

Stata Lab 2 – Dummy, Heteroskedasticity, Specification Issues

1 Does "beauty" help increase wage?

1. Download the datafile "beauty.xlsx" from your EE425 Moodle page.
2. Open the STATA software program. Click on the "Data Editor" icon.
3. Copy the entire dataset from the excel file and paste it onto the STATA's Data Editor page.

Choose "Treat first row as variable names".

4. Save the new STATA dataset.

Choose File -> Save As -> (then name the new dataset "beauty_stata")

5. Open a new do-file and save it.

Choose "New Do-file Editor" icon

On the Do-file's top panel, choose File -> Save As -> (then name the new do-file "second_stata_lab")

• To explore and understand the data

6. type: sum
7. type: codebook
8. type: tab look
9. type: tab look female

• We want to test whether the fraction of above-average-looking women and men are the same

$$abvavg = \beta_1 female$$

$$H_0 : \beta_1 = 0.5$$

$$H_0 : \beta_1 \neq 0.5$$

10. type: gen abvavg = 0
11. type: replace abvavg = 1 if look > 3
12. type: regress abvavg female, nocon

13. Can we reject H_0 ?

- We want to test whether "good look" has a positive impact on wage

14. type: `gen belavg = 0`

15. type: `replace belavg = 1 if look < 3`

16. type: `gen log_wage = log(wage)`

17. `regress log_wage belavg abvavg`

- Seems like we may have the omitted variable bias. Let's take into account other variables.

18. type: `regress log_wage abvavg belavg educ`

19. type: `regress log_wage abvavg belavg educ exper expersq`

20. type: `regress log_wage abvavg belavg educ exper expersq bigcity`

21. type: `regress log_wage abvavg belavg educ exper expersq bigcity black`

22. type: `regress log_wage abvavg belavg educ exper expersq bigcity black union`

23. type: `regress log_wage abvavg belavg educ exper expersq bigcity black union female`

- Check if we have the heteroskedasticity problem. (Let's use the White Test (special case))

24. type: `regress log_wage abvavg belavg educ exper expersq bigcity black union female`

25. type: `predict u_hat, resid`

26. type: `predict y_hat, xb`

27. type: `gen u_hat_sq = u_hat^2`

28. type: `gen y_hat_sq = y_hat^2`

29. type: `regress u_hat_sq y_hat y_hat_sq`

30. Calculate $LM = nR^2$

31. Do we reject H_0 : homoskedasticity at 5% level of confidence?

32. Now, try using the ready-made test by STATA
33. type: regress log_wage abvavg belavg educ exper expersq bigcity black union female
34. type: estat hettest
35. Do we reject H_0 : homoskedasticity at 5% level of confidence?

- **Should we believe that the value of β are the same for female and male?** (Chow Test)

- Chow statistic is a type of F-statistic $F = \frac{SSR_p - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \cdot \frac{[n - 2(k + 1)]}{k + 1}$

36. To get SSR_p : regress log_wage abvavg belavg educ exper expersq bigcity black union
37. To get SSR_1 : regress log_wage abvavg belavg educ exper expersq bigcity black union if female == 0
38. To get SSR_2 : regress log_wage abvavg belavg educ exper expersq bigcity black union if female == 1
39. What is the value of the F-statistic? Can we reject H_0 (can use the same model)?

2 Fixing Heteroskedasticity

1. Download the "GPA1.DTA" dataset from your EE425 Moodle page and open it in STATA.

Choose File -> Open -> (then direct to the location of the file)

2. type: des
3. type: regress colGPA hsGPA ACT skipped PC

- **Check if we have the heteroskedasticity problem. (Let's use the White Test (special case))**

4. type: predict u_hat, resid
5. type: predict y_hat, xb
6. type: gen u_hat_sq = u_hat^2

7. type: `gen y_hat_sq = y_hat^2`
8. type: `regress u_hat_sq y_hat y_hat_sq`
9. Calculate $LM = nR^2$
10. Do we reject H_0 : homoskedasticity at 5% level of confidence?
 - **Now we think the fitted value in #8. is a reasonable candidate for $\hat{h}_i$** So,
11. type: `predict h_hat, xb`
12. To check if $\hat{h}_i$ are all positive type: `sum h_hat`
13. type: `gen sqrt_h = h_hat^0.5`
14. type: `gen wcolGPA = colGPA/sqrt_h`
15. type: `gen whsGPA = hsGPA/sqrt_h`
16. type: `gen wACT = ACT/sqrt_h`
17. type: `gen wskipped = skipped/sqrt_h`
18. type: `gen wPC = PC/sqrt_h`
19. type: `gen w = 1/sqrt_h`
20. type: `regress colGPA hsGPA ACT skipped PC`
21. type: `regress wcolGPA w whsGPA wACT wskipped wPC, nocon`
22. type: `regress wcolGPA w whsGPA wACT wskipped wPC, nocon robust`

3 Labor Force Participation of Female

1. Download the "MORA.DTA" dataset from your EE425 Moodle page and open it in STATA.
Choose File -> Open -> (then direct to the location of the file)
2. type: `des`
3. type: `regress inlf nwifeinc educ exper expersq age kidslt6 kidsge6`
4. type: `estat hettest`
5. type: `regress inlf nwifeinc educ exper expersq age kidslt6 kidsge6, robust`
6. type: `predict y_hat, xb`
7. type: `twoway scatter inlf educ || line y_hat educ`
8. type: `sort educ`

9. type: twoway scatter inlf educ || line y_hat educ

- **We need to "hold other things constant". Suppose nwifeinc = 30, exper = 10, age = 35, kidslt6 = 0, kidsage6 = 0.**

10. type: gen y_new = 0.585 + 30*(-0.0034) + educ*(0.0379) + 10*(0.0395) + 100*(-0.0006) + 35*(-0.0161)

11. type: twoway scatter inlf educ || line y_new educ