

```
-----
-----
name: <unnamed>
log: C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\ex1.log
log type: text
opened on: 8 Nov 2011, 09:29:19
```

```
. * change the path *
.
. *****
. * Load the panel data and tell Stata it is panel data *
. *****
.
. use "`working'\longperson_unbal.dta",replace
```

```
. * Change the path *
.
. * covert a string variable to the numeric variable *
. cap drop id

. destring xwaveid, gen (id)
xwaveid has all characters numeric; id generated as long
```

```
. tsset id wave
panel variable: id (unbalanced)
time variable: wave, 1 to 9, but with gaps
delta: 1 unit
```

```
. describe, short
```

```
Contains data from C:\Users\nw\Documents\HILDA Project\Training\Melbourne
Nov 2011\Exercises\longperson_unbal.dta
```

```
obs: 116,685
vars: 51 8 Nov 2011 09:29
size: 14,352,255 (97.4% of memory free)
Sorted by: id wave
Note: dataset has changed since last saved
```

```
.
. *****
. * Look at the structure of the data using -xtdes - *
. * List the most common 60 patterns of panel participation *
. * is the panel balanced? is it compact ? *
. * How many observations are there per wave? *
. *****
. xt des, patterns(60)
```

```
id: 100001, 100002, ..., 901101 n =
20453
wave: 1, 2, ..., 9 T =
9
Delta(wave) = 1 unit
Span(wave) = 9 periods
(id*wave uniquely identifies each observation)
```

Distribution of T<sub>i</sub>:      min      5%      25%      50%      75%      95%

max

                         1      1      2      7      9      9

9

Freq.	Percent	Cum.	Pattern
7721	37.75	37.75	111111111
1346	6.58	44.33	1.....
854	4.18	48.51	.....1
729	3.56	52.07	11.....
554	2.71	54.78	111.....
526	2.57	57.35	.....11
475	2.32	59.67	.....1111
436	2.13	61.81	.....111
396	1.94	63.74	....11111
374	1.83	65.57	1111.....
369	1.80	67.37	.11111111
313	1.53	68.90	11111111.
310	1.52	70.42	...111111
305	1.49	71.91	111111...
301	1.47	73.38	11111.....
300	1.47	74.85	..1111111
271	1.32	76.17	.1.....
269	1.32	77.49	1111111..
239	1.17	78.66	..1.....
203	0.99	79.65	....1....
189	0.92	80.57	...1.....
187	0.91	81.49	1.1111111
173	0.85	82.34	11.111111
170	0.83	83.17	.....1...
162	0.79	83.96	.....1..
161	0.79	84.75	.....1.
161	0.79	85.53	111.11111
106	0.52	86.05	1111111.1
99	0.48	86.53	1111.1111
96	0.47	87.00	.....11..
90	0.44	87.44	111111.11
86	0.42	87.86	.11.....
83	0.41	88.27	..11.....
78	0.38	88.65	....11...
77	0.38	89.03	.....11.
75	0.37	89.40	11111.111
65	0.32	89.71	11..11111
61	0.30	90.01	...11....
55	0.27	90.28	.111.....
55	0.27	90.55	1.1.....
52	0.25	90.80	.....111.
48	0.23	91.04	1..111111
43	0.21	91.25	..111.....
43	0.21	91.46	1...11111
42	0.21	91.66	11.1.....
41	0.20	91.86	111111..1
39	0.19	92.05	....1111.
33	0.16	92.22	....111..
31	0.15	92.37	111..1111
30	0.15	92.51	..1111...
30	0.15	92.66	111.11...

29	0.14	92.80		.1111....
29	0.14	92.94		1.11.....
27	0.13	93.08		111.1....
24	0.12	93.19		...1111..
24	0.12	93.31		11111..11
23	0.11	93.42		.111111..
22	0.11	93.53		...111...
22	0.11	93.64		...1111.1
22	0.11	93.75		...11111.
1279	6.25	100.00		(other patterns)
-----+-----				
20453	100.00			XXXXXXXXXX

. tab wave

wave	Freq.	Percent	Cum.
1	13,969	11.97	11.97
2	13,041	11.18	23.15
3	12,728	10.91	34.06
4	12,408	10.63	44.69
5	12,759	10.93	55.62
6	12,905	11.06	66.68
7	12,789	10.96	77.64
8	12,785	10.96	88.60
9	13,301	11.40	100.00
-----+-----			
Total	116,685	100.00	

```
.
. *****
. * look at key variables: weekly gross wages&salary(wscei),hours*
. * per week usually worked(jbhruc),and highest qualification *
. * (edhigh). Use summarize with detail option for continuous *
. * variables and tabulate for categorical variables *
. *****
. su wscei, de //have -ve values
```

DV: Current weekly gross wages & salary, all jobs,  
includes estimated from net (

Percentiles	Smallest		
1%	-3	-6	
5%	0	-6	
10%	0	-6	Obs 116685
25%	0	-6	Sum of Wgt. 116685
-----			
50%	184		Mean 463.6091
		Largest	Std. Dev. 635.3012
75%	780	10293	
90%	1245	10293	Variance 403607.6
95%	1592	10293	Skewness 2.626671
99%	2644	10293	Kurtosis 18.62156

. su wscei, de

DV: Current weekly gross wages & salary - all jobs  
(\$) [imputed] [weighted topco

```
-----
```

Percentiles	Smallest		
1%	0	0	
5%	0	0	
10%	0	0	Obs 116685
25%	0	0	Sum of Wgt. 116685
50%	250		Mean 481.1738
		Largest	Std. Dev. 640.7189
75%	800	9687	
90%	1250	9687	Variance 410520.8
95%	1600	9687	Skewness 2.572841
99%	2685	9687	Kurtosis 17.85833

. su jbhruc, de //have -ve values

DV: Hours per week usually worked in all jobs

```
-----
```

Percentiles	Smallest		
1%	-1	-6	
5%	-1	-6	
10%	-1	-6	Obs 116685
25%	-1	-6	Sum of Wgt. 116685
50%	24		Mean 23.14987
		Largest	Std. Dev. 22.28545
75%	40	126	
90%	50	126	Variance 496.6412
95%	60	130	Skewness .3053373
99%	72	140	Kurtosis 1.971426

. tab edhigh,mi

DV: Highest education level   achieved	Freq.	Percent	Cum.
[1] Postgrad - masters or doctorate	3,577	3.07	3.07
[2] Grad diploma, grad certificate	5,429	4.65	7.72
[3] Bachelor or honours	13,977	11.98	19.70
[4] Adv diploma, diploma	9,747	8.35	28.05
[5] Cert III or IV	21,819	18.70	46.75
[6] Cert I or II	1,666	1.43	48.18
[7] Cert not defined	677	0.58	48.76
[8] Year 12	17,546	15.04	63.79
[9] Year 11 and below	42,195	36.16	99.96
[10] Undetermined	52	0.04	100.00
Total	116,685	100.00	

. tab esdtl, mi

DV: Labour force status - detail	Freq.	Percent
[1] Employed FT	50,302	43.11

43.11

63.67	[2] Employed PT	23,988	20.56
65.96	[3] Unemployed, looking for FT work	2,678	2.30
67.20	[4] Unemployed, looking for PT work	1,443	1.24
73.44	[5] Not in the labour force, marginally	7,286	6.24
99.98	[6] Not in the labour force, not margin	30,960	26.53
100.00	7	28	0.02
-----+			
---			
	Total	116,685	100.00

```

.
. *****
. * Do these variables contain any invalid values? *
. * There are several different ways to identify the different *
. * types of invalid cases and their code: *
. * - consult the online dictionary: *
. *http://www.melbourneinstitute.com/hilda/onlinedd/Default.aspx?varnt= *
. * - list the value labels *
. * - tabulate numerical values of invalid cases *
. *****
.
. describe wsce wscei jbhruc edhigh esdtl

```

variable name	storage type	display format	value label	variable label
wsce	int	%30.0g	AWSCE	DV: Current weekly gross wages & salary, all jobs, includes estimated from net (
wscei	int	%30.0g	AWSCEI	DV: Current weekly gross wages & salary - all jobs (\$)
jbhruc	int	%30.0g	AJBHRUC	DV: Hours per week usually worked in all jobs
edhigh	byte	%35.0g	AEDHIGH	DV: Highest education level achieved
esdtl	byte	%52.0g	AESDTL	DV: Labour force status - detail

```

. /*
> label list awscei
> label list ajbhruc
> label list aedhigh
> label list aesdtl
> */
.
. ta wsce if wsce<0,nol

```

DV: Current |  
weekly |  
gross wages |  
& salary, |

```
all jobs, |
includes |
estimated |
from net ( |
```

	Freq.	Percent	Cum.
-6	37	1.36	1.36
-4	263	9.67	11.03
-3	2,405	88.39	99.41
-1	16	0.59	100.00
Total	2,721	100.00	

. ta jbhrc if jbhrc<0, nol

```
DV: Hours |
per week |
usually |
worked in |
all jobs |
```

	Freq.	Percent	Cum.
-6	12	0.03	0.03
-4	21	0.05	0.08
-3	56	0.13	0.21
-1	42,367	99.79	100.00
Total	42,456	100.00	

. tab edhigh, nol

```
DV: Highest |
education |
level |
achieved |
```

	Freq.	Percent	Cum.
1	3,577	3.07	3.07
2	5,429	4.65	7.72
3	13,977	11.98	19.70
4	9,747	8.35	28.05
5	21,819	18.70	46.75
6	1,666	1.43	48.18
7	677	0.58	48.76
8	17,546	15.04	63.79
9	42,195	36.16	99.96
10	52	0.04	100.00
Total	116,685	100.00	

. tab esdtl, nol

```
DV: Labour |
force |
status - |
detail |
```

	Freq.	Percent	Cum.
1	50,302	43.11	43.11
2	23,988	20.56	63.67
3	2,678	2.30	65.96
4	1,443	1.24	67.20

5	7,286	6.24	73.44
6	30,960	26.53	99.98
7	28	0.02	100.00
-----			
Total	116,685	100.00	

```

. * use imputed income *
. *****
. * Assume we aren't interested in why cases are missing *
. * convert all invalid codes to missing(=.) *
. * Warning: STAT also interprets . as infinity! *
. *****
. replace jbhruc=. if jbhruc<0
(42456 real changes made, 42456 to missing)

. replace edhigh=. if edhigh<0
(0 real changes made)

. replace wscei=. if wscei<0
(0 real changes made)

.
. su wscei, de

```

DV: Current weekly gross wages & salary - all jobs  
(\$ [imputed] [weighted topco

Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	116685
25%	0	0	Sum of Wgt.	116685
50%	250		Mean	481.1738
		Largest	Std. Dev.	640.7189
75%	800	9687		
90%	1250	9687	Variance	410520.8
95%	1600	9687	Skewness	2.572841
99%	2685	9687	Kurtosis	17.85833

```

. su jbhruc, de

```

DV: Hours per week usually worked in all jobs

Percentiles		Smallest		
1%	3	0		
5%	8	0		
10%	13	0	Obs	74229
25%	27	0	Sum of Wgt.	74229
50%	40		Mean	36.9658
		Largest	Std. Dev.	16.00244
75%	45	126		
90%	55	126	Variance	256.0782
95%	60	130	Skewness	.0301273
99%	80	140	Kurtosis	3.554237

```

. tab edhigh, mi

```

DV: Highest education level achieved	Freq.	Percent	Cum.
[1] Postgrad - masters or doctorate	3,577	3.07	3.07
[2] Grad diploma, grad certificate	5,429	4.65	7.72
[3] Bachelor or honours	13,977	11.98	19.70
[4] Adv diploma, diploma	9,747	8.35	28.05
[5] Cert III or IV	21,819	18.70	46.75
[6] Cert I or II	1,666	1.43	48.18
[7] Cert not defined	677	0.58	48.76
[8] Year 12	17,546	15.04	63.79
[9] Year 11 and below	42,195	36.16	99.96
[10] Undetermined	52	0.04	100.00
Total	116,685	100.00	

```

.
. *****
. * Transitions in categorical variables. *
. * Our aim is to examine transitions in the categorical vars *
. * As a simple example, create a binary variable from esdtl *
. * to indicate part-time work *
. *
. * Create its lagged values and browse the data to see what *
. * happened if a wave is missing. *
. *****
. cap drop pt lpt

. recode esdtl(2=1) (else=0), gen(pt)
(116685 differences between esdtl and pt)

. gen lpt=1.pt
(23328 missing values generated)

.
. browse id wave jbhruc pt lpt

.
. *****
. * Examine transitions by using xttrans and tabulate. *
. * Are there any differences between the two? *
. * Create an alternative measure of lagged pt work, ignoring *
. * missing wave, and tabulate it against current pt work. *
. * what do you notice? *
. *****
. cap drop lpt1

. xttrans pt, freq //give absolute frequencies as well

RECODE of |
esdtl (DV: |
Labour | RECODE of esdtl (DV:
force | Labour force status -
status - | detail)
detail) | 0 1 | Total
-----+-----+-----
0 | 69,997 6,262 | 76,259

```

	91.79	8.21	100.00
1	6,550	13,423	19,973
	32.79	67.21	100.00
Total	76,547	19,685	96,232
	79.54	20.46	100.00

. tab lpt pt, row //give row percentages as well

```

+-----+
| Key          |
+-----+
| frequency    |
| row percentage |
+-----+

```

RECODE of esdt1 (DV: Labour force status - detail)			
lpt	0	1	Total
0	67,968	5,999	73,967
	91.89	8.11	100.00
1	6,227	13,163	19,390
	32.11	67.89	100.00
Total	74,195	19,162	93,357
	79.47	20.53	100.00

. cap drop lpt1

. by id: gen lpt1=pt[\_n-1] //lpt1=previous obs of pt  
(20453 missing values generated)

. tab lpt1 pt, row

```

+-----+
| Key          |
+-----+
| frequency    |
| row percentage |
+-----+

```

RECODE of esdt1 (DV: Labour force status - detail)			
lpt1	0	1	Total
0	69,997	6,262	76,259
	91.79	8.21	100.00
1	6,550	13,423	19,973
	32.79	67.21	100.00

Total	76,547	19,685	96,232
	79.54	20.46	100.00

```
.
. browse id wave jbhruc pt lpt lpt1
```

```
.
. *****
. * A way to get xttrans to take account of missing wave is *
. * to use fillin to create missing values explicitly where *
. * waves are missing. But note this will increase size of *
. * data set! *
. *****
. preserve //store data in original form
```

```
. describe, short
```

Contains data from C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov 2011\Exercises\longperson\_unbal.dta

```
obs:      116,685
vars:      54                               8 Nov 2011 09:29
size:     15,402,420 (97.2% of memory free)
Sorted by: id wave
Note:     dataset has changed since last saved
```

```
. fillin id wave
```

```
. describe, short
```

Contains data from C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov 2011\Exercises\longperson\_unbal.dta

```
obs:      184,077
vars:      55                               8 Nov 2011 09:29
size:     24,482,241 (95.5% of memory free)
Sorted by: id wave
Note:     dataset has changed since last saved
```

```
.
. xttrans pt, freq
```

```
RECODE of |
esdt1 (DV: |
  Labour | RECODE of esdt1 (DV:
  force | Labour force status -
  status - | detail)
  detail) |      0      1 |      Total
-----+-----+-----+-----+
      0 | 67,968  5,999 | 73,967
      | 91.89   8.11 | 100.00
-----+-----+-----+
      1 |  6,227 13,163 | 19,390
      | 32.11  67.89 | 100.00
-----+-----+-----+
  Total | 74,195 19,162 | 93,357
      | 79.47  20.53 | 100.00
```

```

.
.
. *****
. * Examine transitions in education. Are the patterns *
. * plausible? *
. *****
. cap drop ledhigh

```

```

. gen ledhigh=1.edhigh
(80693 missing values generated)

```

```

. label value ledhigh edhigh

```

```

. xttrans ledhigh,freq

```

		ledhigh					
		1	2	3	4	5	
6	ledhigh	7	8	9	10	Total	
-----+-----							
0	1	2,481	0	0	0	0	0
0.00	0	0	0	0	0	2,481	0.00
		100.00	0.00	0.00	0.00	0.00	0.00
		0.00	0.00	0.00	0.00	100.00	
-----+-----							
0	2	43	3,804	0	0	0	0
0.00	0	0	0	0	0	3,847	0.00
		1.12	98.88	0.00	0.00	0.00	0.00
		0.00	0.00	0.00	0.00	100.00	
-----+-----							
0	3	90	139	9,519	0	0	0
0.00	0	0	0	0	0	9,748	0.00
		0.92	1.43	97.65	0.00	0.00	0.00
		0.00	0.00	0.00	0.00	100.00	
-----+-----							
0	4	6	10	60	6,859	0	0
0.00	0	0	0	0	0	6,935	0.00
		0.09	0.14	0.87	98.90	0.00	0.00
		0.00	0.00	0.00	0.00	100.00	
-----+-----							
0	5	0	12	33	84	14,869	0
0.00	0	0	0	0	0	14,998	0.00
		0.00	0.08	0.22	0.56	99.14	
		0.00	0.00	0.00	0.00	100.00	
-----+-----							
1,060	6	0	1	1	2	48	0
94.73	0	0	7	1	0	1,119	4.29
		0.00	0.09	0.09	0.18	100.00	
		0.00	0.63	0.00	0.00	100.00	
-----+-----							



(n = 20453)

```

.
.
. *****
. * Now look at within VS between variation for the continuous *
. * variable wsce using xtsum *
. * use pre-imputed value
*
. *****
. xtsum wsce

```

Variable		Mean	Std. Dev.	Min	Max	
Observations						
wsce overall		463.6091	635.3012	-6	10293	N = 116685
	between		527.4345	-4.5	7221.8	n = 20453
	within		340.0576	-5782.058	8620.276	T-bar = 5.70503

```

. display r(sd_w)
340.05765

```

```

. display r(sd)
635.30119

```

```

. display r(sd_w)^2/r(sd)^2 //proportion of within variation
.28651395

```

```

. display r(sd_b)^2/r(sd)^2 //proportion of between variation
.68925162

```

```

. *****
. * look at imputed current weekly gross wages and salary *
. *****
. describe wscei

```

variable name	storage type	display format	value label	variable label
wscei	int	%30.0g	AWSCEI	DV: Current weekly gross wages & salary - all jobs (\$)
			[imputed]	[weighted topco]

```

. ta wscei if wscei<0, nol
no observations

```

```

. xtsum wscei

```

Variable		Mean	Std. Dev.	Min	Max
Observations					

```

-----+-----+-----
-----
wscei overall | 481.1738 640.7189 0 9687 | N =
116685
between | 541.8385 0 7131.556 | n =
20453
within | 326.6909 -5945.048 8132.84 | T-bar =
5.70503

```

```

.
.
. *****
. * Do a graphical cohort analysis of age-earnings profiles *
. *****
. cap drop year

. cap drop cohort

. gen year=2000+wave

. gen cohort=year-hgage //derive year of birth from age
(67392 missing values generated)

. tab cohort

```

cohort	Freq.	Percent	Cum.
1909	30	0.03	0.03
1910	33	0.03	0.05
1911	72	0.06	0.12
1912	52	0.04	0.16
1913	69	0.06	0.22
1914	81	0.07	0.29
1915	116	0.10	0.39
1916	106	0.09	0.48
1917	153	0.13	0.61
1918	177	0.15	0.76
1919	233	0.20	0.96
1920	222	0.19	1.15
1921	421	0.36	1.51
1922	482	0.41	1.93
1923	462	0.40	2.32
1924	589	0.50	2.83
1925	549	0.47	3.30
1926	782	0.67	3.97
1927	713	0.61	4.58
1928	738	0.63	5.21
1929	968	0.83	6.04
1930	921	0.79	6.83
1931	898	0.77	7.60
1932	881	0.76	8.35
1933	821	0.70	9.06
1934	1,000	0.86	9.91
1935	950	0.81	10.73
1936	1,044	0.89	11.62
1937	1,097	0.94	12.56
1938	1,090	0.93	13.50
1939	1,111	0.95	14.45

1940		1,077	0.92	15.37
1941		1,449	1.24	16.61
1942		1,286	1.10	17.72
1943		1,211	1.04	18.75
1944		1,381	1.18	19.94
1945		1,509	1.29	21.23
1946		1,359	1.16	22.40
1947		1,846	1.58	23.98
1948		1,893	1.62	25.60
1949		1,710	1.47	27.07
1950		1,720	1.47	28.54
1951		1,620	1.39	29.93
1952		1,704	1.46	31.39
1953		1,887	1.62	33.01
1954		2,066	1.77	34.78
1955		1,929	1.65	36.43
1956		2,044	1.75	38.18
1957		2,260	1.94	40.12
1958		2,285	1.96	42.08
1959		2,351	2.01	44.09
1960		2,289	1.96	46.05
1961		2,259	1.94	47.99
1962		2,714	2.33	50.31
1963		2,319	1.99	52.30
1964		2,475	2.12	54.42
1965		2,118	1.82	56.24
1966		2,191	1.88	58.12
1967		2,148	1.84	59.96
1968		2,347	2.01	61.97
1969		2,078	1.78	63.75
1970		2,258	1.94	65.68
1971		2,196	1.88	67.57
1972		1,982	1.70	69.27
1973		2,054	1.76	71.03
1974		1,955	1.68	72.70
1975		1,996	1.71	74.41
1976		1,929	1.65	76.06
1977		1,657	1.42	77.48
1978		1,505	1.29	78.77
1979		1,728	1.48	80.26
1980		1,747	1.50	81.75
1981		1,904	1.63	83.38
1982		1,939	1.66	85.05
1983		1,877	1.61	86.65
1984		2,091	1.79	88.45
1985		2,237	1.92	90.36
1986		2,079	1.78	92.15
1987		1,952	1.67	93.82
1988		1,673	1.43	95.25
1989		1,576	1.35	96.60
1990		1,252	1.07	97.68
1991		1,198	1.03	98.70
1992		728	0.62	99.33
1993		530	0.45	99.78
1994		256	0.22	100.00

-----+-----

Total		116,685	100.00
-------	--	---------	--------

```

.
. * generate real earnings - deflate by wage price index
. * (financial year, original)
. * (ABS CateNo 6345.0 Dec 2010, base year = 2009)
. * (Series ID: A2705194A, Financial Year Index;
. *      Total hourly rates of pay excluding bonuses;
. *      Australia; Private and Public; All industries)
. cap drop rwsce

. gen      rwsce=wsce/0.742 if year==2001
(170108 missing values generated)

. replace rwsce=wsce/0.766 if year==2002
(13041 real changes made)

. replace rwsce=wsce/0.793 if year==2003
(12728 real changes made)

. replace rwsce=wsce/0.821 if year==2004
(12408 real changes made)

. replace rwsce=wsce/0.852 if year==2005
(12759 real changes made)

. replace rwsce=wsce/0.887 if year==2006
(12905 real changes made)

. replace rwsce=wsce/0.923 if year==2007
(12789 real changes made)

. replace rwsce=wsce/0.962 if year==2008
(12785 real changes made)

. replace rwsce=wsce/1      if year==2009
(13301 real changes made)

.
.
. * convert year of birth to 5-year cohort groups for 1941-1985
.
. recode cohort (-999/1940=.) (1941/1945=1) (1946/1950=2) ///
>      (1951/1955=3) (1956/1960=4) (1961/1965=5) (1966/1970=6) ///
>      (1971/1975=7) (1976/1980=8) (1981/1985=9) (1986/9999=.)
(cohort: 116685 changes made)

. tab cohort

```

cohort	Freq.	Percent	Cum.
1	6,836	7.81	7.81
2	8,528	9.75	17.56
3	9,206	10.52	28.08
4	11,229	12.83	40.91
5	11,885	13.58	54.49
6	11,022	12.60	67.09
7	10,183	11.64	78.73
8	8,566	9.79	88.52
9	10,048	11.48	100.00

```
-----+-----  
Total |      87,503      100.00
```

```
.  
. * use the collapse command to replace dataset by one containing  
. * nominal & real earnings averages for age-cohort groups  
. collapse wsce rwsce, by(cohort hgage) //average earnings within age-  
cohort cells  
  
. keep if hgage>=16&hgage<=65  
(32 observations deleted)  
  
. browse //look at its effect  
  
. * create cohort-specific earnings variables and label them  
. cap drop c1/c9  
  
. forvalues c=1/9 {  
  2. gen e`c'=wsce if cohort==`c'  
  3. }  
(117 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
(114 missing values generated)  
  
. label variable e1 "1941-45"  
  
. label variable e2 "1946-50"  
  
. label variable e3 "1951-55"  
  
. label variable e4 "1956-60"  
  
. label variable e5 "1961-65"  
  
. label variable e6 "1966-70"  
  
. label variable e7 "1971-75"  
  
. label variable e8 "1976-80"  
  
. label variable e9 "1981-85"  
  
. * plot the age-earnings profiles for nominal earnings  
. graph twoway scatter e1-e9 hgage ///  
> , msize(small..) connect(1..) ytitle("earnings") ///  
> yscale(titlegap(1)) xtitle("age") xscale(range(16 65)  
titlegap(1)) ///
```

```

> legend(rows(3))

.
. * now export the graph to a Windows MetaFile for inclusion in a Word
document
.
. graph export "`working'/cohort-age-nomearnings-profiles_v2.wmf",
as(wmf) replace
(file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises/cohort-age-nomearnings-profiles_v2.wmf written in Windows
Metafile format)

.
. * and now do the same for real earnings
. cap drop reale1/reale9

. forvalues c=1/9 {
  2. gen reale`c'=rwsce if cohort==`c'
  3. }
(117 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)
(114 missing values generated)

. label variable reale1 "1941-45"

. label variable reale2 "1946-50"

. label variable reale3 "1951-55"

. label variable reale4 "1956-60"

. label variable reale5 "1961-65"

. label variable reale6 "1966-70"

. label variable reale7 "1971-75"

. label variable reale8 "1976-80"

. label variable reale9 "1981-85"

. graph twoway scatter reale1-reale9 hgage ///
> , msize(small..) connect(1..) ytitle("earnings") ///
> yscale(titlegap(1)) xtitle("age") xscale(range(16 65)
titlegap(1)) ///
> legend(rows(3))

. graph export "`working'/cohort-age-realearnings-profiles_v2.wmf",
as(wmf) replace
(file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises/cohort-age-realearnings-profiles_v2.wmf written in Windows
Metafile format)

```

```

.
. restore
.
.
. *****
. *
. *                               Additional Exercise                               *
. *                               *****                                       *
.
. *****
. * Calculate between and within variation manually and compare *
. * to xtsum. *
. * First calculate the grand mean and individual means, using *
. * the egen command *
. *****
. cap drop gmws imws

. egen gmws=mean(wsce)

. egen imws=mean(wsce), by(id) //mean for each individual

.
. *****
. * Next, calculate the deviations of individual means from *
. * grand means, and of each wsce observation from the *
. * individual-specific mean. Also calculate total deviation *
. *****
. replace gmws=. if wsce==. //don't count missing obs
(0 real changes made)

. replace imws=. if wsce==. //don't count missing obs
(0 real changes made)

. cap drop bdevws wdevws tdevws

. gen bdevws=imws-gmws //between deviation

. gen wdevws=wsce-imws //within deviation

. gen tdevws=wsce-gmws //total deviation

.
. browse id wave wsce gmws imws bdevws wdevws tdevws

.
. *****
. * square and sum the deviations to get between sum of squares, *
. * within sum of squares and total sum of squares. Calculate *
. * proportion due to within variation. *
. *****
. cap drop bssws wssws tssws

. egen bssws= total(bdevws^2)

. egen wssws= total(wdevws^2)

```

```

. egen tssws= total(tdevws^2)

.
. su bssws wssws tssws

```

Variable	Obs	Mean	Std. Dev.	Min	Max
bssws	116685	3.36e+10	0	3.36e+10	3.36e+10
wssws	116685	1.35e+10	0	1.35e+10	1.35e+10
tssws	116685	4.71e+10	0	4.71e+10	4.71e+10

```

. cap drop propw propb

. gen propw=wssws/tssws //proportion of within variation

. gen propb=bssws/tssws //proportion of between variation

```

```

. su propw propb

```

Variable	Obs	Mean	Std. Dev.	Min	Max
propw	116685	.286514	0	.286514	.286514
propb	116685	.7134861	0	.7134861	.7134861

```

.
. *****
. *                               End of additional exercise                               *
. *****
.
. save "`working'\longperson_unbal.dta",replace
file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\longperson_unbal.dta saved

```

```

. set more on

```

```

. log c
name: <unnamed>
log: C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\ex1.log
log type: text
closed on: 8 Nov 2011, 09:29:28
-----
-----
-----

```