Development Economics   Spring 2010

Stata Exam   May 20, 2010

Instructions

1. Please do not begin until given permission.

2. This exam is a closed-book exam. The only reference you can use during the exam
    is the help function inside of Stata. You may use an English dictionary (either
    electronic or paper-based one). You may also use a calculator.

3. There are 10 questions. The points allocated to each question are shown beside the
    question number. The full score is 34.

4. Type your name and student ID in the first line of your do file using * (Remember
    that Stata ignores the line starting with *) or whatever appropriate codes.

5. You need to turn in your *txt* file as well as *do* file at the end of the exam. **The name
    of your *do* file must be exam.do. The name of your *txt* file must be exam.txt.**
    Your *txt* file must be created from your *do* file by using the commands *do* and
    translate. In creating your final *smcl* file (to be translated into your final *txt* file),
    you must literally use the command *do* to run your completed *do* file. Repeatedly
    highlighting a part of a *do* file and creating a final *smcl* file is NOT allowed.

6. You need to write Stata commands that implement the tasks given in the questions.
    You must write your commands in a *do* file in which you must clearly show
    question numbers using *.

7. Please reduce error commands as much as possible in your *do* file. If your *do* file
    contains error commands that can easily be removed, your credits will be lowered
    substantially.

8. If you need a few steps to obtain the requested answer, you must show ALL of your
    work.

9. Some questions are sequential, meaning that, for example, question 6 uses answer
    to question 3. If you cannot answer question 3 but you can answer question 6, type
    your answer to question 6 using *. If your answer to question 6 is correct, I will
    give you the full credit to question 6 no matter whether you are correct in question
    3 or not.

10. **Please work on the desktop of the computer assigned to you. The default
    working directory is set to the desktop of your computer when you open Stata.
    Please do not change the working directory.**

11. You must stop working when you are asked to stop. Literally, no typing any more.
    You must have finished converting your *smcl* file into a *txt* file by the end time of
    the exam.

12. Please neither log out nor shut down your computer when you leave the room. The instructor will shut down your computer after confirming that your *do* and *txt* files have been collected.

13. All students cannot take this exam simultaneously. To make the exam fair to everybody, you must not discuss the contents of this exam with other classmates until the second session of the exam is over.

14. You must follow the above 13 rules. Breaking any single rule automatically reduces your credits to zero as well as potential disciplinary actions taken by the Dean's office.

   You must agree to the above requirements of the exam and sign below.

   Print your student ID number:

   Print your name:

   Print today's date:

   Sign:

Advice: Please save your *do* file as often as possible. Every year there is one or more students who lose their *do* files because they close their *do* files without saving them.

We discussed in class the intra-household allocation of resources. In particular, we learned that there has been accumulated evidence of unequal intra-household allocation of resources against girls in favor of boys in many Asian countries. In this exam, we are interested in examining the hypothesis of preferential allocation of household resources to boys rather than girls, using data from Indonesia. Deaton (1989) proposed an ingenious method of applying household-level expenditure data to the examination of the intra-household resource allocation. In this exam, you are supposed to do some data-preparation work and run an OLS regression that is relevant to the research question.

The data file, data.dta, contains all data necessary to complete this exam. This compiled data set is a part of data collected for the Indonesia Family Life Survey. Survey interviews were fielded in 2000. The data file, data.dta, is household-level (each row represents a single household).

There are 15 variables in data.dta. Below is a description of each variable.
- *hhid00* is household id number.
- *island* is the island group in which the household is located (one of the five island groups: either Sumatra, Java, Nusa Tenggara, Kalimantan, or Sulawesi).
- *urban* is a dummy of residence location, which is equal to one if the household is located in an urban area and zero if the household is located in a rural area.
- *lnhhsize* is logarithm of household size where the base is e.
- *wm0_6* is the ratio of the number of <u>male</u> household members whose <u>ages are 0 or older but less than 6</u>, to the household size.
- *wf0_6* is the ratio of the number of <u>female</u> household members whose <u>ages are 0 or older but less than 6</u>, to the household size.
- *wm6_19* is the ratio of the number of <u>male</u> household members whose <u>ages are 6 or older but less than 19</u>, to the household size.
- *wf6_19* is the ratio of the number of <u>female</u> household members whose <u>ages are 6 or older but less than 19</u>, to the household size.
- *wm19_60* is the ratio of the number of <u>male</u> household members whose <u>ages are 19 or older but less than 60</u>, to the household size.
- *wf19_60* is the ratio of the number of <u>female</u> household members whose <u>ages are 19 or older but less than 60</u>, to the household size.
- *wm60p* is the ratio of the number of <u>male</u> household members whose <u>ages are 60 or older</u>, to the household size.

- *wf60p* is the ratio of the number of <u>female</u> household members whose <u>ages are 60 or older</u>, to the household size.
- *rpce_pl* is real monthly per-capita household expenditure at 2000 Jakarta prices.
- *waltb* is the ratio of the monthly expenditure on alcohol and tobacco, to the total monthly household expenditure.
- *wfdout* is the ratio of the monthly expenditure on food out of home, to the total monthly household expenditure.

1) 2 points

There should be no irrelevant values in any variable in this data set. Check the mean, minimum, and maximum of each variable, to confirm that there are no obvious irrelevant values (such as -999 or -99) in the data set. Just write a Stata command that shows you the mean, minimum, and maximum of each variable. (*hhid00* is id variable for each household. You do not need to worry about irrelevant values for this id variable.)

2) 2 points

Create a dummy variable named *no_prime_female* which is equal to unity if the household contains no females whose ages are 19 or older but less than 60, and is equal to zero otherwise (i.e. if the household contains at least one female member whose age is 19 or older but less than 60).

3) 3 points

Calculate the mean ratio of the monthly expenditure on food out of home, to the total monthly household expenditure, separately for the two types of households defined in 2) above. Which type of households has a higher mean ratio of the monthly expenditure on food out of home, to the total monthly household expenditure? (Type your answer using * or whatever appropriate symbols.)

4) 2 points

Create a new variable named *lnrpce_pl* which is logarithm of real monthly per-capita household expenditure. The base of logarithm should be e.

5) 5 points

Next, create 10 regional dummies as follows. There are five island groups one of which

each household belongs to. Create 10 regional dummies separately for island group and urban/rural. For example, one regional dummy should be unity if the household is located in urban Sumatra and zero otherwise. Another dummy should be unity if the household is located in rural Sumatra and zero otherwise. The names of the dummies are *usumatra* (for urban Sumatra), *rsumatra* (for rural Sumatra) *ujava* (for urban Java), *rjava* (for rural Java), *utenggara* (for urban Nusa Tenggara), *rtenggara* (for rural Nusa Tenggara), *ukalimantan* (for urban Kalimantan), *rkalimantan* (for rural Kalimantan), *usulawesi* (for urban Sulawesi), and *rsulawesi* (for rural Sulawesi).

6) 5 points
Use a mean comparison test to examine whether the mean difference between urban and rural areas in the ratio of the household expenditure on alcohol and tobacco, to total household expenditure is statistically significant or not. Here, the null hypothesis is $H_0: \mu_{urban} = \mu_{rural}$ and the alternative hypothesis is $H_A: \mu_{urban} \neq \mu_{rural}$. Can you reject the null hypothesis against the alternative hypothesis at the 1 % level of significance? Justify your answer using the p-value. (Type your answer using * or whatever appropriate symbols.)

7) 5 points
Following Deaton (1989), we run the following OLS regression.
$$
\begin{aligned}
waltb = \beta_0 &+ \beta_1 lnrpce\_pl + \beta_2 lnhhsize + \beta_3 wm0\_6 + \beta_4 wf0\_6 + \beta_5 wm6\_19 \\
&+ \beta_6 wf6\_19 + \beta_7 wf19\_60 + \beta_8 wm60p + \beta_9 wf60p + \beta_{10} usumatra \\
&+ \beta_{11} rsumatra + \beta_{12} rjava + \beta_{13} utenggara + \beta_{14} rtenggara \\
&+ \beta_{15} ukalimantan + \beta_{16} rkalimantan + \beta_{17} usulawesi + \beta_{18} rsulawesi \\
&+ \varepsilon
\end{aligned}
$$
where wm19_60 and ujava are the reference groups.
Explain the idea behind this regression model. How can we examine the issue of the intra-household allocation of resources with this regression model? (Type your answer using * or whatever appropriate symbols.)

8) 2 points
Run an OLS regression on the model shown in 7).

9) 3 points
Carefully interpret the OLS coefficient $\hat{\beta}_{18}$. You must discuss the magnitude of the coefficient estimate. If you cannot run a regression in 8) above, use $\hat{\beta}_{18} = 0.02$ for this

question. (Type your answer using * or whatever appropriate symbols.)

10) 5 points

Do you find evidence of unequal allocation of household resources against girls in favor of boys? Justify your answer. You do not need to run econometric tests. You can base your argument on the signs and magnitudes of the coefficient estimates. (Type your answer using * or whatever appropriate symbols.)

(End of Stata Exam 2010)