

```

1 *****
2 * A Practical Guide to Using Panel Data
3 * Simonetta Longhi and Alita Nandi
4 * ISER, University of Essex
5 * Chapter 14
6 *****
7
8 version 12
9 clear all
10 set more off
11 capture log close
12
13
14 cd "C:\My Documents\"
15
16 capture mkdir "C:\My Documents\chapter14"
17
18 global analysisdir "C:\My Documents\chapter14"
19
20 log using "$analysisdir\Example_Chapter14.log", replace
21
22 use DatasetR.dta, clear
23
24 * 14.2. Tables of Descriptive Statistics
25 *-----
26
27 * 14.2.1. Using the Collapse Command
28 *-----
29
30 use DatasetR.dta, clear
31
32 recode region2 (1/9 = 1) (10 = 2) (11 = 3) (12 = 4), gen(Country)
33 label var Country "Country of residence"
34 label define Country 1 "England" 2 "Wales" 3 "Scotland" 4 "Northern Ireland"
35 label value Country Country
36
37 * Totals by row (country)
38
39 bysort Country: egen X_NotMarried = total(Married==0)
40 bysort Country: egen X_Married = total(Married==1)
41 egen N_NotMarried = total(Married==0) if Country < .
42 egen N_Married = total(Married==1) if Country < .
43
44
45 keep Country X_* N_*
46 bysort Country: keep if _n==1
47
48 generate T_NotMarried = X_NotMarried / N_NotMarried
49 generate T_Married = X_Married / N_Married
50
51 label var T_NotMarried ///
52     "Proportion of not married or cohabiting people living in country"
53 label var T_Married ///
54     "Proportion of people married or cohabiting people living in country"
55
56 drop if Country == .
57
58 list
59
60 * add one more line of observation and populate it
61 set obs 6
62
63 replace T_Married      = N_Married[1]      if _n==6
64 replace T_NotMarried = N_NotMarried[1] if _n==6
65 replace Country      = 6                  if _n==6
66
67 replace T_Married      = 1 if _n==5
68 replace T_NotMarried = 1 if _n==5
69 replace Country      = 5 if _n==5
70
71 label define Country 5 "Total %" 6 "Observations", modify
72
73 generate CountryString = "England"          if Country == 1
74 replace CountryString = "Wales"            if Country == 2
75 replace CountryString = "Scotland"         if Country == 3

```

```

76  replace CountryString = "Northern Ireland" if Country == 4
77  replace CountryString = "Total"           if Country == 6
78
79  keep Country CountryString T_NotMarried T_Married
80  order Country CountryString T_NotMarried T_Married
81  save "$analysisdir\DescriptiveTableMethod1.dta", replace
82
83  export excel using "$analysisdir\DescriptiveTableMethod1.xls", ///
84      replace firstrow(variable)
85
86
87  * 14.2.2. Using the Saved Results
88  *-----
89
90  use DatasetR.dta, clear
91
92  recode region2 (1/9 = 1) (10 = 2) (11 = 3) (12 = 4), gen(Country)
93  label var Country "Country of residence"
94  label define Country 1 "England" 2 "Wales" 3 "Scotland" 4 "Northern Ireland"
95  label value Country Country
96
97  generate Names = ""
98  replace Names = "England" in 1
99  replace Names = "Wales" in 2
100  replace Names = "Scotland" in 3
101  replace Names = "Northern Ireland" in 4
102  replace Names = "Total %" in 5
103  replace Names = "Observations" in 6
104
105  generate T_NotMarried = .
106  generate T_Married = .
107
108  tabulate Country Married, col matcell(Total)
109  matrix list Total
110
111  local i = 1
112  while `i' <= 4 {
113      replace T_NotMarried = Total[`i',1] in `i'
114      replace T_Married = Total[`i',2] in `i'
115      local i = `i' + 1
116  }
117  * browse Names T_*
118
119  * Proportions by row
120  *-----
121  *generate Percentage = .
122  *generate Total = T_NotMarried + T_Married
123  *foreach var of varlist T_NotMarried T_Married {
124  *    replace `var' = `var' / Total in 1/4
125  *    }
126  *replace Percentage = T_NotMarried + T_Married
127  * browse Names T_* Total Percentage
128
129  * Proportions by column
130  *-----
131  tab Country if Married == 0
132  replace T_NotMarried = r(N) in 5/6
133  generate Tempo = r(N) in 1/5
134  replace T_NotMarried = T_NotMarried / Tempo in 1/5
135  drop Tempo
136  tab Country if Married == 1
137  replace T_Married = r(N) in 5/6
138  generate Tempo = r(N) in 1/5
139  replace T_Married = T_Married / Tempo in 1/5
140  drop Tempo
141  * browse Names T_*
142
143  * Save into table *
144  * If we computed proportions by row:
145  * keep Names T_* Percentage Total
146  * If we computed proportions by column:
147  keep Names T_*
148  keep in 1/6
149  rename Names Country
150  order Country T_NotMarried T_Married

```

```

151 export excel using "$analysisdir\DescriptiveTableMethod2", ///
152     firstrow(varlabels) replace
153
154
155
156 * 14.2.3. Using the collapse command
157 *-----
158 use DatasetR.dta, clear
159
160 recode region2 (1/9 = 1) (10 = 2) (11 = 3) (12 = 4), gen(Country)
161
162 keep if Country<.
163 save temp, replace
164
165 tabstat Married, by(Country) s(mean n)
166
167 use temp, clear
168 collapse (mean) ProportionMarried=Married ///
169     (count) Number_Observations = Married, by(Country)
170 list
171 save temp1, replace
172
173 use temp, clear
174 collapse (mean) ProportionMarried=Married ///
175     (count) Number_Observations = Married
176 generate Country = 5
177 append using temp1
178
179 label var Country "Country of residence"
180 label define Country 5 "UK" 1 "England" 2 "Wales" 3 "Scotland" ///
181     4 "Northern Ireland"
182 label value Country Country
183
184 list
185
186 sort Country
187 order Country ProportionMarried Number_Observations
188 export excel using "$analysisdir\DescriptiveTableMethod3.xls", ///
189     replace firstrow(variable)
190
191 erase temp.dta
192 erase temp1.dta
193
194
195 * 14.3 Graphs of Descriptive Statistics
196 *-----
197
198 * 14.3.1 Bar Graphs
199 *-----
200
201 graph bar ProportionMarried, ///
202     over(Country, label(angle(45))) ytitle("Proportion") ///
203     scheme(slmanual) title("Proportion married - UK countries") ///
204     note("Source: BHPS wave 18") ///
205     saving("$analysisdir\PropMarried1.gph", replace)
206
207 graph export "$analysisdir\PropMarried1.pdf", replace
208
209
210 use DatasetR.dta, clear
211 recode region2 (1/9 = 1) (10 = 2) (11 = 3) (12 = 4), gen(Country)
212 keep if Country<.
213 keep if Female<.
214
215 collapse (mean) ProportionMarried=Married, by(Country Female)
216 label define Country 5 "UK" 1 "England" 2 "Wales" 3 "Scotland" ///
217     4 "Northern Ireland"
218 label value Country Country
219 label define Female 0 "Men" 1 "Women"
220 lab value Female Female
221 list
222
223 graph bar ProportionMarried, ///
224     over(Country, label(angle(45))) over(Female) ///
225     ytitle("Proportion") scheme(slmanual) ///

```

```

226 title("Proportion married across UK countries") ///
227 note("Source: BHPS wave 18") ///
228 saving("$analysisdir\PropMarried2.gph", replace)
229
230
231 use DatasetR.dta, clear
232 recode region2 (1/9 = 1) (10 = 2) (11 = 3) (12 = 4), gen(Country)
233 keep if Country<.
234
235 collapse (mean) Q1 Q2 Q3 Q4 Q5 Q6, by(Country)
236 label define Country 5 "UK" 1 "England" 2 "Wales" 3 "Scotland" ///
237 4 "Northern Ireland"
238 label value Country Country
239
240 graph hbar Q1 Q2 Q3 Q4 Q5 Q6, stack ///
241 over(Country, label(angle(45))) ///
242 ytitle("Proportion") ///
243 title("Composition across UK countries") ///
244 note("Source: BHPS wave 18") ///
245 scheme(slmanual) ///
246 legend(label(1 "1st or higher") ///
247 label(2 "HND,HNC,teaching") ///
248 label(3 "A level") ///
249 label(4 "O Level") ///
250 label(5 "CSE") ///
251 label(6 "None")) ///
252 saving("$analysisdir\PropEdu1.gph", replace)
253
254
255 use DatasetR.dta, clear
256 recode region2 (1/9 = 1) (10 = 2) (11 = 3) (12 = 4), gen(Country)
257 keep if Country<.
258
259 graph hbar Q1 Q2 Q3 Q4 Q5 Q6, stack ///
260 over(Country, relabel(1 "England" 2 "Wales" 3 "Scotland" ///
261 4 "Northern Ireland") label(angle(45))) ///
262 ytitle("Proportion") ///
263 title("Composition across UK countries") ///
264 scheme(slmanual) ///
265 note("Source: BHPS wave 18") ///
266 legend(label(1 "1st or higher") ///
267 label(2 "HND,HNC,teaching") ///
268 label(3 "A level") ///
269 label(4 "O Level") ///
270 label(5 "CSE") ///
271 label(6 "None")) ///
272 saving("$analysisdir\PropEdu2.gph", replace)
273
274
275 label define Female 0 "Men" 1 "Women"
276 lab value Female Female
277
278 graph hbar Q1 Q2 Q3 Q4 Q5 Q6, stack ///
279 over(Country, ///
280 relabel(1 "England" 2 "Wales" 3 "Scotland" 4 "Northern Ireland") ///
281 label(angle(45))) ///
282 by(Female) ytitle("Proportion") ///
283 scheme(slmanual) ///
284 note("Source: BHPS wave 18") ///
285 legend(label(1 "1st or higher") ///
286 label(2 "HND,HNC,teaching") ///
287 label(3 "A level") ///
288 label(4 "O Level") ///
289 label(5 "CSE") ///
290 label(6 "None")) ///
291 saving("$analysisdir\PropEdu3.gph", replace)
292
293
294 * 14.3.2 Time Series Graphs
295 *-----
296
297 use DataFile.dta, clear
298
299 recode region (1/16 = 1) (17 = 2) (18 = 3) (19 = 4), gen(Country)
300 * Note that this is region, not region2 as in the previous data file

```

```

301 label var Country "Country of residence"
302 label define Country 1 "England" 2 "Wales" 3 "Scotland" 4 "Northern Ireland"
303
304 recode mastat (1 2 7 = 1) (3/6 8/10 = 0), gen(Married)
305
306 * Totals and proportions by row (i.e. country)
307 bysort Country wave: egen T_NotMarried = total(Married==0)
308 bysort Country wave: egen T_Married = total(Married==1)
309
310 * Collapse
311 collapse T_*, by(Country wave)
312
313 generate PropNotMarried = T_NotMarried / (T_NotMarried + T_Married)
314 generate PropMarried = T_Married / (T_NotMarried + T_Married)
315
316 label var PropNotMarried "Proportion of people not married or cohabiting"
317 label var PropMarried "Proportion of people married or cohabiting"
318
319 keep wave Country PropNotMarried PropMarried
320 drop if Country == .
321
322 label value Country Country
323 * We have already defined the label above,
324 * here we only need to attach it to this new variable
325
326 generate year = wave + 1990
327
328 tsset Country year, yearly
329
330 twoway (tsline PropMarried, lcolor(black)) ///
331        (tsline PropNotMarried, lpattern(dash)) ///
332        if Country == 1, ///
333        ytitle(Proportion) ytitle(, size(medsmall)) ///
334        ttitle(Year) ttitle(, size(medsmall)) ///
335        scheme(slmanual) ///
336        legend(cols(1) nobox region(lpattern(blank)))
337 graph save Graph "$analysisdir\Graph1.gph", replace
338
339
340 * 14.4. Saving Regression Results
341 *-----
342
343 * 14.4.1. Saving Results of a Linear Regression
344 *-----
345
346 use DatasetR.dta, clear
347
348 regress LnW age age2 Female Married Q1-Q5, vce(robust)
349 estimates store R_OLS1
350
351 regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12, vce(robust)
352 estimates store R_OLS2
353
354 estout R_OLS* ///
355     using "$analysisdir\Wages.out", ///
356     keep (age age2 Female Married Q1 Q2 Q3 Q4 Q5) ///
357     cells(b(star fmt(%9.3f)) se(par fmt(%9.3f))) ///
358     style(tab) stats(r2 N, fmt(%9.3f %9.0g)) ///
359     labels(R2 Observations) label collabels(, none) ///
360     starlevels(+ 0.05 * 0.01) ///
361     postfoot("Robust std.err. in parenthesis + Sign. at 5% * Sign. at 1%") ///
362     replace
363
364 * 14.4.2. Saving Marginal Effects
365 *-----
366
367 recode jbstat (1 2 = 1) (3/10 = 0), gen(Employed)
368 label define Employed 1 "Employed/Self-Employed" 0 "Unemployed or Inactive"
369 label value Employed Employed
370
371 generate Variable = ""
372 replace Variable = "age" in 1
373 replace Variable = "age2" in 3
374 replace Variable = "Married" in 5
375 replace Variable = "Q1" in 7

```

```

376 replace Variable = "Q2" in 9
377 replace Variable = "Q3" in 11
378 replace Variable = "Q4" in 13
379 replace Variable = "Q5" in 15
380 replace Variable = "Log likelihood" in 18
381 replace Variable = "Observations" in 19
382 * We leave cells empty for the standard errors
383
384 * Men
385 probit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0
386 margins, dydx(age age2 Married Q1 Q2 Q3 Q4 Q5)
387
388 generate Men = e(11) in 18
389 mat R = r(table)
390
391 local i = 1
392 local j = 1
393 while `i' <= 8 {
394     * Save the coefficient/marginal effect:
395     replace Men = R[1,`i'] in `j'
396     local j = `j' + 1
397     * Save standard error in row below marginal effect:
398     replace Men = R[2,`i'] in `j'
399     local j = `j' + 1
400     local i = `i' + 1
401 }
402 * Observations
403 mat N = r(_N)
404 replace Men = N[1,1] in 19
405
406 * Women
407 probit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 1
408 margins, dydx(age age2 Married Q1 Q2 Q3 Q4 Q5)
409
410 generate Women = e(11) in 18
411 mat R = r(table)
412
413 local i = 1
414 local j = 1
415 while `i' <= 8 {
416     * Coefficients
417     replace Women = R[1,`i'] in `j'
418     local j = `j' + 1
419     * Standard errors
420     replace Women = R[2,`i'] in `j'
421     local j = `j' + 1
422     local i = `i' + 1
423 }
424 * Observations
425 mat N = r(_N)
426 replace Women = N[1,1] in 19
427
428 * Save into table *
429
430 keep Variable Men Women
431 keep in 1/19
432 export excel using "$analysisdir\MarginalEffects", firstrow(varlabels) replace
433
434
435 * Using margins and estpost with estout and esttab
436 use DatasetR.dta, clear
437 recode jbstat (1 2 = 1) (3/10 = 0), gen(Employed)
438 label define Employed 1 "Employed/Self-Employed" 0 "Unemployed or Inactive"
439 label value Employed Employed
440
441 probit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 1
442 estpost margins, dydx(age age2 Married Q1 Q2 Q3 Q4 Q5)
443 estimates store employed
444
445 estout employed using "$analysisdir\MarginalEffects2.txt", replace
446 esttab employed using "$analysisdir\MarginalEffects3.txt", replace cell("b se")
447
448
449
450 * 14.5 Graphs of Regression Results

```

```

451 *****
452
453 * 14.5.1 The Command parmes
454 *-----
455
456 use DatasetR.dta, clear
457
458 regress LnW age age2 Female Married Q1-Q5, vce(robust)
459 parmes, format(estimate min95 max95) ///
460     label saving("$analysisdir\RegGraph", replace)
461
462 use "$analysisdir\RegGraph", clear
463 drop if parm == "_cons"
464 * Otherwise the graph would look too small
465
466 sencode parm, gen(parmid)
467 eclplot estimate min95 max95 parmid, ///
468     horizontal xline(0) scheme(smanual) ///
469     ylabel(1 "age" 2 "age square" 3 "female" ///
470         4 "married" 5 "Q1" 6 "Q2" 7 "Q3" 8 "Q4" ///
471         9 "Q5", labsize(small)) ///
472     title(Wage Regression)
473
474
475 * 14.5.2 The Command marginsplot
476 *-----
477
478 use DatasetR.dta, clear
479
480 recode jbstat (1 2 = 1) (3/10 = 0), gen(Employed)
481 label define Employed 1 "Employed/Self-Employed" 0 "Unemployed or Inactive"
482 label value Employed Employed
483
484 rename qfachi edu
485
486 label define Female 1 "Women" 0 "Men"
487 label value Female Female
488
489
490 * Models with polynomials
491 probit Married c.age##c.age i.Female ib7.edu ib7.region2 nchild i.Employed ///
492     if age>=18 & age<60, vce(robust)
493
494 margins, at(age = (25(1)45))
495 marginsplot, scheme(smanual) xlabel(25(5)45) ///
496     xtitle("Age as of date of interview") ///
497     note("Source: BHPS wave 18") ///
498     saving("$analysisdir\predprob_age", replace)
499
500
501 margins, dydx(age) at(age = (25(1)45))
502 marginsplot, scheme(smanual) xlabel(25(5)45) ///
503     xtitle("Age as of date of interview") ///
504     note("Source: BHPS wave 18") ///
505     saving("$analysisdir\ame_age", replace)
506
507 quietly margins Female#Employed, at(age = (25(5)45))
508 marginsplot, scheme(smanual) xlabel(25(5)45) ///
509     legend(row(4)) xtitle("Age as of date of interview") ///
510     note("Source: BHPS wave 18") ///
511     title("Predicted probability of being married with 95% CIs") ///
512     saving("$analysisdir\predprob_sexempagel", replace)
513
514
515
516 * 14.5.3 The Command marginsplot and interaction effects
517 *-----
518 * models with interaction effects: one continuous and one categorical variables
519
520
521 probit Married c.age##i.Female i.Employed ib7.edu ib7.region2 nchild ///
522     if age>=18 & age<60, vce(robust)
523 quietly margins Female, at(age = (25(5)45))
524 marginsplot, scheme(smanual) xlabel(25(5)45) ///
525     xtitle("Age as of date of interview") ///

```

```

526     note("Source: BHPS wave 18") ///
527     title("Predicted prob. of being married with 95% CIs") ///
528     saving("$analysisdir\predprob_sex", replace)
529
530
531 * models with interaction effects: two categorical variables
532 probit Married c.age##c.age i.Female##i.Employed ib7.edu ib7.region2 nchild ///
533     if age>=18 & age<60, vce(robust)
534 quietly margins, at(age = (25(5)45)) over(Female Employed)
535
536 marginsplot, scheme(slmanual) ///
537     legend(row(4)) xtitle("Age as of date of interview") ///
538     note("Source: BHPS wave 18") ///
539     title("Predicted probability of being married" "with 95% CIs") ///
540     saving("$analysisdir\predprob_sexempage2", replace)
541
542
543
544 * 14.5.4 Combining multiple graphs
545 *-----
546
547 quietly margins Female if Employed==0, at(age = (25(5)45))
548 marginsplot, scheme(slmanual) name(gr1) ///
549 yscale(r(.3(.2).9)) ylabel(.3(.2).9) title("Not employed")
550
551 quietly margins Female if Employed==1, at(age = (25(5)45))
552 marginsplot, scheme(slmanual) name(gr2) ///
553 yscale(r(.3(.2).9)) ylabel(.3(.2).9) title("Employed")
554
555
556 graph combine gr1 gr2, ///
557 title("Prediced Prob. of Being Married", size(medium)) ///
558 scheme(slmanual) note("Source: BHPS wave 18") ///
559 saving("$analysisdir\predprob_sexempage3", replace)
560
561 grc1leg gr1 gr2, ///
562 title("Predictive Probability of Being Married with 95% CIs", size(medium)) ///
563 scheme(slmanual) note("Source: BHPS wave 18") ///
564 saving("$analysisdir\predprob_sexempage4", replace)
565
566 graph drop _all
567 log close
568

```