Homework

Load the California school data contained in the caschool.txt file on the website into Gauss. The first row contains variable names, the remaining ones data. Separate the data names from the observations. The code would look something like this:

```
load datall[421,6]=caschool1.txt;
vnames = datall[1,.];
$vnames;
```

If you are using Gauss Light, then you'll need to limit your data to the first 100 observations.

```
dat = datall[2:101,.];
```

To clear the contents of the **datall** matrix from memory:

```
clear datall ;
```

Part I

Consider the following model of test scores:

$$testscr = \beta_1 X_1 + \beta_2 X_2 + u_t$$

X_1 contains a constant, *str, el_pct*, and *avginc;* X_2 contains *comp_st* and *expn_stu*.

1. Write a Gauss program to obtain the least squares estimates of β_1 and β_2. Remember, these are vectors.
2. Reestimate the model omitting X_2. Why do you think the estimates of β_1 change so much?
3. Use the orthogonal projection matrices to obtain residuals and predicted values. Compare these to ones you generate in Stata using the same data.

Part II

Consider the following model of test scores:

$$testscr = \beta_1 X_1 + \beta_2 X_2 + u_t$$

X_1 contains a constant, *str, el_pct*, and *avginc;* X_2 contains *comp_st* and *expn_stu*.

1. Using these data and definitions, complete exercise 2.16 in ETM.
2. Although you can verify each result easily using the data, also try to explain the principle that causes each result.

3. Use Stata (or gretl) to replicate (g), (h) and (i) and (j). You will basically be running regressions, saving residuals and predictions and running subsequent regressions. The purpose of this is to help give you some intuition about what these projections actually do.