

POLS 6481 - Methods 2 - Lab 4

Tom Hanna

University of Houston

tlhanna@uh.edu

February 22, 2021

Overview

1 Statistics

- Lab 3 Followup: Statistics
- Lab 4: Statistics Overview

2 Technical Lab

- Lab Survey from last week
- Lab 3 Followup: Coding
- Lab 4: Coding

Nothing to see here!

There were no major questions. I'll cover minor points in the survey answers.

Concepts and comments

- Article reading in ECON 4389
- Testing significance of coefficients
- t-tests and confidence intervals
- single coefficient versus zero
- two-coefficients - same or different
- Variance-covariance matrix
- ANOVA

Things to look for!

- Computing standard errors from the Variance-covariance matrix
- Computing the t-statistic
- Computing R-squared and adjusted R-squared
- Relationship between R-squared and adjusted R-squared
- Relationship between F and t-test

Question 1

Your Name!

Question 2: Favorite foods

- Mine: Pizza because it can be anything
- Several votes for sushi/sashimi, various pasta and noodle dishes, and pizza
- The stumper: Donburi

Question 4

Question

In line 14, the code `"cor(cadta[c(15, 8, 9, 16, 12, 13, 14)])"` produces a ____ matrix. The numbers inside the brackets indicate the _____ in the dataframe `cadta`.

Answer

correlation; columns or variables

Question 5

Question

What is the baseline model's R^2 ? _____. What is the model's residual standard error? _____.

Answer

.707, 10.3

Question 6

Question

What happened to the size of the effect and the significance of the effect of calw_pct when meal _pct was added to the model? (What was the difference for calw_pct between baseplus1 and baseplus2?)

Answer

The effect got smaller and it was no longer significant.

Question 7: Question

Question

There are two related reasons this change from baseplus1 to baseplus2 might have happened based on the relationship between calw_pct and meal_pct, one based in theory and one based in statistics. What are they? You can summarize the statistical reason with a sentence like "calw_pct and meal_pct are highly _____."

Question 7: Answer

Question

From a theory standpoint, both welfare participation and reduced price school lunches are based on family income or economic status. From a statistical perspective, they are highly correlated. *The statistical effect should be expected given the theoretical relationship.*

Question 8

Question

Multicollinearity is often a result of too little "free" _____ in the explanatory variables in a small sample.

Answer

variation

See Section 3.4 in Wooldridge and Section D: Multicollinearity Redux in the worksheet. This will be a nontrivial point for many of you in your own research.

Question 10

The three counties in the subset

- Los Angeles
- Orange
- Ventura

Question Extra Credit

Question

From the worksheet: Line 44 provides a summary of the residual from the 'final' regression. Obviously the mean residual should equal 0, but when you type line 45, R also tells you the standard deviation of the residual which is slightly smaller than the Residual standard error; do you know why?

Answer

"The formulas for RSE and SD of resid are almost equal, except that the denominator inside the root for the SD formula is $n-k$ instead of $n-2$ as in the RSE formula." - Someone in the class

Script issue 1

Had to install package "here."

The code for this was in the updated week 3 script but is "commented" out. It's only necessary to do this one time, so once lab 3 is run with that line uncommented once, you don't need it anymore.

Script issue 2

"I also had to attach the data, which the script didn't show."

Attaching is actually not the preferred method. Instead directly call the dataframe object, for example:

- **cadta**\$somevariablename
- `base <- lm(testscr ~ avginc + el_pct, cadta)`
- `base <- lm(testscr ~ avginc + el_pct, data = cadta)`

Things to look for

- Multiple ways to compute the same statistics
- How to address specific model elements
- Continue addressing specific rows and columns
- Subsetting to eliminate NA results

The End