Multiple Regression II

KNNL – Chapter 7

Extra Sums of Squares

- For a given dataset, the total sum of squares remains the same, no matter what predictors are included;
- As we include more predictors, the regression sum of squares (SSR) increases (technically does not decrease), and the error sum of squares (SSE) decreases;
- SSR + SSE = SSTO, regardless of predictors in model;
- An extra sum of squares, for one or more extra predictor variables, measures the marginal reduction in the SSE when the extra predictor(s) are added to the model, given the other predictors are already in the model.

Extra Sums of Squares

- When a model contains just X_1, denote: $SSR(X_1)$, $SSE(X_1)$
- Model Containing both X_1, X_2: $SSR(X_1, X_2)$, $SSE(X_1, X_2)$
- Predictive contribution of X_2 above that of X_1:

  $SSR(X_2|X_1) = SSE(X_1) – SSE(X_1, X_2)$

  Or equivalently,

  $SSR(X_2|X_1) = SSR(X_1,X_2) – SSR(X_1)$
- Similarly, the predictive contribution of X_1 above that of X_2:

  $SSR(X_1|X_2) = SSE(X_2) – SSE(X_1, X_2) = SSR(X_1,X_2) – SSR(X_2)$

- Extends to any number of predictors

Definitions and Decomposition of SSR

$$SSTO = SSR(X_1) + SSE(X_1) = SSR(X_1, X_2) + SSE(X_1, X_2) = SSR(X_1, X_2, X_3) + SSE(X_1, X_2, X_3)$$

$$SSR(X_1 \mid X_2) = SSR(X_1, X_2) - SSR(X_2) = SSE(X_2) - SSE(X_1, X_2)$$
$$SSR(X_2 \mid X_1) = SSR(X_1, X_2) - SSR(X_1) = SSE(X_1) - SSE(X_1, X_2)$$

$$SSR(X_3 \mid X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3)$$
$$SSR(X_2, X_3 \mid X_1) = SSR(X_1, X_2, X_3) - SSR(X_1) = SSE(X_1) - SSE(X_1, X_2, X_3)$$

$$SSR(X_1, X_2) = SSR(X_1) + SSR(X_2 \mid X_1) = SSR(X_2) + SSR(X_1 \mid X_2)$$

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2 \mid X_1) + SSR(X_3 \mid X_1, X_2)$$
$$SSR(X_1, X_2, X_3) = SSR(X_2) + SSR(X_1 \mid X_2) + SSR(X_3 \mid X_1, X_2)$$
$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2, X_3 \mid X_1)$$

Note that as the # of predictors increases, so does the ways of decomposing SSR

ANOVA – Sequential Sum of Squares

Source of Variation	SS	df	MS
Regression	SSR(X1,X2,X3)	3	MSR(X1,X2,X3)
X1	SSR(X1)	1	MSR(X1)
X2|X1	SSR(X2|X1)	1	MSR(X2|X1)
X3|X1,X2	SSR(X3|X1,X2)	1	MSR(X3|X1,X2)
Error	SSE(X1,X2,X3)	n-4	MSE(X1,X2,X3)
Total	SSTO	n-1	

$$MSR\left(X_1\right) = \frac{SSR\left(X_1\right)}{1} \qquad MSR\left(X_2 \mid X_1\right) = \frac{SSR\left(X_2 \mid X_1\right)}{1}$$

$$MSR\left(X_3 \mid X_1, X_2\right) = \frac{SSR\left(X_3 \mid X_1, X_2\right)}{1}$$

$$MSR\left(X_1, X_2, X_3\right) = \frac{SSR\left(X_1, X_2, X_3\right)}{3}$$

$$MSR\left(X_2, X_3 \mid X_1\right) = \frac{SSR\left(X_2, X_3 \mid X_1\right)}{2}$$

Extra Sums of Squares & Tests of Regression Coefficients (Single β_k)

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \varepsilon_i \sim NID\left(0, \sigma^2\right)$

$H_0 : \beta_3 = 0 \quad H_A : \beta_3 \neq 0 \quad \Rightarrow$ Reduced Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

General Linear Test: $F^* = \dfrac{\left[\dfrac{SSE(R) - SSE(F)}{df_R - df_F}\right]}{\left[\dfrac{SSE(F)}{df_F}\right]}$

Full Model: $SSE(F) = SSE\left(X_1, X_2, X_3\right) \qquad df_F = n - 4$

Reduced Model: $SSE(R) = SSE\left(X_1, X_2\right) \qquad df_R = n - 3$

$SSE(R) - SSE(F) = SSE\left(X_1, X_2\right) - SSE\left(X_1, X_2, X_3\right) = SSR\left(X_3 \mid X_1, X_2\right)$

$df_R - df_F = \left(n-3\right) - \left(n-4\right) = 1$

$$\Rightarrow \quad F^* = \frac{\left[\dfrac{SSR\left(X_3 \mid X_1, X_2\right)}{1}\right]}{\left[\dfrac{SSE\left(X_1, X_2, X_3\right)}{n-4}\right]} = \frac{MSR\left(X_3 \mid X_1, X_2\right)}{MSE\left(X_1, X_2, X_3\right)} \overset{H_0}{\sim} F\left(1, n-4\right)$$

Rejection Region: $F^* \geq F\left(1-\alpha; 1, n-4\right) \qquad P\text{-value} = P\left(F\left(1; n-4\right) \geq F^*\right)$

Extra Sums of Squares & Tests of Regression Coefficients (Multiple β_k)

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \varepsilon_i \sim NID\left(0,\sigma^2\right)$

$H_0 : \beta_2 = \beta_3 = 0 \quad H_A : \beta_2$ and/or $\beta_3 \neq 0 \quad \Rightarrow \quad$ Reduced Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

General Linear Test: $F^* = \dfrac{\left[\dfrac{SSE(R) - SSE(F)}{df_R - df_F}\right]}{\left[\dfrac{SSE(F)}{df_F}\right]}$

Full Model: $SSE(F) = SSE\left(X_1, X_2, X_3\right) \qquad df_F = n-4$

Reduced Model: $SSE(R) = SSE\left(X_1\right) \qquad df_R = n-2$

$SSE(R) - SSE(F) = SSE\left(X_1\right) - SSE\left(X_1, X_2, X_3\right) = SSR\left(X_2, X_3 \mid X_1\right)$

$df_R - df_F = \left(n-2\right) - \left(n-4\right) = 2$

$$\Rightarrow \quad F^* = \frac{\left[\dfrac{SSR\left(X_2, X_3 \mid X_1\right)}{2}\right]}{\left[\dfrac{SSE\left(X_1, X_2, X_3\right)}{n-4}\right]} = \frac{MSR\left(X_2, X_3 \mid X_1\right)}{MSE\left(X_1, X_2, X_3\right)} \overset{H_0}{\sim} F\left(2, n-4\right)$$

Rejection Region: $F^* \geq F\left(1-\alpha; 2, n-4\right) \qquad P$-value $= P\left(F\left(2; n-4\right) \geq F^*\right)$

Extra Sums of Squares & Tests of Regression Coefficients (General Case)

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad \varepsilon_i \sim NID\left(0, \sigma^2\right)$

$H_0 : \beta_q = \ldots = \beta_{p-1} = 0 \quad H_A :$ At least one of $\beta_q \ \ldots \ \beta_{p-1} \neq 0$

$\Rightarrow$ Reduced Model: $Y_i = \beta_0 + \beta_1 X_{i1} \ldots + \beta_{q-1} X_{i,q-1} + \varepsilon_i \qquad (q < p)$

General Linear Test: $$F^* = \frac{\left[\dfrac{SSE(R) - SSE(F)}{df_R - df_F}\right]}{\left[\dfrac{SSE(F)}{df_F}\right]}$$

Full Model: $SSE(F) = SSE\left(X_1, X_2 \ldots, X_{p-1}\right) \qquad df_F = n - p$

Reduced Model: $SSE(R) = SSE\left(X_1, X_2 \ldots, X_{q-1}\right) \qquad df_R = n - q$

$SSE(R) - SSE(F) = SSE\left(X_1, X_2 \ldots, X_{q-1}\right) - SSE\left(X_1, X_2 \ldots, X_{p-1}\right) = SSR\left(X_q, \ldots, X_{p-1} \mid X_1, X_2 \ldots, X_{q-1}\right)$

$df_R - df_F = (n - q) - (n - p) = p - q$

$$\Rightarrow \quad F^* = \frac{\left[\dfrac{SSR\left(X_q, \ldots, X_{p-1} \mid X_1, X_2 \ldots, X_{q-1}\right)}{p - q}\right]}{\left[\dfrac{SSE\left(X_1, X_2 \ldots, X_{p-1}\right)}{n - p}\right]} = \frac{MSR\left(X_q, \ldots, X_{p-1} \mid X_1, X_2 \ldots, X_{q-1}\right)}{MSE\left(X_1, X_2 \ldots, X_{p-1}\right)} \overset{H_0}{\sim} F\left(p - q, n - p\right)$$

Rejection Region: $F^* \geq F\left(1 - \alpha; p - q, n - p\right) \qquad P\text{-value} = P\left(F\left(p - q; n - p\right) \geq F^*\right)$

Other Linear Tests

Suppose firm has two types of advertising:

print $(X_1,\ \text{in } \$1000\text{s})$ and internet $(X_2,\ \text{in } \$1000\text{s})$ as well as promotional expenditures $(X_3,\ \text{in } \$1000\text{s})$:

Let Sales = Y, they vary their expenditures on n periods and observe sales in each (Price is constant)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \varepsilon_i \sim NID\left(0, \sigma^2\right)$$

Test of equal effects of increasing each input by 1 unit (say \$1000s):

$H_0: \beta_1 = \beta_2 = \beta_3 \qquad H_A: H_0$ is False

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad df_F = n-4$

Reduced Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_1 X_{i2} + \beta_1 X_{i3} + \varepsilon_i \quad \Rightarrow \quad Y_i = \beta_0 + \beta_1 \left(X_{i1} + X_{i2} + X_{i3}\right) + \varepsilon_i$

$\Rightarrow \quad Y_i = \beta_0 + \beta_1 W_i + \varepsilon_i \qquad W_i = X_{i1} + X_{i2} + X_{i3} \qquad df_R = n-2$

Test that Mean sales when all inputs=0 is \$10,000 and effect of increasing X_3 by 1 unit is 1:

$H_0: \beta_0 = 10, \beta_3 = 1 \qquad H_A: H_0$ is False (all units are \$1000s)

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad df_F = n-4$

Reduced Model: $Y_i = 10 + \beta_1 X_{i1} + \beta_2 X_{i2} + 1X_{i3} + \varepsilon_i \quad \Rightarrow \quad Y_i - 10 - 1X_{i3} = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

$\Rightarrow \quad U_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad U_i = Y_i - 10 - 1X_{i3} \qquad df_R = n-2$ (no intercept)

Coefficients of Partial Determination-I

Proportion of Variation Explained by 1 or more variables, not explained by others

Regression of Y on X_1: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

Variation Explained: $SSR(X_1)$ Unexplained: $SSE(X_1) = SSTO - SSR(X_1)$

Regression of Y on X_2: $Y_i = \beta_0 + \beta_2 X_{i2} + \varepsilon_i$

Variation Explained: $SSR(X_2)$ Unexplained: $SSE(X_2) = SSTO - SSR(X_2)$

Regression of Y on X_1, X_2: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

Variation Explained: $SSR(X_1, X_2)$ Unexplained: $SSE(X_1, X_2) = SSTO - SSR(X_1, X_2)$

Proportion of Variation in Y, Not Explained by X_1, that is Explained by X_2:

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

Proportion of Variation in Y, Not Explained by X_2, that is Explained by X_1:

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

Coefficients of Partial Determination-II

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

$$= \frac{SSR(X_1,X_2,X_3)-SSR(X_2,X_3)}{SSTO-SSR(X_2,X_3)} = \frac{SSR(X_1 \mid X_2,X_3)}{SSE(X_2,X_3)}$$

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

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

$$R^2_{Y23|1} = \frac{SSE(X_1)-SSE(X_1,X_2,X_3)}{SSE(X_1)} = \frac{SSR(X_1,X_2,X_3)-SSR(X_1)}{SSE(X_1)} =$$

$$= \frac{SSR(X_1,X_2,X_3)-SSR(X_1)}{SSTO-SSR(X_1)} = \frac{SSR(X_2,X_3 \mid X_1)}{SSE(X_1)}$$

Coefficient of Partial Correlation:

$$R_{Y2|1} = \operatorname{sgn}\{\beta_2\}\sqrt{R^2_{Y2|1}} \qquad \operatorname{sgn}\{\beta_2\} = \begin{cases} + \text{ if } \hat{\beta}_2 > 0 \text{ in } \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \\ - \text{ if } \hat{\beta}_2 < 0 \text{ in } \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \end{cases}$$

Multicollinearity

- Consider model with 2 Predictors (this generalizes to any number of predictors) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$
- When X_1 and X_2 are uncorrelated, the regression coefficients b_1 and b_2 are the same whether we fit simple regressions or a multiple regression, and: $SSR(X_1) = SSR(X_1|X_2)$ $\qquad$ $SSR(X_2) = SSR(X_2|X_1)$
- When X_1 and X_2 are highly correlated, their regression coefficients become unstable, and their standard errors become larger (smaller t-statistics, wider CI^s), leading to strange inferences when comparing simple and partial effects of each predictor
- Estimated means and Predicted values are not affected

Standardized Regression Model - I

- Useful in removing round-off errors in computing $(X'X)^{-1}$, which used to be a big problem in early computer days, especially when different predictors have different magnitudes. The standardized regression makes all entries in the $X'X$ matrix to be between -1 and 1.
- Makes easier comparison of magnitude of effects of predictors measured on different measurement scales
- Coefficients represent changes in Y (in standard deviation units) as each predictor increases 1 SD (holding all others constant)
- Since all variables are centered, no intercept term

Standardized Regression Model - II

The usual standardized variables are such that the mean is 0 and st. dev. is 1:

$$Y_i^* = \frac{Y_i - \bar{Y}}{s_Y} \text{ and } X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{s_k}, \qquad k = 1, \dots, p-1$$

In standardized regression, a very similar transformation is done, called Correlation Transformation:

$$Y_i^* = \frac{1}{\sqrt{n-1}} \frac{Y_i - \bar{Y}}{s_Y} \text{ and } X_{ik}^* = \frac{1}{\sqrt{n-1}} \frac{X_{ik} - \overline{X_k}}{s_k}, \qquad k = 1, \dots, p-1$$

It produces same regression coefficients as the usual standardization, but more computationally stable.
The resulting regression model is:

$$Y_i^* = \beta_1^* X_{i1}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^*$$

Connection with regular regression model:

$$\beta_k = \left(\frac{s_Y}{s_k}\right) \beta_k^*, \qquad k = 1, \dots, p-1$$

Standardized Regression Model - III

Standardized Regression Model:

$$Y_i^* = \beta_1^* X_{i1}^* + \ldots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^*$$

$$\underset{n\times(p-1)}{\mathbf{X}^*} = \begin{bmatrix} X_{11}^* & \cdots & X_{1,p-1}^* \\ \vdots & & \vdots \\ X_{n1}^* & \cdots & X_{n,p-1}^* \end{bmatrix} \qquad \underset{n\times 1}{\mathbf{Y}^*} = \begin{bmatrix} Y_1^* \\ Y_1^* \\ \vdots \\ Y_1^* \end{bmatrix} \qquad \underset{(p-1)\times(p-1)}{\mathbf{X}^{*\prime}\mathbf{X}^*} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{21} & 1 & \cdots & r_{1,p-1} \\ \cdots & \cdots & \ddots & \cdots \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix} = \mathbf{r}_{XX} \qquad \underset{(p-1)\times 1}{\mathbf{X}^{*\prime}\mathbf{Y}^*} = \begin{bmatrix} r_{Y1} \\ r_{Y2} \\ \vdots \\ r_{Y,p-1} \end{bmatrix} = \mathbf{r}_{YX}$$

This results from:

$$\sum_i \left(X_{ik}^*\right)^2 = \sum_i \left(\frac{1}{\sqrt{n-1}}\left(\frac{X_{ik}-\overline{X}_k}{s_k}\right)\right)^2 = \left(\frac{1}{s_k^2}\right)\frac{\sum_i \left(X_{ik}-\overline{X}_k\right)^2}{n-1} = \left(\frac{1}{s^2\left\{X_k\right\}}\right)s^2\left\{X_k\right\} = 1$$

$$\sum_i \left(X_{ik}^*\right)\left(X_{ik'}^*\right) = \sum_i \left(\frac{1}{\sqrt{n-1}}\left(\frac{X_{ik}-\overline{X}_k}{s_k}\right)\right)\left(\frac{1}{\sqrt{n-1}}\left(\frac{X_{ik'}-\overline{X}_{k'}}{s_{k'}}\right)\right) = \left(\frac{1}{s_k s_{k'}}\right)\frac{\sum_i \left(X_{ik}-\overline{X}_k\right)\left(X_{ik'}-\overline{X}_{k'}\right)}{n-1} = \frac{s\left\{X_k, X_{k'}\right\}}{s\left\{X_k\right\}s\left\{X_{k'}\right\}} = r_{kk'}$$

$$\sum_i \left(Y_i^*\right)\left(X_{ik}^*\right) = \sum_i \left(\frac{1}{\sqrt{n-1}}\left(\frac{Y_i-\overline{Y}}{s_Y}\right)\right)\left(\frac{1}{\sqrt{n-1}}\left(\frac{X_{ik}-\overline{X}_k}{s_k}\right)\right) = \left(\frac{1}{s_Y s_k}\right)\frac{\sum_i \left(Y_i-\overline{Y}\right)\left(X_{ik}-\overline{X}_k\right)}{n-1} = \frac{s\left\{Y, X_k\right\}}{s\left\{Y\right\}s\left\{X_k\right\}} = r_{Yk}$$

$$\underset{(p-1)\times(p-1)}{\mathbf{X}^{*\prime}\mathbf{X}^*}\ \underset{(p-1)\times 1}{\mathbf{b}^*} = \underset{(p-1)\times 1}{\mathbf{X}^{*\prime}\mathbf{Y}^*} \quad\Rightarrow\quad \mathbf{b}^* = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^* \quad\Rightarrow\quad \mathbf{b}^* = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX}$$

$$b_k = \left(\frac{s_Y}{s_k}\right)b_k^* \qquad k = 1,\ldots,p-1 \qquad b_0 = \overline{Y} - b_1\overline{X}_1 - \ldots - b_{p-1}\overline{X}_{p-1}$$