

Lab 5: making a table using several decennial censuses from IPUMS

Puzzle: How many blacks migrated from the south to the north from 1910 to 1960? How would we find this out?

In this lab we will use a Stata ado file to learn how to quickly make a table with information from several decennial censuses. The table we are making for this lab can be checked with published census data. However, the same procedure could be used to generate tables that do not exist in published form.

What is an ado file? It is a “new” stata command that we have created. Explain:

Example:

```
-----myado.ado-----
```

```
program define myado
local j=1
while `j'<10 {
    di “I wrote a do-file ” `j’
    local j=`j'+1
}
```

```
end
```

In stata, type: myado

Explain why this is useful when using several decennial censuses:

For this lab, I wrote a simple do file that will get the data we need from each census 1910-1960, aggregate it, and save it so we can use it.

Question: how many southern blacks migrated to the north during each decade from 1910-1960?

Answer: we can answer this question using the IPUMS data, aggregated into the tables created by cohort.do, cohort.ado, analyze.do and some basic information on age-specific mortality rates

(<http://www.demog.berkeley.edu/wilmoth/mortality/States/ssa/life.tables/utper.lt.5x10>).

[The historical mortality rate data I found is not race specific, but we will ignore this for the moment.]

Table 2:
 Number of (lifetime) migrants from south to north alive by
 census decade:

. tab year mig [fw=n]

(mean) year	mig		Total
	0	1	
91	9444551	432516	9877067
92	9760750	743822	10504572
94	11400996	1414801	12815797
95	12849209	2191828	15041037
96	16273545	2627057	18900602
Total	59729051	7410024	67139075

Table 3:
 Population by age and census decade:

. tab age2 year [fw=n]

age2	(mean) year					Total
	91	92	94	95	96	
10	2805319	2730434	2834427	3661516	5652341	17684037
20	2149341	2253012	2614565	2549528	3244896	12811342
30	1944323	1995245	2341071	2561616	2390904	11233159
40	1283854	1483564	1968965	2248329	2452656	9437368
50	801013	1048709	1370825	1787076	2047043	7054666
60	515121	559959	870383	1111203	1530681	4587347
70	244593	284388	555296	741270	983906	2809453
80	103276	111313	201253	293714	471082	1180638
90	23176	30276	47367	74580	110661	286060
100	7051	7672	11645	12205	16432	55005
Total	9877067	10504572	12815797	15041037	18900602	67139075

Table 1, Migrant cohorts: Number of blacks born in the south currently living in the Midwest or northeast, by age.

Age category	Decennial Census					Total
	1910	1920	1940	1950	1960	
10	16878	60341	46183	149356	216329	489087
20	47363	88004	149385	236272	328962	849986
30	127960	220740	339970	588317	550392	1827379
40	123433	197918	402054	621623	701214	2046242
50	69773	123929	296829	491184	599793	1581508
60	42076	59134	161644	271827	423366	958047
70	18635	24228	85719	146649	239579	514810
80	8567	9686	31786	58329	100383	208751
90	1008	3531	5357	11644	18922	40462
100	1260	303	909	1988	2790	7250
Total	456953	787814	1519836	2577189	3181730	8523522

In class puzzle: How many blacks migrated from the south to the north between 1910 and 1960? Nicholas Lehman says that between 1910 and 1970 6.5 million blacks migrated from the south to the north (Lehman, 1991, The Promised Land, p.6).

Is that what our data shows?

Where did he get this number?

According to our data, how many blacks migrated from the south to the north (west, Midwest, + northeast) from 1910-1960?

Note: Take mortality into account using table 3

Lab 5 assignment:

- 1) (in groups) Finish the in-class demography puzzle: Figure out how many blacks migrated from the south to the north (west, northeast, & midwest) between 1910-1960.
- 2) Read cohort.do, cohort.ado, and analyze.do, annotate the commands you do not know, and turn in a copy. Do you understand what cohort.ado does? I would like you to understand what I have done in these programs because this process has broad applications in similar situations. I.e., you can apply this general model to make any type of table from the PUMS census data. I will go over the programs at the start of lab and answer any questions you have.
- 3) [You can do this before or during lab on Thursday]. Figure out how to generate a table similar to table 1 except just for black women. Change to file cohort.ado to make this change [hint: it is very easy, ask me if you get stuck], and run cohort.do to do it. Turn in stata output for table 1 (this is the last command of analyze.do).
- 4) The rest of lab is time for you to work on your projects. I will help you translate what you want to do into stata, and answer other questions. Projects are due next Monday.

cohort.do

```
prog drop _all  
clear
```

```
set obs 1  
gen zzz=1  
save cohort, replace
```

```
cohort tedmo039  
cohort tedmo038  
cohort tedmo032  
cohort tedmo034  
cohort tedmo035
```

```
use cohort  
drop if zzz==1
```

```
save, replace
```

```
do analyze
```

1910

1920



1940

1950

1960

runs
cohort.do

NOTE: [FW = perwt]
scaled up to whole pop.

Cohort.ado

```
program define cohort

global name `1'

clear
set more 1

! rm -f ${name}.dat

cd data
! cp -f ${name}.dat.Z ..
cd ..
! uncompress ${name}.dat.Z

infix using ${name}.dct

compress

do recode

* get some basic descriptives so we can check our results later

tab region2 [fw=perwt]

tab bregion [fw=perwt]

tab region2 bregion

tab region2 bregion [fw=perwt]

tab age2 [fw=perwt]

tab age2 bregion [fw=perwt]

* I use x to count everyone
gen x=1

* total population
tab x [fw=perwt]

collapse (mean) year (count) n=x [fw=perwt], by(age2 bregion region2)

sum
sum n
di _N*r(mean)
* check

tab bregion
tab bregion [fw=n]

save $name, replace

append using cohort
```

```
save cohort, replace
```

```
! rm -f ${name}.dat
```

```
end
```

recode.do

```
keep if raceg==2
tab year
```

```
gen samest=bplg==state
tab samest year
tab samest year [w=perwt], col nofreq
```

```
gen illinois=state==17
tab ill year
```

* notice how many people are living in the same state they were born

```
gen age2=autocode(age,10,0,100)
tab age2
```

```
gen occat=occc1950
```

```
recode occat 0/99=1 100/123=2 200/290=3 300/390=4 400/490=5 500/595=6 /*
*/ 600/690=7 700/790=8 800/840=9 900/970=10 980/998=11 999=.
```

* note: 999=not working (blank)

```
lab define occat 1 "professional and technical" 3 "managers" 4 "clerical" /*
*/ 2 "agriculture" 6 "craftsman" 5 "sales" 7 "operatives" 8 "service workers" /*
*/ 9 "farm laborers" 10 "laborers" 11 "misc"
```

```
lab val occat occat
```

```
gen region2=region
recode region2 11/13=1 21/23=2 31/33=3 41/43=4 91/99=9
lab define region2 1 "northeast" 2 "midwest" 3 "south" 4 "west" 9 "unknown"
lab val region2 region2
```

* we want to make a region2 variable for the place of birth

```
sort bplg
merge bplg using bpcodes
tab _merge
replace bpreion=. if _merge==1
label val bpreion region2
drop if _merge==2
```

analyze.do

```
clear
```

```
use cohort
```

```
tab year [fw=n]
```

```
gen mig=bpreregion==3 & (region2==2 | region2==1 | region2==4)
```

```
tab mig
```

```
tab mig [fw=n]
```

```
tab age2 year if mig==1 [fw=n]
```