ECON 626: Empirical Microeconomics

Power simulation

Department of Economics
University of Maryland
Fall 2019

The objective of this activity is to simulate power, and to confirm the empirically simulated power using the analytical formula and the convenience commands.

1. Start a Stata .do file with these lines:

```
clear all
set obs 1600
gen treatment=cond(_n>=800,1,0)
gen betahat=.
gen se=.
gen tstat=.
gen pval=.
```

   This creates a dataset of 1600 observations; the `treatment` variable has the value 1 for half these observations, and 0 for the rest. The `pval`, `tstat`, and `betahat` variables will be used to store the results from the simulation.

2. Write a loop that, for 100 or even 1000 iterations, does the following things:

   - creates (or replaces) a variable `epsilon` with (new) draws from the standard normal distribution (mean zero, variance one).
   - creates (or replaces) a variable `y` so that `y = 0.15*treatment + epsilon`
   - regresses `y` on `treatment`
   - Stores the resulting estimated treatment coefficient, standard error, t-statistic, and p-value to one of the rows of the dataset. One way to do this is to use commands such as these:

```
matrix theseresults=r(table)
qui replace betahat=theseresults[1,1] in ‘iteration’
qui replace se=theseresults[2,1] in ‘iteration’
qui replace tstat=theseresults[3,1] in ‘iteration’
qui replace pval=theseresults[4,1] in ‘iteration’
```

3. After this loop, summarize the four variables you created. In what fraction of cases was the p-value less than 0.05? Are the results what you would expect? Compare to the calculations below:

4. The expected standard error is the standard deviation of the residual divided by (the square root of the sample size times the square root of $p(1-p)$), where p is the fraction of observations treated. How close is this to the typical simulated standard error?

5. The power to reject the null (considering $p < 0.05$ to be rejection) in the presence of the true effect of 0.15 is the mass of the normal distribution to the left of (the true effect divided by the (anticipated) standard error minus the critical value of 1.96). What is this power?

6. Of the p-values you recorded, what fraction are below 0.05?

7. Try the command `power twomeans 0 0.15, n(1600)` - what does it tell you?

8. Try the command `sampsi 0 0.15, sd(1) power(XXXXX)` where `XXXXX` is the power you calculated in one of the previous parts. The result should agree with the rest of the exercise.