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

Coping with Complex Samples 2: Basic analysis using the Complex Samples (CS) Module in SPSS

Having created your CS plan file, you can now use it as a component in your analysis allowing you to check statistical significance whilst correcting for the complex sample design in GUS.

This worksheet illustrates how to run, and check significance in, a simple cross-tabulation using the CS module. The example is cross-sectional using the birth cohort data from sweep 3 (and therefore corresponds with the CS plan file created in the previous worksheet). Note that unless you are testing significance, there is no need to use the CS analysis approach, a simple weighted table or crosstab will do.

1. Open the sweep 3 birth cohort dataset.

2. On the menu tool bar select: Analyze/Complex Samples/Crosstabs
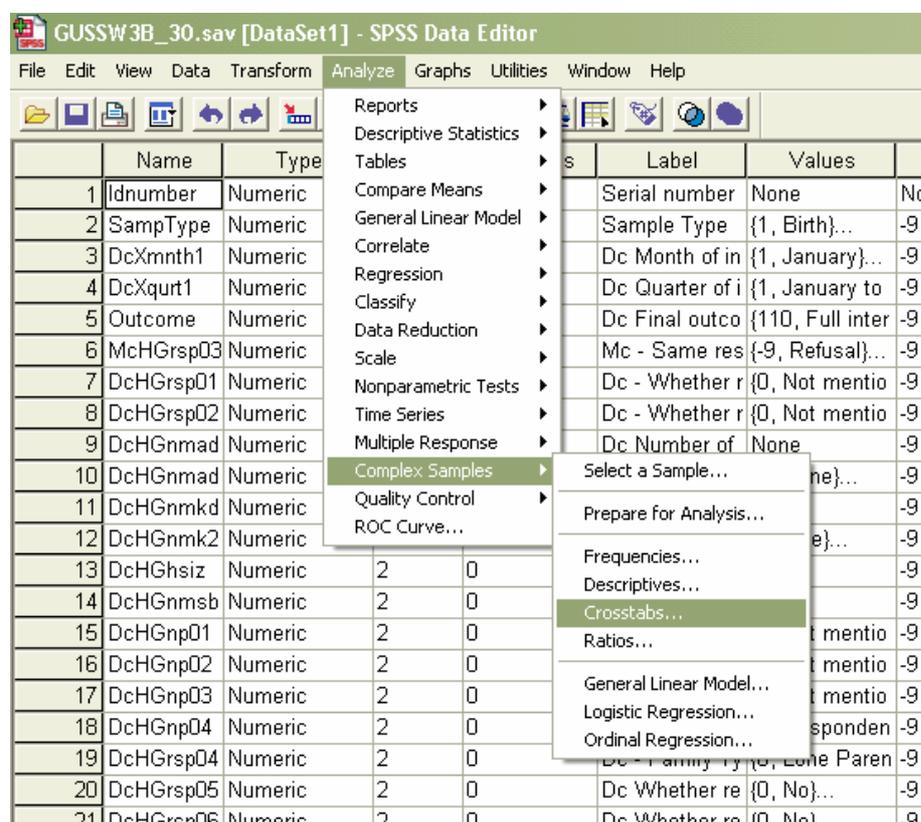
2. In the 'Complex Samples Plan for Crosstabs Analysis' dialog box, the plan file you have just created may already appear in the 'File' box, if not, click 'Browse' and navigate to the correct plan file. [REMEMBER – the plan file you require is dependent on the data you are using and the analysis you are running]. Once you have selected your plan file, click 'Continue' (Ignore 'Joint Probabilities')

3. The 'Complex Samples Crosstabs' dialog box which appears is almost identical to the standard Crosstabs dialog box. Select your Row variables and Column variables (and nested variables under 'Subpopulations') as necessary. For this example we want to compare household income by family type:

- Under the variable list find and select the household income variable (DcWinc01) – once selected click the black arrow next to the 'Column' box.

- Back in the variable list find and select the family type variable (DcHGrsp04)– once selected click the black arrow next to the 'Row' box
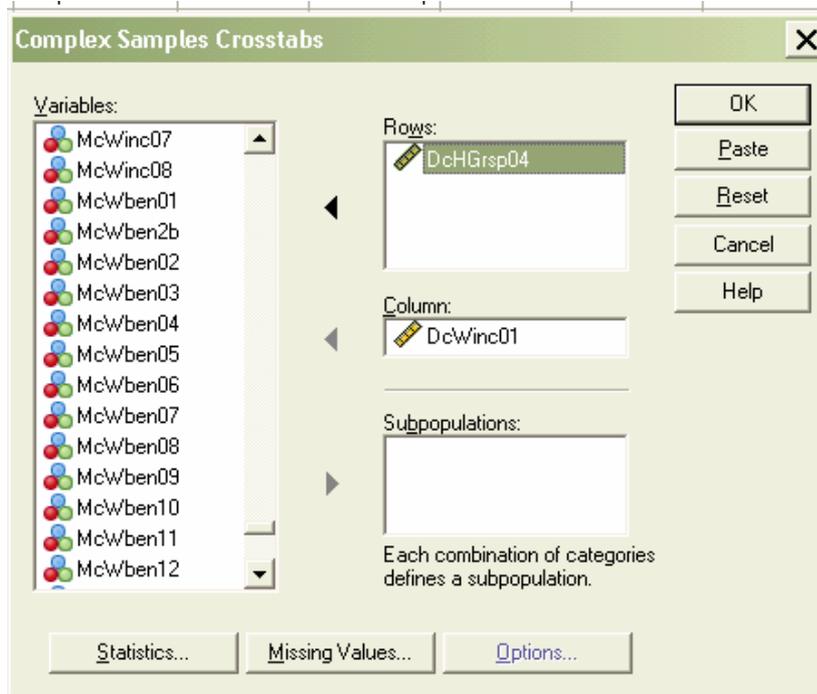
- Click on the 'Statistics' tab. In the 'Statistics' dialogue box, under 'Cells' un-select 'Population Size' and select 'Column' percent. Under 'Statistics', un-select 'Standard Error' and select 'Confidence Interval' (keep the level at 95%). At the bottom, select 'Test of independence of rows and columns'. Click 'Continue'.
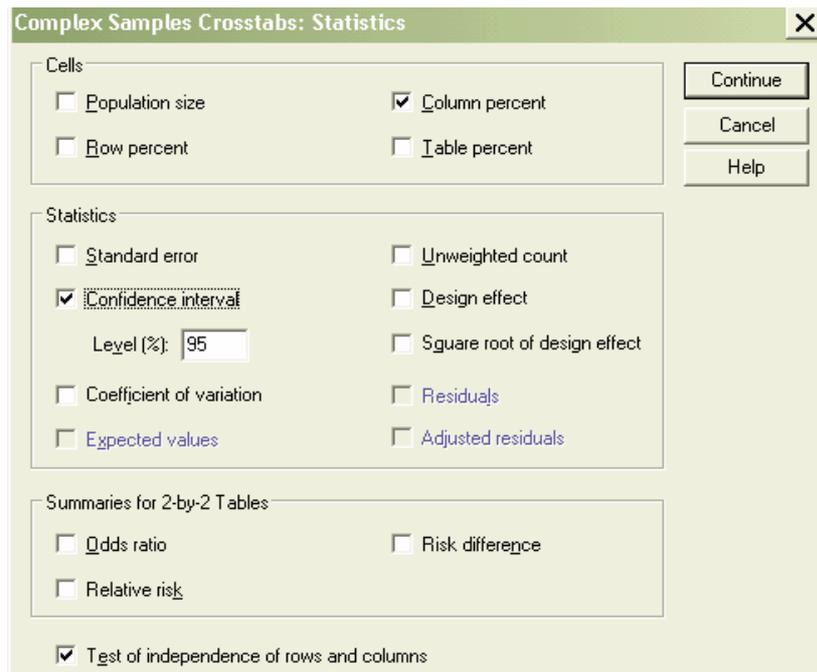
- Back in the 'Crosstabs' dialog box, click 'Paste' to paste your syntax, which should look like this:

```
CSTABULATE
 /PLAN FILE = 'C:\temp\sw3bcxs.csaplan'
 /TABLES VARIABLES = DcHGrsp04 BY DcWinc01
 /CELLS  COLPCT
 /STATISTICS  SE
 /TEST  INDEPENDENCE
     /MISSING SCOPE = TABLE CLASSMISSING = EXCLUDE.
```

4. Select the syntax with your mouse and click on the blue play button on the toolbar or click CTRL-R on your keyboard. SPSS will return the following output:

Dc - Family Type * Dc Household income - banded

Dc - Family Type				Dc Household income - banded				
				Up to £14999 per year	From £15000 to £25999 per year	From £26000 to £43999	£44,000 and above	Total
Lone Parent	% within Dc Household income - banded	Estimate		62.9%	15.5%	2.7%	1.2%	20.2%
		95% Confidence Interval	Lower	57.6%	11.4%	1.4%	.4%	17.8%
			Upper	67.9%	20.6%	5.1%	3.6%	22.9%
Couple Family	% within Dc Household income - banded	Estimate		37.1%	84.5%	97.3%	98.8%	79.8%
		95% Confidence Interval	Lower	32.1%	79.4%	94.9%	96.4%	77.1%
			Upper	42.4%	88.6%	98.6%	99.6%	82.2%
Total	% within Dc Household income - banded	Estimate		100.0%	100.0%	100.0%	100.0%	100.0%
		95% Confidence Interval	Lower	100.0%	100.0%	100.0%	100.0%	100.0%
			Upper	100.0%	100.0%	100.0%	100.0%	100.0%

Tests of Independence

		Chi-Square	Adjusted F	df1	df2	Sig.
Dc Household income - banded * Dc - Family Type	Pearson	463.872	159.212	2.838	184.482	.000
	Likelihood Ratio	452.276	155.232	2.838	184.482	.000

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

5. The significance value is returned in the final column of the 'Tests of Independence' variable. At a significance of $p < 0.001$, we can conclude that family type is significantly associated with household income. And from the data, we can see that lone parent families are considerably more likely to be on lower incomes than are couple families.

To run further crosstabs with significance testing on the *sweep 3 birth cohort* data, all you now need to do is copy and paste the syntax you've created and adjust the row and column variables as necessary.

/TABLES VARIABLES = *DcHGrsp04* BY *DcWinc01*

If you are running analysis on a *different dataset*, or *combination of data*, and have already prepared an appropriate plan file, all you need to do is change the plan file reference in the syntax

/PLAN FILE = 'C:\temp\sw3bcxs.csaplan'