Using the Partial F-test (Sect. 4.10 in text)

Consider our seasonal model, with linear trend, to estimate the calculator sales index at a particular university.

We would like to know if we can ignore the seasonal variation. Specifically, can we drop the dummy variables for quarters 1 ($x_{s1,t}$), 2 ($x_{s2,t}$) and 3 ($x_{s3,t}$)? To answer this, we use a partial F-test.

Complete model: $y_t = \beta_0 + \beta_1 t + \beta_{s1} x_{s1,t} + \beta_{s2} x_{s2,t} + \beta_{s3} x_{s3,t} + \epsilon_t$

Reduced model: $y_t = \beta_0 + \beta_1 t + \epsilon_t$

Our hypotheses are:

$$H_o : \beta_{s1} = \beta_{s2} = \beta_{s3} = 0$$

$$H_a : H_o \text{ not true}$$

We have 2 ways of finding $SSE_R - SSE_C$. The first is to fit the complete model, as shown on our previous Minitab output for this example. We repeat a part of it below:

```
Data: y: Sales Index; t: Time (1, 2,..., 20)
      x_(1s,t): 1 if obs. in quarter 1  (Q1)
      x_(2s,t): 1 if obs. in quarter 2  (Q2)
      x_(3s,t): 1 if obs. in quarter 3  (Q3)

# most output omitted

Source       DF      Seq SS
Time          1      114343
Q1            1       81883 # (81883 + 94610 + 27724): amount
Q2            1       94610 # by which SSE rises if Q1-Q3
Q3            1       27724 # (seasonal terms) dropped from model
```

and use the sequential sums of squares given by the Minitab output:

$$SSE_R - SSE_C = 81883 + 94610 + 27724 = 204217$$

The second method is to use Minitab to fit both possible models. We fit the complete model on our previous handout. Here is some of the output from using the reduced model:

```
The regression equation is
Sales = 303 + 13.1 Time

s = 109.1       R-sq = 34.8%     R-sq(adj) = 31.2%
```

```
Analysis of Variance

SOURCE           DF          SS          MS        F        P
Regression        1      114343      114343     9.60    0.006
Residual Error   18      214356       11909
Total            19      328700
```

Then $SSE_R = 214356$. This gives us

$$SSE_R - SSE_C = 214356 - 10139 = 204217$$

We see both methods give the same result.

What if we had wanted to test if we could use a model which only used time and the dummy variable for the third quarter in our model? This doesn't make much sense, but we'll do it to illustrate a point. We would have

$$H_o : \beta_{s1} = \beta_{s2} = 0$$

The $SSE_R - SSE_C$ could be found from the model output below:

```
#some output deleted

 The regression equation is
 Sales = 315 + 13.3 Time - 54.1 Q3

 s = 109.4       R-sq = 38.1%     R-sq(adj) = 30.8%

 Analysis of Variance

 SOURCE          DF          SS          MS        F        P
 Regression       2      125282       62641     5.24    0.017
 Residual Error  17      203417       11966
 Total           19      328700
```

and find $SSE_R - SSE_C = 203417 - 10139 = 193278$.

However, we could **not** use the sequential sums of squares above, *i.e.* add the `Seq SS` rows that correspond to the variables we want to drop, because we'd find $81883 + 94610 \neq 193278$.

To use the sequential SS values from Minitab, the variables we propose to drop from our model must be entered **last** in our regression model (so you specify them last in the predictor box in Minitab):

```
Source        DF      Seq SS
Time           1      114343
Q3             1       10939
Q1             1       71060 # 71060 + 122218 = amount by which SSE rises
Q2             1      122218 # if Q1 and Q2 dropped from model
```

Then $SSE_R - SSE_C = 71060 + 122218 = 193278$, as desired.