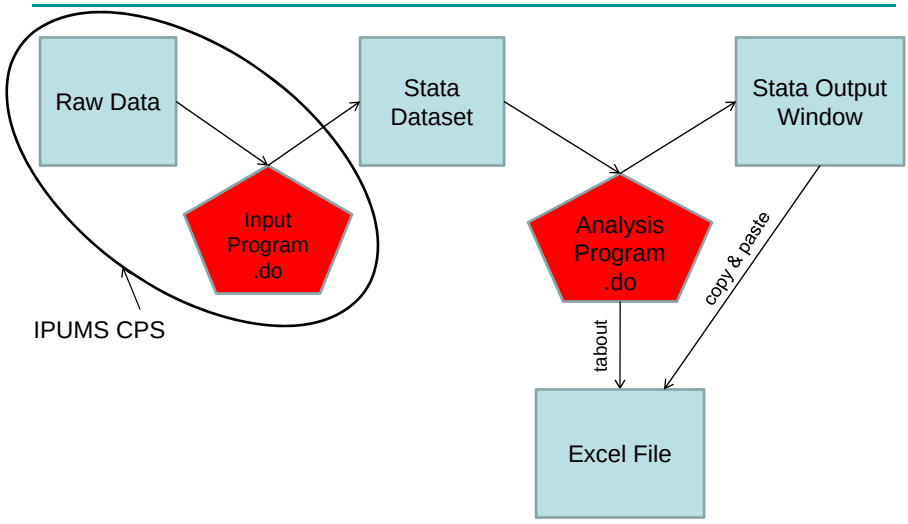
shadac
State Health Access Data Assistance Center
Bridging the gap between research and policy

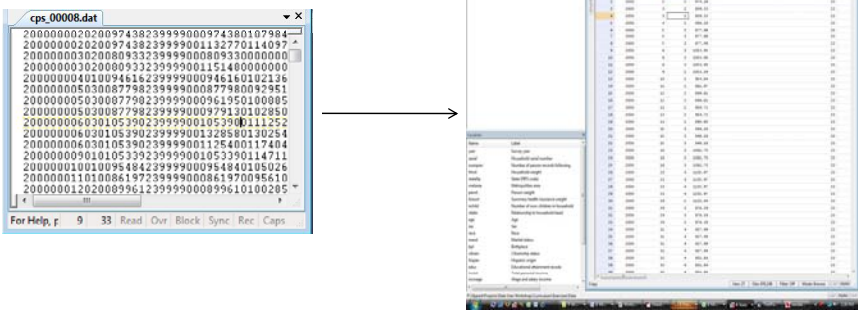
SHADAC Data User Workshop

October 22, 2010

Exercise Outline



From Raw Data to Stata Dataset



From Raw Data to Stata Dataset

- Need an input program
  - pretty time consuming to do manually
  - IPUMS automatically creates one when it exports your raw data
- Run program in stata
  - `infix` reads the fixed format of the raw data
  - `format` provides display formats for some variables
  - `label var` creates names for your variables
  - `label define` creates labels for variable values
  - `label value` assigns the label to the variable

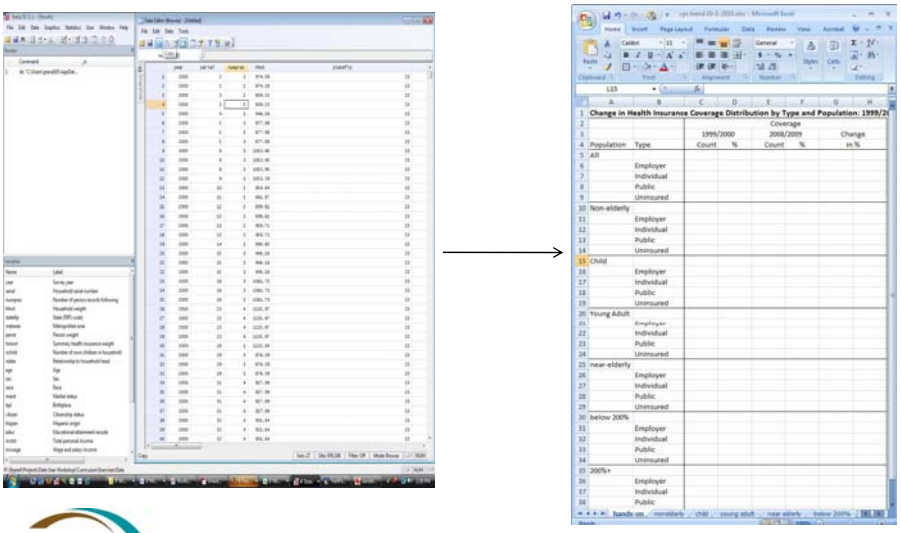
Recoding Variables

- Primary methods to create/recode a variable
  - if
    - `replace agecat=1 if age<=18`
  - and/ or
    - `replace agecat=1 if age>=19 & age<=64`
  - inlist
    - `replace hh_cat=1 if inlist(hhsize,1,2,3)==1`
    - `replace hh_cat=2 if inlist(hhsize,4,5,6)==1`
  - egen
    - `egen agecat=cut(age), at(0,19,100)`

Setting Complex Survey Design Info

- Create strata variable using Davern technique that gives unique code to MSAs and the remainder of the state
  - `gen geocode=.`
  - `replace geocode = metarea if metarea<9997`
  - `replace geocode = (statefip*10000) if geocode==.`
- Set the survey design information
  - `svyset [w=hinswt], psu (serial) strata (geocode)`

From a Stata Dataset to an Excel File



Copy and Paste

- Stata Output window will produce the results from a `svy: mean` or `svy: tabulate` command
- Using your mouse you can highlight the portion you want as use your favorate copy and paste method
  - (ctrl+c, ctrl+p), (Edit>Copy, Edit>Paste), etc.
- Then after doing this for years you can calculate the increased rate of wrist injuries for this method, or....use an output command

Tabout Introduction

- Developed as a user-written .ado file by Ian Watson and reviewed/approved by Stata
- Can be added to your stata by typing
  - `ssc install tabout`
- Many limitations for doing your whole table at once but flexible enough to do in multiple commands
- Unfortunately, additional outputs must be appended to the bottom of a file or output to a new file

Getting the statistic you want with tabout

- Percentage or a mean
  - the percentage will work for categorical variables like Race/ethnicity
  - the mean will work for a continuous variable like income
  - the mean can also work like the percentage if:
    - you have just 2 categories
    - they are coded 1 and 0 with 1 being the percentage you want
- Count or number of observations
  - weighted or unweighted frequency

Using Tabout to produce 1 year of data

- The most basic version
- `tabout hcovany using cps_percent.xls, cell(column) svy f(6) replace`
  - hcovany is the only row variable
  - `using` declares cps_percent.xls as the output file
  - `cell(column)` says the statistic in the cell should be the column percent

    `svy` says it should use the svyset information we set
  - `f(6)` says the column percents should have 6 decimals
  - `replace` says it should replace cps_percent.xls if it already exists (Let's run it and check the output)

Using Tabout to produce multiple years

- `tabout hcovany year using cps_percent.xls, cell(column) svy f(6) replace`
  - now we have added year to the list. Unless we specify `oneway`, tabout assumes we want year as our column variable
- `tabout hinsemp hinspur hcovpub hcovany year using cps_percent.xls, cell(column) svy f(6) replace`
  - now, we have multiple row variables and still have year as our column variable

Using Tabout to produce multiple years

- We just produced code for the percentages but now we need it for a subpopulation
- `tabout hinsemp hinspur hcovpub hcovany year if non_elderly==1 using cps_percent.xls, cell(column) svy f(6) replace h1(Subpopulation: Non_elderly==1)`
  - this code creates a subpopulation of just the non-elderly using `if` and we used `h1` to add a label (replacing our column variable name) to the top so we can see in our output file what the subpopulation it is

Using Tabout to produce multiple years

- Next we add our code to get standard errors for significance testing
- `tabout hinsemp hinspur hcovpub hcovany year if non_elderly==1 using cps_percent.xls, cell(column se) sebnone clab(% SE) svy f(6) replace h1(Subpopulation: Non_elderly==1)`
  - `se` adds the standard error statistic to the output
  - `sebnone` tells it not to put the SE values in brackets
  - `clab(% SE)` will give the column statistics labels
- this is all we need for the first block of our table

Using Tabout to produce multiple years

- Because it is uses the svy information, it cannot produce both the column percent and the count.
- `tabout hinsemp hinspur hcovpub hcovany year if non_elderly==1 using cps_percent.xls, cell(freq) svy f(1) replace h1(Subpopulation: Non_elderly==1)`
- `freq` requests the weighted frequency or count statistic, we don't need the SE code for this. We also only need 1 decimal for rounding
- We can calculate the difference and averages in Excel

Using Tabout to produce multiple years

- Now we need one of these line for each of our 6 subpopulations

- tabout hinsemp hinspur hcovpub hcovany year using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) replace h1(Subpopulation: All)
- tabout hinsemp hinspur hcovpub hcovany year if non_elderly==1 using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) append h1(Subpopulation: _Non_elderly==1)
- tabout hinsemp hinspur hcovpub hcovany year if children==1 using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) append h1(Subpopulation: children==1)
- tabout hinsemp hinspur hcovpub hcovany year if young_adult==1 using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) append h1(Subpopulation: young_adult==1)
- tabout hinsemp hinspur hcovpub hcovany year if near_elderly==1 using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) append h1(Subpopulation: near_elderly==1)
- tabout hinsemp hinspur hcovpub hcovany year if under_200hhspov==1 using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) append h1(Subpopulation: under 200hhspov==1)
- tabout hinsemp hinspur hcovpub hcovany year if under_200hhspov==0 using CPS_percent.xls, svy cell(col se) sebnone clab(% SE) f(6) append h1(Subpopulation: under_200hhspov==0)

Steps to create Excel file from tabout

1) Have a Final Table Layout in Excel
2) Paste output file from the tabout command into a new sheet (sheet1)
   - We have 2 files: CPS_count.xls and CPS_Percent.xls
3) Assess if you have the numbers you need for the table/or could calculate them easily with a formula.
   - We'll create an intermediate one because we have extra rows in ours
4) Copy and paste output file (again) into a new sheet (sheet2)
5) Create a relative formula (ie. "=sheet1!A1") in the second sheet that links to the first sheet and copy and paste this formula over the entire second sheet
6) Move the cells in the second sheet around to look like the final table
7) Copy and Paste any additional headers and formats to this table

Extra: HHS Poverty Guidelines

- SHADAC has created a file that allows you to merge on HHS poverty guidelines for analysis
- Requires Year, State, Family income, Family size
- http://www.ssa.gov/policy/docs/statcomps/supplement/2009/3e.html#table3.e8