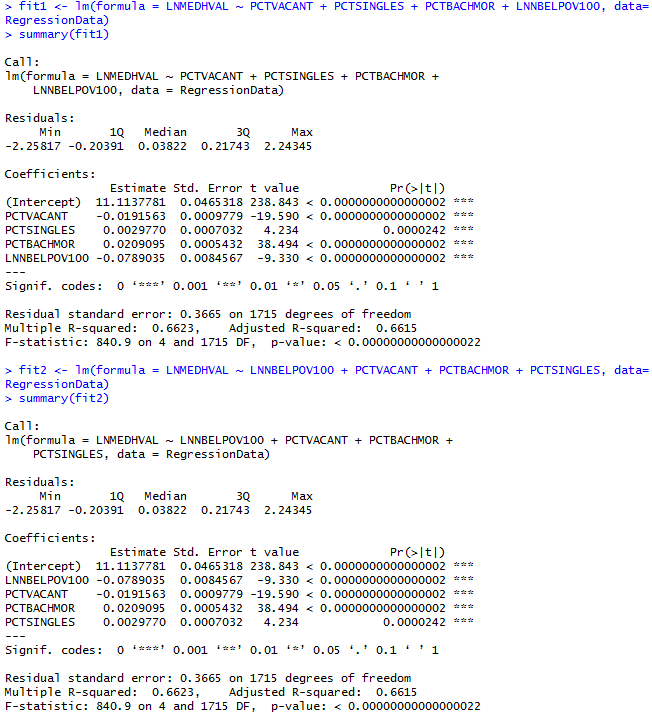
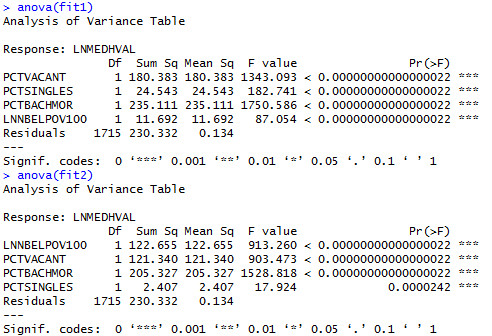
*A note on the order of regression predictors and the* ***anova(fit)*** *output*

Imagine that we regress **LNMEDHVAL** on 4 predictors: **PCTVACANT**, **PCTSINGLES**, **PCTBACHMOR** and **LNNBELPOV100**. Assuming we are NOT dealing with stepwise regression, we know that the coefficient and significance estimates, as well as estimates of the model R2 and F-statistic, shouldn’t depend on the order of the predictors in the **lm** statement. As you can see, **fit1** and **fit2** below yield the same results.



However, what happens if we look at **anova(fit1)** and **anova(fit2)**?



If we examine the results above, we will see that the SS (sum of squares) associated with each predictor will be different for **fit1** and **fit2**! Before moving on to explaining why that is the case, let’s make several important observations:

1) SSE (error sum of squares) is the same for **fit1** and **fit2** (230.332)

2) SSR (regression sum of squares) is the same in both models. In **fit1**, SSR = 180.383 + 24.543 + 235.111 + 11.692 = 451.729. In **fit2**, SSR = 122.655 + 121.340 + 205.327 + 2.407 = 451.729

3) The fact that SSE and SSR are the same for both **fit1** and **fit2** implies that SST = SSE + SSR is the same for both, and so is R2 = SSR/SST. We’ve seen by looking at **summary(fit1)** and **summary(fit2)** above that R2 is 0.6623 in both cases.

For the intents and purposes of this course (1-3) above is all we care about.

However, what is the interpretation of sum of squares for each predictor in the **anova()** output? The command **anova()** returns the so-called [Type I Sum of Squares](http://www.stat.tamu.edu/~hart/652/poly.more.pdf). For type I SS, the sum of squares for each predictor is the amount by which SSR goes up when the predictor in a given line is added into a model containing all predictors on previous lines, but no others. So for **fit1**, the variable on line 2 (**PCTSINGLES**) has an SS of 24.543. That means that – after accounting for the effect of the variable on the previous line (**PCTVACANT**) – entering **PCTSINGLES** into the model raises SSR by 24.543. For **fit2, PCTSINGLES** is on line 4 and its SS is 2.407. It means that, after including predictors on the previous lines (**LNBELPOV100**, **PCTVACANT**, **PCTBACHMOR**), the predictor **PCTSINGLES** has a relatively small contribution to SSR (which we have calculated above to be 451.729).