

Project 1: Ambulatory Care Medical Data

McCourt School of Public Policy, Georgetown University

Week 3: Postestimation commands

Key Ideas:

- test, lincom
- postestimation
- estimation results
- outreg2

Overview

- This week, it is very important to save all project files to a dedicated folder
- We will be creating several output files this week, and they will save to your working directory, wherever that may be
- If you're unsure about the location of your working directory, please ask.

Questions

1.21 Review Question

- Create a new indicator variable for patients who are male.
- Create a second indicator variable, `male_overwt` for patients who are both male and overweight.
- Hint: try to create `male_overwt` using multiplication, not if-statements.
- Run a regression of systolic blood pressure on male, overwt, and male_overwt.
- Test the null hypothesis that overweight and male_overwt are jointly equal to zero.
- Test the null hypothesis that the coefficient on overwt is equal to 2 times the coefficient on male.

1.22 Update Sample

- It's a fact of life, we all make mistakes
- Sometimes you make a choice at the beginning of your analysis that you later regret
- For example, in week 1 we limited our sample to patients age 18 and older
- We have just realized that we want to look at patients age 16 and older
- Fortunately, this mistake is easy to fix.
- Just go back in your do-file and change the relevant line of code.
- Rerun the entire do-file to make sure all of the variables you created are up-to-date with the new sample.
- Verify that your new sample has 3,965 observations.

1.23 Replay Estimates

- The next question is going to require you to run many postestimation tests.
- Sometimes when you're running these tests, you need to review the regression results, but they're already gone from the review window.
- You can "replay" the last estimation results by retyping the command with no variables.
- Type `regress` to re-display the most recent regression results.

1.24 Linear Combinations of Coefficients

- Run the following regression: `regress bpsys current_tobac male wtlb age`
- Suppose we want to know the change in predicted blood pressure for a male tobacco user
- We could just calculate the sum of the coefficients: `display 3.00 + -1.10`
- Or: `display "The combined effect is: " _b[current_tobac] + _b[male]`
- We could also test if the combined change is different from zero: `test current_tobac + male = 0`
- But, what if we want to know the confidence interval around the combination?
- We need to calculate statistics for a "linear combination" of coefficients: `lincom current_tobac + male`
- How do the statistics displayed after `lincom` compare with the previous output from `test` and `display`.
- What is the total change in predicted bpsys from one additional pound of weight and one additional year of age?
- What is the total change from the "freshman 15"? (15 additional pounds of weight and one additional year of age)
- What is the combined change of being male and having 10 additional pounds of weight?
- Is the above combination statistically different from zero?
- What difference in weight would cause a tobacco user and a non-user to have the same predicted value of bpsys?

1.25 Other Postestimation Tests

- We have covered two postestimation commands so far: `test` and `lincom`
- In general, postestimation refers to commands that are run after an estimation command such as `regress`, and use the results from that command.
- You will be learning other estimation commands this semester, in addition to `regress`. (`probit`, `logit`)
- Each estimation command has a slightly different set of postestimation commands available.
- Look at the postestimation commands available after `regress` with: `help regress postestimation`
- You will be using a few of these postestimation commands throughout quant classes and thesis
- Try running the following tests, but don't worry about the meaning of the output for now.

```
hettest
linktest
ovtest
```

1.26 Predicted Values

- Another popular activity after estimation is generating predicted values.
- This can be done manually for parsimonious models (few variables).
- With larger models, the `predict` command is much easier.
- Both methods are demonstrated below:

```
regress bmi wtlb
gen pr_bmi_manual = _b[_cons] + _b[wtlb] * wtlb
predict pr_bmi , xb
twoway scatter bmi pr_bmi wtlb
```

- You can generate the predicted value of y using the `xb` option of the `predict` command, like above.
- Or you could generate residuals by changing the `xb` option to `residuals`
- Generate a new variable containing the residuals from the previous regression.
- Add the new variable to the previous twoway scatter plot. (Make sure "Weight" remains on the x-axis.)

- Summarize actual bmi, predicted bmi, and the residuals.
- Calculate the same summary statistics for the groups defined by every possible combination of `male` and `current_tobac` (hint: use `bysort`).
- Which group had the worst predictions for bmi? (Define "worst" as the largest absolute value of average residuals.)
- Was the average predicted value of bmi higher or lower than the actual average bmi for that group?

1.27 Table of Estimates

- Postestimation commands always apply to the most recently run estimation command
- That means there is only one set of estimates that are "active" at any time
- As you've seen, the active estimates can be "replayed" by typing the estimation command again without any variables
- You can also see the currently active results, exactly as Stata stores them, by typing `ereturn list`
- Some postestimation commands require multiple estimation results at the same time.
- For example, you can make a table out of multiple estimation results.
- In this case, you must first store each set of estimates with a different name using `estimates store`
- Then you can use the `estimates table` command, and list the names of your stored estimates.

```
regress bpsys current_tobac overwt
estimates store basemodel
regress bpsys current_tobac overwt overwt_current_tobac
estimates store bettermodel
regress bpsys current_tobac overwt overwt_current_tobac age age2 bmi_log
estimates store bestmodel
estimates table basemodel bettermodel bestmodel
```

- For more on storing and using estimates, see `help estimates`
- Create a table of estimates displaying coefficients from two or more of your own regressions.

1.28 Outreg2

- There are commands for Stata that take estimates and output them into nice-looking tables.
- One of the most popular is `outreg2`.
- `outreg2` is not an official Stata command, so you have to install it using the command: `ssc install outreg2`
- `ssc` is like Stata's App Store.
- Make sure `outreg2` is installed by looking for the help page: `help outreg2`
- Here is an example of putting regression results into a table using `outreg2`
- Note that the when the first set of estimates is added, the `replace` option is used.
- When subsequent estimates are added, the `append` option is used to add them to the same table.
- The option `excel` save the table in an Excel-readable format. Another option you can try is `word`.

```
regress bpsys overwt current_tobac mftall age age2 if sex==1
outreg2 using myestimates , replace excel
regress bpsys overwt current_tobac mftall age age2 if sex==2
outreg2 using myestimates , append excel
regress bpsys overwt current_tobac mftall age age2
outreg2 using myestimates , append excel
regress bpsys overwt male male_overwt current_tobac mftall age age2
outreg2 using myestimates , append excel
```

- The file, myestimates.xml, is now saved in your current working directory
- If you're on a Mac, you have to first open Excel, then open the myestimates.xml file from within Excel.
- Try to recreate the attached table using outreg2.
- Save the new table output as a new file by changing `myestimates` to a new file name.

1.29 Outreg2 challenge

- The second attached outreg2 table is much more fancy.
- Try to modify your outreg2 command to create the fancy table.
- Hint: You will have to label all relevant variables before the outreg2 command.