

POL 204b: Homework 5

Due November 11, 2014

Notes

This homework we will do a little more practice with regression, apply some F-tests, and will practice our data management. We'll move to interactions next week.

1 General Problems

1. I started this proof on the board at the beginning of class. Finish it.

From the general statement of variance and covariance, derive the variance-covariance matrix of $\hat{\beta}$.

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \text{Var}(\hat{\beta}) &= E[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])'] \\ &= \text{do stuff here...} \\ &= \sigma_{\epsilon}^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

2. Using the following dataset, estimate least-squares regression coefficients and conduct a hypothesis test for $\beta_1 = 0$ and $\beta_2 = 0$, using linear algebra. Do this *by hand*.¹ Then use the L matrix approach to test whether $\beta_1 = \beta_2 = .25$ (you may do this one using R and not paper/pencil if you wish).

After completing the problem, look at the variance-covariance matrix. Explain how to interpret the covariance between the estimates for β_0 , β_1 , and β_2 , and what feature of the dataset produced the values you observed. Again, you should be able to do the first by hand, and the second with linear algebra in R, and finally verify it with `lm(...)`.

¹The inversion is only a 2x2

	y	x1	x2
[1,]	0.0	-1.5	-1.5
[2,]	3.0	-1.5	-0.5
[3,]	0.5	0.0	0.5
[4,]	3.5	0.0	1.5
[5,]	1.0	0.5	1.5
[6,]	2.0	0.5	0.5
[7,]	-2.5	2.0	-0.5
[8,]	-1.5	2.0	-1.5

2 Fox Problems

3. The proof for Fox 9.8 follows a similar flow. So, do it!
4. This week, I really do want you to do 9.10 in Fox.
5. (Optional) Fox 9.11 builds on what you did last week in 9.6 and helps (maybe) to generate some intuition about what is happening in this $(\mathbf{X}'\mathbf{X})^{-1}$ matrix.
6. Fox 9.14: (a) and (c).² **Additional**, (required) parts: Use the dataset available [here](#) to:
 - (a) Program – aka don't use `lm(...)` – the OLS estimator to generate these estimated regression coefficients
 - (b) Generate an object that is the VCOV of $\hat{\beta}$,
 - (c) Use these two pieces to perform a significance test for the individual regressors
 - (d) Test the joint significant of all the regressors using an F-test.
 - (e) Use an L matrix to test, and then interpret the results of, these two hypotheses:
 - i. *women* has no model-level effect
 - ii. *women* and *income* together have no model level effects
 - (f) Create a \mathbf{x}_0 matrix with the information from part (c) (\$12,000 dollars, $\bar{edu} = 13$, 50% women) and do the matrix multiplication to predict the prestige score for the this occupation. Use the same linear algebra to place a 90% confidence interval around your prediction, just as you did in part (c).
 - (g) Now, make a \mathbf{x}_0 matrix that has 1000 rows. Hold all the RHS regressors at some sensible value (their mean, median or mode), but vary income from a low of \$8,000 to a high of \$40,000 (`seq(from = 8000, to = 40000, length.out = 1000)`). Again, perform the matrix multiplication to generate predicted values and confidence intervals for those predictions. *Describe what this is, and how you would interpret it.* Also describe why the same of the confidence interval looks like it does.

²But only do part (c) that doesn't depend on part (b). Confusing description? Yes. Less, work, also yes. The regression results Fox mentions are on p. 91 of the textbook.

7. Fox 7.1. Yes, I do mean to assign this dummy regression question even though we didn't talk long about them in lecture.

3 Data Task

8. Get some data from somewhere and present what you think are interesting results. Write out a *very* simple theory and the hypotheses that you will be testing. Tell us where you got the data, who produced it, etc. For each variable that you use, include a brief (one-sentence) description of the variable along with the variable name so that David has *some* idea of what you're talking about. Include a summary table or plot of *all* of the variables that you use so that we have an understanding of how your data is distributed.

Include a table with the results of your regression. This table should have all of these pieces:

- (a) A title for the table
- (b) The name of the **DV** and the type of model that you used (OLS)
- (c) Human-legible names the variables in your model. YearsEdu is fine. yr.edu is not.
- (d) The estimated relationship between IVs and DV
- (e) The standard error of these estimates
- (f) The t-stat and p-val for a test of each of these against the null hypothesis (which need not be zero)
- (g) The F-test for joint significance of the entire model against a null model and the p-val of an F-statistic of this size
- (h) The number of observations that are included in your estimation
- (i) A caption that describes the key findings that you would want people to walk away with *if they only ever read your table* and not your paper.

Include a plot that tells us something about the estimate that you just produced. This could be a visual way of presenting the table, or the marginal effects of one of your variable on the DV. Make something that looks good.

Bundle all of this – your 1 paragraph theory, hypotheses, plots, model, and table – into something that looks nice. Imagine it to be like a lab report. So...David and I should know everything necessary to identify the question you're asking, how you've answered it, and what you decided at the end. We don't need (or want) to read any more than that.