

Final Exam: Due on 1:45pm Apr 24, 2008

Very Important:

- 1) You should turn in your final exam solution ON TIME (on or before 1:45pm Apr 24, 2008). NO LATE WORK WILL BE ACCEPTED. You may either turn in your solution to my mailbox in Bellamy 601, or hand in to me in person in Bellamy 610.
- 2) You should work on your final exam independently, and produce your own work to be handed in.
- 3) You are encouraged to submit your solution in paper copy. However, if you do submit your works electronically by using email, you should have my confirmation email indicating that I do receive your work; otherwise no score will be given for this final exam.
- 4) Please send me email ONLY when you don't understand what I want you to do (rather than asking me how to do it).

Final Exam Project:

Part I: Prepare Wave I data:

Download dataset transition1 from <http://cdph.fsu.edu/sya6933/transyt1.dta> , and do the following:

- 1). keep only these variables: newpid a1t1 a5t1 a22t1 a23t1 f1t1-f11t1 c1t1-c14t1
- 2). Rename, recode, and label:
 - a) Rename “newpid” to “id”.
 - b) Rename “a1t1” to “sex”, and recode “sex” in a way such that 0=female and 1=male. Please also do the label define and label value job, so that if you tabulate variable “sex” you will get below output:

```
. tab sex
```

gender	Freq.	Percent	Cum.
female	848	47.03	47.03
male	955	52.97	100.00
Total	1,803	100.00	

- c) Rename “a5t1” to “race”, recode “race” and do the label define and label values in a way such that 1=non-Hispanic White; 2=Hispanic White or Black; 3=African-American; and 4=other. For example, if you tabulate variable “race”, you should have below output:

```
. tabulate race
```

racial/ethnic identity	Freq.	Percent	Cum.
non-hispanic white	460	25.54	25.54
hispanic white or black	850	47.20	72.74
african-american	434	24.10	96.84
other	57	3.16	100.00
Total	1,801	100.00	

- d) Rename “a22t1” to “service1”;
 - e) Rename “a23t1” to “belief1”, also recode “belief1” in a way such that higher value indicating higher frequency. If you do not want to do it manually, you may want to find and install an add-on program that can be used to reverse variable value order. In either way, you should have below output if you tabulate “belief1”

```
. tab belief1
```

how often relig. used

to deal w/ daily problems	Freq.	Percent	Cum.
never	223	12.38	12.38
rarely	311	17.27	29.65
sometimes	440	24.43	54.08
often	422	23.43	77.51
always	405	22.49	100.00
Total	1,801	100.00	

- f) recode f1t1-f11t1 these eleven variables in a way such that 0="no", and 1="yes", rather than 5="no" and 1="yes" in its original form (tip: you may use the tip I provided on class to do these jobs in only one line of command, rather than recode these 11 variables one by one). Please also create a new variable "sub1" which is the sum of the values for f1t1-f11t1. For example:

```
. list sub1 f1t1-f11t1 in 1,nol
```

	sub1	f1t1	f2t1	f3t1	f4t1	f5t1	f6t1	f7t1	f8t1	f9t1	f10t1	f11t1
1.	2	0	0	0	1	0	0	1	0	0	0	0

- g) recode c1t1-c14t1 these fourteen variables in a way such that highest value indicating strongly agree, and lowest value indicating strongly disagree.
- h) create a new variable "p_sup1" (positive family support) which is the sum of the variables of c1t1 c2t1 c3t1 c7t1 c8t1 c9t1 c11t1 and c13t1, for example:

```
. list p_sup1 c1t1-c3t1 c7t1-c9t1 c11t1 c13t1 in 1,nol
```

	p_sup1	c1t1	c2t1	c3t1	c7t1	c8t1	c9t1	c11t1	c13t1
1.	37	5	4	4	5	5	5	5	4

- i) create a new variable "n_sup1" (negative family support) which is the sum of the variables of c4t1 c5t1 c6t1 c10t1 c12t1 c14t1, for example:

```
. list n_sup1 c4t1-c6t1 c10t1 c12t1 c14t1 in 1,nol
```

	n_sup1	c4t1	c5t1	c6t1	c10t1	c12t1	c14t1
1.	17	4	3	3	3	2	2

- j) keep only variables id sex race service1 belief1 p_sup1 n_sup1 sub1, and save it with a new data name "transition1.dta".

Please submit all syntax for above jobs, also a Stata output of the data summary by using command: "summarize", so that I know you have created a correct dataset.

Part II: Prepare Wave II data:

Download dataset transition2 from <http://cdph.fsu.edu/sya6933/transyt2.dta>, and do the following:

- 1). Keep only these variables: newpid a1t2 a2t2 a4t2 f1t2-f11t2 a16t2 a17t2 c1t2-c14t2

2). Rename, recode, and label:

- a) Rename "newpid" to "id".

- b) Rename “a2t2” to “birthd”, and use this birthday variable to create a new variable “age”. The age is calculated by March 26, 2001. For example if a person was born on May 50, 1978, then his/her age by March 26, 2001 should be 23. You may also assume there are 365.25 days a year, and if age is not an integer, you should do the round off (for example, round off 23.4 gives 23, and round off 23.5 gives 24).
- c) Rename “a16t2” to “service2”;
- d) Rename “a17t2” to “belief2”, also recode “belief2” in a way such that higher value indicating higher frequency, exactly the same as you did for “belief1” in Part I. You should have below output if you tabulate “belief2”:

```
. tab belief2
```

how often relig. used to deal w/ daily problems	Freq.	Percent	Cum.
never	160	13.29	13.29
rarely	203	16.86	30.15
sometimes	287	23.84	53.99
often	292	24.25	78.24
always	262	21.76	100.00
Total	1,204	100.00	

- e) Recode f1t2-f11t2 these eleven variables in a way such that 0=“no”, and 1=“yes”, rather than 5=“no” and 1=“yes”. Please also create a new variable “sub2” which is the sum of the values for f1t2-f11t2 like you did for “sub1” in Part I.
- f) Recode c1t2-c14t2 these fourteen variables in a way such that highest value indicating strongly agree, and lowest value indicating strongly disagree.
- g) Create a new variable “p_sup2” (positive family support) which is the sum of the variables of c1t2 c2t2 c3t2 c7t2 c8t2 c9t2 c11t2 and c13t2, like you did for “p_sup1” in Part I.
- h) Create a new variable “n_sup2” (negative family support) which is the sum of the variables of c4t2 c5t2 c6t2 c10t2 c12t2 c14t2, like you did for “n_sup1” in Part I.
- i) Keep only variables id age2 service2 belief2 p_sup2 n_sup2 sub2, and save it with a new data name “transition2.dta”.

Please submit all syntax for above jobs, also a Stata output of the data summary by using command: “summarize”, so that I know you have created a correct dataset.

Part III: Merge data:

- 1) Please merge data transition1 with data transition2 by using linking variable “id”, you should keep only those observations from both dataset. If you did it correctly, you final sample should be 1204.
- 2) Drop variable “_merge”, and create a new variable “age1” by subtracting 3 from variable “age2”. For example, if a person has a value 23 for “age2”, then his/her “age1” should be 20.
- 3) Order you variable list by using command:

```
. order id sex race age1 age2 service* belief* p_sup* n_sup* sub*
```
- 4) Reshape you data to long form in a way such that each person has two observations (for two waves), and each observation has records for all variables. For example, after reshape your data, if you use command `. list`, you may have below output in Stata:

```
. list in 1/6,nol
```

```
+-----+
| id  wave  sex  race  age  service  belief  p_sup  n_sup  sub  |
+-----+
```

1.	1	1	1	3	20	3	5	37	17	2
2.	1	2	1	3	23	2	5	38	9	2
3.	5	1	1	3	21	2	5	37	20	1
4.	5	2	1	3	24	2	4	33	23	1
5.	7	1	1	3	20	4	3	40	13	1

6.	7	2	1	3	23	4	5	39	15	1

Please submit all syntax for above jobs, also a Stata output of your final data list in first 10 observations (commands: list in 1/10).

Part IV: Longitudinal/Panel Data Analysis:

- a) Use variables “sex” “race” “age” “service” “belief” “p_sup” and “n_sup” to predict “sub”, please do the xtreg by fitting: between-effects model, fixed-effects model, and random-effects model. Please submit all commands for 3 models and INTERPRET below results from all three models:
- 1) Relevant number of observations/groups (depends on which model you fit);
 - 2) Relevant R-sq;
 - 3) The F or Chi2 statistic tests: for example what is the p-value for the test, and what does it mean;
 - 4) Those coefficients that is significant at level .05.
- b) Please also fit the random-effects model by using the maximum-likelihood estimator AND population-averaged estimator.
- c) Finally, please produce a table to list those coefficients (only significant at level .05) and standard errors from GLS (xtreg, re) estimator, Maximum-likelihood estimator, and Population-averaged estimator, and compare the coefficients and SE difference. The format of the table may look like (please keep 4 decimal places):

Estimator		Coefficient	SE
Sex	GLS		
	ML		
	PA		
Race			
...(those variables that are significant)			