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 SAS 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 to 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 of 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.

➢ The syntax of the SAS procedure is standardized. To avoid repeating several times, the same code with different parameters (e.g.: different analytical variable), you can create a generic SAS macro that allows you to only call the relevant parameter each time you invoke the macro. Find below an example (**code 1**) of the type of SAS macro you could create for this exercise. Note that you can use more parameters to make the SAS macro as generic as possible (**code 2**) and use it for almost any situation where your SAS procedure is needed.

code 1
```
%MACRO avg;
  PROC MEANS DATA=lmp N MEAN MIN MAX;
    &where;
    VAR    &var;
    WEIGHT pweight;
    BY     emp;
  RUN;
%MEND avg;
```

code 2
```
%MACRO avg;
  PROC MEANS DATA=&ds &stat &opt;
    &where;
        &classe ;
    &typ    ;
    VAR &var;
    &wgt;
    &by ;
  RUN;
%MEND avg;
```

Program

```
OPTIONS NOSOURCE NONOTES NOFMTERR NODATE NOCENTER LABEL NONUMBER LS=200
PS=MAX ;

%MACRO avg ;
  PROC MEANS DATA=lmp N MEAN MIN MAX  ;
    &where ;
    VAR     &var ;
    WEIGHT pweight ;
    BY      emp ;
  RUN ;
%MEND avg ;
%MACRO frq ;
  %LET varlist = psecjob psearch pcare pactiv pind ptypewk pfulpar ;
  %LET i = 1 ;
  %DO %UNTIL (&i > 7) ;
     TITLE "%scan(&varlist,&i) - unweighted - all values" ;
     PROC FREQ DATA=lmp ;
       TABLES %scan(&varlist,&i) / MISSPRINT NOROW NOCOL NOCUM MISSING ;
       BY     emp ;
     RUN ;
     TITLE "%scan(&varlist,&i)  - weighted - cleaned values" ;
     PROC FREQ DATA=lmp ;
       WHERE  %scan(&varlist,&i) ne -1   ;
       TABLES %scan(&varlist,&i) / MISSPRINT NOROW NOCOL NOCUM MISSING;
       WEIGHT pweight ;
       BY     emp ;
     RUN ;
    %LET i = %eval(&i+1) ;
  %END ;
%MEND frq ;

DATA lmp ;
 SET &es00p (KEEP=pweight page pclfs pactiv pind ptypewk pfulpar phoursu
pweektl psecjob psearch pcare);
    * Retain Adults ;
    IF (page > 15) ;
    *--- Employment rate ;
    emp = pclfs ;
    IF      pclfs in (-1)      THEN emp = . ;
    ELSE IF (100<=pclfs<=199)  THEN emp = 1 ;
    ELSE IF (200<=pclfs<=999)  THEN emp = 0 ;
RUN ;
PROC SORT DATA=lmp ;
 BY emp ;
RUN ;

TITLE "phours - all Values " ;
%LET where = ; %LET var   = phoursu ; %avg
TITLE "pweektl - all Values " ;
%LET where = ; %LET var   = pweektl ; %avg
TITLE "phours - cleaned Values " ;
%LET where = WHERE (( phoursu > 0) OR (phoursu < 999)) ;
%LET var   = phoursu ;
```

```
%avg
TITLE "pweektl - cleaned Values  " ;
%LET where = WHERE ((pweektl > 0) OR (pweektl < 999)) ;
%LET var  = pweektl ;
%avg

%frq
```

Results

Table 1

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

Table 2

Categorical Variables		Employed	Not Employed
Characteristics of the Individual			
psecjob	# obs = -1	202	6,492
	# missing obs	154	50
	% with multiple jobs	3.14	NA
psearch	# obs = -1	0	0
	# missing obs	155	51
	% looking for a job	8.78	11.6
pcare	# obs = -1	0	0
	# missing obs	155	71
	% caregivers	22.61	21.5
Characteristics of the Job			
pactiv	# obs = -1	0	6,492
	# missing obs	134	50
	% own-account workers	11.52	NA
pind	# obs = -1	0	6,492
	# missing obs	134	50

	% in construction	*11.33*	*NA*
ptypewk	# obs = -1	*202*	*6,492*
	# missing obs	*154*	*50*
	% in public sector	*16.46*	*NA*
pfulpar	# obs = -1	*0*	*6,492*
	# missing obs	*6*	*50*
	% in full-time employment	*87.2*	*NA*

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 (10**xx** 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.