

Section #6

In today's section, we will learn how to download data from the University of Minnesota's Integrated Public Use Microdata Series (IPUMS).

We will download similar variables used in "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," by W. Collins and M. Wanamaker. The authors use linked 1910 and 1930 Census data). As Collins and Wanamaker write in their paper:

"Our analysis focuses on the period from 1910 to 1930 for several reasons. Constructing a dataset that links individuals across census years requires full information on the names of individuals and access to the entire collection of handwritten census manuscripts. Until very recently, 1930 was the latest census year for which this was possible. Linking census data requires access to information (full name) that is confidential in later years. The 1940 census manuscripts have recently been released to the public. These should, in principle, be useful for characterizing the Depression's effects on worker outcomes... A significant limitation of studying [the 1910 to 1930] period is that the census did not collect individual-level income data before 1940." (p. 222)

The 1940 Census is an important census because it is the first Census to include education and income information. A 1% sample of the 1940 census has long been available, but as of December 2014, the 100% census was made available on IPUMS. While names are not publically available on IPUMS to all registered users, researchers can apply for access to names.

In today's exercise, we will work with the 1% cross-sectional sample for illustrative purposes. Because we are using the cross-sectional data, we will not be able to control for selection using previous wages. However, we will have one advantage over Collins and Wanamaker: we will use actual wages rather than the imputed occupational wages. We will rerun similar regression specifications using data from the cross-section, and compare results.

1 Part I: Extract 1940 Census Data from IPUMS

1. **Register.** If you don't already have an IPUMS account, you will need to register for an account. Go to <https://usa.ipums.org/usa-action/users/login>
2. **Select Sample.** Go to Select Data -> Select Samples. Unselect "Default samples from each year." Select "1940 1%". (Note: You can select the 100 % by selecting the top tab "USA Full Count.") Click "Submit sample selections" at the bottom.
3. **Select Variables.** We will need the following variables:
 - (a) Household->Geographic->STATEFIP
 - (b) Household->Geographic->COUNTY

- (c) Person->Demographic->SEX
 - (d) Person->Demographic->AGE
 - (e) Person->Race,Ethnicity and Nativity->RACE
 - (f) Person->Race,Ethnicity and Nativity->BPL
 - (g) Person->Education->SCHOOL
 - (h) Person->Education->HIGRADE
 - (i) Person->Work->EMPSTAT
 - (j) Person->Work->OCC1950
 - (k) Person->Work->IND1950
 - (l) Person->Work->CLASSWKR
 - (m) Person->Work->CLASSWKR
 - (n) Person->Work->HRSWORK1
 - (o) Person->Income->INCWAGE
 - (p) Add some of your own!
4. **Check Out.** Click “View Cart” in the top right corner. Double check you have all the variables you wanted. Click “Create data extract.”
5. **Select cases.** We don’t need all the data (even the 1% sample will have a very big file size.) Let’s restrict the sample in a similar way that the authors do. The authors’ sample is restricted to black male residents of southern states between the ages of 0 and 40.
- (a) Click “Select cases.”
 - (b) Select “Sex,” “Age,” “Race,” and’ “BPL.”
 - (c) Indicate using the menus that you want black males aged 0-40 from Southern States (Alabama, Arkansas, Florida, Georgia, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Texas, Virginia, and West Virginia.
 - (d) Click “Submit” to return to the “Extract Request” menu. On this menu, click “Submit extract”. This brings you to the “Download or Revise Extracts” page. Your data, and a Stata .do file to read the data into Stata, will appear here when the data are ready (usually takes a few minutes).
 - (e) **Download data and Stata file.** when ready. You will need to unzip the data. Then run the do file in Stata.

2 Part II: Regressions

Let’s run a regression similar to the regression in the paper. Specifically:

$$Y_{i,1940} = \alpha + \beta_1 M_{i,1940} + X_i \delta + u_i$$

where:

- $Y_{i,1940}$ is log of income. To construct this variable in Stata, first recode the not in universe values as missing as follows: “replace incwage = . if incwage== 999999 ”. Then create the main dependent variable of interest: “gen ln_incwage = ln(incwage) ”

- $M_{i,1940}$ is an indicator of migration before 1940. Construct this in Stata as: `genmigrated = (bpl! = statefip)`

Let's first run this regression with no controls. The syntax is as follows:

```
regress ln_incwage migrated
```

Next, let's include age as a control. We also might want separate dummies for every grade of education. We can easily do this in Stata with the following syntax:

```
regress ln_incwage migrated age i.higrade
```

Finally, let's include state of migration fixed effects. Let's also turn our linear age control above into age fixed effects. We do this in Stata as follows:

```
reg ln_incwage migrated i.age i.higrade i.bpl
```

Results for these three specifications are shown below:

Table 1: Main Regression-Log Earnings

	(1)	(2)	(3)
	ln_incwage	ln_incwage	ln_incwage
migrated	0.651*** (0.0157)	0.417*** (0.0154)	0.403*** (0.0153)
_cons	5.519*** (0.00897)	4.129*** (0.0418)	3.947*** (0.0763)
N	14813	14813	14813
Controls	No	Yes	Yes
State-of-Birth Fixed Effects	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1. Compare and contrast these results with the results in Collins and Wanamaker.