Multiple Regression II

Paweł Polak

October 23, 2017

Linear Regression Models - Lecture 8

Content: ALRM Book Chapter 7 (Sec. 7.1-7.6)

- Extra Sums of Squares.
- Tests for regression coefficients using extra sums of squares.
- Coefficients of partial determination.
- Standardized Multiple Regression Model.
- Multicollinearity and its effects.

General Linear Model

- *Independent responses* of the form $Y_i \sim N(\mu_i, \sigma^2)$, where

$$\mu_i = \mathbf{X}_i^\top \boldsymbol{\beta}$$

  for some known vector of *explanatory* variables $\mathbf{X}_i^\top = (X_{i1}, \ldots, X_{ip})$.

- Unknown *parameter* vector $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{P-1})^\top$, where $P < N$.

- This is the *linear model* and is usually written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  (in vector notation) where

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1^\top \\ \vdots \\ x_N^\top \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{P-1} \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix},$$

  where $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$, for $i = 1, 2, \ldots, N$.

Sum of Squares Decomposition

- Recall the sums of square decomposition:

$$SST = SSR + SSE,$$
$$\sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2.$$

- Using matrix notation

$$\mathbf{Y}^T(\mathbf{I} - \mathbf{J}/N)\mathbf{Y} = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y} + \mathbf{Y}^T(\mathbf{H} - \mathbf{J}/N)\mathbf{Y}.$$

- As usual SST has $(N-1)$ degrees of freedom associated with it.
- The term SSE has $(N-P)$ degrees of freedom.
- The term SSR has $(P-1)$ degrees of freedom.

Extra sums of squares

- An *extra sums of squares* measures the *marginal reduction* in the *SSE* when one or more explanatory variables are *added* to the regression model.

- They are useful in a variety of tests where the question of interest is whether certain explanatory variables can be dropped from the regression model.

Illustration

- Consider the following two regression models:
  I. $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$
  II. $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

- Model II contains one additional variable compared to Model I.

Sums of squares

- Since $SSE = \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$ and $SSR = \sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2$ are different depending on which explanatory variables are include in the model, we use the following notation:

$SSR(X_1)$ – the SSR for a model with only X_1.
$SSE(X_1)$ – the SSE for a model with only X_1.
$SSR(X_1, X_2)$ – the SSR for a model with X_1 and X_2.
$SSE(X_1, X_2)$ – the SSE for a model with X_1 and X_2.

Sums of squares

- $SST = SSR(X_1) + SSE(X_1)$
- $SST = SSR(X_1, X_2) + SSE(X_1, X_2)$
- We also know that

$$SSR(X_1, X_2) \geq SSR(X_1)$$

and

$$SSE(X_1, X_2) \leq SSE(X_1).$$

Extra sums of squares

- $SSR(X_1)$ measures the contribution by including X_1 alone in the model.

- $SSR(X_2|X_1)$ measures the marginal effect of adding X_2 to a model that already includes X_1.

- The *difference* between the *SSE*'s is called an *extra sum of squares* and is denoted by

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2).$$

- The extra sum of squares $SSR(X_2|X_1)$ can equivalently be viewed as the *marginal increase* in the regression sum of squares:

$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1).$$

Verify at home!

Generalization

- The notation can easily be generalized to include more than two variables.

- For example we can write:

$$
\begin{aligned}
SSR(X_3|X_1, X_2) &= SSR(X_1, X_2, X_3) - SSR(X_1, X_2) \\
&= SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \\
SSR(X_2, X_3|X_1) &= SSR(X_1, X_2, X_3) - SSR(X_1) \\
&= SSE(X_1) - SSE(X_1, X_2, X_3)
\end{aligned}
$$

- When dealing with 3 or more variables *SSR* can be decomposed in a variety of ways. For example:

$$
\begin{aligned}
SSR(X_1, X_2, X_3) &= SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \\
&= SSR(X_2) + SSR(X_1|X_2) + SSR(X_3|X_1, X_2) \\
&= SSR(X_1) + SSR(X_2, X_3|X_1)
\end{aligned}
$$

Decomposition

- Consider the following regression models:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \varepsilon_i \\
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i
\end{aligned}
$$

- Y_i's are the same in both models, hence the two SST's are of course equal but

$$
\begin{aligned}
SST &= SSR(X_1) + SSE(X_1) \\
SST &= SSR(X_1, X_2) + SSE(X_1, X_2)
\end{aligned}
$$

- We can re-express the latter term:

$$
\begin{aligned}
SST &= SSR(X_1) + SSE(X_1) \\
&= SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2)
\end{aligned}
$$

ANOVA Table

- The *ANOVA* table corresponding to the model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

Table: ANOVA Table

Source of Variation	SS	df
Regression	$SSR(X_1, X_2, X_3)$	$P - 1 = 3$
Error	$SSE(X_1, X_2, X_3)$	$N - P = N - 4$
Total	SST	$N - 1$

Table: Modified ANOVA Table

Source of Variation	SS	df
Regression	$SSR(X_1, X_2, X_3)$	3
X_1	$SSR(X_1)$	1
$X_2 \| X_1$	$SSR(X_2 \| X_1)$	1
$X_3 \| X_1, X_2$	$SSR(X_3 \| X_1, X_2)$	1
Error	$SSE(X_1, X_2, X_3)$	$N - 4$
Total	SST	$N - 1$

Mean Squares

- *Mean squares* are constructed as usual. For example:

$$
\begin{aligned}
MSR(X_2|X_1) &= \frac{SSR(X_2|X_1)}{1} \\
MSR(X_2, X_3|X_1) &= \frac{SSR(X_2, X_3|X_1)}{2}
\end{aligned}
$$

Hypothesis Testing

- Extra sums of squares occur in a variety of tests where the question of interest is whether certain explanatory variables can be *dropped* from the regression model.

General Linear Test

The general linear test approach has three parts

- Compute the full model.
- Compute a *reduced model* corresponding to the null hypothesis.
- Compute a test statistic that compares the two models.

Test Statistic

- In the general linear test, the test statistic is

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}}$$

- Here df_R and df_F are the degrees of freedom associated with the reduced and full model error sums of square respectively.
- When H_0 holds $F^* \sim F_{df_R - df_F, df_F}$.

Testing individual coefficients

- Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

- Suppose we want to test: $H_0 : \beta_3 = 0$ versus $H_a : \beta_3 \neq 0$.

- In the full model, $SSE(F) = SSE(X_1, X_2, X_3)$; $df_F = N - 4$.

- In the reduced model

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

$SSE(R) = SSE(X_1, X_2)$; $df_R = N - 3$.

- Thus,

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSR(X_3|X_1,X_2)}{1}}{\frac{SSE(X_1,X_2,X_3)}{N-4}} = \frac{MSR(X_3|X_1,X_2)}{MSE(X_1,X_2,X_3)}$$

- Compare F^* with the $F_{1,N-4}$ distribution.

Testing subsets of coefficients

- Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

- Test $H_0 : \beta_2 = \beta_3 = 0$ versus H_a: *Not Both equal to 0.*

- In the full model $SSE(F) = SSE(X_1, X_2, X_3)$; $df_F = N - 4$.

- In the reduced model $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$, $SSE(R) = SSE(X_1)$; $df_R = N - 2$

- Thus,

$$F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{\frac{SSR(X_3,X_2|X_1)}{2}}{\frac{SSE(X_1,X_2,X_3)}{N-4}} = \frac{MSR(X_2,X_3|X_1)}{MSE(X_1,X_2,X_3)}.$$

- Compare F^* with the $F_{2,N-4}$ distribution.

Comments

- When testing whether a single β_k equals 0, we can use *either* a t-test or a general linear F-test.
- Both tests will give equivalent results.
- When testing whether several β_k equal 0, only the general linear F-test is available.
- $H_0 : \beta_1 = \beta_2 = \ldots = \beta_{P-1} = 0$ – use the F-test.
- $H_0 : \beta_k = 0$ – use either the t-test or the F-test.
- $H_0 : \beta_1 = \beta_2 = 0$ – use the F-test.
- In the special case of simple linear regression the t-test allows for a *one-sided* alternative hypothesis, while the F-test only allows for *two-sided* alternative hypothesis.

Other Tests

- Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

- Suppose we want to test the hypothesis:

$$H_0 : \beta_1 = \beta_2 \quad \text{versus} \quad H_a : \beta_1 \neq \beta_2 .$$

- Compare the full model with a reduced one.
- Full model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$
- Reduced model: $Y_i = \beta_0 + \beta_1 (X_{i1} + X_{i2}) + \beta_3 X_{i3} + \varepsilon_i$
- Use the general linear test statistic with 1 and $N - 4$ degrees of freedom.

Other Tests

- Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

- Suppose we want to test the hypothesis:

$$H_0 : \beta_1 = 2, \beta_2 = 4 \quad \text{versus} \quad H_a : \text{ not both equalities hold.}$$

- Compare the full model with a reduced one.
- Full model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$
- Reduced model: $Y_i - 2X_{i1} - 4X_{i2} = \beta_0 + \beta_3 X_{i3} + \varepsilon_i$
- Use the general linear test statistic with 2 and $N - 4$ degrees of freedom.

Coefficient of Partial Determination

- The *coefficient of multiple determination*, R^2, measures the proportionate reduction in the variation of Y achieved by introducing all of the X variables into the model.

- A *coefficient of partial determination* measures the marginal contribution of a single X variable when all the others are already included in the model.

Illustration

- Consider the following regression model:

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

- The relative marginal reduction in the variation of Y associated with X_1 when X_2 is already in the model is given by:

$$\frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1|X_2)}{SSE(X_2)}.$$

Coefficient of Partial Determination

- Coefficient of partial determination:

$$R^2_{Y1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)}$$

- Entries to the left of $|$ show the response variable and the X variable being added.

- Entries to the right show the X variables already in the model.

- Other examples:

$$\begin{aligned} R^2_{Y1|23} &= \frac{SSR(X_1|X_2, X_3)}{SSE(X_2, X_3)} \\ R^2_{Y4|123} &= \frac{SSR(X_4|X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)} \end{aligned}$$

Coefficient of Partial Correlation

- The *square root* of a coefficient of partial determination is called a coefficient of partial correlation.
- It is given the same sign as the corresponding regression coefficient in the fitted regression function.
- It is used in variable selection algorithms (discussed later).

Numerical Precision Errors

- Numerical precision errors can occur when
  1. the predictor variables have substantially different magnitudes
  2. $(\mathbf{X}^T\mathbf{X})^{-1}$ is poorly conditioned near singular : multicolinearity
- Solutions
  1. Standardized multiple regression (Correlation Transformation).
  2. Regularization (Ridge regression).

Standardized Multiple Regression

- Lack of comparability of regression coefficients

$$\widehat{Y} = 200 + 20000X_1 + 0.2X_2$$

  Y in dollars, X_1 in thousand dollars, X_2 in cents.
  - Which is most important predictor?
  - If X_1 increases $1,000$ dollars, then Y increases $20,000$ dollars
  - If X_2 increases $1,000$ dollars, then Y also increases $20,000$ dollars
- Solution:
  - centering and scaling

$$Y_i^* = \frac{1}{\sqrt{N-1}} \frac{Y_i - \bar{Y}}{s_Y}, \quad \text{where} \quad s_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}},$$

  and, for $k = 1, \ldots, P-1$,

$$X_{i,k}^* = \frac{1}{\sqrt{N-1}} \frac{X_{i,k} - \bar{X}_k}{s_{X_k}}, \text{ where } s_{X_k} = \sqrt{\frac{\sum_{i=1}^{N}(X_{i,k} - \bar{X}_k)^2}{N-1}}.$$

  - it makes all entries in $\mathbf{X}^T\mathbf{X}$ matrix for the transformed variables fall between -1 and 1 inclusive. Hence, it is also called Correlation Transformation.

Standardized Regression Model

- The regression model using the transformed variables:

$$Y_i^* = \beta_1 X_{i,1}^* + \beta_2 X_{i,2}^* + \ldots + \beta_{P-1} X_{i,P-1}^* + \varepsilon_i^*$$

- Notice that there is no need for intercept.
- We can set up a standard linear regression problem ($\mathbf{X}^*$ is without the column of ones)

$$\mathbf{X}^{*T}\mathbf{X}^*\mathbf{b}^* = \mathbf{X}^{*T}\mathbf{Y}^*$$

and solve it to get

$$\mathbf{b}^* = \left(b_1^*, \ldots, b_{P-1}^*\right)^T$$

- $\mathbf{b}^*$ can be related to the solution to the untransformed regression problem through the relationship

$$b_k = (\frac{s_Y}{s_{X_k}})b_k^*, \text{ for } k = 1, \ldots, P-1$$

$$b_0 = \bar{Y} - b_1\bar{X}_1 - \ldots - b_{P-1}\bar{X}_{P-1}.$$

Proof: ($P = 2$ case), take the original regression function: $Y_i = b_{\mathbf{0}} + b_{\mathbf{1}}X_{i,\mathbf{1}}$, then

$$\frac{1}{\sqrt{N-1}}\frac{Y_i - \bar{Y}}{s_Y} = \frac{1}{\sqrt{N-1}s_Y}\underbrace{(b_{\mathbf{0}} - \bar{Y} + b_{\mathbf{1}}\bar{X}_{\mathbf{1}})}_{=\mathbf{0}} + \frac{1}{\sqrt{N-1}}\frac{s_{X_{\mathbf{1}}}}{s_Y}b_{\mathbf{1}}\frac{(X_{i,\mathbf{1}} - \bar{X}_{\mathbf{1}})}{s_{X_{\mathbf{1}}}}.$$

Multicollinearity

- In multiple regression, the hope is that the explanatory variables $\mathbf{X}_k$, for $k = 1, \dots, P-1$, are highly correlated with the response variable $\mathbf{Y}$.
- However, it is not desirable for the explanatory variables $\mathbf{X}_k$, for $k = 1, \dots, P-1$, to be *correlated* with one another.
- *Multicollinearity* exists when two or more of the explanatory variables $\mathbf{X}_k$, for $k = 1, \dots, P-1$, used in the regression model are highly correlated and provide redundant information about the response $\mathbf{Y}$.

Uncorrelated Variables

- Sometimes a set of explanatory variables are *uncorrelated*
- In this case the regression coefficient for X_1 is the *same* whether only X_1 is included in the model or both X_1 and X_2 are included.
- Also, the relationships $SSR(X_1|X_2) = SSR(X_1)$ and $SSR(X_1|X_2) = SSR(X_1)$ hold.
- When the explanatory variables are uncorrelated, the effects ascribed to each one of them is the same no matter which of the other variables are included in the model.

Problems When Variables Are Correlated

Multicollinearity among the explanatory variables leads to the following problems:

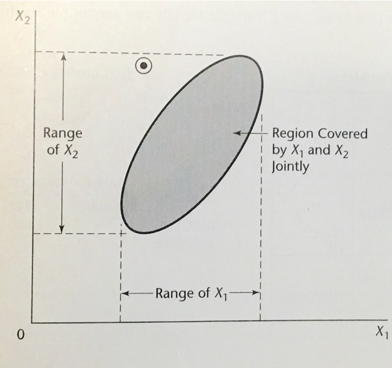
- The parameter estimates are *unstable*.
- The regression coefficients are not *interpretable*.
- The standard deviation of the regression coefficients $s\{\mathbf{b}\} = \sqrt{MSE * \text{diag}((\mathbf{X}^T\mathbf{X})^{-1})}$ is disproportionately *large*.
  - multicollinearity implies that some of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ are close to 0,
  - $\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \left(\mathbf{C}^T\mathbf{D}\mathbf{C}\right)^{-1} = \left(\mathbf{C}^T\mathbf{D}^{-1}\mathbf{C}\right)$ (recall that $\mathbf{C}^{-1} = \mathbf{C}^T$)
  - hence, the same eigenvalues of the inverse matrix are very large.
- The estimated regression coefficient *may not be individually significant* even though they are statistically related to the response.

Signs of Multicollinearity

Some typical signs of multicollinearity:

- There are *large changes* in the estimated regression coefficients when an explanatory variable is added or deleted.
- We obtain *non-significant* results in individual tests on the regression coefficients for important explanatory variables.
- We obtain estimated regression coefficients with an *opposite sign* from what was expected from theoretical considerations or prior experience.

Caution About Hidden Extrapolations



- When one estimates a mean response or predicts a new observation in multiple regression, one needs to be particularly careful that the estimate or prediction does not fall outside the scope of the model.
- It is easy to spot this extrapolation when there are only two predictor variables, but it becomes much more difficult when the number of predictor variables is large and they are correlated.
- We will revisit this topic for more than two predictors variables.