Chapter 6

Multiple Linear Regression

Chanwoo Yoo, Division of Advanced Engineering,
Korea National Open University

Contents

1. Example on IQ and Physical Characteristics

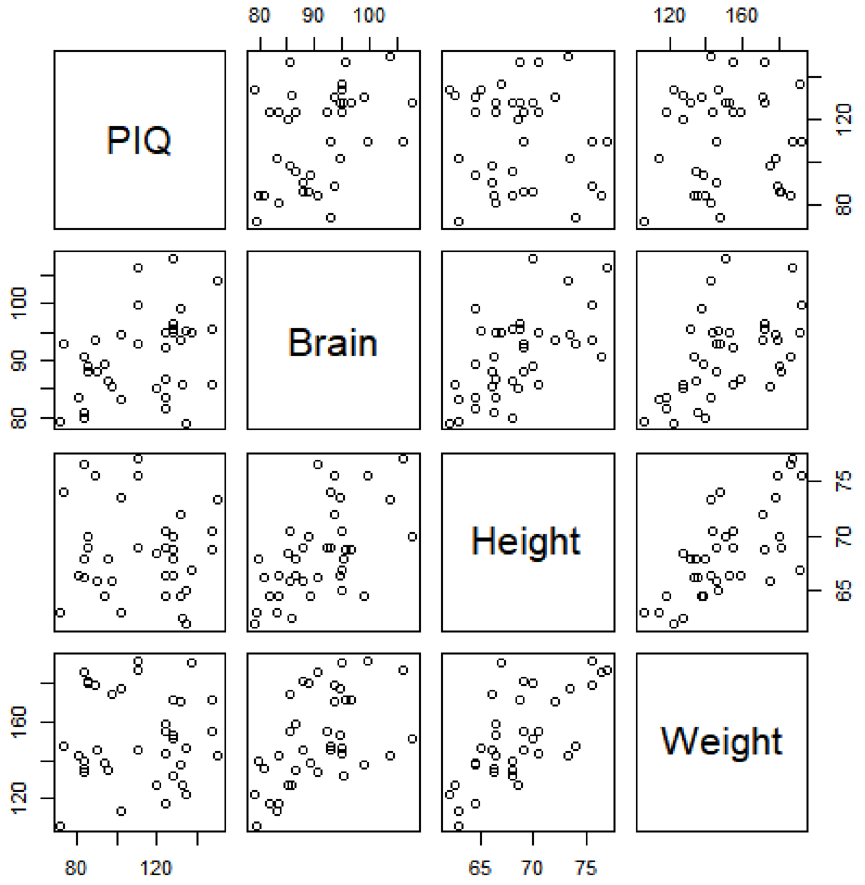
1. Data: IQ Size

- Data: IQ Size
  - y (PIQ): Performance IQ scores from the revised Wechsler Adult Intelligence Scale.
  - x_1 (Brain): Brain size based on the count obtained from MRI scans (given as count/10,000).
  - x_2 (Height): Height in inches.
  - x_3 (Weight): Weight in pounds.

2. Code: IQ Size

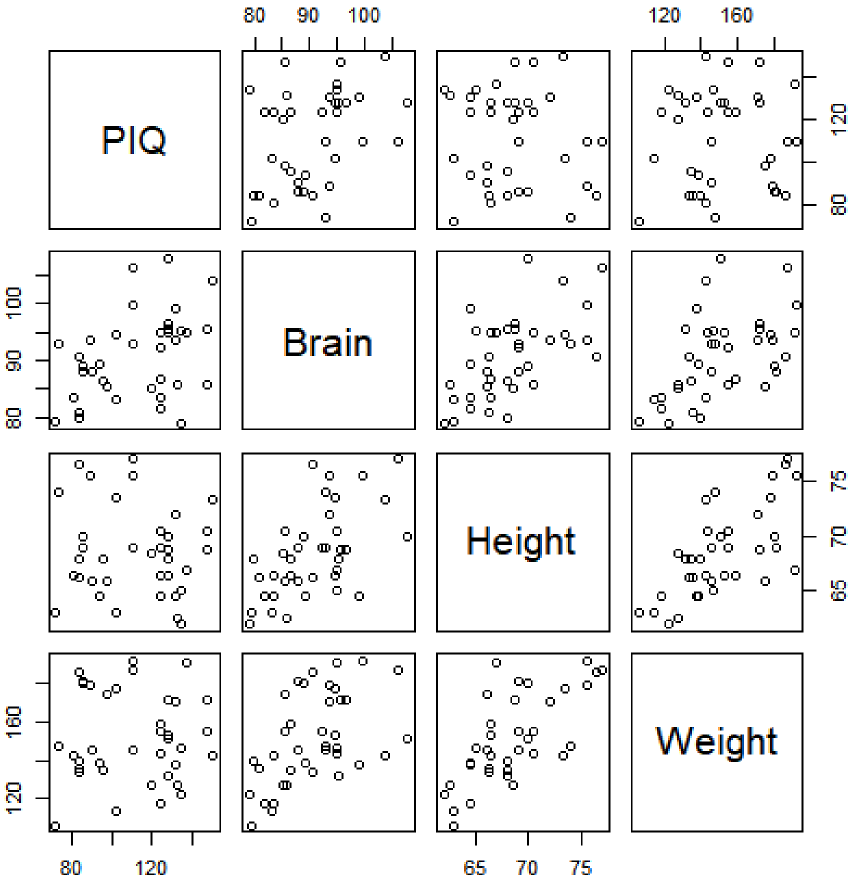
```
iqsize <- read.table("iqsize.txt", header=T)
attach(iqsize)

pairs(cbind(PIQ, Brain, Height, Weight))

model <- lm(PIQ ~ Brain + Height + Weight)
summary(model)
```

3. Scatter Plot Matrix



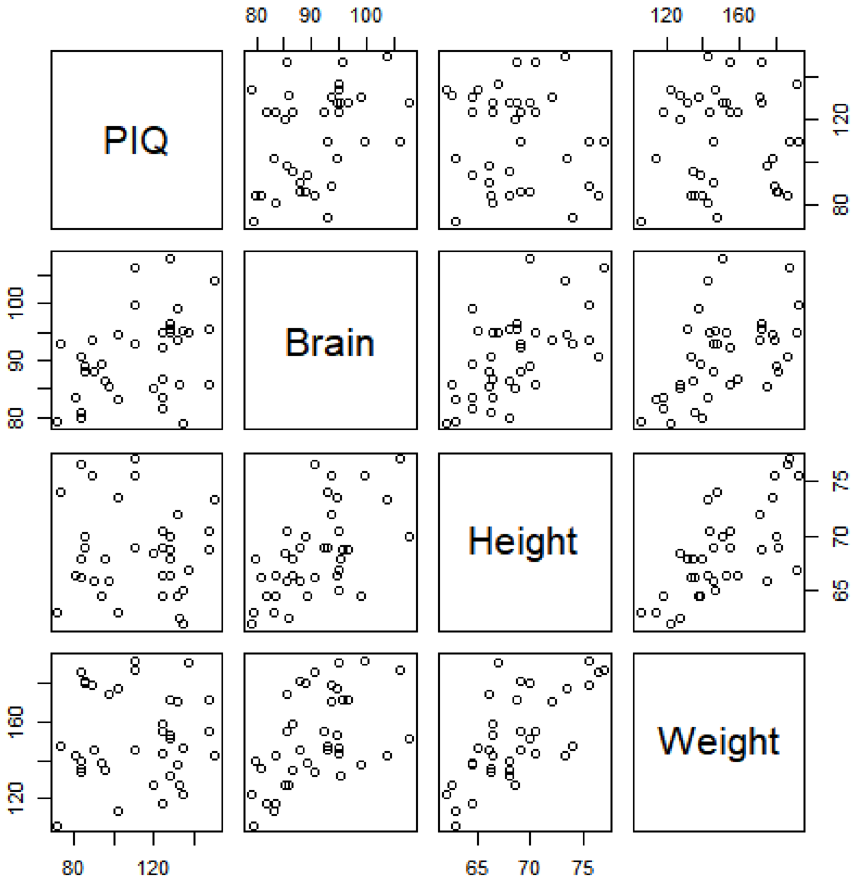
- Scatter plot matrix contains a scatter plot of each pair of variables arranged in an orderly array.

3. Scatter Plot Matrix



- It appears that brain size is the best single predictor of PIQ, but none of the relationships are particularly strong.

3. Scatter Plot Matrix



- In multiple linear regression, the challenge is to see how the response y relates to all three predictors simultaneously.

4. Results: IQ Size

```
> summary(model)

Call:
lm(formula = PIQ ~ Brain + Height + Weight)

Residuals:
   Min      1Q  Median      3Q     Max
-32.74 -12.09  -3.84   14.17   51.69
```

4. Results: IQ Size

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.114e+02  6.297e+01   1.768 0.085979 .
Brain        2.060e+00  5.634e-01   3.657 0.000856 ***
Height      -2.732e+00  1.229e+00  -2.222 0.033034 *
Weight       5.599e-04  1.971e-01   0.003 0.997750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. Results: IQ Size

```
Residual standard error: 19.79 on 34 degrees of freedom
Multiple R-squared:  0.2949,  Adjusted R-squared:  0.2327
F-statistic: 4.741 on 3 and 34 DF,  p-value: 0.007215
```

- The R^2 value is 29.49%. This tells us that 29.49% of the variation in intelligence, as quantified by PIQ, is reduced by taking into account brain size, height and weight.

4. Results: IQ Size

```
Residual standard error: 19.79 on 34 degrees of freedom
Multiple R-squared:  0.2949,   Adjusted R-squared:  0.2327
F-statistic: 4.741 on 3 and 34 DF,  p-value: 0.007215
```

- The adjusted R^2 value is 23.27%. When considering different multiple linear regression models for PIQ, we could use this value to help compare the models.

5. Adjusted r^2

- Adjusted $R^2 = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2)$
  - n: number of samples
  - p: number of parameters

2. Multiple Linear Regression Model

1. Multiple Linear Regression Model

- $y_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}) + \epsilon_i$
  - y_i is the intelligence (PIQ) of student i
  - x_{i1} is the brain size (Brain) of student i
  - x_{i2} is the height (Height) of student i
  - x_{i3} is the weight (Weight) of student i
  - ϵ_i is the independent error terms, which follow a normal distribution with mean 0 and equal variances σ^2.

1. Multiple Linear Regression Model

- $y_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}) + \epsilon_i$
  - Because we have more than one predictor (x) variable, we use slightly modified notation. The x-variables (e.g. x_{i1}, x_{i2} and x_{i3}) are now subscripted with a 1, 2, and 3 as a way of keeping track of the three different quantitative variables. We also subscript the slope parameters with the corresponding numbers (e.g. β_1, β_2 and β_3).

1. Multiple Linear Regression Model

- $y_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}) + \epsilon_i$
  - The "LINE" conditions must still hold for the multiple linear regression model. The linear part comes from the formulated regression function — it is, what we say, "linear in the parameters." This simply means that each beta coefficient multiplies a predictor variable or a transformation of one or more predictor variables.

2. Notation for the Population Model

- $y_i = (\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{p-1} x_{i,p-1}) + \epsilon_i$
  - We assume that the have a normal distribution with mean 0 and constant variance σ^2.
  - The subscript i refers to the ith individual or unit in the population.
  - The model includes p−1 x−variables, but p regression parameters (beta) because of the intercept term β_0.

3. Estimates of the Model Parameters

- The estimates of the β parameters are the values that minimize the sum of squared errors for the sample.
  - Least squares estimation

3. Estimates of the Model Parameters

- The letter b is used to represent a sample estimate of a β parameter. Thus b_0 is the sample estimate of β_0, b_1 is the sample estimate of β_1, and so on.

3. Estimates of the Model Parameters

- $MSE = \frac{SSE}{n-p}$ estimates σ^2, the variance of the errors.
  - n = sample size,
  - p = number of parameters in the model (including the intercept)
  - SSE = sum of squared errors. Notice that for simple linear regression p = 2.

3. Estimates of the Model Parameters

- $S = \sqrt{MSE}$ estimates and is known as the regression standard error or the residual standard error.

4. Interpretation of the Model Parameters

- Each β parameter represents the change in the mean response, E(y), per unit increase in the associated predictor variable when all the other predictors are held constant.
- The intercept term, β_0, represents the estimated mean response, E(y), when all the predictors , x_1, x_2, …, x_{p-1}, are all zero.

5. Predicted Values and Residuals

- A predicted value is calculated as $\hat{y}_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \cdots + b_{p-1} x_{i,p-1}$.
- A residual (error) term is calculated as $e_i = y_i - \hat{y}_i$, the difference between an actual and a predicted value of y.
- A plot of residuals (vertical) versus predicted values (horizontal) ideally should resemble a horizontal random band. Departures from this form indicates difficulties with the model and/or data.

5. Predicted Values and Residuals

- Other residual analyses can be done exactly as we did in simple regression. For instance, we might wish to examine a normal probability plot (NPP) of the residuals.
- Additional plots to consider are plots of residuals versus each x-variable separately. This might help us identify sources of curvature or nonconstant variance.

3. Least Squares Estimation

1. Least Squares Estimation

- $A = \begin{bmatrix} 0.1 & -0.1 \\ 0.2 & 0.1 \\ 0.3 & 0.2 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$
  - $2 = b_0 + 0.1b_1 - 0.1b_2$
  - $3 = b_0 + 0.2b_1 + 0.1b_2$
  - $4 = b_0 + 0.3b_1 + 0.2b_2$

1. Least Squares Estimation

- $2 = b_0 + 0.1b_1 - 0.1b_2$
- $3 = b_0 + 0.2b_1 + 0.1b_2$
- $4 = b_0 + 0.3b_1 + 0.2b_2$

- $\mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} b_0 & 0.1b_1 & -0.1b_2 \\ b_0 & 0.2b_1 & 0.1b_2 \\ b_0 & 0.3b_1 & 0.2b_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 & -0.1 \\ 1 & 0.2 & 0.1 \\ 1 & 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$

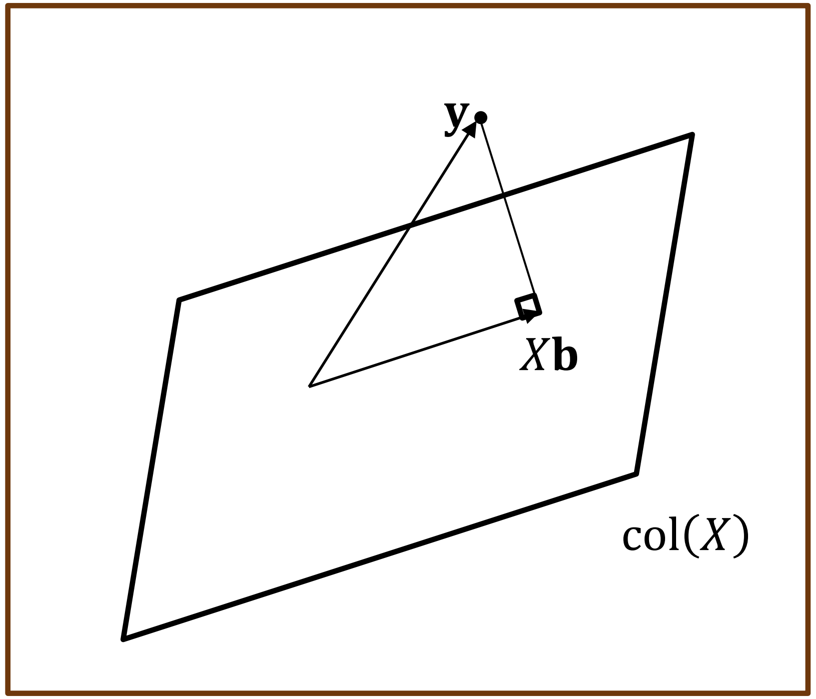
1. Least Squares Estimation

- $\mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} b_0 & 0.1b_1 & -0.1b_2 \\ b_0 & 0.2b_1 & 0.1b_2 \\ b_0 & 0.3b_1 & 0.2b_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 & -0.1 \\ 1 & 0.2 & 0.1 \\ 1 & 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$

- $X = \begin{bmatrix} 1 & 0.1 & -0.1 \\ 1 & 0.2 & 0.1 \\ 1 & 0.3 & 0.2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \rightarrow \mathbf{y} = X\mathbf{b}$

1. Least Squares Estimation

- $\mathbf{y} = X\mathbf{b}$



1. Least Squares Estimation

- Minimize $\|\mathbf{y} - X\mathbf{b}\|^2$

$$L = \|\mathbf{y} - X\mathbf{b}\|^2$$
$$= (\mathbf{y} - X\mathbf{b})^T(\mathbf{y} - X\mathbf{b})$$
$$= (\mathbf{y}^T - \mathbf{b}^T X^T)(\mathbf{y} - X\mathbf{b})$$
$$= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T X\mathbf{b} - \mathbf{b}^T X^T\mathbf{y} + \mathbf{b}^T X^T X\mathbf{b}$$

$$\frac{\partial L}{\partial \mathbf{b}} = -\mathbf{y}^T X - \mathbf{y}^T X + 2\mathbf{b}^T X^T X = 0$$

1. Least Squares Estimation

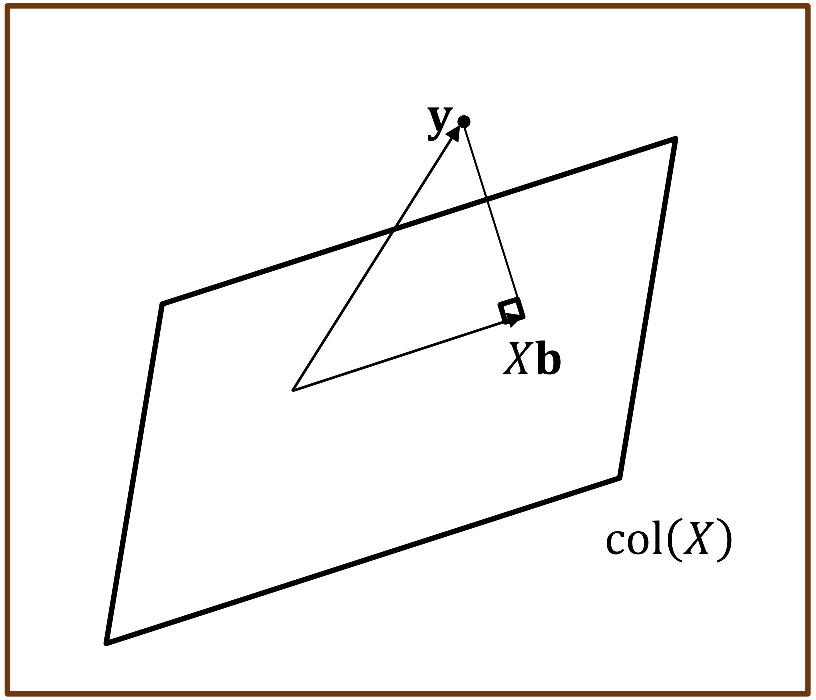
- Minimize $\|\mathbf{y} - X\mathbf{b}\|^2$
  - $\frac{\partial L}{\partial \mathbf{b}} = -\mathbf{y}^T X - \mathbf{y}^T X + 2\mathbf{b}^T X^T X = 0$
  - $2\mathbf{b}^T X^T X = 2\mathbf{y}^T X$
  - $X^T X\mathbf{b} = X^T\mathbf{y}$ (normal equation)
  - $\mathbf{b} = (X^T X)^{-1} X^T\mathbf{y}$

2. Hat Matrix

- $\hat{\mathbf{y}} = X\mathbf{b}$
- $\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}$
- $\hat{\mathbf{y}} = X(X^T X)^{-1} X^T \mathbf{y}$
- Hat Matrix: $X(X^T X)^{-1} X^T$

3. Least Squares Estimation: Another Way

- $X\mathbf{b} = proj_{col(X)}\mathbf{y}$
- $\mathbf{y} - X\mathbf{b} = \mathbf{y} - proj_{col(X)}\mathbf{y}$
- $(\mathbf{y} - proj_{col(X)}\mathbf{y}) \perp col(X)$
- $col(X) = row(X^T), row(X^T) \perp \mathrm{null}(X^T)$
- $(\mathbf{y} - proj_{col(X)}\mathbf{y}) \in \mathrm{null}(X^T)$
- $X^T\left(\mathbf{y} - proj_{col(X)}\mathbf{y}\right) = \mathbf{0}$



3. Least Squares Estimation: Another Way

- $X^T(\mathbf{y} - proj_{col(X)}\mathbf{y}) = \mathbf{0}$
- $X^T(\mathbf{y} - X\mathbf{b}) = \mathbf{0}$
- $X^TX\mathbf{b} = X^T\mathbf{y}$

4. Linear Dependence

- $\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}$
- Estimates $\mathbf{b}$ depends on $(X^T X)^{-1}$.
- The inverse of a square matrix exists only if the columns are linearly independent.
- $\mathbf{b}$ cannot be uniquely determined if some of the columns of X are linearly dependent.

4. Linear Dependence

```
Brain2 = Brain * 2
model <- lm(PIQ ~ Brain + Brain2)
summary(model)
```

4. Linear Dependence

```
> summary(model)

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6519    43.7118   0.106   0.9158
Brain         1.1766     0.4806   2.448   0.0194 *
Brain2            NA         NA      NA       NA
```

4. Linear Dependence

- Don't collect your data in such a way that the predictor variables are perfectly correlated.
- If your software package reports an error message concerning high correlation among your predictor variables, then think about linear dependence and how to get rid of it.

Next

Chapter 7
MLR Model Evaluation