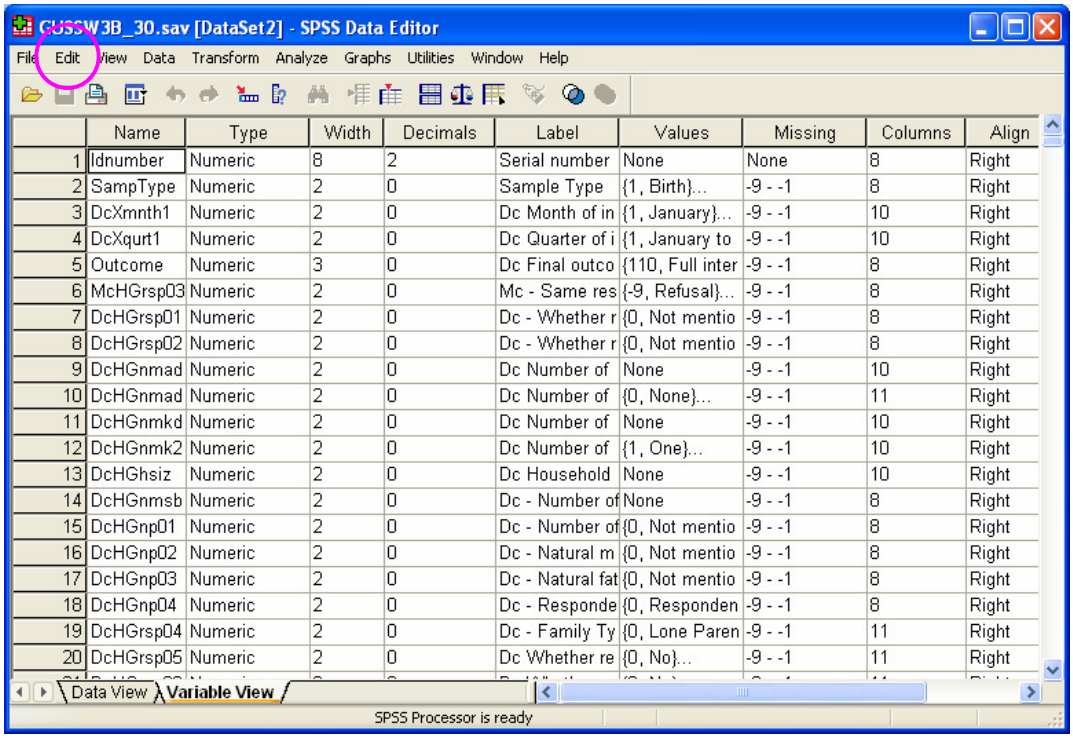
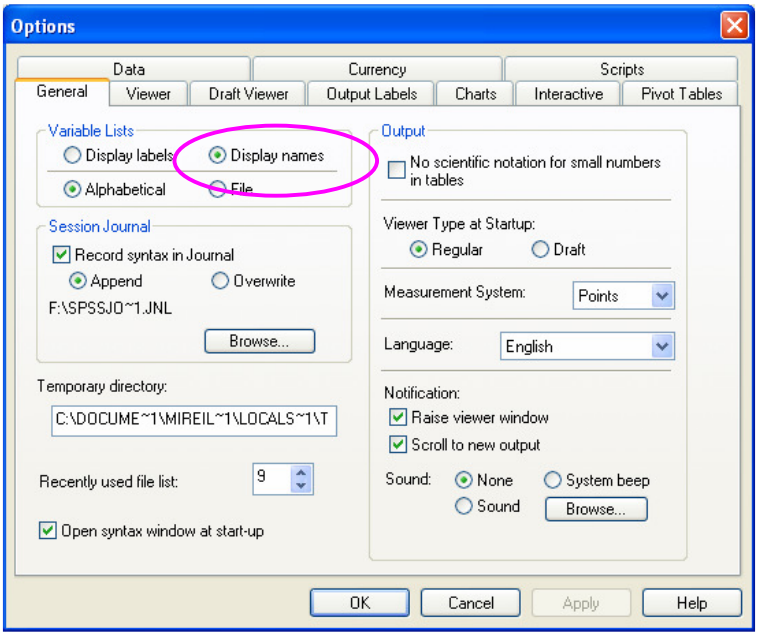
DATA MANAGEMENT IN SPSS

Practical tips on SPSS set-up:

In the menu click on 'Edit / Options':

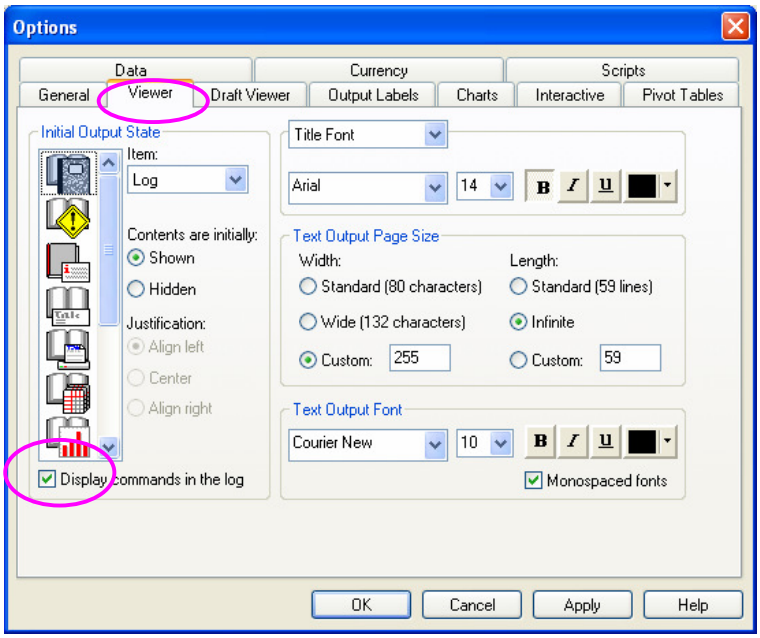


The 'Options' dialog box opens at the 'General' tab:

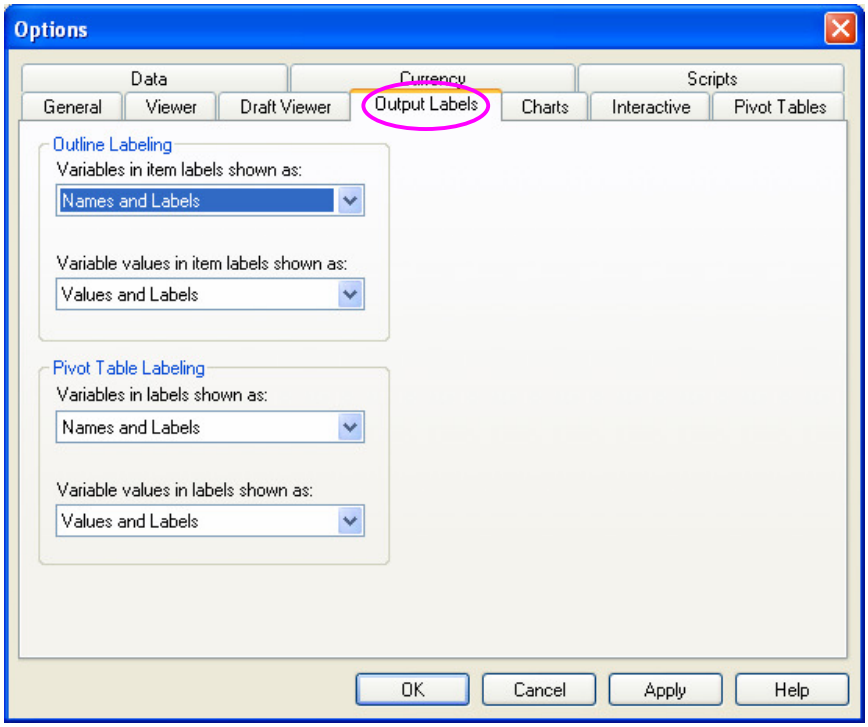


Select the option 'Display names' in the 'Variable Lists'

In the ‘Viewer’ tab tick ‘Display commands in the log’ at the bottom left: if you run some syntax, it will be displayed in the output screen which opens up automatically when a dataset is opened. This helps spot any errors in the syntax you write.

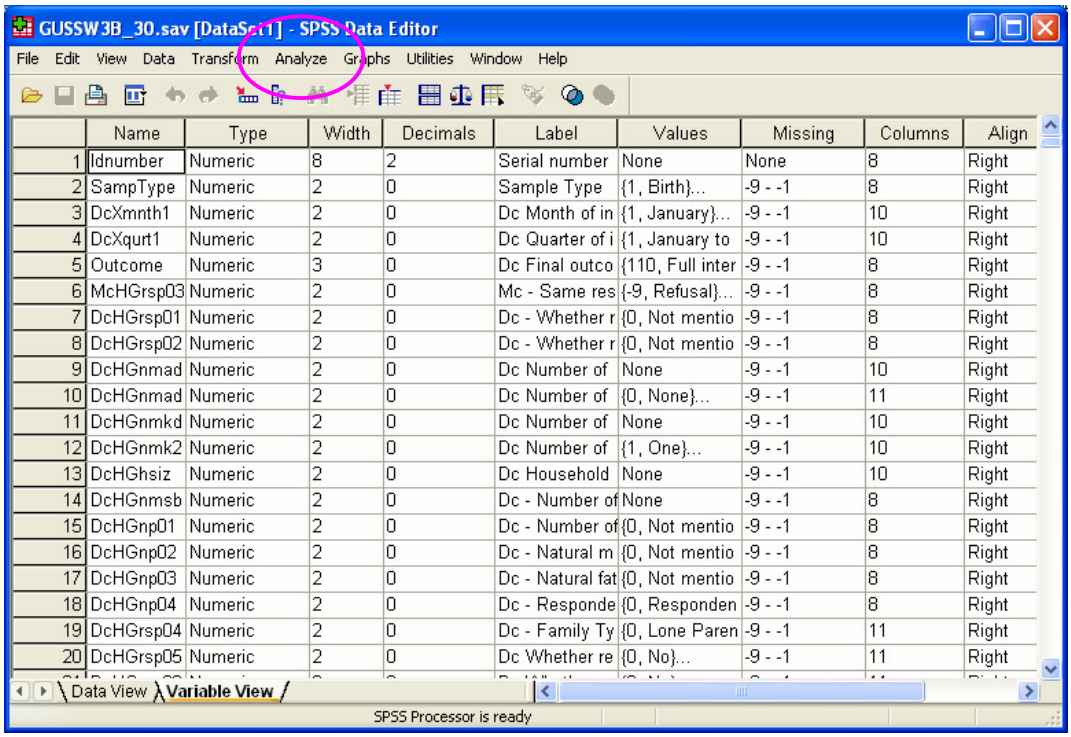


In the ‘Output Labels’ tab select the following options in the drop down lists:

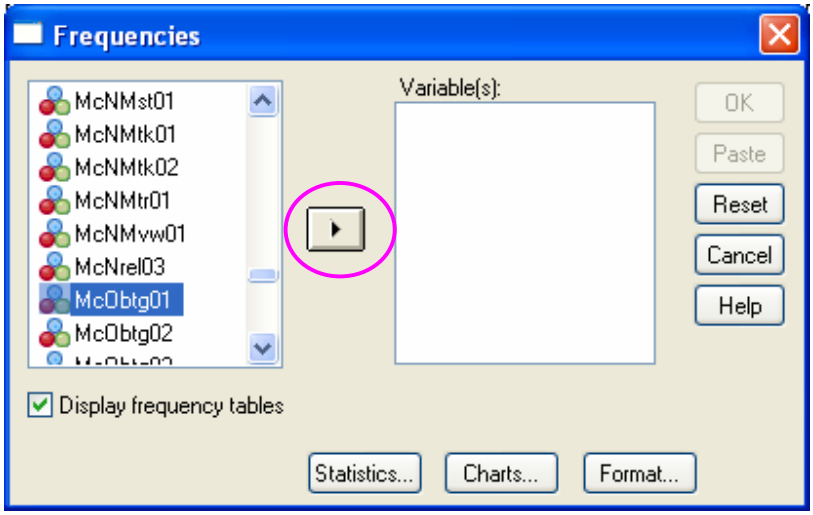


Examples of frequencies & cross-tabs via menu:

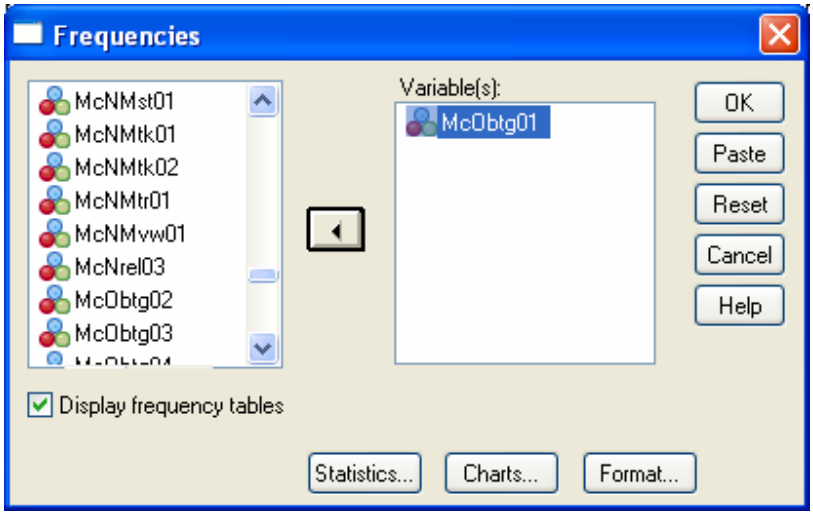
1) Select 'Analyze' in the menu



2) Select 'Descriptive statistics -> Frequencies' in the Analyze drop-down menu: the following dialog box will appear:



3) Browse the list of variables on the left and highlight the one(s) you want and click on the arrow in the middle:

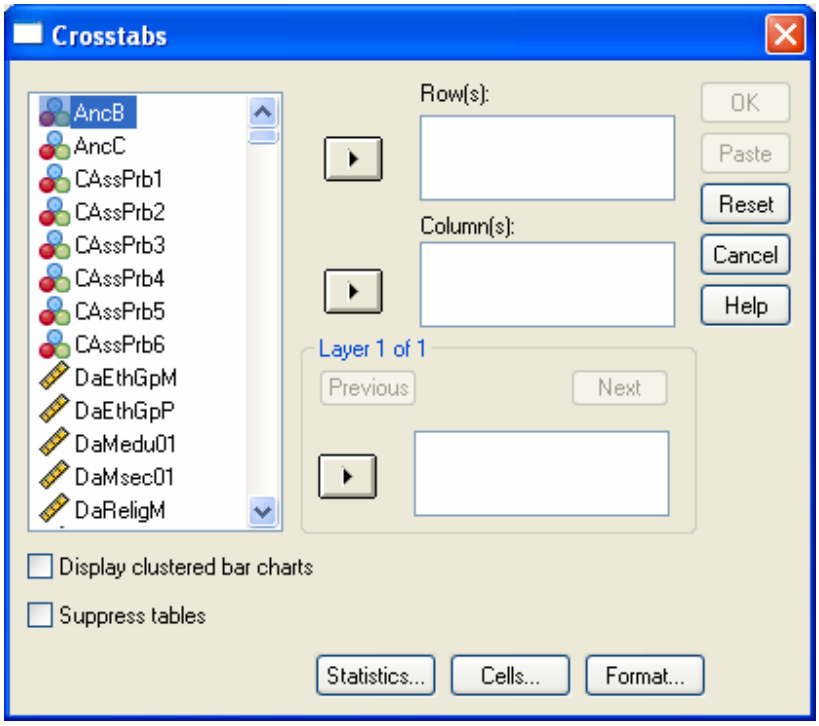


4) Click 'OK': the output will show you the frequency table for that variable

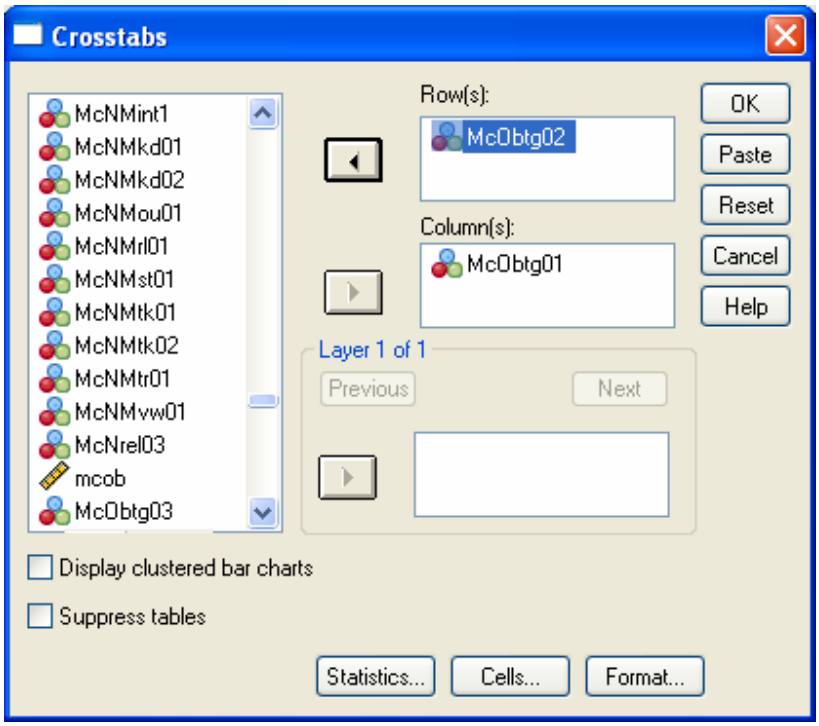
McObtg01 Mc - Attend group in last 12 months

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	568	45.6	45.6	45.6
	2 No	678	54.4	54.4	100.0
	Total	1246	100.0	100.0	

5) To do cross-tabs, use the same menu as for frequencies above but select 'Crosstabs' instead: Analyse → Descriptive statistics → Crosstabs



6) As before browse and highlight the variables you want to cross-reference and move them to the row and column boxes as applicable:



7) Click 'OK': the following output is displayed

McObtg02 Mc - Not group - nothing suitable * McObtg01 Mc - Attend group in last 12 months Crosstabulation

Count

		McObtg01 Mc - Attend group in last 12 months	
		2 No	Total
McObtg02 Mc - Not group - nothing suitable	0 Not mentioned	624	624
	1 Mentioned	54	54
Total		678	678

<u>Merging datasets via menu</u> : *see separate handout*

Recoding variables via syntax

example: study child's general health between Sw2 and Sw3

1) Check frequencies of the original variables MbHgen01 and McHgen01

McHgen01 Mc - Childs general health

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 ...very good,	898	72.1	72.1	72.1
	2 good,	288	23.1	23.1	95.2
	3 fair,	56	4.5	4.5	99.7
	4 bad,	4	.3	.3	100.0
	Total	1246	100.0	100.0	

MbHgen01 Mb - Child's general health

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Very good	829	66.5	67.7	67.7
	2 Good	313	25.1	25.6	93.2
	3 Fair	79	6.3	6.4	99.7
	4 Bad	2	.2	.2	99.8
	5 Very bad	2	.2	.2	100.0
	Total	1225	98.3	100.0	
Missing	**System**	**21**	1.7		
Total		1246	100.0		

21 sysmis in MbHgen01 = cases which were at Sweep 3 but not at Sweep 2 (previous sweep = Sweep 1 for those cases)

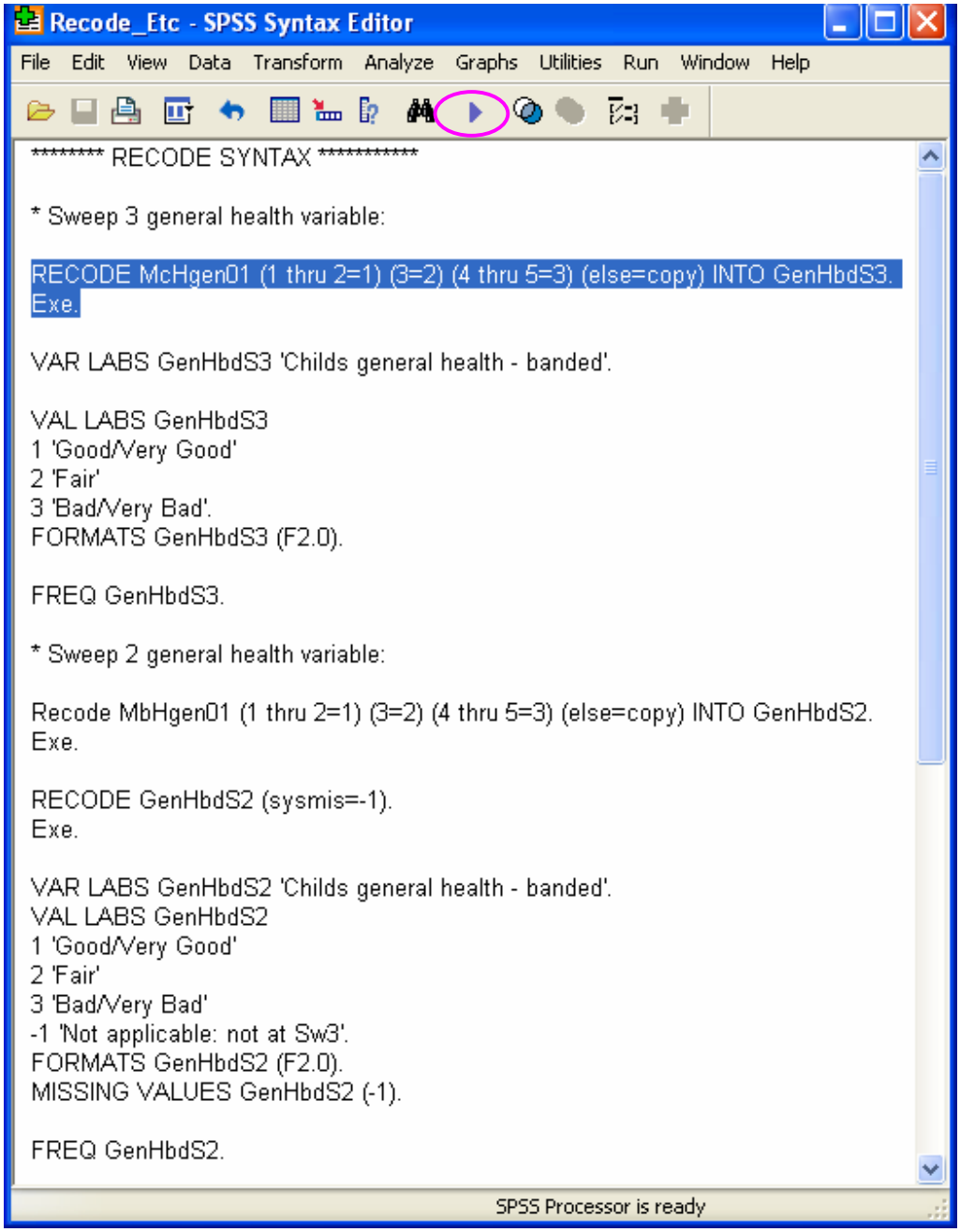
2) Open syntax file: go to menu 'File / Open / Syntax …'

3) Type Recode syntax in the syntax file: we want to group the original variables into answer categories Good (1,2) / Fair (3) and Bad (4,5)

   Example for Sweep 3:

   ```
   RECODE McHgen01 (1 thru 2=1) (3=2) (4 thru 5=3) (else=copy) INTO GenHbdS3.
   Exe.
   ```

4) and run it:



5) Check frequencies of the new variable:

GenHbdS3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1186	95.2	95.2	95.2
	2.00	56	4.5	4.5	99.7
	3.00	4	.3	.3	100.0
	Total	1246	100.0	100.0	

6) Tidy up variable label, value labels and output format directly within the variable row in the dataset 'Variable View' or with additional syntax below:

```
VAR LABS GenHbdS3 'Childs general health - banded'.

VAL LABS GenHbdS3
1 'Good/Very Good'
2 'Fair'
3 'Bad/Very Bad'.
FORMATS GenHbdS3 (F2.0).
```

7) Run the frequencies again:

GenHbdS3 Childs general health - banded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Good/Very Good	1186	95.2	95.2	95.2
	2 Fair	56	4.5	4.5	99.7
	3 Bad/Very Bad	4	.3	.3	100.0
	Total	1246	100.0	100.0	

8) If similar Recode syntax run for Sweep 2 variable:

GenHbdS2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1142	91.7	93.2	93.2
	2.00	79	6.3	6.4	99.7
	3.00	4	.3	.3	100.0
	Total	1225	98.3	100.0	
Missing	**System**	**21**	1.7		
Total		1246	100.0		

9) So adding a line to recode the system missings into a missing code:

```
RECODE GenHbdS2 (sysmis=-1).
Exe.
```

GenHbdS2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1.00	21	1.7	1.7	1.7
	1.00	1142	91.7	91.7	93.3
	2.00	79	6.3	6.3	99.7
	3.00	4	.3	.3	100.0
	Total	1246	100.0	100.0	

10) And running the frequencies again after tidying up variable label, value labels and output format:

```
VAR LABS GenHbdS2 'Childs general health - banded'.
VAL LABS GenHbdS2
1 'Good/Very Good'
2 'Fair'
3 'Bad/Very Bad'
-1 'Not applicable: not at Sw3'.
FORMATS GenHbdS2 (F2.0).
MISSING VALUES GenHbdS2 (-1).
```

GenHbdS2 Childs general health - banded

		Frequency	Percent	**Valid Percent**	Cumulative Percent
Valid	1 Good/Very Good	1142	91.7	93.2	93.2
	2 Fair	79	6.3	6.4	99.7
	3 Bad/Very Bad	4	.3	.3	100.0
	Total	1225	98.3	100.0	
Missing	-1 Not applicable: not at Sw3	21	1.7		
Total		1246	100.0		

Computing a Derived Variable using syntax

example: evolution of study child's general health between Sw2 and Sw3

1) Create a new variable 'GenHS2S3' coded 0 if no change between Sw2 and Sw3, 1 if child's health not as good at Sw3 and 2 if child's health better at Sw3

2) Set up and run Compute syntax:

```
COMPUTE GenHS2S3=-1.
IF (GenHbdS3=GenHbdS2) GenHS2S3=0.
IF (GenHbdS3<GenHbdS2) GenHS2S3=1.
IF (GenHbdS3>GenHbdS2) GenHS2S3=2.
Exe.
```

GenHS2S3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	**-1.00**	**21**	1.7	1.7	1.7
	.00	1119	89.8	89.8	91.5
	1.00	66	5.3	5.3	96.8
	2.00	40	3.2	3.2	100.0
	Total	1246	100.0	100.0	

The same number of Sw2 system missings have been coded as -1 by default

3) Check frequencies, tidy up variable and value labels, output formats:

```
VAR LABS GenHS2S3 'Childs general health evolution Sw2 to Sw3'.
VAL LABS GenHS2S3
0 'No change'
1 'better at Sw3'
```

2 ' not as good at Sw3'
-1 'Not Applicable: no data at Sw2'.
FORMATS GenHS2S3 (F2.0).
MISSING VALUES GenHS2S3 (-1).

GenHS2S3 Childs general health evolution Sw2 to Sw3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No change	1119	89.8	91.3	91.3
	1 better at Sw3	66	5.3	5.4	96.7
	2 not as good at Sw3	40	3.2	3.3	100.0
	Total	1225	98.3	100.0	
Missing	-1 Not Applicable: no data at Sw2	21	1.7		
Total		1246	100.0		