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 crosstabulation 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

🏭 GUSSW3B_30.sav [DataSet1] - SPSS Data Editor											
File Edit	View Data	Transform	Analyze	Graphs	Utilities	٧	Vindow	Help			
B	🖹 🖭 📥	•	Repor Descri	ts ptive Stat	istics 🕨	2	1	<u>s</u>			
	Name	Туре	Tables	;	•	s		Label	V	alues	
1	ldnumber	Numeric	Compa	are Means	; 🕨	Г	Seri	al number	None		No
2	SampType	Numeric	Gener	al Linear M	Model 🕨	Г	San	nple Type	{1, Bir	th}	-9 -
3	DcXmnth1	Numeric	Correl	ate		Г	Dc I	Month of in	{1, Ja	nuary}	-9 -
	DcXqurt1	Numeric	Classif	ssion :o		Γ	Dc (Quarter of i	{1, Ja	nuary to	-9 -
5	Outcome	Numeric	Data P	,γ Reduction		Г	Dc I	Final outco	{110,	Full inter	-9 -
E	McHGrsp03	Numeric	Scale			Г	Mo	- Same res	{-9, R	efusal}	-9 -
7	DcHGrsp01	Numeric	Nonpa	rametric ⁻	Tests 🕨	Г	Dc -	Whether r	{0, No	t mentio	-9 -
6	DcHGrsp02	Numeric	Time Series			Г	Dc -	Whether r	{0, No	-9 -	
9	DcHGnmad	Numeric	Multipl	e Respon	se 🕨		Dc I	Number of	None	1	-9 -
10	DcHGnmad	Numeric	Compl	ex Sample	es 🕨		Select	a Sample		ne}	-9 -
11	DcHGnmkd	Numeric	Quality	y Control	•		Prepar	e for Analysis			-9 -
12	DcHGnmk2	Numeric	ROCO	urve			F			e}	-9 -
13	DcHGhsiz	Numeric	2	0			Descrip	ncles			-9 -
14	DcHGnmsb	Numeric	2	0			Crossta	ahs			-9 ·
15	DcHGnp01	Numeric	2	0			Ratios.			t mentio	-9 -
18	DcHGnp02	Numeric	2	0			~	1		t mentio	-9 ·
17	DcHGnp03	Numeric	2	0			Genera	al Linear Mode : Dogrossion	1	t mentio	-9 -
18	DcHGnp04	Numeric	2	0			Ordinal	Regression		sponden	-9 -
19	DcHGrsp04	Numeric	2	0		_		паншу ту	10, E0	ne Paren	-9 -
20	DcHGrsp05	Numeric	2	0			Dc ۱	Whether re	{0, No	}	-9 -
21	IDeHGren06	Numeric	2	n			De Y	Whether re	(O No	1	-9.

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

· ·		· · · · ·	
Complex Samples	Crosstabs		×
Variables:	-	Ro <u>w</u> s: DcHGrsp04	OK Paste <u>R</u> eset Cancel
McWbenU2 McWben03 McWben04 McWben05 McWben06 McWben07 McWben08 McWben08	4	<u>C</u> olumn:	Help
McWben10 McWben11 McWben12	-	Each combination of categories defines a subpopulation.	
<u>S</u> tatistics	<u>M</u> issing	Values Options	

 Click on the 'Statistics' tab. In the 'Statistics' dialogue box, under 'Cells' unselect 'Population Size' and select 'Column' percent. Under 'Statistics', unselect 'Standard Error' and select 'Confidence Interval' (keep the level at 95%). At the bottom, select 'Test of independence of rows and columns'. Click 'Continue'.

Complex Samples Crosstabs: Statis	stics	×
Cells Population size <u>R</u> ow percent	 ☑ <u>C</u>olumn percent ☑ <u>I</u>able percent 	Continue Cancel Help
Statistics ☐ <u>S</u> tandard error ✔ <u>Confidence interval</u> Le <u>v</u> el (%): 95 ☐ Coefficient of variation	<u>U</u> nweighted count <u>D</u> esign effect Sguare root of design effect Residua <u>ls</u>	
Expected values Summaries for 2-by-2 Tables Qdds ratio Relative risk	Adjusted residuals Risk difference	

 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					
				Up to £14999 per vear	From £15000 to £25999 per vear	From £26000 to £43999	£44,000 and above	Total
Lone Parent	% within Dc Household income -	Estimate		62.9%	15.5%	2.7%	1.2%	20.2%
	banded	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 -	Estimate		37.1%	84.5%	97.3%	98.8%	79.8%
	banded	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%

Dc - Family Type * Dc Household income - banded

Tests of Independence

		Chi-Square	Adjusted F	df1	df2	Sig.
Dc Household	Pearson	463.872	159.212	2.838	184.482	.000
Dc - Family Type	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 = <u>DcHGrsp04</u> BY <u>DcWinc01</u>

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'