STAT 423/523 Statistical Methods for Engineers and Scientists

Lecture 11: Multiple Linear Regression

Chencheng Cai

Washington State University

Multiple Linear Regression

In cases when we have more than one predictor variable, we can extend the simple linear regression model to a **multiple linear regression model**:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \beta_p x_{ki} + \epsilon_i,$$

where

- y_i is the response variable,
- x_{ji} is the jth predictor variable for the ith observation
- $\epsilon_i \sim N(0, \sigma^2)$ is the error term.

Multiple Linear Regression

In cases when we have more than one predictor variable, we can extend the simple linear regression model to a **multiple linear regression model**:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \beta_p x_{ki} + \epsilon_i,$$

where

- y_i is the response variable,
- x_{ji} is the jth predictor variable for the ith observation
- $\epsilon_i \sim N(0, \sigma^2)$ is the error term.

The predictors could be:

- additional covariates in the dataset
- interactions between predictors
- nonlinear functions of predictors

Ordinary Least Squares

We follow the same principle as in simple linear regression and minimize the residual sum of squares (RSS):

$$\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k = \underset{\beta_0, \beta_1, \ldots, \beta_k}{\arg\min} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2$$

Ordinary Least Squares

We follow the same principle as in simple linear regression and minimize the residual sum of squares (RSS):

$$\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k = \underset{\beta_0, \beta_1, \ldots, \beta_k}{\arg\min} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2$$

We compute the partial derivatives of the RSS with respect to each β_j:

$$\frac{\partial \mathrm{RSS}}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})$$

$$\frac{\partial \mathrm{RSS}}{\partial \beta_j} = -2 \sum_{i=1}^{n} x_{ji} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki}), \ j = 1, \ldots, k$$

Ordinary Least Squares

The OLS estimators can be obtained by setting the partial derivatives to zero:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki}) = 0$$

$$\sum_{i=1}^{n} x_{1i}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki}) = 0$$

$$\sum_{i=1}^{n} x_{2i}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki}) = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} x_{ki}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki}) = 0$$

Ordinary Least Squares

This is a linear system of equations in the unknowns $\beta_0, \beta_1, \ldots, \beta_k$.

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

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

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

$$\vdots$$

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

Ordinary Least Squares

We can write it in matrix form:

$$
\begin{bmatrix}
n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{ki} \\
\sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \cdots & \sum_{i=1}^n x_{1i}x_{ki} \\
\sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{2i}x_{1i} & \sum_{i=1}^n x_{2i}^2 & \cdots & \sum_{i=1}^n x_{2i}x_{ki} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{ki}x_{1i} & \sum_{i=1}^n x_{ki}x_{2i} & \cdots & \sum_{i=1}^n x_{ki}^2
\end{bmatrix}
\begin{bmatrix}
\beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k
\end{bmatrix}
=
\begin{bmatrix}
\sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i}y_i \\ \sum_{i=1}^n x_{2i}y_i \\ \vdots \\ \sum_{i=1}^n x_{ki}y_i
\end{bmatrix}
$$

Ordinary Least Squares

We can write it in matrix form:

$$\begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{ki} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \cdots & \sum_{i=1}^n x_{1i}x_{ki} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{2i}x_{1i} & \sum_{i=1}^n x_{2i}^2 & \cdots & \sum_{i=1}^n x_{2i}x_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{ki}x_{1i} & \sum_{i=1}^n x_{ki}x_{2i} & \cdots & \sum_{i=1}^n x_{ki}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i}y_i \\ \sum_{i=1}^n x_{2i}y_i \\ \vdots \\ \sum_{i=1}^n x_{ki}y_i \end{bmatrix}$$

more compactly, we can write it as:

$$\begin{bmatrix} S_{x_0x_0} & S_{x_0x_1} & S_{x_0x_2} & \cdots & S_{x_0x_k} \\ S_{x_1x_0} & S_{x_1x_1} & S_{x_1x_2} & \cdots & S_{x_1x_k} \\ S_{x_2x_0} & S_{x_2x_1} & S_{x_2x_2} & \cdots & S_{x_2x_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{x_kx_0} & S_{x_kx_1} & S_{x_kx_2} & \cdots & S_{x_kx_k} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} S_{x_0y} \\ S_{x_1y} \\ S_{x_2y} \\ \vdots \\ S_{x_ky} \end{bmatrix}$$

where $S_{x_jx_l} = \sum_{i=1}^n x_{ji}x_{li}$ and $S_{x_jy} = \sum_{i=1}^n x_{ji}y_i$ with $x_{0i} = 1$.

Ordinary Least Squares

The OLS estimators can be computed using matrix algebra:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} S_{x_0x_0} & S_{x_0x_1} & S_{x_0x_2} & \cdots & S_{x_0x_k} \\ S_{x_1x_0} & S_{x_1x_1} & S_{x_1x_2} & \cdots & S_{x_1x_k} \\ S_{x_2x_0} & S_{x_2x_1} & S_{x_2x_2} & \cdots & S_{x_2x_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{x_kx_0} & S_{x_kx_1} & S_{x_kx_2} & \cdots & S_{x_kx_k} \end{bmatrix}^{-1} \begin{bmatrix} S_{x_0y} \\ S_{x_1y} \\ S_{x_2y} \\ \vdots \\ S_{x_ky} \end{bmatrix}$$

Ordinary Least Squares

We can verify the solution is compatible with the simple linear regression case.

Ordinary Least Squares

We can verify the solution is compatible with the simple linear regression case. When $k = 1$, we have:

$$\begin{aligned}
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \begin{bmatrix} S_{x_0x_0} & S_{x_0x_1} \\ S_{x_1x_0} & S_{x_1x_1} \end{bmatrix}^{-1} \begin{bmatrix} S_{x_0y} \\ S_{x_1y} \end{bmatrix} \\
&= \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \\
&= \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \\
&= \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ n \sum x_i y_i - \sum x_i \sum y_i \end{bmatrix} \\
&= S_{xx}^{-1} \begin{bmatrix} \bar{y} S_{xx} - \bar{x} S_{xy} \\ S_{xy} \end{bmatrix}
\end{aligned}$$

Ordinary Least Squares

For the variance component, we have:

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{RSS}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}$$

where

Ordinary Least Squares

For the variance component, we have:

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{RSS}(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}$$

where

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}$ is the **predicted** or **fitted** value of y_i
- The degrees of freedom is $n-k-1$ because we have estimated $k+1$ parameters $(\beta_0, \beta_1, \ldots, \beta_k)$ from the data.

Ordinary Least Squares

The OLS estimators are **unbiased**:

$$\mathbb{E}[\hat{\beta}_j] = \beta_j, \;\; j = 0, 1, \ldots, k$$

Ordinary Least Squares

The OLS estimators are **unbiased**:

$$\mathbb{E}[\hat{\beta}_j] = \beta_j, \; j = 0, 1, \ldots, k$$

Let $s_{\hat{\beta}_j}$ be the estimated standard error of $\hat{\beta}_j$. Then

$$\frac{\hat{\beta}_j}{s_{\hat{\beta}_j}} \sim t_{n-k-1}$$

which is a t-distribution with $n - k - 1$ degrees of freedom.

Confidence interval and t-test

The $(1 - \alpha)$ confidence interval for β_j is given by:

$$\hat{\beta}_j \pm t_{\alpha/2, n-k-1} s_{\hat{\beta}_j}.$$

Confidence interval and t-test

The $(1-\alpha)$ confidence interval for β_j is given by:

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1} s_{\hat{\beta}_j}.$$

Consider the hypothesis test:

$$H_0 : \beta_j = 0 \text{ vs. } H_a : \beta_j \neq 0$$

We reject H_0 if:

Confidence interval and t-test

The $(1-\alpha)$ confidence interval for β_j is given by:

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1} s_{\hat{\beta}_j}.$$

Consider the hypothesis test:

$$H_0 : \beta_j = 0 \text{ vs. } H_a : \beta_j \neq 0$$

We reject H_0 if:

- The CI does not contain 0.
- The t-statistic

  $$t = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}$$

  has absolute value greater than $t_{\alpha/2,n-k-1}$.
- The p-value

  $$p = 2(1 - F_{t,n-k-1}(|t|))$$

  is less than α.

Confidence interval and t-test

- The standard error of $\hat{\beta}_j$ can be read from the output of the regression models in R and Python.
- Sometimes the p-values are reported in the output as well.

Confidence interval and t-test

- The standard error of $\hat{\beta}_j$ can be read from the output of the regression models in R and Python.
- Sometimes the p-values are reported in the output as well.
- A covariate $x_{ji}, i = 1, \ldots, n$ is **significant** if the null hypothesis $H_0 : \beta_j = 0$ is rejected.
- A covariate $x_{ji}, i = 1, \ldots, n$ is **insignificant** if the null hypothesis $H_0 : \beta_j = 0$ is not rejected.

Confidence interval and t-test

- The standard error of $\hat{\beta}_j$ can be read from the output of the regression models in R and Python.
- Sometimes the p-values are reported in the output as well.
- A covariate $x_{ji}, i = 1, \ldots, n$ is **significant** if the null hypothesis $H_0 : \beta_j = 0$ is rejected.
- A covariate $x_{ji}, i = 1, \ldots, n$ is **insignificant** if the null hypothesis $H_0 : \beta_j = 0$ is not rejected.
- Insignificant covariates can be removed from the model to simplify the model.

Example

We consider the **mtcars** dataset in R and run a linear regression model of `mpg` (miles per gallon) on `disp` (displacement), `hp` (gross horsepower), and `wt` (weight of car).

```
Call:
lm(formula = mpg ~ disp + hp + wt, data = mtcars)

Residuals:
Min      1Q Median     3Q    Max
-3.891 -1.640 -0.172  1.061  5.861

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.105505   2.110815  17.579  < 2e-16 ***
disp        -0.000937   0.010350  -0.091  0.92851
hp          -0.031157   0.011436  -2.724  0.01097 *
wt          -3.800891   1.066191  -3.565  0.00133 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.639 on 28 degrees of freedom
Multiple R-squared:  0.8268,Adjusted R-squared:  0.8083
F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

Example

- The estimated intercept is $\hat{\beta}_0 = 37.11$.
- The estimated slope for `disp` is $\hat{\beta}_1 = -0.000937$.
- The estimated slope for `hp` is $\hat{\beta}_2 = -0.03116$.
- The estimated slope for `wt` is $\hat{\beta}_3 = -3.8009$.

Example

- The estimated intercept is $\hat{\beta}_0 = 37.11$.
- The estimated slope for `disp` is $\hat{\beta}_1 = -0.000937$.
- The estimated slope for `hp` is $\hat{\beta}_2 = -0.03116$.
- The estimated slope for `wt` is $\hat{\beta}_3 = -3.8009$.
- The intercept, `hp`, and `wt` are significant at $\alpha = 0.05$ level.
- The `disp` is insignificant at $\alpha = 0.05$ level.

Example

- The estimated intercept is $\hat{\beta}_0 = 37.11$.
- The estimated slope for `disp` is $\hat{\beta}_1 = -0.000937$.
- The estimated slope for `hp` is $\hat{\beta}_2 = -0.03116$.
- The estimated slope for `wt` is $\hat{\beta}_3 = -3.8009$.
- The intercept, `hp`, and `wt` are significant at $\alpha = 0.05$ level.
- The `disp` is insignificant at $\alpha = 0.05$ level.
- fitted model is

$$\text{mpg} = 37.11 - 0.0009 \times \text{disp} - 0.0312 \times \text{hp} - 3.801 \times \text{wt} + \epsilon \quad \text{with } \epsilon \sim N(0, 2.639^2)$$

Example

A direct improvement of the model is to remove `disp` from the model and refit the model:

```
Call:
lm(formula = mpg ~ hp + wt, data = mtcars)

Residuals:
Min      1Q Median      3Q     Max
-3.941 -1.600 -0.182  1.050  5.854

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.22727    1.59879  23.285  < 2e-16 ***
hp          -0.03177    0.00903  -3.519  0.00145 **
wt          -3.87783    0.63273  -6.129 1.12e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.593 on 29 degrees of freedom
Multiple R-squared:  0.8268,Adjusted R-squared:  0.8148
F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

Model Comparison

Consider two **nested** models:

- The **full model**: (all subscript i are removed for simplicity)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q + \beta_{q+1} x_{q+1} + \cdots + \beta_k x_k + \epsilon$$

- The **reduced model**: (all subscript i are removed for simplicity)

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q + \epsilon$$

Model Comparison

Consider two **nested** models:

- The **full model**: (all subscript i are removed for simplicity)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q + \beta_{q+1} x_{q+1} + \cdots + \beta_k x_k + \epsilon$$

- The **reduced model**: (all subscript i are removed for simplicity)

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q + \epsilon$$

- The reduced model is a special case of the full model with $\beta_{q+1} = \cdots = \beta_k = 0$.
- Comparing the two models is equivalent to testing the null hypothesis:

$$H_0 : \beta_{q+1} = \cdots = \beta_k = 0$$

Model Comparison

$$H_0 : \beta_{q+1} = \cdots = \beta_k = 0$$

In order to compare the nested models, we can use the **F-test**:

$$F = \frac{(SSE_{reduced} - SSE_{full})/(k-q)}{SSE_{full}/(n-k-1)}$$

Model Comparison

$$H_0 : \beta_{q+1} = \cdots = \beta_k = 0$$

In order to compare the nested models, we can use the **F-test**:

$$F = \frac{(SSE_{reduced} - SSE_{full})/(k-q)}{SSE_{full}/(n-k-1)}$$

reject null if

- $F > F_{\alpha,k-q,n-k-1}$
- The p-value:

$$1 - F_{F,k-q,n-k-1}(F)$$

  is less than α.

Example

Recall the **mtcars** dataset, we compare the following two models:

```
> model1 = lm(mpg~disp+hp+wt, mtcars)
> model2 = lm(mpg~disp, mtcars)
```

Example

Recall the **mtcars** dataset, we compare the following two models:

```
> model1 = lm(mpg~disp+hp+wt, mtcars)
> model2 = lm(mpg~disp, mtcars)
```

The F-test result can be read from anova function:

```
> anova(model2, model1)
Analysis of Variance Table

Model 1: mpg ~ disp
Model 2: mpg ~ disp + hp + wt
Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1      30 317.16
2      28 194.99  2    122.17 8.7715 0.001102 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Comparison

- R^2 is a metric for the goodness of fit of the model.
- But we **cannot** use R^2 to compare two models with different number of predictors, because **adding more predictors will always increase** R^2.

Model Comparison

- R^2 is a metric for the goodness of fit of the model.
- But we **cannot** use R^2 to compare two models with different number of predictors, because **adding more predictors will always increase** R^2.
- We can use the **adjusted** R^2:

$$R^2_{adj} = 1 - \frac{n-1}{n-k-1}\frac{\text{SSE}}{\text{SST}}$$

Model Comparison

- R^2 is a metric for the goodness of fit of the model.
- But we **cannot** use R^2 to compare two models with different number of predictors, because **adding more predictors will always increase** R^2.
- We can use the **adjusted** R^2:

$$R^2_{adj} = 1 - \frac{n-1}{n-k-1}\frac{\text{SSE}}{\text{SST}}$$

- The adjusted R^2 adds a penalty for the number of predictors in the model.
- The adjusted R^2 is always less than or equal to R^2.

Example

Recall part of the output of the `mtcars` example:

```
Residual standard error: 2.639 on 28 degrees of freedom
Multiple R-squared:  0.8268,Adjusted R-squared:  0.8083
F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

Example

Recall part of the output of the mtcars example:

```
Residual standard error: 2.639 on 28 degrees of freedom
Multiple R-squared:  0.8268,Adjusted R-squared:  0.8083
F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

- The R^2 is 0.8268, which means 82.68% of the variability in mpg can be explained by the model.
- The adjusted R^2 is 0.8083.

Example

Recall part of the output of the `mtcars` example:

```
Residual standard error: 2.639 on 28 degrees of freedom
Multiple R-squared:  0.8268,Adjusted R-squared:  0.8083
F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

- The R^2 is 0.8268, which means 82.68% of the variability in `mpg` can be explained by the model.
- The adjusted R^2 is 0.8083.
- The F-statistic and the p-value are for the following hypothesis test:

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

- The p-value is very small, which means at least one of the predictors is significant in the model or the model is significant.

Example

However, if we consider a linear regression model of `mpg` on `disp`, `hp`, and `cyl`.

```
Call:
lm(formula = mpg ~ disp + hp + cyl, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0889 -2.0845 -0.7745  1.3972  6.9183 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 34.18492    2.59078  13.195 1.54e-13 ***
disp        -0.01884    0.01040  -1.811   0.0809 .  
hp          -0.01468    0.01465  -1.002   0.3250    
cyl         -1.22742    0.79728  -1.540   0.1349    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.055 on 28 degrees of freedom
Multiple R-squared:  0.7679,Adjusted R-squared:  0.743 
F-statistic: 30.88 on 3 and 28 DF,  p-value: 5.054e-09
```

Example

However, if we consider a linear regression model of `mpg` on `disp`, `hp`, and `cyl`.

```
Call:
lm(formula = mpg ~ disp + hp + cyl, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0889 -2.0845 -0.7745  1.3972  6.9183

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 34.18492    2.59078  13.195 1.54e-13 ***
disp        -0.01884    0.01040  -1.811   0.0809 .
hp          -0.01468    0.01465  -1.002   0.3250
cyl         -1.22742    0.79728  -1.540   0.1349
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.055 on 28 degrees of freedom
Multiple R-squared:  0.7679,Adjusted R-squared:  0.743
F-statistic: 30.88 on 3 and 28 DF,  p-value: 5.054e-09
```

None of the covariates are significant at $\alpha = 0.05$ level. But they are jointly significant.

Multicollinearity

The **multicollinearity** is a problem when two or more predictors are highly correlated with each other.

Multicollinearity

The **multicollinearity** is a problem when two or more predictors are highly correlated with each other.

- It can cause the estimated coefficients to be unstable and have large standard errors.
- Individual covariates may not be significant, but the model is significant.

Multicollinearity

The **multicollinearity** is a problem when two or more predictors are highly correlated with each other.

- It can cause the estimated coefficients to be unstable and have large standard errors.
- Individual covariates may not be significant, but the model is significant.

To verify it, we can check the correlation matrix of the predictors in prevous example:

```
> cor(mtcars[,c("disp", "hp", 'cyl')])
          disp        hp       cyl
disp 1.0000000 0.7909486 0.9020329
hp   0.7909486 1.0000000 0.8324475
cyl  0.9020329 0.8324475 1.0000000
```

Multicollinearity

To measure the multicollinearity, we can use the **variance inflation factor** (VIF):

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R^2 of the regression of x_j on all other predictors.

Multicollinearity

To measure the multicollinearity, we can use the **variance inflation factor** (VIF):

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R^2 of the regression of x_j on all other predictors.

- If $VIF_j > 10$, we consider x_j is highly correlated with other predictors.
- If $5 < VIF_j < 10$, we consider x_j is correlated with other predictors.
- If $1 < VIF_j < 5$, we consider x_j is lightly correlated with other predictors.
- If $VIF_j = 1$, we consider x_j is not correlated with other predictors.

Example

We can use the `vif` function in R to compute the VIF for each predictor:

```
> library(car)
> model = lm(mpg~disp+hp+cyl, mtcars)
> vif(model)
disp       hp       cyl
5.521460 3.350964 6.732984
```

Example

We can use the `vif` function in R to compute the VIF for each predictor:

```
> library(car)
> model = lm(mpg~disp+hp+cyl, mtcars)
> vif(model)
disp       hp       cyl
5.521460 3.350964 6.732984
```

We should consider removing `cyl` from the model.