

```
*****
* A Practical Guide to Using Panel Data
* Simonetta Longhi and Alita Nandi
* ISER, University of Essex
* Chapter 5
*****
```

```
-----
name: <unnamed>
log: C:\My Documents\\Example_Chapter5.log
log type: text
opened on: 30 Jul 2014, 09:49:29
```

```
.
.
. * Section 5.3.1
```

```
. use pid rhid rage rhgsex rhgemp rhoh rxewght using "$datadir/rindall", clear
```

```
. describe rhgemp
```

variable name	storage type	display format	value label	variable label
rhgemp	byte	%8.0g	rhgemp	employment status - hh grid

```
. label list rhgemp
```

```
rhgemp:
-9 missing or wild
-8 inapplicable
-2 refused
-1 don't know
0 under 16 years
1 yes
2 no
```

```
. generate NotEmp = (rhgemp ~= 1) if rhgemp>=0
(106 missing values generated)
```

```
. tab2 NotEmp rhgemp, missing
```

```
-> tabulation of NotEmp by rhgemp
```

NotEmp	employment status - hh grid				Total
	don't kno	under 16	yes	no	
0	0	0	9,668	0	9,668
1	0	4,113	0	6,290	10,403
.	106	0	0	0	106
Total	106	4,113	9,668	6,290	20,177

```
. bysort rhid: egen numNotEmp = sum(NotEmp)
```

```
. bysort rhid: egen empmiss = sum(NotEmp==.)
```

```
. replace numNotEmp=. if empmiss>0
(237 real changes made, 237 to missing)
```

```
. bysort rhid: egen numNotEmp2 = sum(rhgemp ~= 1 & rhgemp>=0)
```

```
. bysort rhid: egen empmiss2 = sum(rhgemp<0)
```

```
. replace numNotEmp2=. if empmiss2>0
(237 real changes made, 237 to missing)
```

```
. summarize pid if numNotEmp~=numNotEmp2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pid	0				

```
. tabulate numNotEmp numNotEmp2, missing
```

numNotEmp	numNotEmp2						Total
	0	1	2	3	4	5	
0	4,198	0	0	0	0	0	4,198
1	0	5,399	0	0	0	0	5,399
2	0	0	5,995	0	0	0	5,995
3	0	0	0	2,496	0	0	2,496
4	0	0	0	0	1,060	0	1,060
5	0	0	0	0	0	471	471
6	0	0	0	0	0	0	177
7	0	0	0	0	0	0	76
8	0	0	0	0	0	0	17
9	0	0	0	0	0	0	21
10	0	0	0	0	0	0	14
12	0	0	0	0	0	0	16
.	0	0	0	0	0	0	237
Total	4,198	5,399	5,995	2,496	1,060	471	20,177

numNotEmp	numNotEmp2						Total
	6	7	8	9	10	12	
0	0	0	0	0	0	0	4,198
1	0	0	0	0	0	0	5,399
2	0	0	0	0	0	0	5,995
3	0	0	0	0	0	0	2,496
4	0	0	0	0	0	0	1,060
5	0	0	0	0	0	0	471
6	177	0	0	0	0	0	177
7	0	76	0	0	0	0	76
8	0	0	17	0	0	0	17
9	0	0	0	21	0	0	21
10	0	0	0	0	14	0	14
12	0	0	0	0	0	16	16
.	0	0	0	0	0	0	237
Total	177	76	17	21	14	16	20,177

numNotEmp	numNotEmp2	Total
	.	
0	0	4,198
1	0	5,399
2	0	5,995
3	0	2,496
4	0	1,060
5	0	471
6	0	177
7	0	76
8	0	17
9	0	21
10	0	14
12	0	16
.	237	237
Total	237	20,177

```
. tabulate empmiss empmiss2, missing
```

empmiss	empmiss2				Total
	0	1	2	3	4

0	19,940	0	0	0	0	19,940
1	0	141	0	0	0	141
2	0	0	33	0	0	33
3	0	0	0	45	0	45
4	0	0	0	0	18	18
Total	19,940	141	33	45	18	20,177

```

.
. label list rhoh rhgsex
rhoh:
    -9 missing or wild
    -8 inapplicable
    1 head of household
    2 not head
rhgsex:
    -9 missing or wild
    -8 inapplicable
    1 male
    2 female

. bysort rhid: egen femaleHoH = sum(rhoh == 1 & rhgsex == 2)

.
. bysort rhid: egen numkids = sum(rage >= 0 & rage < 16)

.
. save tempIND, replace
(note: file tempIND.dta not found)
file tempIND.dta saved

.
.
. * Section 5.3.2
. *-----
.
. use rhid rfihhmn rfieqfcb rhhsz rregion2 using "$datadir/rhhresp", clear

.
. bysort rhid: gen uniqueHH = _N

. tabulate uniqueHH

    uniqueHH |      Freq.      Percent      Cum.
-----+-----
          1 |      8,144       100.00       100.00
-----+-----
        Total |      8,144       100.00

.
. duplicates report rhid

Duplicates in terms of rhid

-----
    copies | observations      surplus
-----+-----
          1 |          8144              0
-----

. isid rhid

.
.
. sort rhid

. save tempHH, replace
(note: file tempHH.dta not found)
file tempHH.dta saved

.

```

```
. use tempIND, clear
```

```
. sort rhid
```

```
. duplicates report rhid
```

Duplicates in terms of rhid

copies	observations	surplus

1	2103	0
2	5622	2811
3	4107	2738
4	4888	3666
5	2315	1852
6	786	655
7	196	168
8	72	63
9	27	24
10	20	18
11	11	10
14	14	13
16	16	15

```
. duplicates report pid
```

Duplicates in terms of pid

copies	observations	surplus

1	20177	0

```
.
. merge m:1 rhid using tempHH
```

Result	# of obs.

not matched	0
matched	20,177 (_merge==3)

```
. drop _merge
```

```
.
.
. * Ssection 5.3.3
. *-----
.
. generate equivHHinc = rfihhmn/rfieqfcb if rfihhmn>=0 & rfihhmn<.
(1308 missing values generated)
```

```
. summarize equivHHinc [aweight = rxewght], detail
```

equivHHinc				

	Percentiles	Smallest		
1%	323.3831	0		
5%	820.5641	0		
10%	1036.76	0	Obs	10065
25%	1459.49	0	Sum of Wgt.	11023.2248
50%	2276.995		Mean	2703.679
		Largest	Std. Dev.	2038.301
75%	3355.301	34906.02		
90%	4787.225	34906.02	Variance	4154669
95%	5946.25	34906.02	Skewness	4.980602
99%	9390.515	42690.16	Kurtosis	61.3869

```
. return list
```

```
scalars:
```

```

      r(N) = 10065
      r(sum_w) = 11023.22475796938
      r(mean) = 2703.679091031838
      r(Var) = 4154668.930140617
      r(sd) = 2038.300500451446
      r(skewness) = 4.980602030204397
      r(kurtosis) = 61.38689852694022
      r(sum) = 29803262.29386631
      r(min) = 0
      r(max) = 42690.1640625
      r(p1) = 323.3830871582031
      r(p5) = 820.5641479492188
      r(p10) = 1036.760131835938
      r(p25) = 1459.489990234375
      r(p50) = 2276.994873046875
      r(p75) = 3355.30126953125
      r(p90) = 4787.22509765625
      r(p95) = 5946.25048828125
      r(p99) = 9390.5146484375

```

```
. generate PovertyLine = 0.60*r(p50)
```

```
.
. generate poor = (equivHHinc<PovertyLine) if equivHHinc<.
(1308 missing values generated)
```

```
.
. tabulate poor
```

poor	Freq.	Percent	Cum.
0	14,607	77.41	77.41
1	4,262	22.59	100.00
Total	18,869	100.00	

```
.
. summarize equivHHinc [aweight = rxewght], detail
```

equivHHinc				
Percentiles		Smallest		
1%	323.3831	0		
5%	820.5641	0		
10%	1036.76	0	Obs	10065
25%	1459.49	0	Sum of Wgt.	11023.2248
50%	2276.995		Mean	2703.679
		Largest	Std. Dev.	2038.301
75%	3355.301	34906.02		
90%	4787.225	34906.02	Variance	4154669
95%	5946.25	34906.02	Skewness	4.980602
99%	9390.515	42690.16	Kurtosis	61.3869

```
. global PovertyLine = 0.60*r(p50)
```

```
. generate poor2 = (equivHHinc<$PovertyLine) if equivHHinc<.
(1308 missing values generated)
```

```
.
. tabulate poor poor2, missing
```

poor	poor2		.	Total
	0	1		
0	14,607	0	0	14,607
1	0	4,262	0	4,262
.	0	0	1,308	1,308
Total	14,607	4,262	1,308	20,177

```

.
. save PovertyWave_r, replace
(note: file PovertyWave_r.dta not found)
file PovertyWave_r.dta saved

.
.
. * Section 5.3.5
. *-----
.
. import excel using "HHtenure2011UKcensus.xls", clear cellrange(A11:E20)

.
. rename B HH_all

. rename D HH_social_rented

. rename E pr_social_rented

.
. generate rregion2=1 if A=="North East"
(9 missing values generated)

. replace rregion2=2 if A=="North West"
(1 real change made)

. replace rregion2=3 if A=="Yorkshire and The Humber"
(1 real change made)

. replace rregion2=4 if A=="East Midlands"
(1 real change made)

. replace rregion2=5 if A=="West Midlands"
(1 real change made)

. replace rregion2=6 if A=="East"
(1 real change made)

. replace rregion2=7 if A=="London"
(1 real change made)

. replace rregion2=8 if A=="South East"
(1 real change made)

. replace rregion2=9 if A=="South West"
(1 real change made)

. replace rregion2=10 if A=="Wales"
(1 real change made)

. label define rregion2 1 "North East" 2 "North West" ///
>                                     3 "Yorkshire and The Humber" ///
>                                     4 "East Midlands" 5 "West Midlands" 6 "East" 7 "London" ///
>                                     8 "South East" 9 "South West" 10 "Wales"

. label value rregion2 rregion2

.
. fre rregion2

```

rregion2

			Freq.	Percent	Valid	Cum.
Valid	1	North East	1	10.00	10.00	10.00
	2	North West	1	10.00	10.00	20.00
	3	Yorkshire and The Humber	1	10.00	10.00	30.00
	4	East Midlands	1	10.00	10.00	40.00
	5	West Midlands	1	10.00	10.00	50.00
	6	East	1	10.00	10.00	60.00
	7	London	1	10.00	10.00	70.00

8	South East	1	10.00	10.00	80.00
9	South West	1	10.00	10.00	90.00
10	Wales	1	10.00	10.00	100.00
Total		10	100.00	100.00	

```

.
. drop A C

. sort rregion2

. save temp, replace
(note: file temp.dta not found)
file temp.dta saved

.

. use PovertyWave_r, clear

. sort rregion2

. merge m:1 rregion2 using temp
(note: variable rregion2 was byte, now float to accommodate using data's values)
(label rregion2 already defined)

```

Result	# of obs.	
not matched	6,844	
from master	6,844	(<u>_merge</u> ==1)
from using	0	(<u>_merge</u> ==2)
matched	13,333	(<u>_merge</u> ==3)

```

.
. fre rregion2 if _merge==1

```

rregion2 -- government office region

		Freq.	Percent	Valid	Cum.
Valid	-9 missing or wild	136	1.99	1.99	1.99
	11 scotland	3509	51.27	51.27	53.26
	12 northern ireland	3199	46.74	46.74	100.00
	Total	6844	100.00	100.00	

```

.
. keep if _merge==3
(6844 observations deleted)

```

```

. drop _merge

```

```

.
. save PovertyWave_r, replace
file PovertyWave_r.dta saved

```

```

.
.
. * Section 5.4
. *-----
.
. use PovertyWave_r, clear

```

```

.
. foreach var in rhgsex numkids numNotEmp femaleHoH rhhsiz {
2.     tabulate `var' poor, row nofreq chi2
3. }

```

sex - hh grid	poor 0	1	Total
male	79.87	20.13	100.00
female	77.35	22.65	100.00

Total	78.58	21.42	100.00
-------	-------	-------	--------

Pearson chi2(1) = 11.6071 Pr = 0.001

numkids	poor 0	1	Total
0	80.36	19.64	100.00
1	82.86	17.14	100.00
2	78.31	21.69	100.00
3	63.39	36.61	100.00
4	41.95	58.05	100.00
5	48.86	51.14	100.00
6	100.00	0.00	100.00
7	0.00	100.00	100.00
Total	78.58	21.42	100.00

Pearson chi2(7) = 365.1868 Pr = 0.000

numNotEmp	poor 0	1	Total
0	95.95	4.05	100.00
1	81.81	18.19	100.00
2	77.55	22.45	100.00
3	67.03	32.97	100.00
4	57.58	42.42	100.00
5	21.90	78.10	100.00
6	24.05	75.95	100.00
7	40.00	60.00	100.00
8	0.00	100.00	100.00
10	100.00	0.00	100.00
12	100.00	0.00	100.00
Total	78.78	21.22	100.00

Pearson chi2(10) = 1.5e+03 Pr = 0.000

femaleHoH	poor 0	1	Total
0	82.57	17.43	100.00
1	66.94	33.06	100.00
Total	78.58	21.42	100.00

Pearson chi2(1) = 340.0191 Pr = 0.000

number of persons in household	poor 0	1	Total
1	68.76	31.24	100.00
2	80.07	19.93	100.00
3	84.10	15.90	100.00
4	82.79	17.21	100.00
5	71.38	28.62	100.00
6	62.90	37.10	100.00
7	53.85	46.15	100.00
8	50.00	50.00	100.00
9	0.00	100.00	100.00
10	0.00	100.00	100.00
14	100.00	0.00	100.00
16	100.00	0.00	100.00
Total	78.58	21.42	100.00

Pearson chi2(11) = 407.3110 Pr = 0.000

```
. tabstat poor, by(rregion2) format(%9.2f)
```


Summary for variables: poor
by categories of: rregion2 (government office region)

rregion2	mean
north east	0.20
north west	0.24
yorkshire and th	0.21
east midlands	0.29
west midlands	0.22
east of england	0.21
london	0.15
south east	0.15
south west	0.19
wales	0.24
Total	0.21

```
.
. tabstat pr_social_rented, by(poor) format(%9.2f)
```

Summary for variables: pr_social_rented
by categories of: poor

poor	mean
0	17.00
1	16.99
Total	17.00

```
.
.
. * Section 5.6
. *-----
.
. use PovertyWave_r, clear

. bysort rhid: keep if _n == 1
(7954 observations deleted)

. duplicates report rhid
```

Duplicates in terms of rhid

copies	observations	surplus
1	5379	0

```
.
. use PovertyWave_r, clear

. bysort rhid: keep if _n == _N
(7954 observations deleted)

. duplicates report rhid
```

Duplicates in terms of rhid

copies	observations	surplus
1	5379	0

```
.
. use PovertyWave_r, clear
```

```
. bysort rhid (rage): keep if _n == _N
(7954 observations deleted)
```

```
. duplicates report rhid
```

Duplicates in terms of rhid

```
-----
      copies | observations      surplus
-----+-----
           1 |          5379           0
-----
```

```
.
.
. * Section 5.7.1
. *-----
.
. foreach w in q r {
2.      use `w'hid `w'fihhmn `w'fieqfcb using "$datadir/`w'hhresp", clear
3.      generate equivHHinc = `w'fihhmn/`w'fieqfcb if `w'fihhmn>=0
4.      sort `w'hid
5.      save temp, replace
6.
.      use pid `w'hid `w'xewght using "$datadir/`w'indall", clear
7.      sort `w'hid
8.      merge m:1 `w'hid using temp
9.      drop _merge
10.     summarize equivHHinc [aweight = `w'xewght], detail
11.     global PovertyLine = 0.60*r(p50)
12.     generate `w'poor = (equivHHinc<$PovertyLine) if equivHHinc<.
13.     keep `w'poor pid
14.     sort pid
15.     save PoorWave_`w', replace
16. }
(487 missing values generated)
file temp.dta saved
```

```
Result                                     # of obs.
-----
not matched                                0
matched                                  20,715  (_merge==3)
-----
```

```

equivHHinc
-----
Percentiles      Smallest
1%      262.6667      0
5%      790.5507      0
10%     985.6281      0      Obs      10350
25%     1424.412      0      Sum of Wgt. 10333.1036

50%     2176.722
75%     3262.754      Largest      Mean      2579.894
90%     4586.558      20402.16      Std. Dev. 1794.254
95%     5632.917      23907.1      Variance 3219347
99%     8988.595      26974.85      Skewness 3.156949
Kurtosis 25.85613
(1175 missing values generated)
(note: file PoorWave_q.dta not found)
file PoorWave_q.dta saved
(546 missing values generated)
file temp.dta saved
```

```
Result                                     # of obs.
-----
not matched                                0
matched                                  20,177  (_merge==3)
-----
```

```

equivHHinc
-----
Percentiles      Smallest
```

```

1%      323.3831      0
5%      820.5641      0
10%     1036.76      0      Obs      10065
25%     1459.49      0      Sum of Wgt. 11023.2248

50%     2276.995      Mean      2703.679
                        Largest Std. Dev. 2038.301
75%     3355.301      34906.02
90%     4787.225      34906.02      Variance      4154669
95%     5946.25      34906.02      Skewness      4.980602
99%     9390.515      42690.16      Kurtosis      61.3869
(1308 missing values generated)
(note: file PoorWave_r.dta not found)
file PoorWave_r.dta saved

```

```

.
. use PoorWave_r, clear
. merge 1:1 pid using PoorWave_q

```

Result	# of obs.	
not matched	2,416	
from master	939	(_merge==1)
from using	1,477	(_merge==2)
matched	19,238	(_merge==3)

```

. keep if _merge==3
(2416 observations deleted)

```

```

. drop _merge

```

```

.
.
. * Section 5.7.2
. *-----
.
. egen poorch=concat(qpoor rpoor) if qpoor<. & rpoor<.
(1302 missing values generated)

```

```

. tabulate poorch

```

poorch	Freq.	Percent	Cum.
00	12,828	71.52	71.52
01	1,521	8.48	80.00
10	1,119	6.24	86.24
11	2,468	13.76	100.00
Total	17,936	100.00	

```

.
.
. * erase all temporary files
. *-----
.
. erase tempIND.dta
. erase tempHH.dta
. erase temp.dta
. erase PoorWave_q.dta
. erase PoorWave_r.dta
. erase PovertyWave_r.dta
. log close
name: <unnamed>
log: C:\My Documents\\Example_Chapter5.log

```

log type: text
closed on: 30 Jul 2014, 09:49:33
