

```

*****.
*** Longitudinal Data Analysis for Social Science Researchers
**
**
** ESRC Researcher Development Initiative training programme:
**
**
** Training materials lab 2:
** PANEL SURVEY DATA ANALYSIS AND DATA MANAGEMENT .
**
**
** www.longitudinal.stir.ac.uk
** Paul Lambert / Vernon Gayle, 26 August 2007
*****.

**** SPSS VERSION *****

*****.
*****.
** The file below covers examples of data management and data analysis
** primarily using data from the British Household Panel Survey 1991-2004:
**
*** Exercise 1: Reviewing the bhps data resources
*** Exercise 2: Deriving 'long format' panel data from bhps component files
*** Exercise 3: Deriving 'wide format' panel data from bhps component files
*** Exercise 4: Translating between wide and long format panel data
*** Exercise 5: Understanding the bhps: sample origins, and using weights
*** Exercise 6: Dealing with fixed-in-time bhps data
*** Exercise 7: Examples using the bhps youth files
*** Exercise 8: Further examples in matching panel data files
*** Exercise 9: Long format data: descriptive analyses
*** Exercise 10: Panel data analysis: looking at transitions / sequences
*** Exercise 11: Long format data: panel data models
*** Exercise 12: Panel data models: looking at estimation issues
**
*****.

*****.
** GENERAL INSTRUCTIONS ON THESE FILES
**
** Work through this file in the interactive do-file editor, replicating
** the STATA do-file commands. Further help on working with STATA is
** available from the LDA web site.
**
**
** This lab file assumes you have a number of files downloaded to your
** machine. You will need the following:
**
**
** 1) EITHER supplied by the workshop session instructors (taught workshops)
** OR as a product of carrying out the lab2 stata commands:
**
* bhltol5_long.sav
* bhltol5_wide.sav
**
** (derived BHPS data files obtained from merging source BHPS files in lab 2 below).
**
**
** 2) Downloadable from the UK Data Archive:
**
* -All BHPS Waves 1-15 component files in SPSS format (UK Data Archive Study number
* 5151 (June 2007 release) (extracted from the zip file 5151SPSS.ZIP)
* +warning - these are a large volume of files, 143 different files, ~ 600MB)
**
*
*- All SPSS format 'episode' files from the BHPS Derived life history files
* (UKDA study number 3954) (some cover waves 1-10 only) (you want to access

```

```

* the 3 files data files on the top directory of the '3954.zip' archive,
* called newpan.por, xleme.por, and xljobe.por, plus the 3 files in the
* 'episode' subfolder of the '3954.zip' archive, called l*.por)*

*- All BHPS derived net income files (UKDA study number 3909) (cover waves 1-12 only)
*
*
*
** 3) Downloadable from the website:
* varlabstonew.sps
* (syntax command batch files for running specific groups of commands which will
* be used later in the lab).
*
*****.
** .

*****.
** NOTIFICATION OF FILE LOCATIONS / DIRECTORIES
**
**
** i) File location declarations:
*** For the commands below to work, you should begin by running the following
** macros, which tell SPSS where to look for the relevant data files (mentioned
** above) on your machine : .
** NOTE: EDIT THE PATHS BELOW TO EQUIVALENTS APPROPRIATE TO YOUR MACHINE .
*.

define !path1 () 'd:\lda\work\' !enddefine.
* (the location of your working directory - where you will save
* newly created data files and output) .

define !path2 () 'd:\data\lda\' !enddefine.
* (the location of a folder where you have saved the
* WEBCT sourced data files mentioned above) .

define !path3 () 'd:\data\bhps\wltol5\' !enddefine.
* (the location of a folder where you have saved the original BHPS
* panel data files mentioned above) .

define !path3b () 'd:\data\bhps\derived\' !enddefine.
* (the location of a folder where you will (or already have) saved the BHPS derived
* data files bhpsltol5_long.dta and bhpsltol5_wide.dta mentioned above) .

define !path3c () 'd:\data\bhps\inc\' !enddefine.
* (the location of a folder where you have saved the BHPS derived
* income data files [study number 3909] mentioned above) .

define !path3d () 'd:\data\bhps\lifehist\' !enddefine.
* (the location of a folder where you have saved the BHPS derived
* life history data files [study number 3954] mentioned above) .

define !path8 () 'd:\lda\macros\' !enddefine.
* (the location of a folder where you have saved the SPSS command file
* macros described above [available for download from the webpages]) .

define !path9 () 'd:\temp\' !enddefine.
* (a location of a temporary folder where you can save intermediate files) .

*****.

*****.
*
* Reminder: other support materials in working with SPSS in the context
* of longitudinal survey datasets are available from the LDA website,
* http://www.longitudinal.stir.ac.uk/SPSS_support.html

```

```

*
*****

** ..finally, here is the lab exercise...

*****

***** EXERCISE 1) REVIEWING THE BHPS DATA RESOURCES .
*****

* BEFORE YOU DO ANYTHING ELSE...
* - Check the online BHPS documentation at http://iserwww.essex.ac.uk/ulsc/bhps/doc/
* (see especially volume B)

** BHPS : Getting started in SPSS : .

***** i) Core files: Opening an indresp file :.

* Most BHPS individual data is on the 'indresp' files for adult interviews in each wave.
get file=!path3+"nindresp.sav".
descriptives var=all.
** This is a huge dataset.
descriptives var=pid nhid npno.
** these are key index variables.

fre var=ndoim ndoid .
* Most of the time these variables are ignored - but in some analysis the wide window
* of observation could be important.

fre var=nsex.
fre var=nregion.
* Note the unusual distribution: the BHPS includes regional 'booster' samples.
* If you want to do a nationally representative analysis within a particular
* individual year, use an appropriate cross-sectional weight :.
weight by nxrwtkl .
fre var=nregion.
weight off.

** Rather than use the whole file, it is better to make subset extractions: .
get file=!path3+"nindresp.sav" /keep=pid nhid nsex nregion nxrwtkl .
descriptives var=all.
* (Get the variable names from http://iserwww.essex.ac.uk/ulsc/bhps/doc/ ).

***** ii) Some different individual level files :.
get file=!path3+"nindresp.sav" /keep=pid nage nivfio .
descriptives var=pid nage.
fre var=nivfio.
* 15791 adults who were given some sort of interview in 2004 wave.
get file=!path3+"nindall.sav" /keep=pid nage nivfio .
descriptives var=pid nage.
fre var= nivfio .
* 22127 people in BHPS households in w14 (most non-adult interviews are children).
get file=!path3+"nyouth.sav" /keep=pid nypdoby4 .
descriptives var=pid nypdoby4.
* 1397 youths in BHPS samples are interviewed in the youth sample.

***** iii) Core files: Household files :.

get file=!path3+"nhhresp.sav" .
descriptives var=nhid.
* 8897 different households contacted in wave 14.
fre var=nhstype nhspri .
* These household files contain a variety of household level data.

```

```

**** iv) Core files: Event oriented files :.
* (we will return to these event oriented files in lab 3).
get file=!path3+"njobhist.sav".
sort cases by pid.
descriptives var=pid.
compute first=1.
if (pid=lag(pid) ) first=0.
fre var=first.
** Each record here is an employment spell over the last 12 months .
** There are 4418 records of spells, covering 3289 different individuals.
** In fact, these are the additional spells of only people who have changed jobs
* in the last 12 months - ie, people in the same situation as 12 months ago don't have
** a record.

```

```

**** v) Core files: cross-wave files :.
get file=!path3+"xwaveid.sav".
sort cases by pid.
descriptives var=pid .
fre var=sex aivfio livfio nivfio .
cro aivfio by nivfio.
** There's one record per person who has ever been enumerated in the BHPS (n=39987),
** but many people are only interviewed in some and not other waves.
** The cross-wave files are usually used to get summary information on people over time.

```

```

*** vi) Subsidiary files : Combined life history files (Study number 3954).
* (we will return to these event oriented files in lab 3).
import file=!path3d+"ljembe.por".
sort cases by pid date.
descriptives var=all.
fre var=spellno.
cro spellno by stemp.
* This file has a single record for every recorded distinct employment history event in each
* BHPS respondent's life history (assuming the data was collected for them).
* The length of the record is derived from information on the starting and ending time.
* There are 84011 events in all, covering 21815 individuals (spellno=1).
* Spellno is the number of event for each individual .
sort cases by pid spellno.
aggregate outfile=!path9+"mtchl.sav" /break=pid /topsp=max(spellno).
get file=!path9+"mtchl.sav" .
fre var=topsp.
* Most people have at most 10 spells, but a few have more, with the highest being
* a respondent with 41 distinctive employment status events recorded.

```

```

***** vii) Subsidiary files: Derived income files (Study number 3909).

import file=!path3c+"l_nethh.por".
descriptives var=all.
** this is a household level file with information on the household income circumstances
** additional to that released in the main survey (after derivations accounting, for instance,
* for local tax rates).
** (Derived income files currently cover waves 1-12 only).

```

```

*****
*****
*****

```

```

***** EXERCISE 2) DERIVING 'LONG FORMAT' PANEL DATA FROM BHPS COMPONENT FILES .
*****

```

```

** [THIS EXERCISE GENERATES THE COMBINED FILES BHPS1015_WIDE.SAV AND
**   BHPS1015_LONG.SAV WHICH ARE USED IN OTHER PARTS OF THESE LABS]

```

```

*****.
*** This Exercise concerns extracting related data from multiple
**   BHPS Waves (1-15 individual level + 1-15 household level data)

```

```

* (The first bit of syntax just produces 15 smaller wave-specific extracts,
*   which are then used, in the next section, to derived merged longitudinal files).

```

```

*****
**** Segment 2.1) Make 15 successive temporary extract files:
*****.

```

```

** BHPS wave 1 : .
get file=!path3+"ahhsamp.sav" /keep=ahid apsu astrata .
sort cases by ahid .
sav out=!path9+"mtchl.sav".
get file=!path3+"aindresp.sav" /keep=pid ahid apno asex aage aregion aqfedhi avote
  axrwght axewght ajbcssm ajbrgsc afimm ahlgql .
sort cases by ahid.
match files file=* /table=!path9+"mtchl.sav" /by=ahid .
compute wave=1.
compute year=1991 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"aindresp_extl.sav".

```

```

** BHPS wave 2 : .
get file=!path3+"bhhsamp.sav" /keep=bhid bpsu bstrata .
sort cases by bhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"bindresp.sav" /keep=pid bhid bpno bsex bage bregion bqfedhi bvote4
  bxrwght bxewght blrwght blewght bjbcssm bjbrgsc bfimm bhlghql /rename(bvot4=bvot4) .
sort cases by bhid.
match files file=* /table=!path9+"mtchl.sav" /by=bhid .
compute wave=2.
compute year=1992 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"bindresp_extl.sav".

```

```

** BHPS wave 3 : .
get file=!path3+"chhsamp.sav" /keep=chid cpsu cstrata .
sort cases by chid .
sav out=!path9+"mtchl.sav".
get file=!path3+"cindresp.sav" /keep=pid chid cpno csex cage cregion cqfedhi cvote
  cxrwght cxewght clrwght clewght cjbcssm cjbrgsc cfimm chlghql .
sort cases by chid.
match files file=* /table=!path9+"mtchl.sav" /by=chid .
compute wave=3.
compute year=1993 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"cindresp_extl.sav".

```

```

** BHPS wave 4 : .
get file=!path3+"dhhsamp.sav" /keep=dhid dpsu dstrata .
sort cases by dhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"dindresp.sav" /keep=pid dhid dpno dsex dage dregion dqfedhi dvote
  dxrwght dxewght dlrwght dlewght djbcssm djbrgsc dfimm dhlghql .
sort cases by dhid.
match files file=* /table=!path9+"mtchl.sav" /by=dhid .
compute wave=4.
compute year=1994 .

```

```

sort cases by pid.
descriptives var=all .
sav out=!path9+"dindresp_extl.sav".

```

```

** BHPS wave 5 : .
get file=!path3+"ehhsamp.sav" /keep=ehid epsu estrata .
sort cases by ehid .
sav out=!path9+"mtchl.sav".
get file=!path3+"eindresp.sav" /keep=pid ehid epno esex eage eregion eqfedhi evote
  exrwght exewght elrwght elewght ejbcssm ejbrgsc efimm ehlgql .
sort cases by ehid.
match files file=* /table=!path9+"mtchl.sav" /by=ehid .
compute wave=5.
compute year=1995 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"eindresp_extl.sav".

```

```

** BHPS wave 6 : .
get file=!path3+"fhhsamp.sav" /keep=fhid fpsu fstrata .
sort cases by fhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"findresp.sav" /keep=pid fhid fpno fsex fage fregion fqfedhi fvote
  fxrwght fxewght flrwght flewght fjbcssm fjbrgsc fimm fhlgql .
sort cases by fhid.
match files file=* /table=!path9+"mtchl.sav" /by=fhid .
compute wave=6.
compute year=1996 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"findresp_extl.sav".

```

```

** BHPS wave 7 : .
get file=!path3+"ghhsamp.sav" /keep=ghid gpsu gstrata .
sort cases by ghid .
sav out=!path9+"mtchl.sav".
get file=!path3+"gindresp.sav" /keep=pid ghid gpno gsex gage gregion gqfedhi gvote
  gxrwght gxewght glrwght glewght gjbcssm gjbrgsc gimm ghlgql .
sort cases by ghid.
match files file=* /table=!path9+"mtchl.sav" /by=ghid .
compute wave=7.
compute year=1997 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"gindresp_extl.sav".

```

```

** BHPS wave 8 : .
get file=!path3+"hhhsamp.sav" /keep=hhid hpsu hstrata .
sort cases by hhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"hindresp.sav" /keep=pid hhid hpno hsex hage hregion hqfedhi hvote
  hxrwght hxewght hlrwght hlewght hjbcssm hjbrgsc hfimm hhlghql .
sort cases by hhid.
match files file=* /table=!path9+"mtchl.sav" /by=hhid .
compute wave=8.
compute year=1998 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"hindresp_extl.sav".

```

```

** BHPS wave 9 : .
get file=!path3+"ihhsamp.sav" /keep=ihid ipsu istrata .
sort cases by ihid .
sav out=!path9+"mtchl.sav".
get file=!path3+"iindresp.sav" /keep=pid ihid ipno isex iage iregion iqfedhi ivote
  ixrwght ixewght ilrwght ilewght ijbcssm ijbrgsc ifimm ihlghql .
sort cases by ihid.
match files file=* /table=!path9+"mtchl.sav" /by=ihid .
compute wave=9.
compute year=1999 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"iindresp_extl.sav".

```

```

** BHPS wave 10 : .
get file=!path3+"jhhsamp.sav" /keep=jhid jpsu jstrata .
sort cases by jhid .

```

```

sav out=!path9+"mtchl.sav".
get file=!path3+"jindresp.sav" /keep=pid jhid jpno jsex jage jregion jqfedhi jvote
  jxrwght jxewght jlrwght jlewght jjbcssm jjbrgsc jfimm jhlghql .
sort cases by jhid.
match files file=* /table=!path9+"mtchl.sav" /by=jhid .
compute wave=10.
compute year=2000 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"jindresp_extl.sav".

** BHPS wave 11: .
get file=!path3+"khhsamp.sav" /keep=khid kpsu kstrata .
sort cases by khid .
sav out=!path9+"mtchl.sav".
get file=!path3+"kindresp.sav" /keep=pid khid kpno ksex kage kregion kqfedhi kvote
  kxrwght kxewght klrwght klewght kjbcssm kjbrgsc kfimm khlgghql .
sort cases by khid.
match files file=* /table=!path9+"mtchl.sav" /by=khid .
compute wave=11.
compute year=2001 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"kindresp_extl.sav".

** BHPS wave 12: .
get file=!path3+"lhhsamp.sav" /keep=lhid lpsu lstrata .
sort cases by lhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"lindresp.sav" /keep=pid lhid lpno lsex lage lregion lqfedhi lvote
  lxrwght lxewght llrwght llewght ljbcssm ljbrgsc lfimm lhlghql .
sort cases by lhid.
match files file=* /table=!path9+"mtchl.sav" /by=lhid .
compute wave=12.
compute year=2002 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"lindresp_extl.sav".

** BHPS wave 13: .
get file=!path3+"mhhsamp.sav" /keep=mhid mpsu mstrata .
sort cases by mhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"mindresp.sav" /keep=pid mhid mpno msex mage mregion mqfedhi mvote
  mxrwght mxewght mlrwght mlewght mjbcssm mjbrgsc mfimm mhlghql .
sort cases by mhid.
match files file=* /table=!path9+"mtchl.sav" /by=mhid .
compute wave=13.
compute year=2003 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"mindresp_extl.sav".

** BHPS wave 14: .
get file=!path3+"nhhsamp.sav" /keep=nhid npsu nstrata .
sort cases by nhid .
sav out=!path9+"mtchl.sav".
get file=!path3+"nindresp.sav" /keep=pid nhid npno nsex nage nregion nqfedhi nvote
  nxrwght nxewght nlrwght nlewght njbcssm njbrgsc nfimm nhlghql .
sort cases by nhid.
match files file=* /table=!path9+"mtchl.sav" /by=nhid .
compute wave=14.
compute year=2004 .
sort cases by pid.
descriptives var=all .
sav out=!path9+"nindresp_extl.sav".

** BHPS wave 15: .
get file=!path3+"ohhsamp.sav" /keep=ohid opsu ostrata .
sort cases by ohid .
sav out=!path9+"mtchl.sav".
get file=!path3+"oindresp.sav" /keep=pid ohid opno osex oage oregion oqfedhi ovote
  oxrwght oxewght olrwght olewght ojbcssm ojbrgsc ofimm ohlghql .
sort cases by ohid.
match files file=* /table=!path9+"mtchl.sav" /by=ohid .
compute wave=15.
compute year=2005 .

```

```

sort cases by pid.
descriptives var=all .
sav out=!path9+"oindresp_extl.sav".

```

* We have produced 15 extract files.

*** Segment 2.2) Merging Panel data in long format .

** Via the extract files created above (note the need to rename variables
* so that they have the same name on the combined file).

```

get file=!path9+"aindresp_extl.sav"
/rename (ahid apsu astrata apno asex aage aregion aqfedhi avote
  axrwght axewght ajbcssm ajbrgsc afimm ahlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl.sav" .
get file=!path9+"bindresp_extl.sav"
/rename (bhid bpsu bstrata bpno bsex bage bregion bqfedhi bvote
  bxrwght bxewght blrwght blewght bjbcssm bjbrgsc bfimm bhlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch2.sav" .
get file=!path9+"cindresp_extl.sav"
/rename (chid cpsu cstrata cpno csex cage cregion cqfedhi cvote
  cxrwght cxewght clrwght clewght cjbcssm cjbrgsc cfimm chlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch3.sav" .
get file=!path9+"dindresp_extl.sav"
/rename (dhid dpsu dstrata dpno dsex dage dregion dqfedhi dvote
  dxrwght dxewght dlrwght dlewght djbcssm djbrgsc dfimm dhlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch4.sav" .
get file=!path9+"eindresp_extl.sav"
/rename (ehid epsu estrata epno esex eage eregion eqfedhi evote
  exrwght exewght elrwght elewght ejbcssm ejbrgsc efimm ehlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch5.sav" .
get file=!path9+"findresp_extl.sav"
/rename (fhid fpsu fstrata fpno fsex fage fregion fqfedhi fvote
  fxrwght fxewght flrwght flewght fjbcssm fjbrgsc ffimm fhlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch6.sav" .
get file=!path9+"gindresp_extl.sav"
/rename (ghid gpsu gstrata gpno gsex gage gregion gqfedhi gvote
  gxrwght gxewght glrwght glewght gjbcssm gjbrgsc gfimm ghlgghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch7.sav" .
get file=!path9+"hindresp_extl.sav"
/rename (hhid hpsu hstrata hpno hsex hage hregion hqfedhi hvote
  hxrwght hxewght hlrwght hlewght hjbcssm hjbrgsc hfimm hhlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch8.sav" .
get file=!path9+"iindresp_extl.sav"
/rename (ihid ipsu istrata ipno isex iage iregion iqfedhi ivote
  ixrwght ixewght ilrwght ilewght ijbcssm ijbrgsc ifimm ihlghql =
  zhid zpsu zstrata zpno zsex zage zregion zqfedhi zvote
  zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtch9.sav" .
get file=!path9+"jindresp_extl.sav"
/rename (jhid jpsu jstrata jpno jsex jage jregion jqfedhi jvote

```

```

jxrwght jxewght jlrwght jlewght jbcssm jbrgsc jfimm jhlghql =
zhid zpsu zstrata zpnz zsex zage zregion zqfedhi zvte
zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl0.sav" .
get file=!path9+"kindresp_extl.sav"
/renam (khid kpsu kstrata kpno ksex kage kregion kqfedhi kvote
kxrwght kxewght klrwght klewght kjbcssm kjbrgsc kfimm khlghql =
zhid zpsu zstrata zpnz zsex zage zregion zqfedhi zvte
zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl1.sav" .
get file=!path9+"lindresp_extl.sav"
/renam (lhid lpsu lstrata lpno lsex lage lregion lqfedhi lvote
lxrwght lxewght llrwght llewght ljbcssm ljbrgsc lfimm lhlghql =
zhid zpsu zstrata zpnz zsex zage zregion zqfedhi zvte
zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl2.sav" .
get file=!path9+"mindresp_extl.sav"
/renam (mhid mpsu mstrata mpno msex mage mregion mqfedhi mvote
mxrwght mxewght mlrwght mlewght mjbcssm mjbrgsc mfimm mhlghql =
zhid zpsu zstrata zpnz zsex zage zregion zqfedhi zvte
zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl3.sav" .
get file=!path9+"nindresp_extl.sav"
/renam (nhid npsu nstrata npno nsex nage nregion nqfedhi nvote
nxrwght nxewght nlrwght nlewght njbcssm njbrgsc nfimm nhlghql =
zhid zpsu zstrata zpnz zsex zage zregion zqfedhi zvte
zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl4.sav" .
get file=!path9+"oindresp_extl.sav"
/renam (ohid opsu ostrata opno osex oage oregion oqfedhi ovote
oxrwght oxewght olrwght olewght ojbcssm ojbrgsc ofimm ohlghql =
zhid zpsu zstrata zpnz zsex zage zregion zqfedhi zvte
zxrwght zxewght zlrwght zlewght zjbcssm zjbrgsc zfimm zhlghql ).
sav out=!path9+"mtchl5.sav" .

**** .

get file=!path9+"mtchl5.sav".
add files file=* /file=!path9+"mtchl.sav" /file=!path9+"mtch2.sav"
/file=!path9+"mtch3.sav" /file=!path9+"mtch4.sav"
/file=!path9+"mtch5.sav" /file=!path9+"mtch6.sav"
/file=!path9+"mtch7.sav" /file=!path9+"mtch8.sav"
/file=!path9+"mtch9.sav" /file=!path9+"mtchl0.sav"
/file=!path9+"mtchl1.sav" /file=!path9+"mtchl2.sav"
/file=!path9+"mtchl3.sav" /file=!path9+"mtchl4.sav" .
fre var=wave year.

* (comment: the order of the files in the add files doesn't matter much,
* but the w15 file is used first in order to prioritise value labels from w14).

sort cases by pid wave .

* Observe: the BHPS is an UNBALANCED PANEL - not everyone
* is present in every wave .

** To obtain summary data on the file:.

compute cons=1.
aggregate outfile=!path9+"ml.sav" /break=pid
/nrec=sum(cons) /firstw=min(wave) /lastw=max(wave) .
match files file=* /table=!path9+"ml.sav" /by=pid .
fre var=nrec firstw lastw .
* These tables show summary information on the panel - but not on individual people.
* There are a total of 178695 records across the survey.
* The most common first contacts for people occurred in waves 1 and 9.
* The most common last contact for people occurred in wave14.

sort cases by pid wave.
compute first=1.
if (lag(pid)=pid) first=0.
fre var=first.
* This shows 30844 different people were in this panel data.

delete variables cons first nrec firstw lastw .

```

```
descriptives var=all.
```

```
sort cases by pid wave.
sav out=!path3b+"bhltol5_long.sav" .
```

```
*[THIS DERIVED FILE IS USED AT SEVERAL POINTS IN THESE LAB EXERCISES].
```

```
*****.
```

```
*****.
```

```
*****.
```

```
*** EXERCISE 3) DERIVING 'WIDE FORMAT' PANEL DATA FROM BHPS COMPONENT FILES .
*****.
```

```

** 'Wide' format involves one case per person, with entries on that case
* from multiple time points.
** Because of this particular design, it is preferable to avoid having large
** numbers of wave specific variables, as if so, the file will expand horizontally
** to a point where it becomes difficult to manage .

```

```
** To prepare wide format data from 'first principles':.
```

```

get file=!path3+"aindresp.sav" /keep=pid asex aage avote .
sort cases by pid.
sav out=!path9+"mtchl.sav".
get file=!path3+"bindresp.sav" /keep=pid bsex bage bvte4 .
sort cases by pid
sav out=!path9+"mtch2.sav".
get file=!path3+"cindresp.sav" /keep=pid csex cage cvote .
sort cases by pid .
sav out=!path9+"mtch3.sav".
get file=!path3+"dindresp.sav" /keep=pid dsex dage dvote .
sort cases by pid .
sav out=!path9+"mtch4.sav".
get file=!path3+"eindresp.sav" /keep=pid esex eage evote .
sort cases by pid .
sav out=!path9+"mtch5.sav".
get file=!path3+"findresp.sav" /keep=pid fsex fage fvote .
sort cases by pid .
sav out=!path9+"mtch6.sav".
get file=!path3+"gindresp.sav" /keep=pid gsex gage gvote .
sort cases by pid .
sav out=!path9+"mtch7.sav".
get file=!path3+"hindresp.sav" /keep=pid hsex hage hvote .
sort cases by pid .
sav out=!path9+"mtch8.sav".
get file=!path3+"iindresp.sav" /keep=pid isex iage ivote .
sort cases by pid .
sav out=!path9+"mtch9.sav".
get file=!path3+"jindresp.sav" /keep=pid jsex jage jvote .
sort cases by pid .
sav out=!path9+"mtchl0.sav".
get file=!path3+"kindresp.sav" /keep=pid ksex kage kvote .
sort cases by pid .
sav out=!path9+"mtchl1.sav".
get file=!path3+"lindresp.sav" /keep=pid lsex lage lvote .
sort cases by pid .
sav out=!path9+"mtchl2.sav".
get file=!path3+"mindresp.sav" /keep=pid msex mage mvote .
sort cases by pid .
sav out=!path9+"mtchl3.sav".
get file=!path3+"nindresp.sav" /keep=pid nsex nage nvote .
sort cases by pid .
sav out=!path9+"mtchl4.sav".
get file=!path3+"oindresp.sav" /keep=pid osex oage ovote .

```

```

sort cases by pid .
sav out=!path9+"mtch15.sav".

get file=!path9+"mtch1.sav".
sort cases by pid.
match files file=* /file=!path9+"mtch2.sav" /in=w2inf
/file=!path9+"mtch3.sav" /in=w3inf /file=!path9+"mtch4.sav" /in=w4inf
/file=!path9+"mtch5.sav" /in=w5inf /file=!path9+"mtch6.sav" /in=w6inf
/file=!path9+"mtch7.sav" /in=w7inf /file=!path9+"mtch8.sav" /in=w8inf
/file=!path9+"mtch9.sav" /in=w9inf /file=!path9+"mtch10.sav" /in=w10inf
/file=!path9+"mtch11.sav" /in=w11inf /file=!path9+"mtch12.sav" /in=w12inf
/file=!path9+"mtch13.sav" /in=w13inf /file=!path9+"mtch14.sav" /in=w14inf
/file=!path9+"mtch15.sav" /in=w15inf /by=pid.
descriptives var=all.
fre var=w2inf w3inf w4inf w5inf w6inf w7inf w8inf w9inf
w10inf w11inf w12inf w13inf w14inf w15inf .

** Wide format data is interesting because it tells us about transitions:.

fre var=avote bvote4 lvote ovote .
cro avote by ovote .

** Time constant data :.
* (This can be a problem for wide format data construction, since the information
* my be supplied in some years and not others).
* (A simpler way to get time constant data on the BHPS is to use the xwav files).

fre var=asex bsex csex dsex esex fsex gsex hsex isex jsex ksex lsex msex nsex .
* For gender, we assume no changes over time, and code to the first valid value.
compute sex=max(asex,bsex,csex,dsex,esex,fsex,gsex,hsex,isex,jsex,ksex,lsex,msex,nsex ).
add value labels sex 1 "Male" 2 "Female".
cro sex by asex bsex csex dsex esex fsex gsex hsex isex jsex ksex lsex msex nsex .
* (the crosstab confirm nobody changes gender).

fre var=aage bage cage dage eage fage gage hage iage jage kage lage mage nage.
* For age data alone it is reasonable to subtract 1 year for a wave to wave difference
* (although remember that sometimes the gap between interviews is less or more).

compute age=-999.
if (oage ge 15) age=oage-14.
if (nage ge 15) age=nage-13.
if (mage ge 15) age=mage-12.
if (lage ge 15) age=lage-11.
if (kage ge 15) age=kage-10.
if (jage ge 15) age=jage-9.
if (iage ge 15) age=iage-8.
if (hage ge 15) age=hage-7.
if (gage ge 15) age=gage-6.
if (fage ge 15) age=fage-5.
if (eage ge 15) age=eage-4.
if (dage ge 15) age=dage-3.
if (cage ge 15) age=cage-2.
if (bage ge 15) age=bage-1.
if (aage ge 15) age=aage.
*(running the 'if's in this order gives priority to the earliest records).
variable label age "Age in 1991".
fre var=age.

sort cases by pid.
descriptives var= pid sex age
avote bvote4 cvote dvote evote fvote gvote hvote ivote jvote kvote lvote mvote nvote ovote.

save out=!path3b+"bh1to15_wide.sav"
/keep=pid sex age
avote bvote4 cvote dvote evote fvote gvote hvote ivote jvote kvote lvote mvote nvote ovote.

** NOTE THAT THIS FILE IS USED IN SEVERAL OTHER PARTS OF THESE LABS.

*****.

```

```

*****.
*** EXERCISE 4) TRANSLATING BETWEEN WIDE AND LONG FORMAT PANEL DATA .
*****.

```

```

** Unlike Stata, SPSS doesn't have an easily implemented translation mechanism
* to move between long and wide format data types, though it is possible with a
* bit of programming.

```

```

** Example (i) : Long to wide.

```

```

get file=!path3b+"bh1to15_long.sav".
sort cases by pid wave.
sav out=!path9+"temp.sav".
get file=!path9+"temp.sav".
select if (wave=1).
sav out=!path9+"a.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=ajbcssm).
get file=!path9+"temp.sav".
select if (wave=2).
sav out=!path9+"b.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=bjbcssm).
get file=!path9+"temp.sav".
select if (wave=3).
sav out=!path9+"c.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=cjbcssm).
get file=!path9+"temp.sav".
select if (wave=4).
sav out=!path9+"d.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=djbcssm).
get file=!path9+"temp.sav".
select if (wave=5).
sav out=!path9+"e.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=ejbcssm).
get file=!path9+"temp.sav".
select if (wave=6).
sav out=!path9+"f.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=fjbcssm).
get file=!path9+"temp.sav".
select if (wave=7).
sav out=!path9+"g.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=gjbcssm).
get file=!path9+"temp.sav".
select if (wave=8).
sav out=!path9+"h.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=hjbcssm).
get file=!path9+"temp.sav".
select if (wave=9).
sav out=!path9+"i.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=ijbcssm).
get file=!path9+"temp.sav".
select if (wave=10).
sav out=!path9+"j.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=jjbcssm).
get file=!path9+"temp.sav".
select if (wave=11).
sav out=!path9+"k.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=kjbcssm).
get file=!path9+"temp.sav".
select if (wave=12).
sav out=!path9+"l.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=ljbcssm).
get file=!path9+"temp.sav".
select if (wave=13).
sav out=!path9+"m.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=mjbcssm).
get file=!path9+"temp.sav".
select if (wave=14).
sav out=!path9+"n.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=njbcssm).
get file=!path9+"temp.sav".
select if (wave=15).
sav out=!path9+"o.sav" /keep=pid zsex zage zjbcssm /rename (zjbcssm=ojbcssm).

match files file=!path9+"a.sav" /file=!path9+"b.sav" /file=!path9+"c.sav"
/file=!path9+"d.sav" /file=!path9+"e.sav" /file=!path9+"f.sav" /file=!path9+"g.sav"
/file=!path9+"h.sav" /file=!path9+"i.sav" /file=!path9+"j.sav" /file=!path9+"k.sav"
/file=!path9+"l.sav" /file=!path9+"m.sav" /file=!path9+"n.sav" /file=!path9+"o.sav"
/by=pid.
descriptives var=all.
missing values ajbcssm to ojbcssm (lo thru 0).
variable labels ojbcssm "Cambridge scale for job in 1991" /ojbcssm "Cambridge scale for job in
2005".
graph /scatterplot=ajbcssm with ojbcssm.

```

```

** Example (ii) : Wide to Long .
* (we illustrate with just the waves from general election years, though more could be added if
required).

get file=!path3b+"bhltol5_wide.sav".
fre var=sex.
compute year=1992.
sav out=!path9+"y92.sav" /keep=pid sex bvot4 year /rename(bvot4=zvote).
compute year=1997.
sav out=!path9+"y97.sav" /keep=pid sex gvot4 year /rename(gvot4=zvote).
compute year=2001.
sav out=!path9+"y01.sav" /keep=pid sex jvot4 year /rename(jvot4=zvote).
add files file=!path9+"y92.sav" /file=!path9+"y97.sav" /file=!path9+"y01.sav" .
fre var=year.
missing values zvote (lo thru 0).
cro zvot4 by year /cells=count col.

** .

*****.
*****.

*****.
*** EXERCISE 5: UNDERSTANDING THE BHPS: SAMPLE ORIGINS, AND USING WEIGHTS
*****.

*****.
** i) Weighting :.

* Most BHPS individual data is on the 'indresp' files for adult interviews in each wave.
get file=!path3+"lindresp.sav" .
fre var=lsex.
fre var=lregion.
* Note the unusual distribution: the BHPS includes regional 'booster' samples.
* If you want to do a nationally representative analysis within a particular
* individual year, use an appropriate cross-sectional weight :.
weight by lxrwtuk1 .
fre var=lregion.
weight off.
** there are actually a lot of different weights - see the BHPS manual -
** but the main relevant ones are: .

/*
wxrwght : individual files, cross-sectional weight for BHPS main sample only
wxrwtk1 : individual files, cross-sectional weight incorporates BHPS boost sample data
(waves 9 -> only)
wxrwtk2 : individual files, cross-sectional weight for within boost regions
(waves 9 -> only, ie, to allow national level analysis on Wales, Scotland or N
Irel)
wxewght : individual files, cross-sectional weight for BHPS main sample for all enumerated
adults (not just for all interviewed adults)
wlrwght : individual files (except wave A), weighting for balanced panel longitudinal analysis
(only if starting at wave A and following original adult respondents)
whhwght : household files, weighting for households to national population
wxhwtk1 : household files, weighting for households to national population + boost population
wxhwtk2 : household files, cross-sectional weight for within boost regions
(waves 9 -> only, ie, to allow national level analysis on Wales, Scotland or N
Irel)
*/

** Another example : compare Scotland, Wales and N Ireland:
get file=!path3+"lindresp.sav" .
fre var=lsex ljbrgsc .

```

```

fre var=lregion.
* Note the unusual distribution: the BHPS includes regional 'booster' samples.
* If you want to do a nationally representative analysis within a particular
* individual year, use an appropriate cross-sectional weight :.
weight by lxrwtuk2 .
fre var=lregion.
temp.
select if (lregion ge 17).
cro ljbrgsc by lregion /cells=count col .
weight off.

*****.
** ii) Sample membership.

** The complication with the BHPS is that different respondents
** have different status's as members of the sample - see the
** associated handout, Table 2.

**.
** The syntax below simply illustrates a way of deriving the component
** information of that table, from the source BHPS data :

*****.
*** Table showing membership of the BHPS, waves 1-12 :.

get file=!path3+"xwlsten.sav" /keep=pid memorig.
sort cases by pid.
fre var=memorig.
sav out=!path9+"mtchl.sav".

get file=!path3+"aindall.sav".
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav".
fre var=one.
select if (one=1).
compute vfo=aivfio.
recode vfo (1,2=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
cro vfo by memorig .
*.
get file=!path3+"bindall.sav" /rename (bsampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=bivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .

*.
get file=!path3+"cindall.sav" /rename (csampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=civfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"dindall.sav" /rename (dsampst=sampst).

```

```

sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=divfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"eindall.sav" /rename (esampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=eivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"findall.sav" /rename (fsampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=fivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"gindall.sav" /rename (gsampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=givfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"hindall.sav" /rename (hsampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=hivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"iindall.sav" /rename (isampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=iivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .

```

```

compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"jindall.sav" /rename (jsampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=jivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"kindall.sav" /rename (ksampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=kivfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.
get file=!path3+"lindall.sav" /rename (lsampst=sampst).
sort cases by pid.
match files file=* /in=one /file=!path9+"mtchl.sav" /by=pid .
fre var=one.
select if (one=1).
compute vfo=livfio.
recode vfo (1,2,3=1) (else=2).
add value labels vfo 1 "Adult Interview" 2 "Other" .
compute sampst2=sampst.
recode sampst2 (5=2) (else=1).
add value labels sampst2 1 "OSM" 2 "TSM".
cro vfo by memorig.
cro vfo by memorig by sampst2 .
*.

```

*****.

*****.
*****.

*****.
*** Exercise 6: DEALING WITH FIXED-IN-TIME BHPS DATA .
*****.

** A common problem in working with the BHPS is that many variables are
* only asked once to respondents (on the assumption that they are fixed in time)
* The result is that an individual's data on that variable may not be within the yearly
* file that you are looking at.
** Increasingly, the BHPS data suppliers have been addressing this issue by collating
* more and more fixed in time variables onto the 'xwav' data files from the core data
* release. Nevertheless, there remain several example variables where fixed in time
* data still needs to be merged manually.


```

** The manual solution involves knowing which variables may be affected in this
** way (there will be a footnote in the documentation on the variable - high numbers of
** 'inapplicables' is the give-away) and treating those variables separately, as follows :.

```

```

*****.
*** Example fixed-in-time variable: Father's occupation.
* Tassk : get fathers occupational info to the wave 12 individual file .
* (note that this example uses a macro as in the LDA WebCT site (SPSS help section)).
*****.

```

```

** Naive mistake : .
get file=!path3+"lindresp.sav" /keep=pid ljbgrsc lpargsc .
fre var=ljbgrsc lpargsc.
* It seems odd that so few people have fathers occ info ; the reason is
* for most people, it was collected on a different wave.

```

```

*****.
** Solution: Collate all reports of fathers occ info between BHPS surveys :.
* (father's occupational data is collected at wave 1 then w7 onwards).

```

```

get file=!path3+"aindresp.sav"
/keep=pid apasoc apaseg apargsc apacssf apacssm apagold apahgs
      apaboss apamngr apasemp      amasoc amaseg amargsc
      amacssf amacssm amagold amahgs      amaboss amamngr amasemp .
sort cases by pid.
sav out=!path9+"mtch1.sav".
get file=!path3+"hindresp.sav"
/keep=pid hpasoc hparseg hpargsc      hpacssf hpacssm hpagold hpahgs
      hpaboss hpamngr hparsemp      hmasoc hmaseg hmargsc
      hmacssf hmacssm hmagold hmahgs      hmboss hmanmgr hmasemp .
sort cases by pid.
sav out=!path9+"mtch2.sav".
get file=!path3+"iindresp.sav"
/keep=pid ipasoc ipaseg ipargsc      ipacssf ipacssm ipagold ipahgs
      ipaboss ipamngr ipasemp      imasoc imaseg imargsc
      imacssf imacssm imagold imahgs      imaboss imamngr imasemp .
sort cases by pid.
sav out=!path9+"mtch3.sav".
get file=!path3+"jindresp.sav"
/keep=pid jpasoc jparseg jpargsc      jpacssf jpacssm jpagold jpahgs
      jpaboss jpamngr jparsemp      jmasoc jmaseg jmargsc
      jmacssf jmacssm jmagold jmahgs      jmboss jmanmgr jmasemp .
sort cases by pid.
sav out=!path9+"mtch4.sav".
get file=!path3+"kindresp.sav"
/keep=pid kpasoc kparseg kpargsc      kpacssf kpacssm kpagold kpahgs
      kpaboss kpmamngr kparsemp      kmasoc kmaseg kmargsc
      kmacssf kmacssm kmagold kmahgs      kmaboss kmamngr kmasemp
      kpasoc00 kmasoc00 .
sort cases by pid.
sav out=!path9+"mtch5.sav".
get file=!path3+"lindresp.sav"
/keep=pid lpasoc lpaseg lpargsc      lpacssf lpacssm lpagold lpahgs
      lpaboss lpamngr lpasemp      lmasoc lmaseg lmargsc
      lmacssf lmacssm lmagold lmahgs      lmaboss lmanmgr lmasemp
      lpasoc00 lmasoc00 .
sort cases by pid.
sav out=!path9+"mtch6.sav".

```

```

match files file=!path9+"mtch1.sav"
      /file=!path9+"mtch2.sav" /file=!path9+"mtch3.sav"
      /file=!path9+"mtch4.sav" /file=!path9+"mtch5.sav"
      /file=!path9+"mtch6.sav" /by=pid.

```

```

descriptives var=all.
recode all (missing, sysmis, lo thru -1=-999).
missing values all (-777).
descriptives var=all.
compute pasoc=max(apasoc,hpasoc,ipasoc,jpasoc,kpasoc,lpasoc).
compute paseg=max(apaseg,hparseg,ipaseg,jparseg,kparseg,lpaseg).
compute pargsc=max(apargsc,hpargsc,ipargsc,jpargsc,kpargsc,lpargsc).
compute pacssf=max(apacssf,hpacssf,ipacssf,jpacssf,kpacssf,lpacssf).
compute pacssm=max(apacssm,hpacssm,ipacssm,jpacssm,kpacssm,lpacssm).
compute pagold=max(apagold,hpagold,ipagold,jpagold,kpagold,lpagold).
compute pahgs= max(apahgs,hpahgs,ipahgs,jpahgs,kpahgs,lpahgs).

```

```

compute masoc= max(amasoc,hmasoc,imasoc,jmasoc,kmasoc,lmasoc).
compute maseg= max(amaseg,hmaseg,imaseg,jmaseg,kmaseg,lmaseg).
compute margsc=max(amargsc,hmargsc,imargsc,jmargsc,kmargsc,lmargsc).
compute macssf= max(amacssf,hmacssf,imacssf,jmacssf,kmacssf,lmacssf).
compute macssm= max(amacssm,hmacssm,imacssm,jmacssm,kmacssm,lmacssm).
compute magold= max(amagold,hmagold,imagold,jmagold,kmagold,lmagold).
compute mahgs=max(amahgs,hmahgs,imahgs,jmahgs,kmahgs,lmahgs).
compute paboss=max(apaboss,hpaboss,ipaboss,jpaboss,kpaboss,lpaboss).
compute pamngr=max(apamngr,hpamngr,ipamngr,jpamngr,kpamngr,lpamngr).
compute pasemp=max(apasemp,hpasemp,ipasemp,jpasemp,kpasemp,lpasemp).
compute maboss=max(amaboss,hmaboss,lmaboss,jmaboss,kmaboss,lmaboss).
compute mamngr=max(amamngr,hmamngr,lmamngr,jmamngr,kmamngr,lmamngr).
compute masemp=max(amasemp,hmasemp,imasemp,jmasemp,kmasemp,lmasep).
compute pasoc00=max(kpasoc00,lpasoc00).
compute masoc00=max(kmasoc00,lmasoc00).

```

```

* Value labels using macro from wave i or k :.

```

```

include file=!path8+"varlabstnew.sps".
!aplab ipasoc tovars=pasoc.
!aplab ipaseg tovars=paseg.
!aplab ipargsc tovars=pargsc.
!aplab ipacssf tovars=pacssf.
!aplab ipacssm tovars=pacssm.
!aplab ipagold tovars=pagold.
!aplab ipahgs tovars=pahgs.
!aplab ipaboss tovars=paboss.
!aplab ipamngr tovars=pamngr.
!aplab ipasemp tovars=pasemp.
!aplab kpasoc00 tovars=pasoc00
!aplab imasoc tovars=masoc.
!aplab imaseg tovars=maseg.
!aplab imargsc tovars=margsc.
!aplab imacssf tovars=macssf.
!aplab imacssm tovars=macssm.
!aplab imagold tovars=magold.
!aplab imahgs tovars=mahgs.
!aplab imaboss tovars=maboss.
!aplab imamngr tovars=mamngr.
!aplab imasemp tovars=masemp.
!aplab kmasoc00 tovars=masoc00.

```

```

fre var=pasoc paseg pargsc
      pagold paboss pamngr pasemp pasoc00
      masoc maseg margsc
      magold maboss mamngr masemp masoc00 .
descriptives var=pacssf pacssm pahgs macssf macssm mahgs .
sort cases by pid.
sav out=!path9+"parentoccs12.sav".

```

```

** Now return to the individual file but match against the cross-wave data : .

```

```

get file=!path3+"lindresp.sav" /keep=pid ljbgrsc lpargsc .
fre var=ljbgrsc lpargsc.
sort cases by pid.
match files file=* /in=w12 /file=!path9+"parentoccs12.sav"
      /keep=pid ljbgrsc lpargsc pargsc /by=pid .
select if (w12=1).
fre var=ljbgrsc lpargsc pargsc .

```

```

* Much higher coverage on this variable now - though still not complete .

```

```

*****.

```

```

*****.
*****.

```

```

*****.
*** Exercise 7: EXAMPLES USING THE BHPS YOUTH FILES
*****.

```

```

** One particular attraction of the BHPS youth files is the
** potnetial to link youth responses with later adult entries :.

```

```

**** Exercise: Link data from the 1994 BHPS youth samples (11-15yrs)
**** with data from the 2004 adult response (21-25 yrs) .

get file=!path3+"dyouth.sav" /keep=pid dypcomp dypsex .
sort cases by pid .
descriptives var=all.
sav out=!path9+"mtchl.sav".

get file=!path3+"nindresp.sav" /keep=pid ngfedhi njbrgsc nsex nage .
sort cases by pid.
match files file=* /in=w14 /file=!path9+"mtchl.sav" /in=w4y /by=pid .

fre var=w14 w4y .
cro w14 by w4y .
** The merge data here shows :
** 15320 adults in wave n (2004) but not in d youth (1994)
** 302 kids in d youth but not in 2004 wave
** 471 are kids in d youth and also present in 2004 adult .

cro dypsex by nsex .
select if (w14=1 & w4y=1) .
fre var=nage.
fre var=ngfedhi .
fre var=dypcomp .
missing values ngfedhi njbrgsc dypcomp (-9 thru 0).
recode dypcomp (3=2).
cro ngfedhi by dypcomp /cells=count col /statistics=phi .
* The quals profile of comp users seems slightly better, but little difference .

cro njbrgsc by dypcomp /cells=count col.
compute profmang=(njbrgsc=1 | njbrgsc=2).
variable label profmang "Advantaged current occupation".
cro profmang by dypcomp /cells=count col /statistics=phi .
graph /bar=mean(profmang) by nsex by dypcomp
/title="Proportion 21-25 year olds with advantaged job in 2004"
/subtitle="by family access to a computer in 1994".
* The occupational position of comp users seems slightly better

```

*****.

*****.
*** Exercise 8: FURTHER EXAMPLES IN MATCHING PANEL DATA FILES
*****.

*** Other materials : The latter sections of the lab 0 file
** (introduction to SPSS) included several data file matching
** examples. One of those involved 'relationships between cases'
** file matching on the BHPS - reproduced below.
*****.

* There are lots of circumstances where you might want to link together
** information from different cases in a sample .
** Sticking with the BHPS, this example shows how you could link information on
** an individual with their spouse if present .
** Many surveys include information allowing you to link data between cases, whilst
* many also include additional derived variables where some basic related cases information
* has already been added in .

** BHPS files from 2002 for individuals .
** The extract2 file adds a variable lsppid which is needed to link information
** on different spouses.

```

get file=!path3+"lindresp.sav"
/keep=pid lhid lsex lage ljbstat lsppid.
* The variable lsppid is the personal identifier number of the spouse, if present.
* Deal only with those who give a pid for a spouse.
* In addition, in order to distinguish male and female spouses, split by gender :.
select if (lsppid gt 0 & lsex=2).
descriptives var=all.
sort cases by lsppid.
sav out=!path9+"wifl.sav" /keep=lsppid lsex lage ljbstat
/rename(lsppid lsex lage ljbstat=pid wifsex wifage wifjbst) .

* Match the women's data back into the original, for men only :.
get file=!path3+"lindresp.sav"
/keep=pid lhid lsex lage ljbstat lsppid.
select if (lsppid gt 0 & lsex=1).
sort cases by pid.
match files file=* /in=hus /file=!path9+"wifl.sav" /in=wif /by=pid .
fre var=hus wif.
cro hus by wif.
** Most of the non 1-1 cases are males or females whose spouse didn't give
** an interview in the individual file, for whatever reason .
* However, any homosexual spouses will be lost in this missing data here.
select if (hus=1 & wif=1).
descriptives var=all.

```

* Some illustrative analyses :.

```

variable label lsex "Husband sex" /wifsex "Wife's sex"
/lage "Husband age" /wifage "Wife's age"
/ljbstat "Husband economic activity" /wifjbst "Wife's economic activity" .

```

cro ljbstat by wifjbst.

```

graph /scatterplot=lage with wifage .
* (some non-overlap suggests gender imbalance in age differences).
compute agedif=-999.
if (lage gt 16 & wifage gt 16) agedif=lage - wifage.
missing values agedif (-999).
graph / histogram=agedif .
* Average gap in ages between spouses is men 2.3 years older than their spouses.

```

*****.

*****.

*****.
*** Exercise 9: LONG FORMAT DATA: DESCRIPTIVE ANALYSES .
*****.

*** 9.1): Describing patterns within cases .

*** Summary data on the panel structure (also in Exercise 4).

```

get file=!path3b+"bhltol5_long.sav".

```

```

descriptives var=all.

```

```

sort cases by pid wave.
compute first=1.
if (lag(pid)=pid) first=0.
fre var=first.
* 29815 different people in this file, spanning 178695 records.

compute cons=1.
aggregate outfile=!path9+"ml.sav" /break=pid
/nrec=sum(cons) /firstw=min(wave) /lastw=max(wave) .
match files file=* /table=!path9+"ml.sav" /by=pid .
fre var=nrec firstw lastw .
* These tables show summary information on the panel - but not on individual people.
* There are a total of 178695 records across the survey.
* The most common first contacts for people occurred in waves 1 and 9.
* The most common last contact for people occurred in wave15.

**** Context of individual level responses:.

missing values zvote (lo thru 0).
fre var=zvote.
* The 'trouble' is that the same people contribute multiple records.
* SPSS doesn't have extensive solutions to this, but we can try:.

** Use only the first record per person.
temp.
select if (first=1).
fre var=zvote.

** Weight the frequency by inverse of the number of records per person.
compute one=1.
sort cases by pid wave.
aggregate out=!path9+"mtch1.sav" /break=pid /nrec=sum(one).
match files file=* /table=!path9+"mtch1.sav" /by=pid.
fre var=nrec.
compute invnrec=1/nrec.
fre var=zvote.
weight by invnrec.
fre var=zvote.
weight off.

* (notice that the proportion favouring the big parties goes down when
* we discount the influence of repeated measures).

*****
*** 9.2): Describing patterns, adjusting for individual clustering .

fre var=zvote.
compute zvote2=zvote.
recode zvote2 (1=1) (2=2) (3=3) (4 thru 9=4) (else=-999).
missing values zvote2 (-999).
add value labels zvote2 1 "Conservative" 2 "Labour" 3 "Liberal" 4 "Other" .
fre var=zvote2.
cro zsex by zvote2 /cells=count row /statistics=chisq phi.

compute ghq=zhlgghq1.
missing values ghq zjbrgsc (lo thru -1).
graph /errorbar=ghq by zjbrgsc by zsex .
examine variables=ghq by zsex /nottotal /plot=none .

** There is really a clustering effect here - the Chi-square and CI intervals aren't correct .

** Many packages have inbuilt controls which allow correction for within-group
* clustering effects, but SPSS does not have much functionality in this regard.

** Weighting by inverse of sample probability is however a reasonable start:.

cro zsex by zvote2 /cells=count row /statistics=chisq phi.
graph /errorbar=ghq by zjbrgsc by zsex .
examine variables=ghq by zsex /nottotal /plot=none .

weight by invnrec.
cro zsex by zvote2 /cells=count row /statistics=chisq phi.
graph /errorbar=ghq by zjbrgsc by zsex .

```

```

examine variables=ghq by zsex /nottotal /plot=none .
weight off.

** Note that in all cases, the treatment for clustering makes the results more
* conservative, ie, larger confidence intervals and smaller chisquare.
** Generally the core patterns aren't changed much by this adjustment - but some
* differences are more noticable, eg, overlap of CI's in the two graphs.

*****
** 9.3): Using the panel data to get year by year summaries : .

get file=!path3b+"bhltol5_long.sav".
descriptives var=all.
fre var=zvote.
* Do the age profiles of voters change over time?.
select if (zage >= 18 & (zvote=1 | zvote=2 | zvote=3)).
graph /errorbar=zage by year by zsex.
graph /errorbar=zage by year by zvote.

* (women and men age at the same rate, but conservatives age
* disproportionately quickly compared to other parties).

sort cases by zsex zvote.
aggregate outfile=!path9+"age_by_sexvote.sav" /break=zsex zvote /meanage=mean(zage).
get file=!path9+"age_by_sexvote.sav".
descriptives var=all.

* This has created a time series dataset - male and female average
* ages for different types of voting preference .

*****
***** 9.4) Describing wide v's long format data .

*** Wide format is handy for the analysis of transitions and sequences : .

get file=!path3b+"bhltol5_wide.sav".
descriptives var=all.
fre var=avote.

* Three national election years in period were 1992, 1997, 2001.
recode bvote4 gvote kvote (-9,-8,-7,-2,-1, 4 thru hi = -999).
missing values bvote4 gvote kvote (-999).
cro bvote4 by gvote /cells=count row /statistics=phi.
cro bvote4 by kvote /cells=count row /statistics=phi.
cro gvote by kvote /cells=count row /statistics=phi.

compute acon=(avote=1).
compute bcon=(bvote4=1).
compute ccon=(cvote=1).
compute dcon=(dvote=1).
compute econ=(evote=1).
compute fcon=(fvote=1).
compute gcon=(gvote=1).
compute hcon=(hvote=1).
compute icon=(ivote=1).
compute jcon=(jvote=1).
compute kcon=(kvote=1).
compute lcon=(lvote=1).
compute mcon=(mvote=1).
compute ncon=(nvote=1).

compute probcon=mean(acon to ncon).
compute ncon=sum(acon to ncon).
descriptives var=probcon .
fre var=ncon.

```

```

*** Long format is better suited to looking at adjacent transitions
**      or longer term within person stabilities.

get file=!path3b+"bhltol5_long.sav".
descriptives var=all.

compute convot=(zvvote=1).
if (zvvote < 1) convot=-999.
missing values convot (-999).
variable label convot "Supported conservatives".
fre var=convot.

sort cases by pid wave.
compute lconvot=-999.
if pid=lag(pid) lconvot=lag(convot).
missing values lconvot (-999).
cro lconvot by convot /cells=count col /statistics=phi.
* (Voting now compared to last year).

compute l6convot=-999.
if pid=lag(pid,6) l6convot=lag(convot,6).
missing values l6convot (-999).
cro l6convot by convot /cells=count row /statistics=phi .
* (Voting now compared to 6 years ago).

* Long format data is more suited to adding further explanatory variables
* to an analyses through panel modelling techniques
*****.

*****.

*****.
*** EXERCISE 10: PANEL DATA ANALYSIS: LOOKING AT TRANSITIONS / SEQUENCES .
*****.

*****.
***** 1(i) : wave 1 (1991) to wave 12 (2002) voting differences.

get file=!path3+"aindresp.sav" /keep=pid asex aage avote .
sort cases by pid.
sav out=!path9+"mtchl.sav".
get file=!path3+"lindresp.sav" /keep=pid lsex lage lvote .
sort cases by pid .
sav out=!path9+"mtchl2.sav" .

match files file=!path9+"mtchl.sav" /in=w1 /file=!path9+"mtchl2.sav" /by=pid.
descriptives var=all.
fre var=w1 .
select if (w1=1).
descriptives var=all.
fre var=avote lvote .
missing values avote lvote (lo thru -1).
cro avote by lvote / cells=count col /statistics=phi .

** Make some recodes to keep simple : .
compute avote4=avote.
compute lvote4=lvote.
recode avote4 lvote4 (1=1) (2=2) (3 thru 8=3) (10, 11=4) (else=-999).
missing values avote4 lvote4 (-999).
add value labels avote4 lvote4 1 "Conservative" 2 "Labour" 3 "Other party" 4 "None" .

```

```

cro avote4 by lvote4 / cells=count col /statistics=phi .

cro avote4 by lvote4 / cells=count row /statistics=phi .
* The row percents show what happened to people given their 1991 view

* Particular interest might be in 1991 conservatives who switch to
* 2002 labour, cf those who stay 2002 conservative .

compute con2lab=-999.
if (avote=1 & lvote=1) con2lab=0.
if (avote=1 & lvote=2) con2lab=1.
add value labels con2lab 0 "Stable conservative" 1 "Con1999 to labour2002" .
missing values con2lab (-999).
fre var=con2lab .

descriptives var=aage .
fre var=asex .
compute afem=(asex=2).

logistic regression var=con2lab /method=enter afem aage .
** Suggests that Tories who switch votes are disproportionately younger and women .

*****.
***** 1(ii) : wave d youth to wave e youth .

** Compare 11-15 year olds in 1994 and 1995 by household transitions, and check those
** against happiness measures .

** Sample of individuals from 1994 and 1995 :.

get file=!path3+"dyouth.sav" .
fre var= dypsad dyphlf dypsex .
cro dypsad by dyphlf.
compute dsad=((dypsad =3 | dypsad=4) | (dyphlf ge 5) | (dypsad=2 & (dyphlf=3 | dyphlf=4) ) ).
sort cases by pid.
sav out=!path9+"mtchl.sav" /keep=pid dsad dypsex .
get file=!path3+"eyouth.sav" .
fre var= eypsad eyphlf eypsex .
cro eypsad by eyphlf.
compute esad=((eypsad =3 | eypsad=4) | (eyphlf ge 5) ).
sort cases by pid.
sav out=!path9+"mtch2.sav" /keep=pid esad eypsex .

* indall files for age and gender .
get file=!path3+"dindall.sav" / keep=pid dhid dage .
sort cases by pid.
sav out=!path9+"mtch3.sav".
get file=!path3+"eindall.sav" / keep=pid ehid eage .
sort cases by pid.
sav out=!path9+"mtch4.sav".

** Match all these individual files together :.
match files file=!path9+"mtchl.sav" /in=dy /file=!path9+"mtch2.sav" /in=ey
/file=!path9+"mtch3.sav" /file=!path9+"mtch4.sav" /by=pid.
fre var=dy ey.
cro dy by ey.
select if (dy=1 & ey=1).
*(selects only the 580 youths present in both waves d and e).
descriptives var=all.
sav out=!path9+"mtch5.sav" .

** Match in household composition information on both waves.
get file=!path3+"dhhresp.sav" /keep=dhid dhhsz.
sort cases by dhid.
sav out=!path9+"mtch6.sav" .
get file=!path3+"ehhresp.sav" /keep=ehid ehhsz.
sort cases by ehid.
sav out=!path9+"mtch7.sav" .

get file=!path9+"mtch5.sav" .
sort cases by dhid.
match files file=* /table=!path9+"mtch6.sav" /by=dhid.
sort cases by ehid.

```

```

match files file=* /table=!path9+"mtch7.sav" /by=ehid.
descriptives var=all.

cro dhhszsize by ehhszsize.
compute hchange=(dhhszsize ne ehhszsize).
variable label hchange "Wave 5 household isn't exactly equiv people to wave 4" .
fre var=hchange.
* (most households are stable over 12 months, but 81 youths were in a changing household).
* (with more work it would be possible to know how and why the household is changing,
*   eg parental splits v's sibling moves).

** Do household transitors have different subjective welfare experiences? .

cro hchange by dsad esad /cells=count row /statistics=chisq phi .
** there is a trend though it's not quite significant.
compute sadchang=-999.
if (esad=0 & dsad=0) sadchang=1.
if (esad=0 & dsad=1) sadchang=2.
if (dsad=0 & esad=1) sadchang=3.
if (dsad=1 & esad=1) sadchang=4.
add value label sadchang 1 "Not miserable wave 5" 2 "Cheers up wave 5"
3 "Becomes miserable wave 5" 4 "Stays miserable wave 5" .
fre var=sadchang.
cro hchange by sadchang /cells=count row /statistics=chisq phi .
* Again, there is a suggestion of a data pattern, but not confirmed as significant.

** Comment : lots of fairly arbitrary variables defined above - could have experimented a
** bit more with them. Also, the data analysis here probably suffers from relatively low
** sample sizes : the above could be repeated for later waves of the BHPS and a pooled
** sample would increase the analytical sample size (but would need to beware of multiple
** records from the same individuals).

*****.

*****.

*****.

*** EXERCISE 11: LONG FORMAT DATA: PANEL DATA MODELS
*****.

get file=!path3b+"bh1to15_long.sav".
descriptives var=all.

compute ghq=zhlgghq1.
missing values ghq (10 thru -1).

compute lninc=-999.
if (zfimm >= 100) lninc = ln(zfimm).
compute fem=(zsex=2).
compute age=zage.
missing values age (10 thru 15) .
fre var=zqfedhi .
compute hied=-999.
if (zqfedhi >= 1) hied=(zqfedhi ge 1 & zqfedhi le 4).
compute noed=-999.
if (zqfedhi >= 1) noed=(zqfedhi >=12).

compute convot=-999.
if (zvot ge 1) convot=(zvot=1).
compute labvot=-999.
if (zvot ge 1) labvot=(zvot=2).
missing values lninc hied noed convot labvot (-999).

descriptives var= ghq lninc fem age hied noed convot labvot .
correlate variables= ghq lninc fem age hied noed convot labvot .

```

```

***** (a) Metric regression illustration:

** Linear regression on ghq : not appropriate because records aren't independent.

regression /dependent=ghq
/method=enter lninc fem age hied noed convot labvot .

** Random effects panel on ghq - just controls for clustering

mixed ghq with lninc fem age hied noed convot labvot
/fixd=lninc fem age hied noed convot labvot | sstype(3)
/method=reml
/print=corb solution r
/random=intercept | subject(pid) covtype(ID) .

** Panel on ghq with lag effects
sort cases by pid wave .
compute lagghq=-999.
if (pid=lag(pid)) lagghq = lag(ghq).
missing values lagghq (-999).

descriptives var= ghq lagghq lninc fem age hied noed convot labvot .
correlate variables= ghq lagghq lninc fem age hied noed convot labvot .

regression /dependent=ghq
/method=enter lagghq lninc fem age hied noed convot labvot .

mixed ghq with lagghq lninc fem age hied noed convot labvot
/fixd=lagghq lninc fem age hied noed convot labvot | sstype(3)
/method=reml
/print=corb solution r
/random=intercept | subject(pid) covtype(ID) .

** However, this is about as far as SPSS can go with panel models .

*****.

*****.

*****.

*****.

*****.

**** EOF .

```