

Person- and Job-specific Characteristics of Employment

Goal

There are a number of employment characteristics that may interest researchers. This exercise introduces you to a few of the person- and job-specific variables that are available from LIS.

Activity

For Spain in 2000, create a Stata dataset containing *page*, *pweight*, *pclfs*, and the following:

- Characteristics of the Individual
 - continuous: *phoursu*, *pweektl*
 - categorical: *psecjob*, *psearch*, *pcare*
- Characteristics of the Primary Job
 - categorical: *pactiv*, *pind*, *ptypewk*, *pfulpar*

This exercise uses a sample of adults (16 and over). From *pclfs* create a dummy variable (**emp**) that will be coded 0 for individuals that are not employed and 1 for individuals that are employed. Using the characteristics of the individual (continuous variables *phoursu*, *pweektl* as well as the newly created employment dummy variable, **emp**), find the weighted mean, minimum, and maximum. Repeat the same calculations (for the same variables) after you “clean” the data (see the Guidelines for instructions).

Finally, tabulate the dummy variable (**emp**) with each categorical variable. Run this tabulation twice: (1) without weights and with those individuals who are coded -1; (2) with weights and excluding those who are coded -1.

Use the information from your output and/or the documentation to answer the following questions:

1. What are the minimum and maximum numbers of usual weekly hours for the employed?

Explain these results.

What are the implications for interpreting the average hours worked reported above?

2. Why does *phoursu* equal 0 for all observations of those not employed? When is it possible for *phoursu* to equal 0 for some of those employed?

3. How is it possible that some of those who are not employed reported 52 weeks of employment?

4. Which variables have no observations equal to -1 among the employed and all observations equal to -1 among those not employed?

What does this tell us about the universe of those variables?

5. Why does *ptypewk* have some observations equal to -1 among the employed?

6. Which variables have no observations equal to -1 for either group?

What does this tell us about the universe of those variables?

Guidelines

- When creating your employment dummy variable, be sure to include only those you can clearly identify as being either employed or not employed. In other words, be careful what you do with the 900s codes in *pclfs*.
- The variables *phoursu* and *pweektl* (as well as all other time variables), are peculiar in that they may be continuous and discrete at the same time: all the values between 0 and 999 are normal continuous values, representing the duration of time (either hours or weeks), but on top of those there may be some discrete values, notably:
 - the value -9 for *phoursu* and *phoursa* stands for “hours vary”;
 - all the values greater or equal to 1000 indicate partial or incomplete information, where the 1000 is a flag and stands for “at least xx number of hours” (see Section D of the introduction of the Labour Market Variables Guidelines).

As a consequence, none of these values should be used as such (as they do not represent actual numbers of hours / weeks), and they should be excluded from the sample unless recoded to a valid value.

- For the continuous variables, you will need two different commands for each variable: (1) for the entire sample; and (2) for the restricted (or cleaned) sample. In both cases, use the **bysort [varname]**: suffix to separate the values for the employed from the not employed.

- For the categorical variables, you will need two different commands for each variable. In order to get the sample frequencies of a specific value of a variable, remember not to weight the results. To get the correct estimated percentages among the valid observations, you will need to carry out a weighted tabulation excluding -1 from the valid observations. In both cases, use the **bysort [varname]**: suffix to separate the values for the employed from the not employed separately.
- Stata reminder: It is not possible to use analytical weights with the **tab1** command. Instead, use a **foreach** statement combined with a normal **tabulate** to get the weighted frequencies (and percentages). Your command will look like this:

```
foreach var in pactiv pind ptypewk pfulpar psecjob psearch pcare {  
    bysort emp: tab `var' [aw=pweight] if !(`var'==-1), mi  
}
```

Program

```
di "*** LABOUR MARKET - Exercise 18 ***"

global id "pweight page pclfs"
global cont "phoursu pweektl"
global cat "psecjob psearch pcare pactiv pind ptypewk pfulpar "

use $id $cont $cat if inrange(page,16,.) using $es00p, clear
recode pclfs (100/199=1) (200/999=0) (-1=.), gen(emp)

foreach var in $cont {
    display "`var' - unweighted - all values"
    bysort emp: sum `var'
    display "`var' - weighted - clean values"
    bysort emp: sum `var' [w=pweight] if inrange(`var',0,999)
}

foreach var in $cat {
    display "`var' - unweighted - all values"
    bysort emp: tab `var', mi
    display "`var' - weighted - clean values"
    bysort emp: tab `var' [aw=pweight] if !(`var'==-1), mi
}
```

Results**Table 1**

Continuous variables	# observations	Mean	Min	Max
All values				
Employed				
<i>phoursu</i>	5,599	66.2	-9	1,072
<i>pweektl</i>	5,350	51.8	0	1,048
Not Employed				
<i>phoursu</i>	6,489	0	0	0
<i>pweektl</i>	6,446	6.8	0	1,048
“Clean” values only				
Employed				
<i>phoursu</i>	5,410	41.3	2	96
<i>pweektl</i>	5,317	46.1	0	52
Not Employed				
<i>phoursu</i>	6,489	0	0	0
<i>pweektl</i>	6,421	3.1	0	52

Table 2

Categorical Variables	Employed	Not Employed
Characteristics of the Individual		
<i>psecjob</i>	# obs = -1	202
	# missing obs	154
	% with multiple jobs	3.14
<i>psearch</i>	# obs = -1	0
	# missing obs	155
	% looking for a job	8.78
<i>pcare</i>	# obs = -1	0
	# missing obs	155
	% caregivers	22.61
Characteristics of the Job		
<i>pactiv</i>	# obs = -1	0
	# missing obs	134
	% own-account workers	11.52
<i>pind</i>	# obs = -1	0
	# missing obs	134
	% in construction	11.33
<i>ptypewk</i>	# obs = -1	202
	# missing obs	154
	% in public sector	16.46
<i>pfulpar</i>	# obs = -1	0
	# missing obs	6
	% in full-time employment	87.2

Answers to questions 1 to 6:

1. The minimum number of usual weekly hours is -9 and the maximum one is 1072.
-9 is the standard LIS value indicating the employee works irregular hours, while the 1000 is a flag for partial information on hours (10xx stands for “at least xx hours”).
The average hours worked by the employed are underestimated if the -9 values are included in the averages, while they are significantly overestimated if the values above 1000 are included. Negative values and values greater than or equal to 1000 should be excluded (or recoded) when calculating averages.
2. The employment dummy created in this exercise counts even those on leave and in marginal jobs as employed. Therefore, we would expect “usual” hours for the employed to include 0 hours (e.g., those currently on leave), but no one counted as not employed should have worked even one hour.
3. This is due to the different reference periods of the variables. The variable *pclfs* reports the labour force status (LFS) at the time of the interview, and *pweektl* reports total weeks worked during 2000. Looking at the country-specific survey documentation shows that the interview takes place between October and December of 2001, so a person could have worked the entire previous year, but stopped in between the end of 2000 and the interview.
4. Variables *pactiv*, *pind* and *pfulpar* have no observations equal to -1 among the employed and all observations equal to -1 among those not employed.
This means that the universe of those variables (i.e., the individuals to whom this question is relevant) is exactly all employed persons as reported in *pclfs*.
5. Different from the other variables about primary job characteristics, the universe of *ptypewk* is smaller than all employed persons. The documentation tells us that the universe contains only persons who currently work at least 15 hours per week, not all employed individuals.
6. Variables *psearch* and *pcare* have no observations equal to -1 for either group.
This means that the universe of those variables is at least all adults with known LFS, independently from whether they are employed or not.