Chapter 2 Multiple Regression
(Part 2)

1 Analysis of Variance in multiple linear regression

Recall the model again

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, n \]

For the fitted model \( \hat{Y}_i = b_0 + b_1 X_{i1} + \ldots + b_p X_{ip} \),

\[ Y_i = \hat{Y}_i + e_i \quad i = 1, \ldots, n \]

\[ \frac{Y_i - \bar{Y}}{\text{Total deviation}} = \frac{\hat{Y}_i - \bar{Y}}{\text{Deviation due to regression}} + \frac{e_i}{\text{Deviation due to the error}} \]

<table>
<thead>
<tr>
<th>obs</th>
<th>( Y_i )</th>
<th>( \hat{Y}<em>i = b_0 + b_1 X</em>{i1} + \ldots + b_p X_{ip} )</th>
<th>( e_i = Y_i - \hat{Y}_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( Y_1 - \bar{Y} )</td>
<td>( \hat{Y}_1 - \bar{Y} )</td>
<td>( e_1 - \bar{e} = e_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( Y_2 - \bar{Y} )</td>
<td>( \hat{Y}_2 - \bar{Y} )</td>
<td>( e_2 - \bar{e} = e_2 )</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>n</td>
<td>( Y_n - \bar{Y} )</td>
<td>( \hat{Y}_n - \bar{Y} )</td>
<td>( e_n - \bar{e} = e_n )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum of squares</th>
<th>( \sum_{i=1}^{n} (Y_i - \bar{Y})^2 )</th>
<th>( \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 )</th>
<th>( \sum_{i=1}^{n} e_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sum of squares (SST)</td>
<td>Sum of squares due to regression (SSR)</td>
<td>Sum of squares of error/residuals (SSE)</td>
<td></td>
</tr>
</tbody>
</table>

We have

\[ \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} e_i^2 \]
\[\begin{align*}
\sum_{i=1}^{n} (Y_i - \bar{Y})^2 &= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y} + Y_i - \hat{Y}_i)^2 \\
&= \sum_{i=1}^{n} [(\hat{Y}_i - \bar{Y})^2 + (Y_i - \hat{Y}_i)^2 + 2(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)] \\
&= \text{SSR} + \text{SSE} + 2\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})e_i \\
&= \text{SSR} + \text{SSE} + 2\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})e_i
\end{align*}\]

where \(\sum_{i=1}^{n} \hat{Y}_i e_i = 0\) and \(\sum_{i=1}^{n} e_i = 0\) are used, which follow from the Normal equations.

- \(\text{SST} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = Y'Y - \frac{1}{n} Y'JY = Y'(I - \frac{1}{n} J)Y\)
  Degree of freedom? \(n-1\) (with \(n\) being the number of observations)

- \(\text{SSE} = \sum_{i=1}^{n} e_i^2 = e'e = (Y - Xb)'(Y - Xb) = Y'(I - H)Y\)
  Degree of freedom? \(n-p-1\) (with \(p+1\) being the number of coefficients)

- Let \(H = X(X'X)^{-1}X\) and and \(J = 11'/n\). Note that
  \[\hat{Y} = HY\]
  and by the fact \(\sum_{i=1}^{n} e_i = 0\) (see the normal equations),
  \[\bar{Y} = \bar{Y} = 1'Y/n.\]
  Thus
  \[\text{SSR} = (\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y}) = Y'(H - J/n)'(H - J/n)'Y\]
  \[= Y'(H - J/n)Y.\]
  Degree of freedom? \(p\) (the number of variables).

[Another Proof\footnote{Please ignore this proof}]
\[\hat{Y} - \bar{Y} = HY - 1'/nY = (H - J/n)Y.\]
Write $X = (1 : X_1)$. Then

$$H(1 : X_1) = X(X^TX)^{-1}X^TX = X = (1 : X_1)\top$$

Thus

$$H(1 : X_1) = 1$$

Similarly, $1'H = 1'. Thus 

$$(H - J/n)'(H - J/n)' = H - J/nH - HJ/n + J/n = H - J/n$$

It follows that

$$SST = SSR + SSE$$

We further define

$$MSR = \frac{SSR}{p} \text{ called regression mean square}$$

$$MSE = \frac{SSE}{n - p - 1} \text{ called error mean square (or mean squared error)}$$

2 ANOVA table

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>$SSR = Y'(H - J/n)Y$</td>
<td>$p$</td>
<td>$MSR = \frac{SSR}{p}$</td>
<td>MSR/MSE</td>
</tr>
<tr>
<td>Error</td>
<td>$SSE = Y'(I - H)Y$</td>
<td>$n - p - 1$</td>
<td>$MSE = \frac{SSE}{n - p - 1}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SST = Y'(I - J/n)Y$</td>
<td>$n - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 $F$ test for regression relation

- $H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$ versus $H_a :$ not all $\beta_k (k = 1, ..., p)$ equal zero
- Under $H_0$, the reduced model: $Y_i = \beta_0 + \varepsilon_i$

$$SSE(R) = SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

degrees of freedom $n - 1$
• Full model: $Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \varepsilon_i$

$$SSE(F) = SSE = e'e = (Y - Xb)'(Y - Xb)$$

degrees of freedom $n - p - 1$

• $F$ test statistic (also called F-test for the model)

$$F^* = \frac{(SSE(R) - SSE(F))/(df(R) - df(F))}{SSE(F)/df(F)} = \frac{SSR/p}{SSE/(n - p - 1)}$$

• If $F^* \leq F(1 - \alpha; p, n - p - 1)$, conclude(accept) $H_0$

IF $F^* > F(1 - \alpha; p, n - p - 1)$, conclude $H_a$ (reject $H_0$)

4 $R^2$ and the adjusted $R^2$

• $SSR = SST - SSE$ is the part of variation explained by regression model

• Thus, define coefficient of multiple determination

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

which is the proportion of variation in the response that can be explained by the regression model (or that can be explained by the predictors $X_1, ..., X_p$ linearly)

• $0 \leq R^2 \leq 1$

• with more predictor variables, SSE is smaller and $R^2$ is larger. To evaluate the contribution of the predictors fair, we define the adjusted $R^2$:

$$R^2_a = 1 - \frac{SSE}{\frac{n-p-1}{n-1} SST} = 1 - \left( \frac{n-1}{n-p-1} \right) \frac{SSE}{SST}$$

More discussion will be given later about $R^2_a$.

• For two models with the same number of predictor variables, $R^2$ can be used to indicate which model is better.

• If model A include more predictor variables than model B, then the $R^2$ of A must be equal or greater than that of model B. In that case, it is better to use the adjusted $R^2$.
5 Dwaine studios example

- $Y$-sales, $X_1$- number of persons aged 16 or less, $X_2$- income
- $n = 21, p = 3$
- $\text{SST} = 26,196.21$, $\text{SSE} = 2,180.93$, $\text{SSR} = 26,196.21 - 2,180.93 = 24,015.28$
- $F^* = \frac{24,015.28/2}{2,180.93/18} = 99.1$
  
  For $H_0: \beta_1 = \beta_2 = 0$ with $\alpha = 0.05$, $F(0.95; 2, 18) = 3.55$. because

  $$F^* > F(0.95; 2, 18)$$

  we reject $H_0$

- $R^2 = \frac{24,015.28}{26,196.21} = 0.917, \quad R^2_a = 0.907$

Writing a fitted regression model

| Coefficients: Estimate Std. Error t value Pr(>|t|) |
|-------------|-----------|----------|-------|-----------|
| (Intercept)  | -68.8571  | 60.0170  | -1.147 | 0.2663    |
| x1           | 1.4546    | 0.2118   | 6.868  | 2e-06 *** |
| x2           | 9.3655    | 4.0640   | 2.305  | 0.0333 *  |

Residual standard error: 11.01 on 18 degrees of freedom
Multiple R-squared: 0.9167, Adjusted R-squared: 0.9075
F-statistic: 99.1 on 2 and 18 DF, p-value: 1.92e-10

The fitted model is

$$\hat{Y} = -68.86 + 1.45X_1 + 9.937X_2$$

$R^2 = 0.9167, \quad R^2_a = 0.9075, \quad F\text{-statistic: 99.1 on 2 and 18 DF,}$