Linear Statistical Modeling Methods with SAS

Coefficients of Partial Determination and Multicollinearity

Xuemao Zhang
East Stroudsburg University

March 11, 2024

Outline

- Extra Sums of Squares
- Tests Using Extra Sums of Squares
- Coefficients of Partial Determination
- Multicollinearity

Extra Sums of Squares

An extra sum of squares measures the marginal reduction in the error sum of squares when one or several predictor variables are added to the regression model, given that other predictor variables are already in the model. Equivalently, one can view an extra sum of squares as measuring the marginal increase in the regression sum of squares when one or several predictor variables are added to the regression model.

Let

$$A = \{\text{set of predictor variables}\};$$
$$B = \{\text{set of predictors that include all those in A plus more}\};$$

If we fit the two MLR models, then

$$\begin{aligned} SST =& SSR(A) + SSE(A) \\ =& SSR(B) + SSE(B) \end{aligned}$$

$SST = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$ does not depend on the predictors and therefore

$$SSR(B) \geq SSR(A).$$

So whenever we add predictors to the model, SSR goes up (or stays the same) and SSE goes downs (or stays the same).

Extra Sums of Squares

Recall The **Adjusted Coefficient of Multiple Determination**, R_a^2, is a measure that adjusts R^2 for the number of regressors in the model. It is defined as

$$R_a^2 = 1 - \frac{\mathrm{SSE}/(n-1-k)}{SS_{total}/(n-1)} = 1 - \left(\frac{n-1}{n-1-k}\right)\frac{SSE}{SST}.$$

Notice that $R_a^2 \leq R^2$ since $k \geq 1$.

Now let

$$A = \{\text{a set of predictor variables}\};$$

$$C = \{\text{another set of predictors }\};$$

with $A \cap C = \emptyset$ (no predictors in common). Then

$$SSR(A \cup C) \geq SSR(A).$$

That is, more variability among the Y_i's is explained by the larger model which includes both set.

Partitioning of Variablity

SSR can be partitioned into meaningful portions that help assess the worth of a single predictor variable in the traditional sense. $SSR = SSR(X_1, \dots, X_k)$ can be partitioned into **sequential** (SAS produces Type I SS) regression sums of squares as follows:

$$
\begin{aligned}
SSR(X_1, \dots, X_k) =& SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \\
& + \cdots + SSR(X_k|X_1, X_2, \dots, X_{k-1}).
\end{aligned}
$$

The notation $SSR(\cdot|\cdot)$ implies "regression explained by ..." withe the vertical line denoting "in the presence of". For example, $SSR(X_2|X_1)$ is the increase in the SSR when the predictor variable X_2 is added to a model involving only X_1 and the constant term.

The decomposition can be summarized in the following ANOVA table

Extra Sums of Squares

Table 1: ANOVA Table with Decomposition of SSR

Source	df	SS	MS
Regression	k	SSR	SSR/k
X_1	1	$SSR(X_1)$	$SSR(X_1)$
$X_2 \mid X_1$	1	$SSR(X_2 \mid X_1)$	$SSR(X_2 \mid X_1)$
$X_3 \mid X_1, X_2$	1	$SSR(X_3 \mid X_1, X_2)$	$SSR(X_3 \mid X_1, X_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_k \mid X_1, X_2, \ldots, X_{k-1}$	1	$SSR(X_k \mid X_1, X_2, \ldots, X_{k-1})$	$SSR(X_k \mid X_1, X_2, \ldots, X_{k-1})$
Error	n-1-k	SSE	$SSE/(n-1-k)$
Total	n-1	SST	

where

$$\begin{aligned} SSR = & SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \\ & + \cdots + SSR(X_k|X_1, X_2, \ldots, X_{k-1}). \end{aligned}$$

Tests Using Extra Sums of Squares: Test if a Single $\beta_i = 0$

When We wish to test whether the term $\beta_i x_i$ can be dropped from a multiple regression model, we are interested in testing

$$H_0 : \beta_i = 0$$
$$H_a : \beta_i \neq 0$$

Recall that the test statistic appropriate for this test is

$$t = \frac{\widehat{\beta}_i - 0}{s.e.(\widehat{\beta}_i)},$$

with $s.e.(\widehat{\beta}_i) = \sqrt{MSE}\sqrt{C_{ii}}$, where C_{ii} is the $(i+1)$th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$.

Equivalently, We can use the general linear test approach. For example, to test

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0,$$

Tests Using Extra Sums of Squares: Test if a Single $\beta_i = 0$

We consider the full model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \qquad \text{full model}$$

and the reduced model

$$E(Y) = \beta_0 + 0 + \beta_2 x_2 + \cdots + \beta_k x_k, \qquad \text{reduced model}$$

We fit the full model and obtain the error sum of squares $SSE(F)$,

$$SSE(F) = SSE(X_1, X_2, \ldots, X_k).$$

And the degrees of freedom associated with $SSE(F)$ are $n - 1 - k$.

We next fit the reduced model (R) and obtain $SSE(R)$,

$$SSE(R) = SSE(X_2, \ldots, X_k).$$

And the degrees of freedom associated with $SSE(R)$ are $n - 1 - (k - 1) = n - k$.

Tests Using Extra Sums of Squares: Test if a Single $\beta_i = 0$

Apply the general linear test statistic

$$F = \frac{SSE(R) - SSE(F)}{df_R - df_F} \Big/ \frac{SSE(F)}{df_F}$$

and we obtain

$$\begin{aligned} F =& \frac{SSE(X_2, \dots, X_k) - SSE(X_1, X_2, \dots, X_k)}{n - k - (n - 1 - k)} \Big/ \frac{SSE(X_1, X_2, \dots, X_k)}{n - 1 - k} \\ =& \frac{SSR(X_1 | X_2, \dots, X_k)}{1} \Big/ \frac{SSE(X_1, X_2, \dots, X_k)}{n - 1 - k} \\ =& \frac{MSR(X_1 | X_2, \dots, X_k)}{MSE(X_1, X_2, \dots, X_k)} \end{aligned}$$

since $SSE(X_2, \dots, X_k) - SSE(X_1, X_2, \dots, X_k) = SSR(X_1 | X_2, \dots, X_k)$.

Remark. Again, the t-test statistic is the square root of this F test statistic.

Tests Using Extra Sums of Squares: Testing Sets of Parameters

Consider the hypothesis test problem of testing $H_0 : \beta_{r+1} = \beta_{r+2} = \cdots = \beta_k = 0$ versus H_a : At least one of the $\beta_i, i = r+1, \ldots, k$ differs from 0.

We define two models:

- Model R (Reduced model):

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$$

- Model F (Full model):

$$\begin{aligned} E(Y) = \beta_0 &+ \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r X_r \\ &+ \beta_{r+1} X_{r+1} + \beta_{r+2} X_{r+2} + \cdots + \beta_k X_k \end{aligned}$$

If $x_{r+1}, x_{r+2}, \ldots, x_k$ contribute a substantial quantity of information for the prediction of Y that is not contained in the variables $x_1, x_2, \ldots, x_r$ (that is, H_0 is rejected and at least one of the parameters $\beta_{r+1}, \beta_{r+2}, \ldots, \beta_k$ differs from zero), what would be the relationship between SSE_R and SSE_F?

Tests Using Extra Sums of Squares: Testing Sets of Parameters

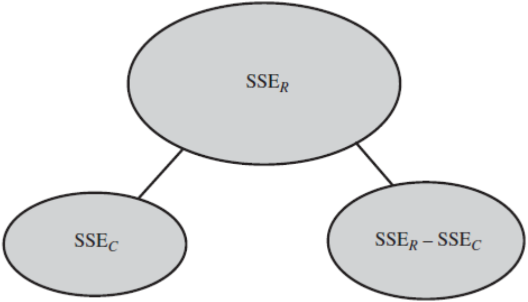
we use the test statistic

$$F^* = \frac{(SSE_R - SSE_F)/(k-r)}{MSE_F},$$

where F^* is based on F-distribution of $(df_1 = k - r, df_2 = n - 1 - k)$.

The rejection region for the test is identical to other analysis of variance F tests. Given significance level α, H_0 is rejected only if $F^* \geq F_{df_1, df_2, \alpha}$.

Partitioning SSE_R



Example

Example. Table 7.1 in contains data for a study of the relation of body fat (Y) to triceps skinfold thickness (X_1),thigh circumference (X_2), and midarm circumferences (X_3) based on a sample of 20 healthy females 25-34 years old. Note that the triceps skinfold thicknesses X_1 and thigh circumferences X_2 for these subjects are highly correlated. The following is the data set.

```
data ch7tab01;
input X1 X2 X3 Y;
label x1 = 'Triceps'
      x2 = 'Thigh cir.'
      x3 = 'Midarm cir.'
        y = 'body fat';
cards;
19.5  43.1  29.1  11.9  24.7  49.8  28.2  22.8
30.7  51.9  37.0  18.7  29.8  54.3  31.1  20.1
19.1  42.2  30.9  12.9  25.6  53.9  23.7  21.7
31.4  58.5  27.6  27.1  27.9  52.1  30.6  25.4
22.1  49.9  23.2  21.3  25.5  53.5  24.8  19.3
31.1  56.6  30.0  25.4  30.4  56.7  28.3  27.2
18.7  46.5  23.0  11.7  19.7  44.2  28.6  17.8
14.6  42.7  21.3  12.8  29.5  54.4  30.1  23.9
27.7  55.3  25.7  22.6  30.2  58.6  24.6  25.4
22.7  48.2  27.1  14.8  25.2  51.0  27.5  21.1
;
run;
```

Example

- Let's consider the following 4 models:

```
proc reg data = ch7tab01;
model y = x1;
model y = x2;
model y = x1 x2;
model y = x1 x2 x3;
run;
```

- X_1 and X_2 are highly correlated, so one of them should be removed from the model. Furthermore, when we fit the full model, it seems that X_3 does not contribute to the model significantly.

Example

- In the following, `Test1` is for the testing one variable, and `Test2` is for the testing two variables at once.

```
proc reg data = ch7tab01;
model y = x1 x2 x3/ss1;
  test1: test x3 = 0;
  test2: test x2=x3=0;
run;
```

Coefficients of Partial Determination

Recall that the **Coefficient of Multiple Determination** is defined as

$$R^2 = \frac{SSR}{SST}.$$

R^2 is the fraction of the variability among the Y_i's that is explained by the MLR model.

Consider a full model of

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

and a reduced model

$$E(Y) = \beta_0 + \beta_1 x_1.$$

A natural question is if the predictor variable X_2 can be eliminated from the full model.

$SSE(X_1)$ measures the variation in Y when X_1 is included in the model. $SSE(X_1, X_2)$ measures the variation in Y when both X_1 and X_2 are included in the model. Hence, the relative marginal reduction in the variation in Y associated with X_2 when X_1 is already in the model is:

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

Coefficients of Partial Determination

This measure is the **coefficient of partial determination** between Y and X_2, given that X_1 is in the model. We denote this measure by $R^2_{Y2|1}$ or $R^2_{Y2\cdot 1}$:

$$R^2_{Y2|1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)} = 1 - \frac{SSE(X_1, X_2)}{SSE(X_1)} = \frac{SSR(X_2|X_1)}{SSE(X_1)}.$$

$R^2_{Y2|1}$ thus measures the proportionate reduction in the variation of Y remaining gained by including X_2 in the model when X_1 is already in the model.

The generalization of coefficients of partial determination to three or more X variables in the model is immediate. For instance:

$$R^2_{Y1|23} = \frac{SSR(X_1|X_2, X_3)}{SSE(X_2, X_3)}$$

$$R^2_{Y2|13} = \frac{SSR(X_2|X_1, X_3)}{SSE(X_1, X_3)}$$

$$R^2_{Y3|12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)}$$

$$R^2_{Y4|123} = \frac{SSR(X_4|X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)}$$

Coefficients of Partial Determination

Coefficients 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 that of the corresponding **regression coefficient** in the fitted regression function. For example, $r_{Y3|12}$ has the same sign as $\widehat{\beta}_3$

Coefficients of partial correlation are frequently used in practice, although they do not have as clear a meaning as coefficients of partial determination. One use of partial correlation coefficients is in computer routines for finding the best predictor variable to be selected next for inclusion in the regression model.

Example

- In the following, $r^2_{y1|2}$ and $r^2_{y2|1}$ are from the first model. $r^2_{y3|12}$ is coefficient from X_3 in the second model.

```
proc reg data = ch7tab01;
  model y = x1 x2 / pcorr2;
  model y = x1 x2 x3 / pcorr2;
run;
```

Multicollinearity

Standardized Multiple Regression Model:

Recall that the data are in the form of

$\mathbf{Y}$	$\mathbf{X}_1$	$\cdots$	$\mathbf{X}_k$
y_1	x_{11}	$\cdots$	x_{k1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_n	x_{1n}	$\cdots$	x_{kn}

Multicollinearity

Standardized Multiple Regression Model:

Now consider the a standardization procedure of the response variable Y and the predictor variables $X_1, \ldots, X_k$

$$\frac{Y_j - \overline{Y}}{\sqrt{n-1}S_Y}, j = 1, \ldots, n$$

$$\frac{x_{ij} - \overline{X_i}}{\sqrt{n-1}S_{X_i}}, j = 1, \ldots, n, i = 1, \ldots, k.$$

where $\overline{Y}$ and $\overline{X_i}$ are the sample means of the variable Y and X_i, respectively. S_Y and S_{X_i} are the sample standard deviations of the variable Y and X_i, respectively.

$$S_Y = \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2/(n-1)}$$

$$S_{X_i} = \sqrt{\sum_{j=1}^{n} (x_{ij} - \overline{X_i})^2/(n-1)}$$

Multicollinearity

Standardized Multiple Regression Model:

This centering and scaling results in $\mathbf{X}'\mathbf{X}$ being a correlation matrix. That is,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & r_{12} & \cdots & r_{1,k-1} & r_{1k} \\ 0 & r_{21} & 1 & \cdots & r_{2,k-1} & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & r_{k-1,1} & r_{k-1,2} & \cdots & 1 & r_{k-1,k} \\ 0 & r_{k1} & r_{k2} & \cdots & r_{k,k-1} & 1 \end{bmatrix}$$

where r_{ij} is the simple correlation coefficient between the predictor X_i and X_j.

The transformation is called **correlation transformation**.

correlation transformation

$$Y_j^* = \frac{1}{\sqrt{n-1}} \frac{Y_j - \overline{Y}}{S_Y}, j = 1, \ldots, n$$

$$x_{ij}^* = \frac{1}{\sqrt{n-1}} \frac{x_{ij} - \overline{X}_i}{S_{X_i}}, j = 1, \ldots, n, i = 1, \ldots, k.$$

Multicollinearity

Standardized Multiple Regression Model:

The regression model with the transformed variables Y^* and X_i^* as defined by the correlation transformation is called a **standardized regression model** and is given by

$$E(Y_i^*) = \beta_1^* x_1^* + \beta_2^* x_2^* + \cdots + \beta_k^* x_k^*.$$

or

$$E(\mathbf{Y}^*) = \mathbf{X}^* \boldsymbol{\beta}^*,\ \boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_k^*]'.$$

Note that there is no intercept parameter since the least square methods always produce an estimated intercept term of zero for the model.

Now $\mathbf{X}^{*'}\mathbf{X}^*$ is a $k \times k$ correlation matrix. That is,

$$\mathbf{X}^{*'}\mathbf{X}^* = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,k-1} & r_{1k} \\ r_{21} & 1 & \cdots & r_{2,k-1} & r_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1,1} & r_{k-1,2} & \cdots & 1 & r_{k-1,k} \\ r_{k1} & r_{k2} & \cdots & r_{k,k-1} & 1 \end{bmatrix}$$

Multicollinearity

Standardized Multiple Regression Model:

The least squares estimator of β^*

$$\widehat{\beta^*} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{Y}^*$$

is reduced to

$$\widehat{\beta^*} = \mathbf{r_{XX}}^{-1}\mathbf{r_{YX}},$$

where $\mathbf{r_{XX}} = \mathbf{X}^{*'}\mathbf{X}^*$ and $\mathbf{r_{YX}}$ is a vector containing the coefficients of simple correlation between the response variable Y and each of the X variables, denoted by r_{Yl}, r_{Y2}, etc.:

$$\mathbf{r_{YX}} = \begin{bmatrix} r_{Yl} \\ r_{Y2} \\ \vdots \\ r_{Yk} \end{bmatrix}$$

Multicollinearity and Its Effects

A very desirable condition in a set of regression data is to have predictors that are not "moving with each other" in the data set. Linear dependencies render it more difficult to sort out the impact of each predictor on the response.

Consider a data set with the following two predictors ($n = 8$)

x_1	10	10	10	10	15	15	15	15
x_2	10	10	15	15	10	10	15	15

For this data set,

$$\mathbf{X}^{*'}\mathbf{X}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (\mathbf{X}^{*'}\mathbf{X}^*)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$Var(\widehat{\beta}_1^*) = Var(\widehat{\beta}_2^*) = \sigma^2.$$

Multicollinearity and Its Effects

Consider another data set with the same size $n = 8$.

x_1	10.0	11.0	11.9	12.7	13.3	14.2	14.7	15.0
x_2	10.0	11.4	12.2	12.5	13.2	13.9	14.4	15.0

For this data set,

$$\mathbf{X}^{*'}\mathbf{X}^{*} = \begin{bmatrix} 1 & 0.99215 \\ 0.99215 & 1 \end{bmatrix}, \quad (\mathbf{X}^{*'}\mathbf{X}^{*})^{-1} = \begin{bmatrix} 63.94 & -63.44 \\ -63.44 & 63.94 \end{bmatrix}$$

and

$$Var(\widehat{\beta}_1^*) = Var(\widehat{\beta}_2^*) = 63.94\sigma^2.$$

Notice how the variances of the coefficients were inflated due to collinearity.

In the first data set, the predictors are said to be **orthogonal**; the condition of orthogonality is a very desirable property of an experimental design when one has the capability to control the independent variables.

Multicollinearity and Its Effects

Multicollinearity

Multicollinearity simply occurs when there are **near linear dependencies** among the $\mathbf{x}_i^*$, the columns of $\mathbf{X}^*$. That is, there is a set of constants (not all 0) for which

$$\sum_{i=1}^{k} c_i \mathbf{x}_i^* \approx 0$$

In later chapters, we return the issues of detecting multicollinearity, the removal of predictors and other methods of minimizing the effects of multicollinearity.

Example

```
proc sql;
  create table temp as
  select *, /*Selects all columns from the original dataset.*/
  ( y - mean(y) )/( std(y)*( sqrt( count(y)-1 ) ) ) as yprime,
  ( x1 - mean(x1) )/( std(x1)*( sqrt( count(x1)-1 ) ) ) as x1prime,
  ( x2 - mean(x2) )/( std(x2)*( sqrt( count(x2)-1 ) ) ) as x2prime,
  ( x3 - mean(x3) )/( std(x3)*( sqrt( count(x3)-1 ) ) ) as x3prime
  from ch7tab01;
quit;
```

Example

```
proc means data = temp mean std ;
var y x1 x2 x3;
var yprime x1prime x2prime x3prime;
run;
```

```
proc print data = temp;
var yprime x1prime x2prime x3prime;
run;
```

```
proc reg data = temp corr;
model yprime = x1prime x2prime x3prime;
run;
```

License