**SOCY7706: Longitudinal Data Analysis**

**Instructor: Natasha Sarkisian**

**Describing Longitudinal Data**

Next, we examine some tools that allow us to describe change using longitudinal data. We will use an example from HRS data that focuses on employment and caregiving.

. use <http://sarkisian.net/socy7706/hrs_hours.dta>

. reshape long r@workhours80 r@poorhealth r@married r@totalpar r@siblog h@childlg r@allparhelptw, i(hhid pn) j(wave)

(note: j = 1 2 3 4 5 6 7 8 9)

Data wide -> long

-----------------------------------------------------------------------------

Number of obs. 6591 -> 59319

Number of variables 75 -> 20

j variable (9 values) -> wave

xij variables:

r1workhours80 r2workhours80 ... r9workhours80->rworkhours80

r1poorhealth r2poorhealth ... r9poorhealth-> rpoorhealth

r1married r2married ... r9married -> rmarried

r1totalpar r2totalpar ... r9totalpar -> rtotalpar

r1siblog r2siblog ... r9siblog -> rsiblog

h1childlg h2childlg ... h9childlg -> hchildlg

r1allparhelptw r2allparhelptw ... r9allparhelptw->rallparhelptw

-----------------------------------------------------------------------------

. tab wave

wave | Freq. Percent Cum.

------------+-----------------------------------

1 | 6,591 11.11 11.11

2 | 6,591 11.11 22.22

3 | 6,591 11.11 33.33

4 | 6,591 11.11 44.44

5 | 6,591 11.11 55.56

6 | 6,591 11.11 66.67

7 | 6,591 11.11 77.78

8 | 6,591 11.11 88.89

9 | 6,591 11.11 100.00

------------+-----------------------------------

Total | 59,319 100.00

To keep things simpler for now, we will keep only two time points, but use preserve to return to the full data.

. preserve

. keep if wave<3

(46137 observations deleted)

Stata provides a number of tools for analyzing panel data. The commands all begin with the prefix xt. To use these commands, we first need to tell Stata that our dataset is a panel dataset. We need to have a variable that identifies the units (for example, a country or person id) and a time variable. To set a dataset as a panel, we need to use xtset command:

. xtset hhidpn wave

panel variable: hhidpn (strongly balanced)

time variable: wave, 1 to 2

delta: 1 unit

Stata thinks the dataset is strongly balanced, meaning all units are observed at all time points (at the same time and equal number of times). But it is not true – we just have empty rows that were created when we went from wide to long format.

. xtdes

hhidpn: 10003020, 10004010, ..., 99564010 n = 6591

wave: 1, 2, ..., 2 T = 2

Delta(wave) = 1 unit

Span(wave) = 2 periods

(hhidpn\*wave uniquely identifies each observation)

Distribution of T\_i: min 5% 25% 50% 75% 95% max

2 2 2 2 2 2 2

Freq. Percent Cum. | Pattern

---------------------------+---------

6591 100.00 100.00 | 11

---------------------------+---------

6591 100.00 | XX

Xtdes also thinks all cases are complete. We will now delete those empty records to have a more accurate picture. Note that those rows are not completely empty – time-invariant variables still have values there, but the time-variant ones are empty. So we will only specify time-varying variables in the egen command:

. egen miss=rowmiss( rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw)

. tab miss

miss | Freq. Percent Cum.

------------+-----------------------------------

0 | 11,327 85.93 85.93

1 | 1,017 7.72 93.64

2 | 115 0.87 94.52

3 | 90 0.68 95.20

4 | 7 0.05 95.25

5 | 3 0.02 95.27

6 | 364 2.76 98.04

7 | 259 1.96 100.00

------------+-----------------------------------

Total | 13,182 100.00

. drop if miss==7

(259 observations deleted)

. xtset hhidpn wave

panel variable: hhidpn (unbalanced)

time variable: wave, 1 to 2

delta: 1 unit

This is more accurate now, and xtdes also shows that there are missing observations at time 2.

. xtdes

hhidpn: 10003020, 10004010, ..., 99564010 n = 6591

wave: 1, 2, ..., 2 T = 2

Delta(wave) = 1 unit

Span(wave) = 2 periods

(hhidpn\*wave uniquely identifies each observation)

Distribution of T\_i: min 5% 25% 50% 75% 95% max

1 2 2 2 2 2 2

Freq. Percent Cum. | Pattern

---------------------------+---------

6332 96.07 96.07 | 11

259 3.93 100.00 | 1.

---------------------------+---------

6591 100.00 | XX

Next, let’s examine change in a continuous variable.

. xtsum rworkhours80

Variable | Mean Std. Dev. Min Max | Observations

-----------------+--------------------------------------------+----------------

rwork~80 overall | 29.53971 22.79859 0 80 | N = 12477

between | 21.33473 0 80 | n = 6580

within | 8.392351 -10.46029 69.53971 | T-bar = 1.8962

Here we see overall standard deviation along with between and within standard deviations – between indicates the amount of variation across individuals (cross-sectional variation, or differences among individuals that are stable over time), and within indicates change over time within individuals (temporal variation). Between variation is essentially variation of average values for individuals over time, and within variation is variation in differences between values at each time point and averages for a given individual (i.e. individual’s deviation from their own overall mean). That is why the minimum and maximum differ from those for overall and between, and can be negative. Moreover, the way they calculate minimum and maximum is such that these are not just differences from the individual’s mean, but such differences plus the overall mean (in this case, 29.5). So the person who has 69.5 (maximum value) in fact only differs from his or her own mean by 69.5-29.5=40 hours. And the minimum value, -10.5, is in fact -10.5-29.5=-40. So it is fairly symmetric, which is what we would expect. Observation column shows that there are 12477 records, 6580 individuals, and an average of 1.8962 time points per person.

So this output allows us to decompose the variance in the variable we are describing into variance components -- into within-group and between-group variance (although they are expressed as standard deviations – to get variances, we’d have to square them). That does not explain anything, but it allows us to evaluate whether there is variation in group means (here, person-specific means), and how much of it. That’s why it is always a good idea to run this basic model when starting the analyses – it’s the null model of our regression analysis. If we find that there is no significant variation across individuals, then there is no need to adjust for the fact that clusters of observations come from the same individuals because all individuals are pretty much the same.

The proportion of variance due to group-level variation in means can be calculated as

ρ = S2between / (S2between + S2within)

It can be interpreted as the proportion of variance due to differences across individuals. It can also be interpreted as the average correlation between two randomly chosen time points that are in the same unit; therefore, it is also known as intra-class correlation. Here, we get:

. di 21.33473^2 / (21.33473^2 + 8.392351^2)

.86599838

So 87% of the total variance in hours of work is due to person-specific effects; the rest is due to changes that individuals experience over time.

To examine change in categorical variables, we can use both xttab and xttrans.

. xttab rmarried

Overall Between Within

rmarried | Freq. Percent Freq. Percent Percent

----------+-----------------------------------------------------

0 | 2662 21.20 1532 23.24 92.13

1 | 9895 78.80 5300 80.41 97.73

----------+-----------------------------------------------------

Total | 12557 100.00 6832 103.66 96.47

(n = 6591)

Here we can see that overall, out of all records in the data, 78.8% indicate that the person is currently married, and 21.2% indicate that the person is currently single. Between percent indicates that 80.41% of all individuals in the data were married at some point during the study (or in this case that means that they were married at either wave 1 or wave 2), and 23.24% of individuals were single at some point during the study period. The total is larger than 100 because any person who experienced both marriage and singlehood over this time period will be counted twice. Within percent indicates that among those individuals that were married at some point, they were married 97.73% of all of their data points, and among those who were single at some point, they were single 92.13% of all of their data points. The total for within is a weighted average – the number of people with at least one 0 multiplied by the proportion of 0s among these people’s records + the number of people with at least one 1 multiplied by the proportion of 1s among these people’s records, all divided by the total of those with at least one 1 and those with at least one 0. So here:

. di (1532\*.9213+5300\*.9773)/(1532+5300)

.96474262

. xttrans rmarried, freq

| rmarried

rmarried | 0 1 | Total

-----------+----------------------+----------

0 | 1,130 87 | 1,217

| 92.85 7.15 | 100.00

-----------+----------------------+----------

1 | 154 4,595 | 4,749

| 3.24 96.76 | 100.00

-----------+----------------------+----------

Total | 1,284 4,682 | 5,966

| 21.52 78.48 | 100.00

Xttrans shows transitions among statuses: so here we see that among those who were married at time point 1, 96.76% were still married at time point 2, while 3.24% were no longer married. Of those who were single at time 1, 92.85% were still single at time 2 and 7.15% were no longer single.

If we have more than 2 time points, xttab and xttrans put them all together. For example, let’s go back to all 9 waves.

. restore

Let’s once again get rid of those “empty” observations (with no data for a given wave):

. egen miss=rowmiss( rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw)

. tab miss

miss | Freq. Percent Cum.

------------+-----------------------------------

0 | 30,546 51.49 51.49

1 | 15,030 25.34 76.83

2 | 1,435 2.42 79.25

3 | 143 0.24 79.49

4 | 7 0.01 79.50

5 | 3 0.01 79.51

6 | 7,512 12.66 92.17

7 | 4,643 7.83 100.00

------------+-----------------------------------

Total | 59,319 100.00

. drop if miss==7

(4643 observations deleted)

. xtset hhidpn wave

panel variable: hhidpn (unbalanced)

time variable: wave, 1 to 9, but with gaps

delta: 1 unit

Let’s save this file:

. save hrs\_hours\_long.dta

And now we can describe the data:

. xtdes

hhidpn: 10003020, 10004010, ..., 99564010 n = 6591

wave: 1, 2, ..., 9 T = 9

Delta(wave) = 1 unit

Span(wave) = 9 periods

(hhidpn\*wave uniquely identifies each observation)

Distribution of T\_i: min 5% 25% 50% 75% 95% max

1 3 9 9 9 9 9

Freq. Percent Cum. | Pattern

---------------------------+-----------

5540 84.05 84.05 | 111111111

154 2.34 86.39 | 11.......

137 2.08 88.47 | 1........

84 1.27 89.74 | 1111.....

81 1.23 90.97 | 11111....

73 1.11 92.08 | 11111111.

69 1.05 93.13 | 111......

55 0.83 93.96 | 1111111..

49 0.74 94.70 | 111111...

349 5.30 100.00 | (other patterns)

---------------------------+-----------

6591 100.00 | XXXXXXXXX

. xtsum rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw

Variable | Mean Std. Dev. Min Max | Observations

-----------------+--------------------------------------------+----------------

rwork~80 overall | 19.67817 22.46339 0 80 | N = 46661

between | 17.10868 0 80 | n = 6587

within | 15.5263 -48.89326 90.78928 | T-bar = 7.0838

| |

rpoorh~h overall | .2340638 .4234167 0 1 | N = 47141

between | .3401185 0 1 | n = 6591

within | .2765936 -.6548251 1.122953 | T-bar = 7.15233

| |

rmarried overall | .7463865 .4350836 0 1 | N = 47115

between | .3947182 0 1 | n = 6591

within | .1958791 -.1425024 1.635275 | T-bar = 7.14838

| |

rtotal~r overall | .8884476 .8662165 0 4 | N = 46830

between | .6585949 .1111111 4 | n = 6591

within | .6045806 -2.111552 4.138448 | T-bar = 7.10514

| |

rsiblog overall | 1.616775 .6256698 0 3.555348 | N = 54595

between | .6154972 0 3.218876 | n = 6588

within | .1870047 -.0368364 3.663518 | T-bar = 8.28704

| |

hchildlg overall | 1.126299 .5481416 0 2.944439 | N = 44219

between | .5389137 0 2.876082 | n = 6272

within | .1319155 -.4992974 2.794222 | T-bar = 7.05022

| |

rallpa~w overall | 1.652933 4.103339 0 19.23077 | N = 32727

between | 2.651108 0 19.23077 | n = 6588

within | 3.228803 -14.14377 18.74695 | T-bar = 4.96767

. for var rpoorhealth rmarried : xttab X

-> xttab rpoorhealth

Overall Between Within

rpoorhe~h | Freq. Percent Freq. Percent Percent

----------+-----------------------------------------------------

0 | 36107 76.59 5995 90.96 82.10

1 | 11034 23.41 3285 49.84 50.82

----------+-----------------------------------------------------

Total | 47141 100.00 9280 140.80 71.02

(n = 6591)

-> xttab rmarried

Overall Between Within

rmarried | Freq. Percent Freq. Percent Percent

----------+-----------------------------------------------------

0 | 11949 25.36 2404 36.47 70.58

1 | 35166 74.64 5455 82.76 89.72

----------+-----------------------------------------------------

Total | 47115 100.00 7859 119.24 83.87

(n = 6591)

. for var rpoorhealth rmarried : xttrans X

-> xttrans rpoorhealth

rpoorhealt | rpoorhealth

h | 0 1 | Total

-----------+----------------------+----------

0 | 89.35 10.65 | 100.00

1 | 28.72 71.28 | 100.00

-----------+----------------------+----------

Total | 76.09 23.91 | 100.00

-> xttrans rmarried

| rmarried

rmarried | 0 1 | Total

-----------+----------------------+----------

0 | 95.77 4.23 | 100.00

1 | 3.49 96.51 | 100.00

-----------+----------------------+----------

Total | 26.10 73.90 | 100.00

It could also be helpful to continue looking at specific transitions, either taking two waves at a time or as sequences of transitions over a number of waves. The latter approach led to a technique called sequence analysis (or social sequence analysis when used in social sciences). If interested, see, for example, “Social Sequence Analysis” book by Benjamin Cornwall (2015). To see some basic sequences, let’s use a special package:

. net search sequence

(contacting http://www.stata.com)

33 packages found (Stata Journal and STB listed first)

------------------------------------------------------

st0244 from http://www.stata-journal.com/software/sj12-1

SJ12-1 st0244. Scrambled Halton sequences in Mata / Scrambled Halton

sequences in Mata / by Stanislav Kolenikov, University of Missouri,

Columbia, USA / Support: skolenik@gmail.com

st0111 from http://www.stata-journal.com/software/sj6-4

SJ6-4 st0111. Sequence analysis with Stata / Sequence analysis with Stata

/ by Christian Brzinsky-Fay, Wissenschaftszentrum Berlin / Ulrich Kohler,

Wissenschaftszentrum Berlin / Magdalena Luniak, Wissenschaftszentrum

Berlin / Support: brzinsky-fay@wz-berlin.de, kohler@wz-berlin.de,

st0103 from http://www.stata-journal.com/software/sj6-2

SJ6-2 st0103. Generating Halton sequences using Mata / Generating Halton

sequences using Mata / by David Drukker, StataCorp / Richard Gates,

StataCorp / Support: rgates@stata.com / After installation, type help

sj\_halton

dm55 from http://www.stata.com/stb/stb43

STB-43 dm55. Generating sequences and patterns of numeric data. / STB

insert by R. Mark Esman, Stata Corporation. / Support: mesman@stata.com /

After installation, see help fill.

dm44 from http://www.stata.com/stb/stb37

STB-37 dm44. Sequences of integers. / STB insert by Nicholas J. Cox,

University of Durham, UK. / Support: n.j.cox@durham.ac.uk / After

installation, see help seq.

sadi from http://teaching.sociology.ul.ie/sadi

Sequence Analysis Distance Measures / {bf: Brendan Halpin, Dept of

Sociology, Univer

--Break--

r(1);

We will install st0111 from http://www.stata-journal.com/software/sj6-4

. sqset rmarried hhidpn wave

Note: Some sequences contains gaps

Consider option -keeplongest-

Note: Some sequences have missings at the end

Consider option -rtrim-

element variable: rmarried, 0 to 1, and missings

identifier variable: hhidpn, 10003020 to 99564010

order variable: wave, 1 to 9

. sqtab

Sequence-Pa |

ttern | Freq. Percent Cum.

------------+-----------------------------------

111111111 | 2,283 41.74 41.74

000000000 | 508 9.29 51.03

111 | 209 3.82 54.85

1 | 206 3.77 58.62

11 | 194 3.55 62.17

11111 | 165 3.02 65.19

1111 | 157 2.87 68.06

11111111 | 155 2.83 70.89

111111 | 120 2.19 73.08

1111111 | 117 2.14 75.22

111111110 | 93 1.70 76.92

00 | 75 1.37 78.30

0 | 73 1.33 79.63

111000000 | 70 1.28 80.91

111110000 | 62 1.13 82.04

111111100 | 59 1.08 83.12

111111000 | 57 1.04 84.17

110000000 | 54 0.99 85.15

0000 | 52 0.95 86.10

00000 | 50 0.91 87.02

111100000 | 49 0.90 87.91

100000000 | 48 0.88 88.79

00000000 | 45 0.82 89.61

000 | 41 0.75 90.36

0000000 | 39 0.71 91.08

011111111 | 36 0.66 91.74

000000 | 24 0.44 92.17

000011111 | 22 0.40 92.58

10 | 11 0.20 92.78

110 | 11 0.20 92.98

11111110 | 11 0.20 93.18

…[output omitted]…

111110110 | 1 0.02 99.98

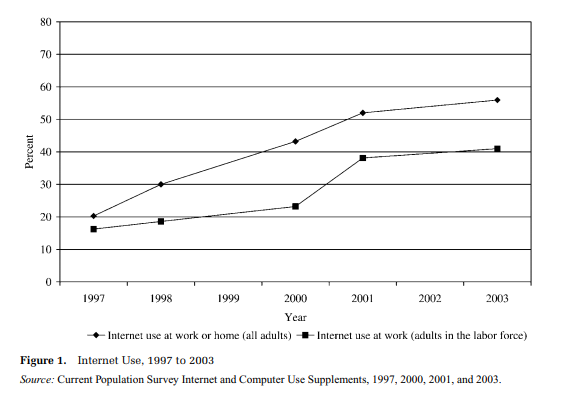
11111101 | 1 0.02 100.00

------------+-----------------------------------

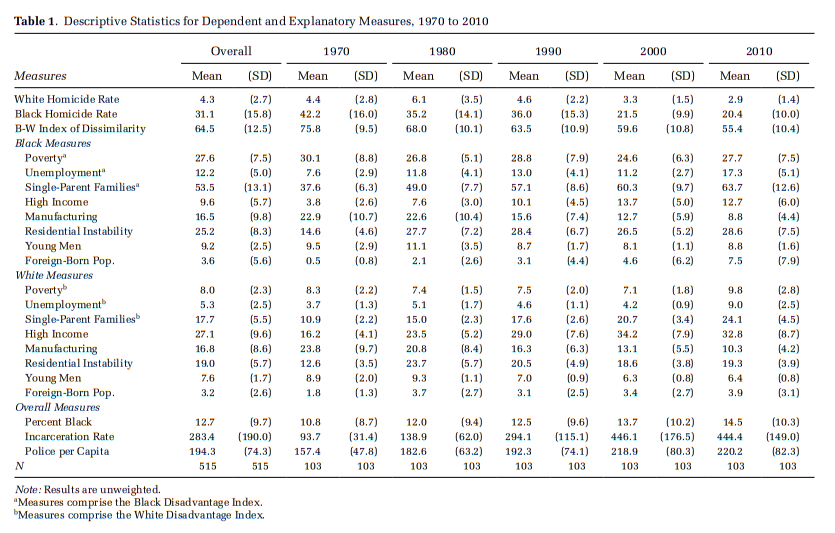
Total | 5,469 100.00

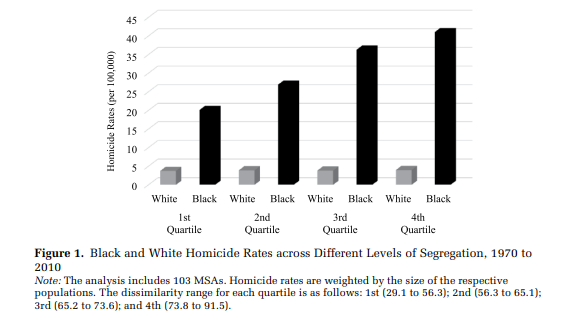
While all of these tools are helpful to better understand the nature of change in your data, such analyses rarely directly appear in articles using longitudinal data – typically, tables of descriptive statistics take one of the following approaches, or some combination:

1. Show means (and standard deviations) for each year, or some select years (e.g., the first and last year of the time sequence), or select ages – this approach aims to show how various variables in the study changed over time, but does not target change within individuals.
2. Show means and standard deviations for pooled data (entire long dataset, where units are person-years or country-years, etc.). This does not focus on change over time (either overall or individual) but rather aims at describing the dataset overall.
3. Show averages or frequencies of changes, transitions, or trajectories of individuals.

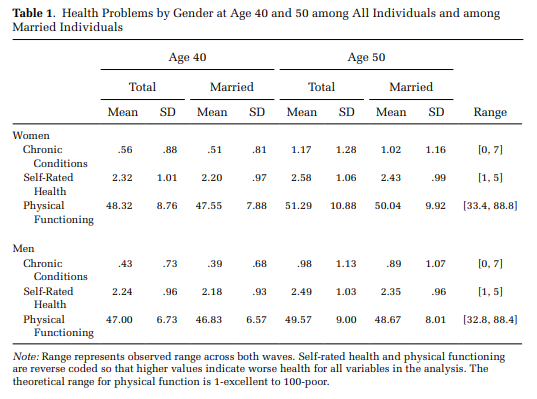
Examples:

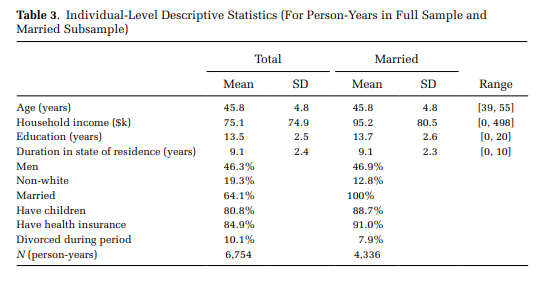
From: DiMaggio, Paul, and Bart Bonikowski. 2008. “Make Money Surfing the Web? The Impact of Internet Use on the Earnings of U.S. Workers.” *American Sociological Review* 73: 227–250.



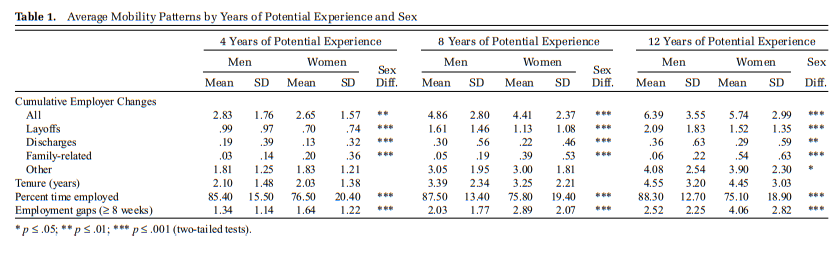


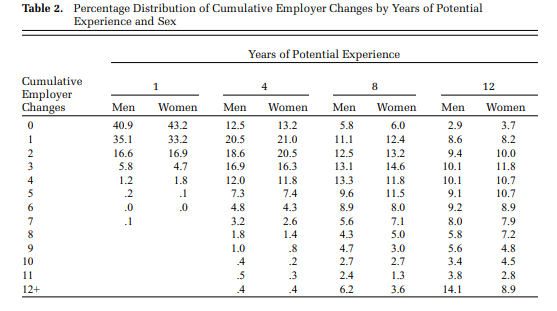
From: Light, Michael T. and Julia T. Thomas. 2019. “Segregation and Violence Reconsidered: Do Whites Benefit from Residential Segregation?” *American Sociological Review*, 84(4): 690–725.





From: Homan, Patricia. 2019. “Structural Sexism and Health in the United States: A New Perspective on Health Inequality and the Gender System.” *American Sociological Review*, 84(3): 486–516.





From: Fuller, Sylvia. 2008. “Job Mobility and Wage Trajectories for Men and Women in the United States.” *American Sociological Review* 73: 158–183.

Let’s look at similar descriptive stats and graphs for our data.

Pooled means (only for time-varying variables):

. mean rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw

Mean estimation Number of obs = 30,546

---------------------------------------------------------------

| Mean Std. Err. [95% Conf. Interval]

--------------+------------------------------------------------

rworkhours80 | 23.16339 .1316364 22.90538 23.42141

rpoorhealth | .2183592 .0023638 .213726 .2229924

rmarried | .8119885 .0022356 .8076066 .8163704

rtotalpar | 1.270019 .0044539 1.261289 1.278749

rsiblog | 1.676201 .0035233 1.669295 1.683107

hchildlg | 1.121192 .0031265 1.115064 1.127321

rallparhelptw | 1.620912 .02321 1.575419 1.666404

---------------------------------------------------------------

Time-specific means (we could then decide which waves to report):

. mean rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw, over(wave)

Mean estimation Number of obs = 30,546

1: wave = 1

2: wave = 2

3: wave = 3

4: wave = 4

5: wave = 5

6: wave = 6

7: wave = 7

8: wave = 8

9: wave = 9

---------------------------------------------------------------

Over | Mean Std. Err. [95% Conf. Interval]

--------------+------------------------------------------------

rworkhours80 |

1 | 30.61686 .2907706 30.04694 31.18679

2 | 28.20681 .3175137 27.58447 28.82915

3 | 25.67603 .3440231 25.00173 26.35033

4 | 23.1508 .3639189 22.43751 23.8641

5 | 20.20652 .3977902 19.42684 20.98621

6 | 16.09536 .4030355 15.3054 16.88533

7 | 13.44678 .4264197 12.61098 14.28258

8 | 11.57035 .4698255 10.64947 12.49123

9 | 9.932146 .50947 8.933563 10.93073

--------------+------------------------------------------------

rpoorhealth |

1 | .1954613 .0051042 .1854568 .2054658

2 | .1988658 .0054884 .1881083 .2096233

3 | .2022571 .005976 .1905439 .2139703

4 | .2434294 .0068561 .2299911 .2568677

5 | .2226708 .0073329 .2082981 .2370435

6 | .2274583 .0080763 .2116284 .2432882

7 | .246778 .0094216 .2283112 .2652448

8 | .2493719 .0108468 .2281117 .270632

9 | .2765055 .0130316 .2509631 .3020479

--------------+------------------------------------------------

rmarried |

1 | .8189498 .0049563 .8092353 .8286643

2 | .8200378 .0052823 .8096843 .8303913

3 | .8012835 .0059366 .7896475 .8129194

4 | .803011 .006354 .7905568 .8154651

5 | .810559 .0069067 .7970216 .8240964

6 | .8181818 .007431 .8036168 .8327468

7 | .8128878 .0085227 .796183 .8295927

8 | .8121859 .0097917 .7929939 .831378

9 | .7989822 .0116765 .7760957 .8218686

--------------+------------------------------------------------

rtotalpar |

1 | 1.606427 .0098604 1.5871 1.625754

2 | 1.405104 .0111955 1.38316 1.427048

3 | 1.285904 .0117704 1.262833 1.308974

4 | 1.215106 .0120179 1.19155 1.238661

5 | 1.15528 .0127173 1.130353 1.180206

6 | 1.036364 .0131864 1.010518 1.06221

7 | .9923628 .0144016 .964135 1.020591

8 | .923995 .0157687 .8930877 .9549023

9 | .8710772 .0172318 .8373021 .9048523

--------------+------------------------------------------------

rsiblog |

1 | 1.673055 .0081439 1.657092 1.689017

2 | 1.680897 .0086198 1.664002 1.697792

3 | 1.691382 .0091705 1.673407 1.709356

4 | 1.666379 .0096965 1.647374 1.685385

5 | 1.678038 .0106761 1.657112 1.698963

6 | 1.680657 .0116672 1.657788 1.703525

7 | 1.678899 .0131027 1.653217 1.704581

8 | 1.656552 .014932 1.627284 1.685819

9 | 1.652234 .0177086 1.617525 1.686944

--------------+------------------------------------------------

hchildlg |

1 | 1.107862 .0070198 1.094103 1.121621

2 | 1.118057 .0074358 1.103482 1.132631

3 | 1.109619 .0081834 1.093579 1.125659

4 | 1.124737 .0087356 1.107614 1.141859

5 | 1.131209 .0096739 1.112248 1.15017

6 | 1.129443 .0106084 1.10865 1.150236

7 | 1.140981 .0119127 1.117632 1.16433

8 | 1.136373 .0136879 1.109544 1.163202

9 | 1.13422 .0158981 1.103059 1.165381

--------------+------------------------------------------------

rallparhelptw |

1 | .6292256 .0373714 .555976 .7024751

2 | 1.228155 .0464493 1.137113 1.319198

3 | 1.584609 .0566464 1.473579 1.695638

4 | 1.896702 .0668625 1.765648 2.027755

5 | 1.763599 .0711459 1.62415 1.903048

6 | 2.296546 .0912181 2.117755 2.475338

7 | 2.563283 .1127419 2.342304 2.784261

8 | 2.68663 .1313707 2.429138 2.944122

9 | 2.63579 .1508951 2.340029 2.93155

---------------------------------------------------------------

Means of time-invariant variables:

. mean raedyrs female age white black latino otherrace minority if wave==1

Mean estimation Number of obs = 6,588

--------------------------------------------------------------

| Mean Std. Err. [95% Conf. Interval]

-------------+------------------------------------------------

raedyrs | 12.27277 .0389866 12.19634 12.3492

female | .4845173 .0061577 .4724462 .4965884

age | 55.45835 .0381301 55.3836 55.53309

white | .7328476 .0054518 .7221602 .743535

black | .153309 .0044392 .1446068 .1620113

latino | .0910747 .003545 .0841253 .0980241

otherrace | .0227687 .0018379 .0191658 .0263716

minority | .2671524 .0054518 .256465 .2778398

--------------------------------------------------------------

A graph of means of work hours over time by gender:

. reg rworkhours80 i.female##i.wave

Source | SS df MS Number of obs = 46,661

-------------+---------------------------------- F(17, 46643) = 580.49

Model | 4111549.35 17 241855.844 Prob > F = 0.0000

Residual | 19433266.7 46,643 416.638439 R-squared = 0.1746

-------------+---------------------------------- Adj R-squared = 0.1743

Total | 23544816 46,660 504.603859 Root MSE = 20.412

------------------------------------------------------------------------------

rworkhours80 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

1.female | -13.77821 .5047102 -27.30 0.000 -14.76745 -12.78897

|

wave |

2 | -2.897701 .5115941 -5.66 0.000 -3.900433 -1.894969

3 | -6.28488 .5194919 -12.10 0.000 -7.303092 -5.266668

4 | -10.22427 .5278551 -19.37 0.000 -11.25887 -9.189664

5 | -14.65338 .5361013 -27.33 0.000 -15.70415 -13.60261

6 | -19.6585 .5421223 -36.26 0.000 -20.72107 -18.59593

7 | -22.98268 .5511448 -41.70 0.000 -24.06293 -21.90242

8 | -25.93671 .5614959 -46.19 0.000 -27.03725 -24.83617

9 | -27.23357 .5705184 -47.73 0.000 -28.35179 -26.11534

|

female#wave |

1 2 | .9457776 .7320558 1.29 0.196 -.4890627 2.380618

1 3 | 1.572734 .7425156 2.12 0.034 .1173924 3.028075

1 4 | 3.481562 .7532003 4.62 0.000 2.005278 4.957846

1 5 | 5.25893 .7644375 6.88 0.000 3.760621 6.757239

1 6 | 6.813259 .7722584 8.82 0.000 5.299621 8.326897

1 7 | 8.23079 .7824995 10.52 0.000 6.69708 9.764501

1 8 | 9.026886 .7952218 11.35 0.000 7.46824 10.58553

1 9 | 9.106228 .8059588 11.30 0.000 7.526536 10.68592

|

\_cons | 37.42107 .3516128 106.43 0.000 36.7319 38.11023

------------------------------------------------------------------------------

. margins, at(wave=(1(1)8) female=(0 1))

Adjusted predictions Number of obs = 46,661

Model VCE : OLS

Expression : Linear prediction, predict()

1.\_at : female = 0

wave = 1

2.\_at : female = 0

wave = 2

3.\_at : female = 0

wave = 3

4.\_at : female = 0

wave = 4

5.\_at : female = 0

wave = 5

6.\_at : female = 0

wave = 6

7.\_at : female = 0

wave = 7

8.\_at : female = 0

wave = 8

9.\_at : female = 1

wave = 1

10.\_at : female = 1

wave = 2

11.\_at : female = 1

wave = 3

12.\_at : female = 1

wave = 4

13.\_at : female = 1

wave = 5

14.\_at : female = 1

wave = 6

15.\_at : female = 1

wave = 7

16.\_at : female = 1

wave = 8

------------------------------------------------------------------------------

| Delta-method

| Margin Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

\_at |

1 | 37.42107 .3516128 106.43 0.000 36.7319 38.11023

2 | 34.52337 .371614 92.90 0.000 33.795 35.25174

3 | 31.13619 .3824137 81.42 0.000 30.38665 31.88572

4 | 27.1968 .3936997 69.08 0.000 26.42514 27.96846

5 | 22.76769 .4046887 56.26 0.000 21.97449 23.56088

6 | 17.76257 .4126318 43.05 0.000 16.9538 18.57133

7 | 14.43839 .4244161 34.02 0.000 13.60653 15.27025

8 | 11.48436 .4377739 26.23 0.000 10.62632 12.3424

9 | 23.64286 .3620785 65.30 0.000 22.93318 24.35254

10 | 21.69093 .3782544 57.34 0.000 20.94955 22.43232

11 | 18.93071 .3877586 48.82 0.000 18.1707 19.69072

12 | 16.90015 .396962 42.57 0.000 16.1221 17.6782

13 | 14.24841 .4072582 34.99 0.000 13.45018 15.04664

14 | 10.79761 .4139875 26.08 0.000 9.986193 11.60904

15 | 8.890971 .421241 21.11 0.000 8.065332 9.71661

16 | 6.733036 .4312764 15.61 0.000 5.887728 7.578344

------------------------------------------------------------------------------

. marginsplot, noci ytitle("Average hours of paid work") xtitle("Wave") title("Paid Work by Gender Trends")

Variables that uniquely identify margins: wave female



Examining the nature of changes wave to wave:

. gen diff\_hours=d.rworkhours80

(16,141 missing values generated)

. mean diff\_hours, over(wave)

Mean estimation Number of obs = 38,535

2: wave = 2

3: wave = 3

4: wave = 4

5: wave = 5

6: wave = 6

7: wave = 7

8: wave = 8

9: wave = 9

--------------------------------------------------------------

Over | Mean Std. Err. [95% Conf. Interval]

-------------+------------------------------------------------

diff\_hours |

2 | -2.818721 .2218065 -3.253468 -2.383975

3 | -3.307249 .2321794 -3.762327 -2.852172

4 | -3.213501 .2324172 -3.669044 -2.757957

5 | -3.753209 .2310037 -4.205982 -3.300436

6 | -4.402008 .2398089 -4.872039 -3.931976

7 | -2.631661 .2151473 -3.053355 -2.209967

8 | -2.50271 .1999011 -2.894522 -2.110899

9 | -1.550472 .1853217 -1.913708 -1.187237

--------------------------------------------------------------

. gen diff\_hours\_cat=(diff\_hours>0) if diff\_hours<.

(16,141 missing values generated)

. replace diff\_hours\_cat=-1 if diff\_hours<0

(9,729 real changes made)

. tab diff\_hours\_cat, m

diff\_hours\_ |

cat | Freq. Percent Cum.

------------+-----------------------------------

-1 | 9,729 17.79 17.79

0 | 22,835 41.76 59.56

1 | 5,971 10.92 70.48

. | 16,141 29.52 100.00

------------+-----------------------------------

Total | 54,676 100.00

. tab diff\_hours\_cat wave, col

+-------------------+

| Key |

|-------------------|

| frequency |

| column percentage |

+-------------------+

diff\_hours | wave

\_cat | 2 3 4 5 6 7 | Total

-----------+------------------------------------------------------------------+----------

-1 | 1,704 1,622 1,421 1,320 1,245 946 | 9,729

| 28.90 30.15 27.88 27.33 27.17 21.09 | 25.25

-----------+------------------------------------------------------------------+----------

0 | 2,922 2,739 2,741 2,782 2,752 2,968 | 22,835

| 49.55 50.91 53.79 57.60 60.06 66.18 | 59.26

-----------+------------------------------------------------------------------+----------

1 | 1,271 1,019 934 728 585 571 | 5,971

| 21.55 18.94 18.33 15.07 12.77 12.73 | 15.50

-----------+------------------------------------------------------------------+----------

Total | 5,897 5,380 5,096 4,830 4,582 4,485 | 38,535

| 100.00 100.00 100.00 100.00 100.00 100.00 | 100.00

diff\_hours | wave

\_cat | 8 9 | Total

-----------+----------------------+----------

-1 | 823 648 | 9,729

| 19.40 16.11 | 25.25

-----------+----------------------+----------

0 | 2,969 2,962 | 22,835

| 69.97 73.64 | 59.26

-----------+----------------------+----------

1 | 451 412 | 5,971

| 10.63 10.24 | 15.50

-----------+----------------------+----------

Total | 4,243 4,022 | 38,535

| 100.00 100.00 | 100.00