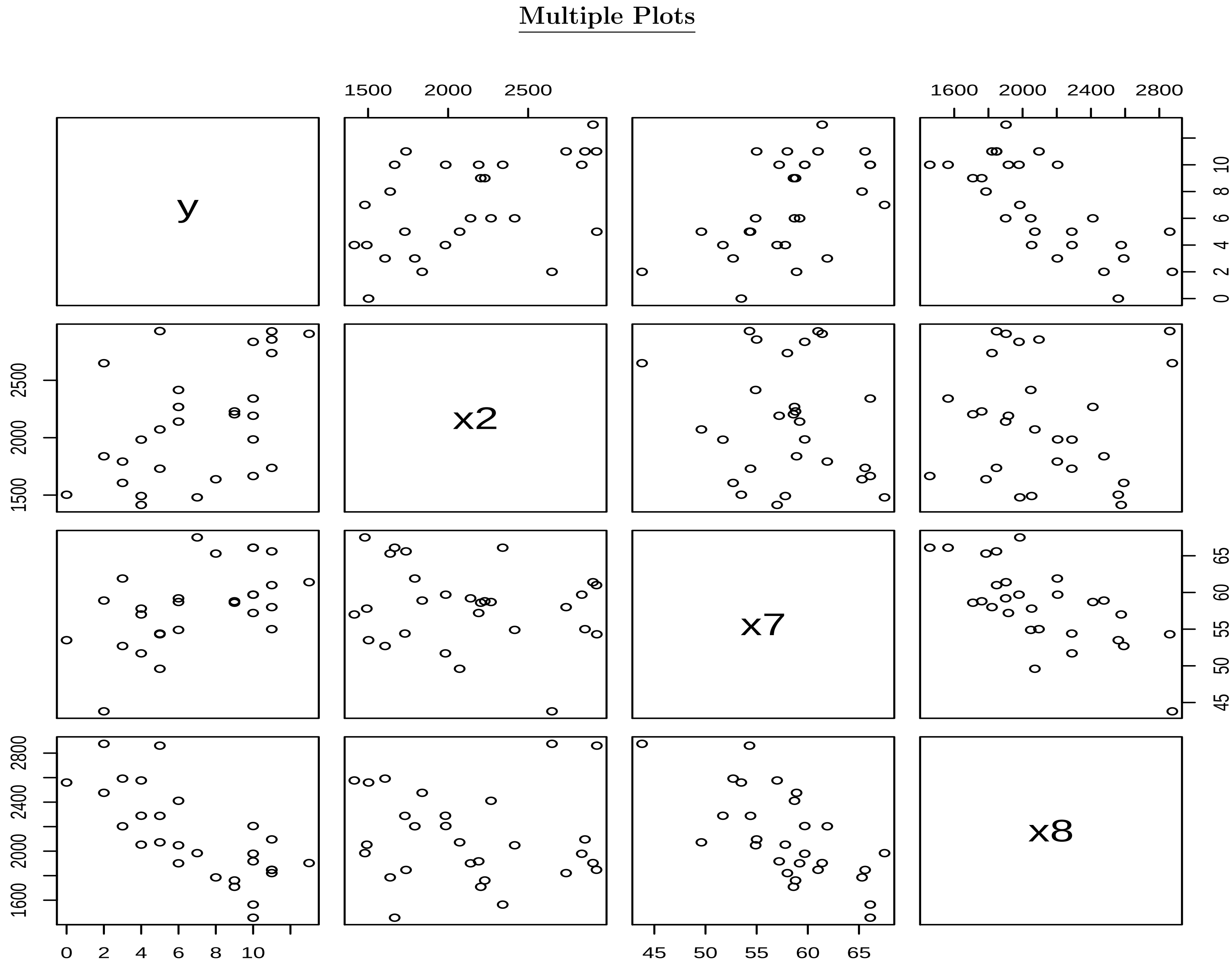
$\hat{\beta}_8$ - If the opponents rushing yards goes up by 100 yards then the number of game won is expected to go down by 0.48. (as it is negative).

If you look at the other parts of the output, we see that

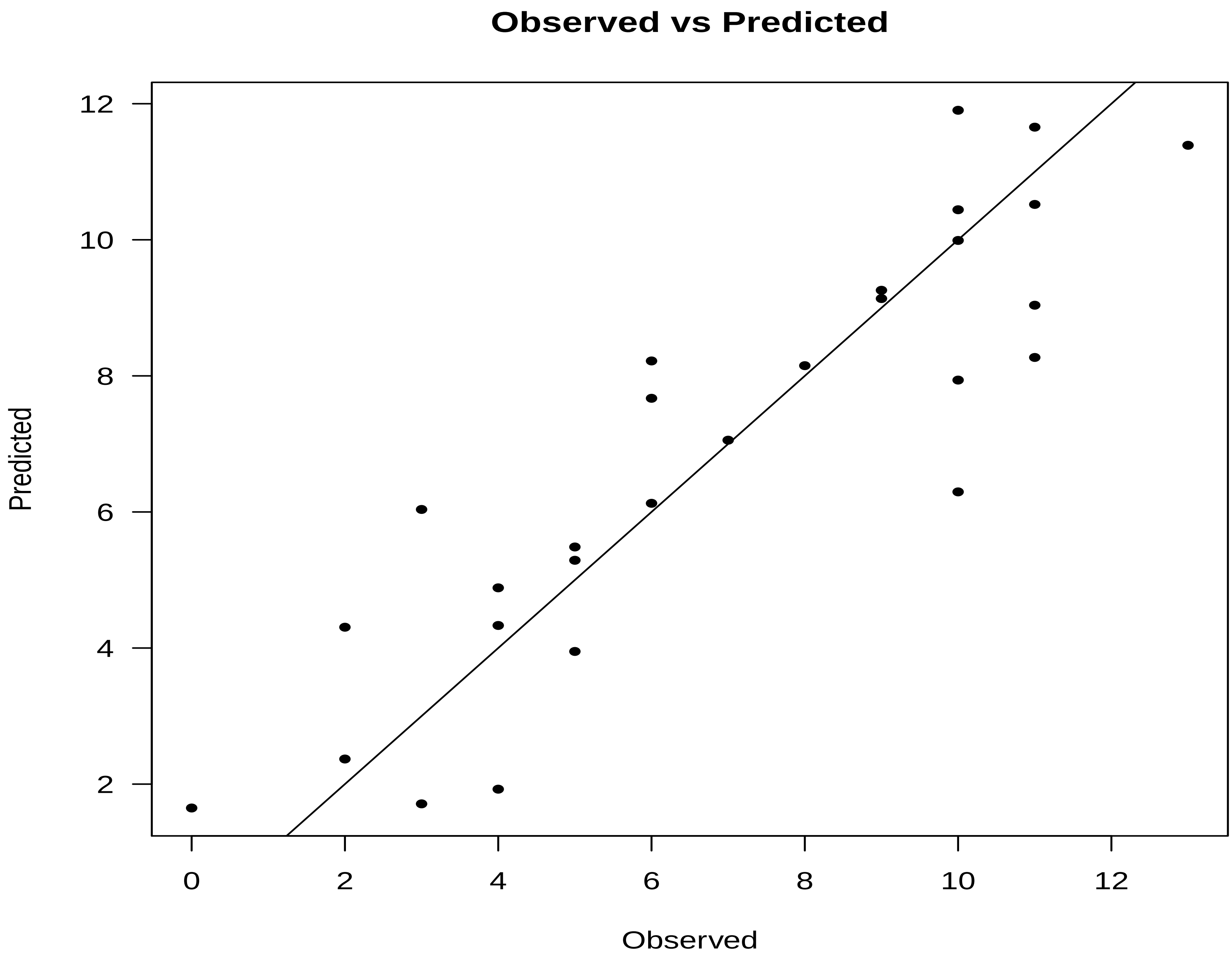
$\hat{\sigma} = 1.706 =$ Residual standard error (in R - output).

If you look at the SAS output (at the end), you can observe almost the same numbers. Note that SAS labeled the $\hat{\sigma}$ as Root MSE and also it gives many other information.

One of the first thing to do is to plot all the variables against each other.



Sometime it is good idea to plot observed against predicted values.



Summary - What you should learn

1. What is MLR model and why it is an extension of SLR model.
2. The matrix form of MLR model.
3. The X & H -Matrix and Y,β,e vectors.
4. Basic Run of MLR model by R.

Output from R - Problem 3.1

```
Call:
lm(formula = y ~ x2 + x7 + x8, data = nfldata)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0370 -0.7129 -0.2043  1.1101  3.7049 



Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.808372   7.900859  -0.229 0.820899    
x2           0.003598   0.000695   5.177 2.66e-05 ***
x7           0.193960   0.088233   2.198 0.037815 *  
x8          -0.004816   0.001277  -3.771 0.000938 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 


Residual standard error: 1.706 on 24 degrees of freedom
Multiple R-squared:  0.7863, Adjusted R-squared:  0.7596 
F-statistic: 29.44 on 3 and 24 DF,  p-value: 3.273e-08 


 (Intercept)           x2           x7           x8 
-1.808372059  0.003598070  0.193960210 -0.004815494 
```

Output from SAS - Problem 3.1

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: y y

Number of Observations Read	28
Number of Observations Used	28

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	257.09428	85.69809	29.44	<.0001
Error	24	69.87000	2.91125		
Corrected Total	27	326.96429			

Root MSE	1.70624	**R-Square**	0.7863
Dependent Mean	6.96429	**Adj R-Sq**	0.7596
Coeff Var	24.49984		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	-1.80837	7.90086	-0.23	0.8209
x2	x2	1	0.00360	0.00069500	5.18	<.0001
x7	x7	1	0.19396	0.08823	2.20	0.0378
x8	x8	1	-0.00482	0.00128	-3.77	0.0009