Growing Up in Scotland Study - Data Workshops - Dec 2011/Jan 2012

Coping with Complex Samples 1: Creating a complex samples plan file in SPSS

NOTE: This process requires the Complex Samples Module in SPSS. This module is not supplied in the standard SPSS package. The module is not available on versions of SPSS prior to version 15.

It is common practise to weight your data to account for the non-response bias in your sample, which usually results in disadvantaged groups being underrepresented. If the sample in your data was generated using clustering and stratification and is not a simple random sample, then you should account for the sampling design as you weight your data. If you do not account for the sample design, SPSS assumes it is dealing with a simple random sample. In doing so it will underestimate the real amount of error (the standard error), affecting the confidence intervals and ultimately provide a potentially inaccurate result in significance tests. By not accounting for the complex sample design, SPSS will return differences which are not significant as significant.

To resolve this, and to account for the complex sample design, SPSS requires you to create a complex samples plan file (these files end with '.csaplan').

1. You need three components to create a plan file
   a) A weight variable
   b) A stratification variable
   c) A cluster variable

   The precise variables to select will depend on the data incorporated in your analysis (i.e. whether it is cross-sectional or longitudinal analysis and which sweeps of data are involved)

   The example below illustrates the creation of a plan file for cross-sectional analysis of birth cohort data from sweep 3.

2. To weight for cross-sectional analysis for the birth cohort at sweep 3 you need the following:
   a) Weight = DcWTbrth
   b) Stratification variable (Strata) = Dcstrat
   c) Cluster variable = Dcpsu

3. On the menu tool bar follow the command: Analyze//Complex samples/prepare for analysis
   → Select: Create a plan file
   → Select a destination to save your plan file: your temporary folder, and give your file a name (you'll find using something descriptive incorporating the sweep and weight used is useful)
   → Click NEXT
   → Insert the relevant variables as shown below in the screenshot, click NEXT
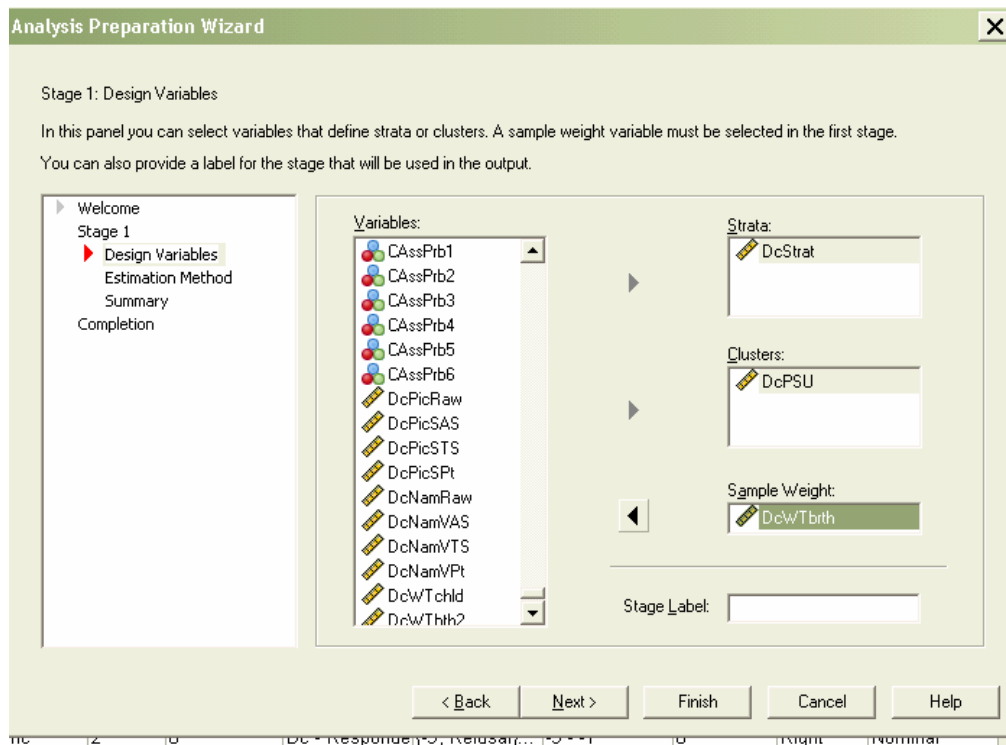
→ Now select the WR option (sampling with replacement) and UNSELECT the option "Use finite population correction" (FPC). Click Next.

→The summary page shows you the selections you have made. Click Next

→ Select 'Paste the syntax generated' and click 'Finish' to see it in your syntax window.

→ You should now be able to see the following syntax in your Syntax Window. Select it with the runner and click on the blue play button on the toolbar or click CTRL-R on your keyboard.

You can now use this syntax as the basis for any plan file you need to create.

All you need to change is:

```
* Analysis Preparation Wizard.
CSPLAN ANALYSIS
 /PLAN FILE='C:\temp\sw3bcxs.csaplan'
 /PLANVARS ANALYSISWEIGHT=DcWTbrth
 /SRSESTIMATOR TYPE=WR
 /PRINT PLAN
 /DESIGN  STRATA= DcStrat  CLUSTER= DcPSU
   /ESTIMATOR TYPE=WR.
```

- the filename
- the weight variable
- the strata variable (dastrat, dbstrat, dcstrat etc)
- the cluster variable (dapsu, dbpsu, dcpsu etc)

To suit the datasets you are using and analysis you are running.

Your *csplan* file is ready. You can use this every time you intend to do cross-sectional analysis of birth cohort data from sweep 3. SPSS will prompt you to select a plan file you have already constructed when you use any of the commands available through the *Analyze/Complex Samples* commands.

If you wish to look at the child cohort you will have to use the appropriate weight and create another csplan file, and you will also have to create a different csplan file for analysis between sweeps 1, 2 and 3, and so on.   However, as suggested above, you can do this by directly editing the above syntax, as well as via the menus.

It is good practise to keep the syntax of your plan files, and select intuitive names for these files so you can identify them instantly.

Identifying the correct GUS weight

Sweep 1 Weight
- Use for any cross-sectional analysis of SWEEP 1 DATA ONLY
    o Birth cohort: **dawtbrth**
    o Child cohort: **dawtchld**

Sweep 2 Main interview weight
- Use for any cross-sectional analysis of SWEEP 2 DATA ONLY
- Use for any longitudinal analysis involving Sweep 1 AND Sweep 2 data
    o Birth cohort: **dbwtbrth**
    o Child cohort: **dbwtchld**

Sweep 2 Partner interview weight
- Use for any analysis of Partner interview data
    o Birth cohort: **dbwtbrtp**
    o Child cohort: **dbwtchlp**

Sweep 3 Longitudinal weight
- Use for any analysis of more than one sweep of data, when data from sweep 3 is included (e.g. data from sweep 1 and sweep 3, or from sweep 2 and sweep 3, or from sweeps 1, 2 and 3)
    o Birth cohort: **dcwtbth2**
    o Child cohort: **dcwtchd2**

Sweep 3 Cross-sectional weight
- Use for any cross-sectional analysis of SWEEP 3 DATA ONLY
    o Birth cohort: **dcwtbrth**
    o Child cohort: **dcwtchld**

Sweep 4 Longitudinal weight
- Use for any analysis of more than one sweep of data, when data from sweep 4 is included (e.g. data from sweep 1 and sweep 4, or from sweep 2 and sweep 4, or from sweeps 1, 2, 3 and 4)
    o Birth cohort: **ddwtbth2**
    o Child cohort: **ddwtchd2**

Sweep 4 Cross-sectional weight
- Use for any cross-sectional analysis of SWEEP 4 DATA ONLY
  - o Birth cohort: **ddwtbrth**
  - o Child cohort: **ddwtchld**

Sweep 5 Longitudinal weight
- Use for any analysis of more than one sweep of data, when data from sweep 4 is included (e.g. data from sweep 1 and sweep 5, or from sweep 3 and sweep 5, or from sweeps 3, 4 and 5)
  - o Birth cohort: **dewtbth2**
  - o Child cohort: **dewtchd2**

Sweep 5 Cross-sectional weight
- Use for any cross-sectional analysis of SWEEP 5 DATA ONLY
  - o Birth cohort: **dewtbrth**