Homework II

Significance Testing, Imputation, and Simulations

Fernando Rios-Avila

Instructions

Submit a document containing your answers to the following questions and a do file with the code used in `Stata` to produce the answers. Ensure your program works as submitted and produces the reported answers for full credit.

Part I: Significance Testing (15pts)

In this section, use simulation to demonstrate: - How rejecting the null hypothesis probability increases with the number of tests. - Benferroni correction application to control for multiple testing. - Joint testing usage to control for multiple testing.

Tasks

Refer to the file `hw2.do` for a template program simulating data for N individuals and K variables. Modify this program to:

1. Test the null hypothesis that the mean of each variable is zero with a significance level of 10%. Indicate if you reject the null for any variable.
2. Repeat the test in 1 using a modified significance level of 10%/K (Benferroni correction). Indicate if you reject the null for any variable.
3. Conduct a joint test that the mean of all K variables is zero, using a 10% significance level.

Repeat the above tasks 1000 times and report the proportion of times you reject the null hypothesis in each case (summary tables). Provide an explanation for your results.

For the parameters K and N, use the following combinations:

- K = 2, N = 20
- K = 5, N = 20

- K = 2, N = 100
- K = 5, N = 100

Using the code, you need to change the parameters K and N to run each simulation. Create a table with the results for each combination of K and N, and explain.

Part II: Imputation and Statistical Matching (30pts)

Imputation and Statistical Matching are methods for handling missing data. Use any of these methods on the dataset `cps_imput_miss.dta`, to address the problem of missing wages.

The dataset focuses on couple-households with or without children (below 15 years old) and no other family members. Their only source of income comes from wages.

The dataset includes variables ending in `_1` for the husband and `_2` for the wife. Household-level variables and the variable `cutoff` indicating the poverty line are also provided

The problem

The problem on this dataset is that in 50% of the households, husband's wages are missing. And we are interested in estimating the poverty rate and for the sample.

Tasks

- Impute missing wages for husbands and estimate the poverty rate for the sample.
-
- Write a report explaining the steps you have taken for imputation, including model specification. Provide a table with poverty rates for the entire sample, by husband education level, and race. Offer a brief explanation of the results.

  Note: Include a do-file with the code for imputation and summary statistics.

Part III: Micro-Simulations (55pts)

Micro-simulations are a valuable tool for studying the effects of policy changes. In this section, use micro-simulations to explore the impact of an Employer of Last Resort (ELR) program on the labor market of Tanga-Mandapio, a small Pacific country similar to the United States.

Considerations

- The ELR program guarantees a job to anyone between 18 and 65 years old who is not employed.
- The program offers different wages based on education levels, based on the following schedule:

Table 1: Wages per hour offered by the ELR program

Education	Wage
Less than High School	10
High School	15
Some College	20
College	25
Graduate School	30

- All jobs offered by the program are full time (40 hours per week), full year (48 weeks per year).

- There is a budget of 100 Million dollars per year for the program.

Tasks

Use the following steps to simulate the impact of the ELR program:

- Determine eligibility for the ELR program.
- Estimate a logit model to predict employment probability considering only those employed and those who are eligible for the program.
- Use model predictions to assign jobs until the budget is exhausted, prioritizing those with higher employment probability.
- Recalculate total household income accounting for new jobs.

Write a report explaining the simulation steps, including model specification, and discuss the program's impact:

- How many people are eligible for the program by education level?
- What would the the budget requirement be to guarantee a job to all unemployed individuals.
- Jobs created, impact on the unemployment rate by education.
- Impact on poverty and inequality using `cutoff` and Income Percapita (`hhincome`/`hhsize`).

Provide summary statistics for poverty before and after the simulation, by gender and race.

Note: Include a do-file with the code for the simulation and summary statistics.

Remarks

- The data file is tanga_mandapio. Assume each person has a weight of 1.
- `hhincome` is total household income, and `incwage` is individual wage income.
- `cutoff` is the poverty line for a household.
- `educ_g` is the aggregated education level.
- Explore the dataset for a better understanding of available variables.

Hints on How to proceed

1. Create a dummy `emp` that is equal to 1 if employed and 0 otherwise only for people age = 18 to 65.
2. Estimate your logit model using `emp` as dependent variable, as a function of characteristics.
3. Using predicted Probabilities, assign jobs to those eligible for the program, using those with highest predicted probability first.

   1. First for the unemployed, sort data by predicted probability (highest to lowest) and assign jobs until the budget is exhausted.(each persons anual wage is **wage**x40x48, where wages depends on his education level)

4. To recalculate total household income, you can simply create `gen new_hhincome = hhincome + new_wage_elr` where `new_wage_elr` the sum of ELR wages for each household. Some households may have more than one person employed by the ELR program.
5. Households are identified with the variable `serial`.