Stat 501 Lab 09

1 Sequential (or extra) sums of squares

This exercise reviews the concept of "**sequential** (or **extra**) **sums of squares**." Sequential sums of squares are useful, because they can be used to test:

- whether one slope parameter is 0 (for example, $H_0: \beta_1 = 0$)
- whether a subset (more than two, but less than all) of the slope parameters are 0 (for example, $H_0: \beta_2 = \beta_3 = 0$)

Again, **what is a sequential sum of squares**? It can be viewed in either of two ways:

- It is the *reduction* in the error sum of squares (SSE) when one or more predictor variables are added to the model.
- Or, it is the *increase* in the regression sum of squares (SSR) when one or more predictor variables are added to the model.

1.1 Brain size and body size study

Recall that the **iqsize.txt** data set contains data on the intelligence based on the performance IQ ($y = PIQ$) scores from the revised Wechsler Adult Intelligence Scale, brain size ($x_1 = brain$) based on the count from MRI scans (given as count/10000), and body size measured by height in inches ($x_2 = height$) and weight in pounds ($x_3 = weight$) on 38 college students.

1. Fit the linear regression model with $x_1 = brain$ as the only predictor.

- Take note of the error sum of squares, and since x_1 is the only predictor in the model, denote this value as $SSE(X_1)$. What is the value of $SSE(X_1)$?
- Take note of the regression sum of squares, and since x_1 is the only predictor in the model, denote this value as $SSR(X_1)$. What is the value of $SSR(X_1)$?
- Take note of the total sum of squares, and since x_1 is the only predictor in the model, denote this value as $SSTO(X_1)$. What is the value of $SSTO(X_1)$?

2. Now, fit the linear regression model with the predictors (in order) $x_1 = brain$ and $x_2 = height$ in the model.

- Take note of the error sum of squares, and since x_1 and x_2 are predictors in the model, denote this value as $SSE(X_1, X_2)$. What is the value of $SSE(X_1, X_2)$?
- Take note of the regression sum of squares, and since x_1 and x_2 are predictors in the model, denote this value as $SSR(X_1, X_2)$. What is the value of $SSR(X_1, X_2)$?
- Take note of the total sum of squares, and since x_1 and x_2 are predictors in the model, denote this value as $SSTO(X_1, X_2)$. What is the value of $SSTO(X_1, X_2)$?

3. Now, let's use the above definitions to calculate the sequential sum of squares of adding X_2 to the model in which X_1 is the only predictor. We denote this quantity as $SSR(X_2|X_1)$. (The bar "|" is read as "given.") According to the alternative definitions:

- $SSR(X_2|X_1)$ is the *reduction in the error sum of squares* when X_2 is added to the model in which X_1 is the only predictor. That is, $SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2)$. What is the value of $SSR(X_2|X_1)$ calculated this way?

- Alternatively, we can think of the $SSR(X_2|X_1)$ as the *increase in the regression sum of squares* when X_2 is added to the model in which X_1 is the only predictor. That is, $SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1)$. What is the value of $SSR(X_2|X_1)$ calculated this way? Did you get the same answer as above? (You should, ignoring small round-off error).

4. Note that because you fit a multiple regression model, Minitab automatically displays a column of sequential sum of squares named "Seq SS". The sequential sums of squares you get depends on the order in which you enter the predictors in the model.

- Since you entered $x_1 = brain$ first, the number Minitab displays for the Seq SS for *brain* is $SSR(X_1)$. What is the value Minitab displays for $SSR(X_1)$, and is it consistent with the value of $SSR(X_1)$ you obtained in question (1)? In words, how would you describe the sequential sum of squares $SSR(X_1)$?
- Since you entered, $x_2 = height$ second, the number Minitab displays for Seq SS for *height* is $SSR(X_2|X_1)$. What is the value Minitab displays for $SSR(X_2|X_1)$, and is it consistent with the value of $SSR(X_2|X_1)$ you obtained in question (3)?

5. Let's make sure we see how the sequential sums of squares that we get depends on the order in which we enter the predictors in the model. Fit the linear regression model with the two predictors in the reverse order. That is, when fitting the model, indicate $x_2 = height$ first and $x_1 = brain$ second.

- Since you entered $x_2 = height$ first, the number Minitab displays for the Seq SS for *height* is $SSR(X_2)$. What is the value Minitab displays for $SSR(X_2)$?
- Since you entered $x_1 = brain$ second, the number Minitab displays for the Seq SS for *brain* is $SSR(X_1|X_2)$. What is the value Minitab displays for $SSR(X_1|X_2)$?

You can (and should!) verify the values Minitab displays for $SSR(X_2)$ and $SSR(X_1|X_2)$ by fitting the linear regression model with $x_2 = height$ as the only predictor.

- Is the value of $SSR(X_2)$ determined this way consistent with the value you obtained under the Seq SS column?
- Calculate $SSR(X_1|X_2)$ using either of the two definitions. Is your calculation consistent with the value Minitab displays under the Seq SS column?

6. Sequential sum of squares can be obtained for any number of predictors that are added sequentially to the model. To see this, now fit the linear regression model with the predictors (in order) $x_1 = brain$ and $x_2 = height$ and $x_3 = weight$. First:

- Take note of the error sum of squares, and since x_1 and x_2 and x_3 are predictors in the model, denote this value as $SSE(X_1, X_2, X_3)$. What is the value of $SSE(X_1, X_2, X_3)$?
- Take note of the regression sum of squares, and since x_1 and x_2 and x_3 are predictors in the model, denote this value as $SSR(X_1, X_2, X_3)$. What is the value of $SSR(X_1, X_2, X_3)$?
- Take note of the total sum of squares, and since x_1 and x_2 and x_3 are predictors in the model, denote this value as $SSTO(X_1, X_2, X_3)$. What is the value of $SSTO(X_1, X_2, X_3)$?

Now, consider the sequential sums of squares Minitab displays:

- The first two values, $SSR(X_1)$ and $SSR(X_2|X_1)$, should be consistent with their previous values, because you entered $x_1 = brain$ first and $x_2 = height$ second. Are they?
- Since you entered $x_3 = weight$ third, the number Minitab displays for the Seq SS for *weight* is $SSR(X_3|X_1, X_2)$. What is the value Minitab displays for $SSR(X_3|X_1, X_2)$? Calculate $SSR(X_3|X_1, X_2)$ using either of the two definitions. Is your calculation consistent with the value Minitab displays under the Seq SS column?

7. All of the sequential sums of squares we considered so far are **"one-degree-of-freedom sequential sums of squares,"** because we have only considered the effect of adding one additional predictor variable to a model. We could, however, want to quantify the effect of adding two additional predictor variables to a model. For example, we might want to know the effect of adding X_2 and X_3 to a model that already contains X_1 as a predictor. The sequential sum of squares $SSR(X_2, X_3|X_1)$ quantifies this effect. It is a **"two-degree-of-freedom sequential sum of squares,"** because it quantifies the effect of adding two additional predictor variables to the model. One-degree-of-freedom sequential sums of squares are used in testing one slope parameter such as $H_0 : \beta_1 = 0$, where as two-degree-of-freedom sequential sums of squares are used in testing two slope parameters, such as $H_0 : \beta_1 = \beta_2 = 0$.

There are two ways of obtaining two-degree-of-freedom sequential sums of squares – by the original definition of a sequential sum of square or by adding the proper one-degree of freedom sequential sums of squares.

- Use either of the two definitions to calculate $SSR(X_2, X_3|X_1)$. That is, calculate $SSR(X_2, X_3|X_1)$ by $SSR(X_1, X_2, X_3) - SSR(X_1)$ or by $SSE(X_1) - SSE(X_1, X_2, X_3)$.
- Calculate $SSR(X_2, X_3|X_1)$ by adding the proper one-degree of freedom sequential sum of squares, that is, $SSR(X_2|X_1) + SSR(X_3|X_1, X_2)$. Do you get the same answer?

Incidentally, you can use the same concepts to get three-degree-of-freedom sequential sum of squares, four-degree-of-freedom sequential sum of squares, and so on.

2 Decomposition of regression sum of squares

This exercise is intended to illustrate how the regression sums of squares can be decomposed into a sum of sequential sum of squares. We can use a decomposition to quantify how important a predictor variable is ("marginally") in reducing the variability in the response (in the presence of the other variables in the model).

2.1 Brain size and body size study

We'll use the **iqsize.txt** data set again, with the response $y = PIQ$ and three possible predictors, $x_1 = brain$ and $x_2 = height$ and $x_3 = weight$.

1. Fit the linear regression model with $y = PIQ$ and (in order) $x_1 = brain$ and $x_2 = height$. Verify that the regression sum of squares obtained, $SSR(X_1, X_2)$, is the sum of the two sequential sum of squares $SSR(X_1)$ and $SSR(X_2|X_1)$. This illustrates how $SSR(X_1, X_2)$ is "decomposed" into a sum of sequential sum of squares.

2. A regression sum of squares can be decomposed in more than way. To see this, fit the linear regression model with $y = PIQ$ and (in order) $x_2 = height$ and $x_1 = brain$. Verify that the regression sum of squares obtained, $SSR(X_1, X_2)$, is now the sum of the two sequential sum of squares $SSR(X_2)$ and $SSR(X_1|X_2)$. That is, we've now decomposed $SSR(X_1, X_2)$ in a different way.

3. Now, fit the linear regression model with $y = PIQ$ and (in order) $x_1 = brain$ and $x_2 = height$ and $x_3 = weight$. Verify that the regression sum of squares obtained, $SSR(X_1, X_2, X_3)$, is the sum of the three sequential sum of squares $SSR(X_1)$ and $SSR(X_2|X_1)$ and $SSR(X_3|X_1, X_2)$.

3 Hypothesis tests for the slope parameters

The exercise in this section is designed to review the hypothesis tests for the slope parameters, as well as to give you more practice on models with a three-group qualitative variable. We consider tests for:

- whether one slope parameter is 0 (for example, $H_0 : \beta_1 = 0$)
- whether a subset (more than two, but less than all) of the slope parameters are 0 (for example, $H_0 : \beta_2 = \beta_3 = 0$ against the alternative $H_A : \beta_2 \neq 0$ or $\beta_3 \neq 0$ or both)