



Teaching students quantitative methods using resources from the British Birth Cohorts

Measuring signs of psychological distress or depression in teenagers and adults

The Malaise Inventory

MalaiseExercises.doc: The SPSS exercises detailed here use data from the 1970 British Cohort Study (BCS70). Students could replicate these exercises using data from the 1958 National Child Development Study (NCDS) for cross-cohort comparisons. After a short summary of the two data files, *b034malaise.sav* and *n042malaise.sav* the document provides a number of step-by-step exercises for students to work through and additional exercises to complete. The exercises use the drop down menus but the SPSS syntax to run the commands also included. The **RutterBehaviourExercises.doc** contains additional exercises. By replacing the Rutter behaviour score variables with the Malaise score variables these exercises could be easily adapted for students using either the *b034malaise.sav* or *n042malaise.sav* data files. The techniques covered in this document range from introductory descriptive statistics to multiple and logistic regression and include:

- a) Frequencies
- b) Means
- c) Cross-tabulations
- d) Constructing summary scores
- e) Recoding continuous into dichotomous variables
- f) Correlations
- g) Linear regression
- h) Logistic regression.

SPSS Data Files: BCS70 data 'b034malaise.sav', NCDS data 'n042malaise.sav'

BCS70 data

This longitudinal dataset includes information from the 1970 British Cohort Study (BCS70) for looking at the relationship between socio-economic circumstances in childhood and mental well-being in adulthood over time. Data from birth are provided by the cohort members' mother; data collected at age 5, age 10 and age 16 are provided by the mother (or father) and the cohort member (child) and data from collected at age 26, 30 and 34 are provided by the cohort member themselves. There are 18,732 cohort members (individuals) in the dataset, 9,740 are men and 8,984 are women. Eight babies did not have a sex recorded at their premature birth.

NCDS data

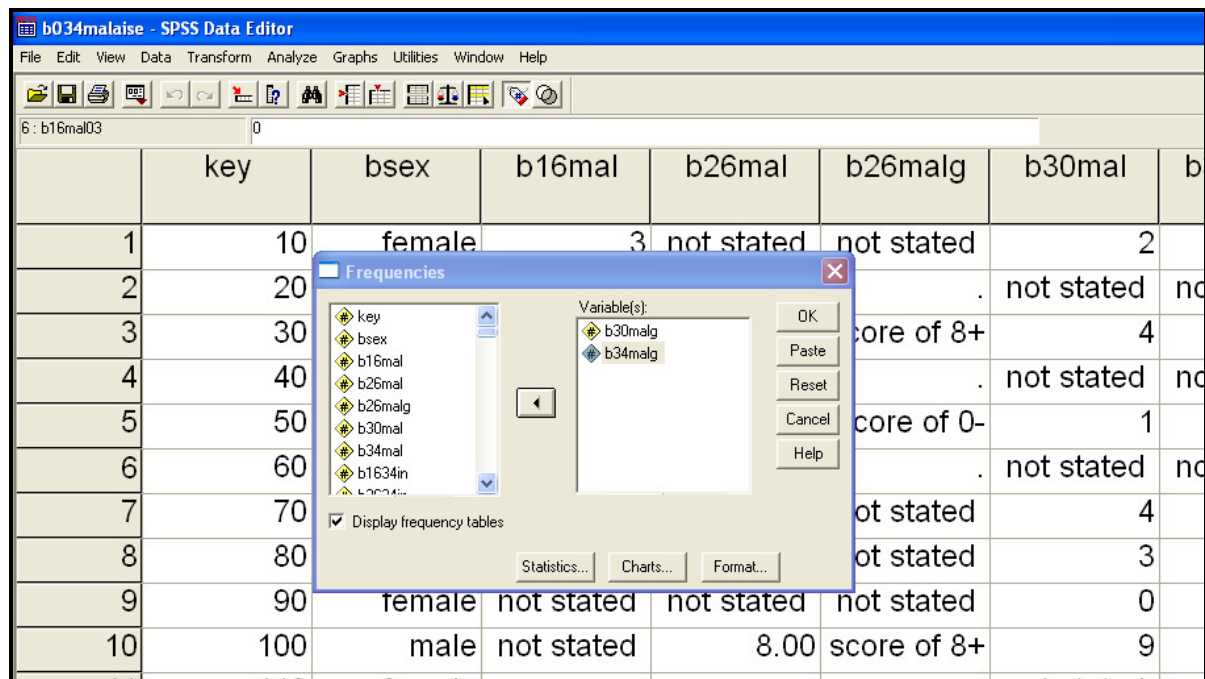
This longitudinal dataset includes information from the 1958 National Child Development Study (NCDS) for looking at the relationship between socio-economic circumstances in childhood and mental well-being in adulthood over time. Information included in this file which was collected at birth is provided by the cohort members' mother; information collected at age 7, age 11 and age 16 is provided from the mother (or father) and at age 23, 33 and 42 the cohort member themselves provided the information. There are 18,560 cases in the dataset, 8,960 are women and 9,593 are men. Seven babies did not have a sex recorded at their premature birth.

Step-by-Step Exercises using SPSS

Exploring the data

1. How many cohort members were depressed in their thirties? Were more depressed at age 30 or age 34? (Hint: frequency of b30malg and b34malg variables).

From the Analyse drop-down menu select Descriptive Statistics and then Frequencies. Select b30malg b34malg and click on the ► button to move the two variables into the Variable(s) box. Click on OK.

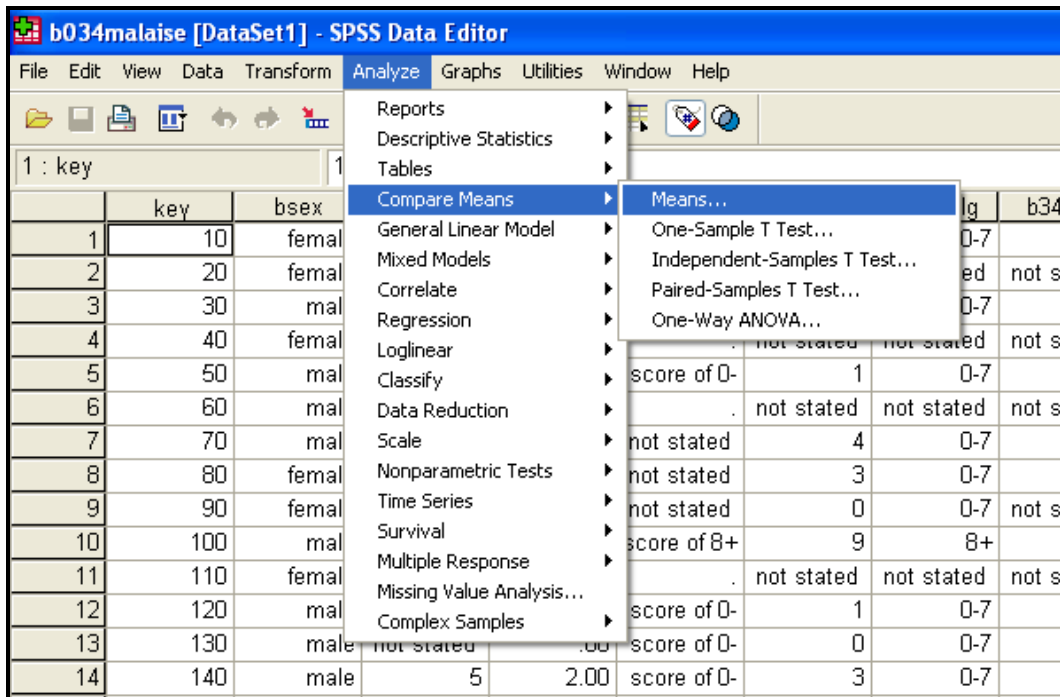


If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

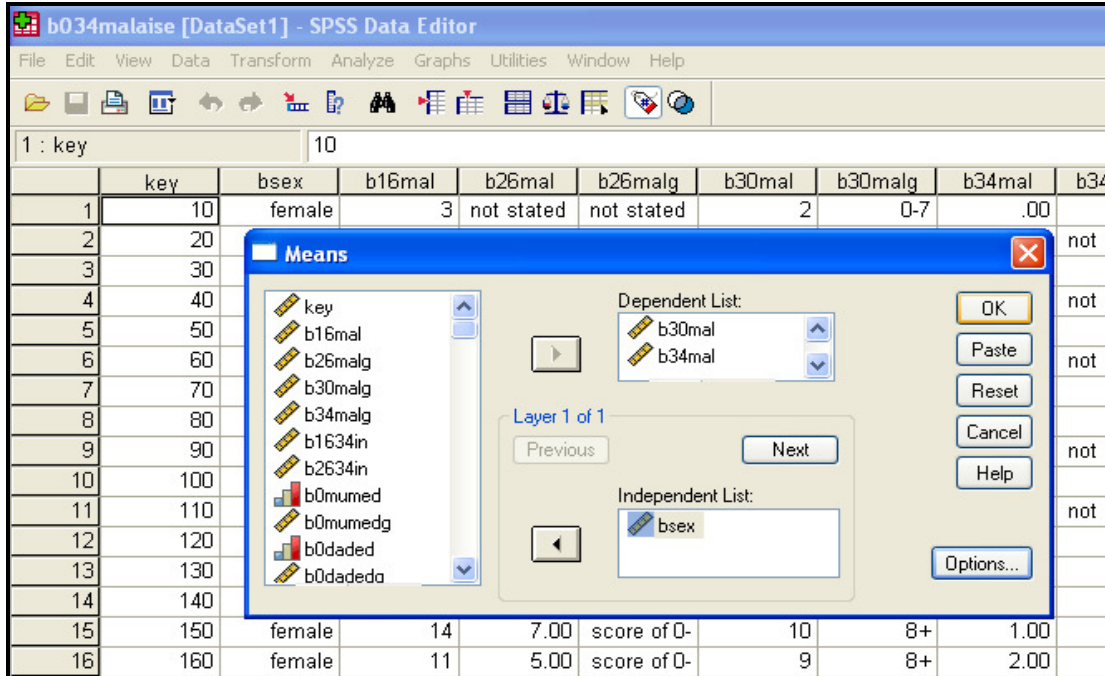
```
FREQUENCIES
  VARIABLES=b30malg b34malg
  ORDER= ANALYSIS .
```

2. Do men or women have higher mean malaise scores in adulthood? Is the mean score consistently higher for men or women at age 26, 30 and 34? (Hint: Means of b26mal b30mal b34mal by bsex)

From the Analyse drop-down menu select Compare Means and then Means.



Click on b26mal and then b30mal and b34mal and move the variables via the ► button into the 'Dependent list' box. Click on bsex and move into the 'Independent list' box. Click on OK.



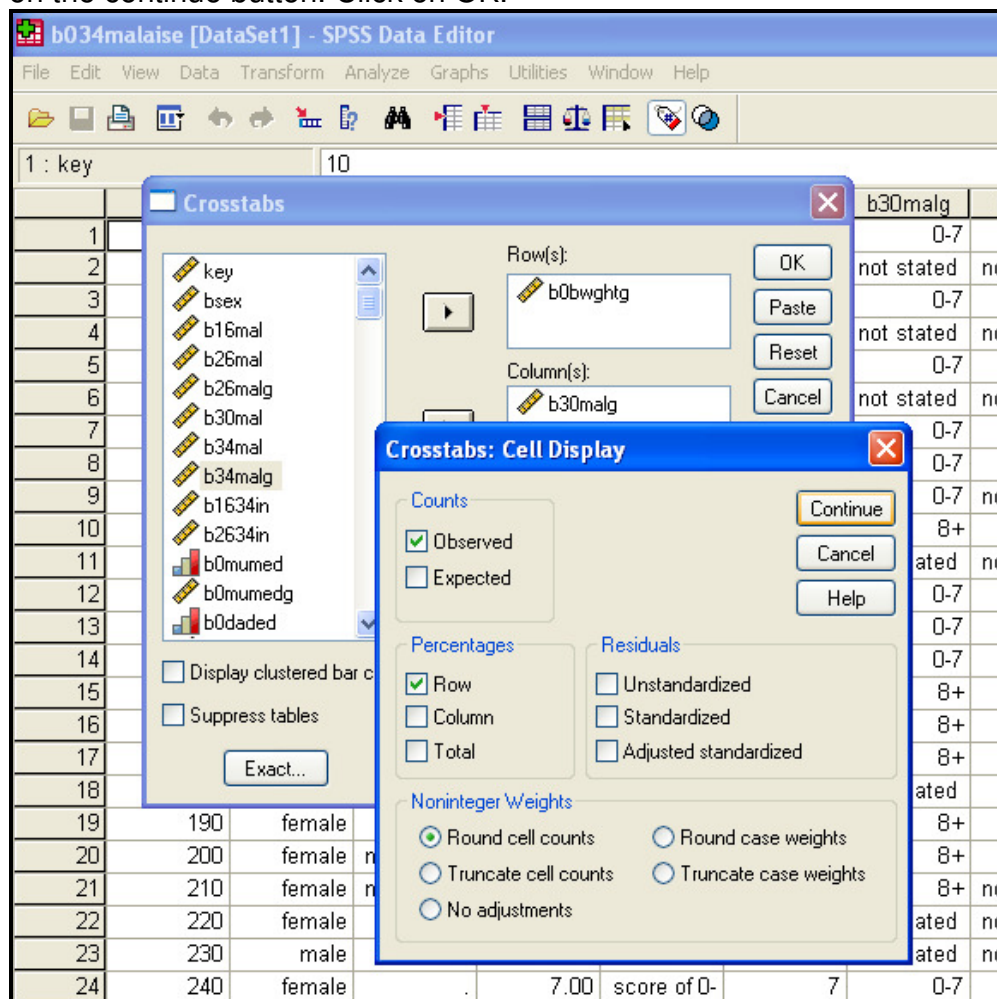
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

MEANS

```
TABLES=b26mal b30mal b34mal By bsex
/CELLS MEAN COUNT STDDEV.
```

- Were cohort members with a low birth weight (<2515 grams) more or less likely than those with a normal birth weight (2515 grams +) to be depressed at age 30? (Hint: cross-tab of b0bwghtg and b30malg).

From the Analyse drop-down menu select Descriptive Statistics and then Crosstabs. Select b0bwghtg and click the ► button to place the variable into the Row(s) box. Click on b30malg and click the ► button to place into column(s) box. Click on the 'Cells' button and select 'row' under the percentage heading. Click on the continue button. Click on OK.



If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
CROSSTABS  
  /TABLES=b0bwghtg BY b30malg  
  /FORMAT= AVALUE TABLES  
  /CELLS= COUNT ROW  
  /COUNT ROUND CELL.
```

4. Is there a relationship between family social class and mental well-being? Are men and women whose father worked in a professional occupation when they were born more or less likely than those with a father who worked in an unskilled manual job when they were born to be depressed at age 30? (Hint: the same cross-tab procedure as above using b0fsoc and b30malg variables).
5. Does academic achievement relate to mental well-being? Are boys or girls who have no academic qualifications by age 34 the most likely to be depressed at age 34? Is the relationship between academic achievement and mental well-being stronger for men or women? Who is the least likely of all to be depressed at age 34? (Hint: cross-tab of b34hq5 by b34malg by bsex).

From the Analyse drop-down menu select Descriptive Statistics and then Crosstabs. Select b34hq5 and click the ► button to place the variable into the Row(s) box. Click on b34malg and click the ► button to place into the column(s) box. Click on bsex and click on the bottom ► button and put into the 'Layer' box. Click on the 'Cells' button and select 'row' under the percentage heading. Click on the continue button. Now click on OK.

The screenshot shows the SPSS Data Editor window with a data view of 24 rows and 10 columns. A 'Crosstabs' dialog box is open, showing the following configuration:

- Row(s): b34hq5
- Column(s): b34malg
- Layer 1 of 1: bsex
- Display clustered bar charts:
- Suppress tables:
- Buttons: Exact..., Statistics..., Cells..., Format...

If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
CROSSTABS
/TABLES=b34hq5 BY b34malg BY bsex
/FORMAT= AVALUE TABLES
/CELLS= COUNT ROW
/COUNT ROUND CELL.
```

Making comparable summary scores at age 30 and age 34

When all 24 Malaise questions are asked the cut off point to indicate that someone is experiencing depression is a score of 8+, that is 'yes' was answered to at least 8 questions. In the survey that took place in 2004 when cohort members were age 34, just 9 of the 24 Malaise questions were asked. A score of 4+ is the cut off point to indicate that someone is experiencing depression.

In the survey that took place in 2000 when cohort members were age 30 the full 24 questions were asked. A way to test the validity of only using the 9 questions

at age 34 is therefore to see if the same percentage of cohort members are identified as depressed at age 30 from their responses to the 9 questions as they are from responses to all 24 questions (b30malg).

6. Make an overall score variable from 9 questions at age 30 to match the overall score variable at age 34. Are the same percentage of men and women identified as depressed by the reduced number of questions? This can be done through the drop down menus or altering some existing syntax. (Hint: in the syntax below replace the variables starting with b34... with variables starting with b30.... Variable names correspond directly. For example b34mal02 is the same question as b30mal02)

This is the syntax used to construct the overall malaise score at age 34 (b34mal) and the grouped variable (b34malg).

```
count b34mal = b34mal02 b34mal03 b34mal05 b34mal09 b34mal12 b34mal14
b34mal16 b34mal20 b34mal21 (1).
```

The syntax below maximises the number of cohort members included in the variables by only excluding those with enough missing values to give them a 'high' (4+) malaise score.

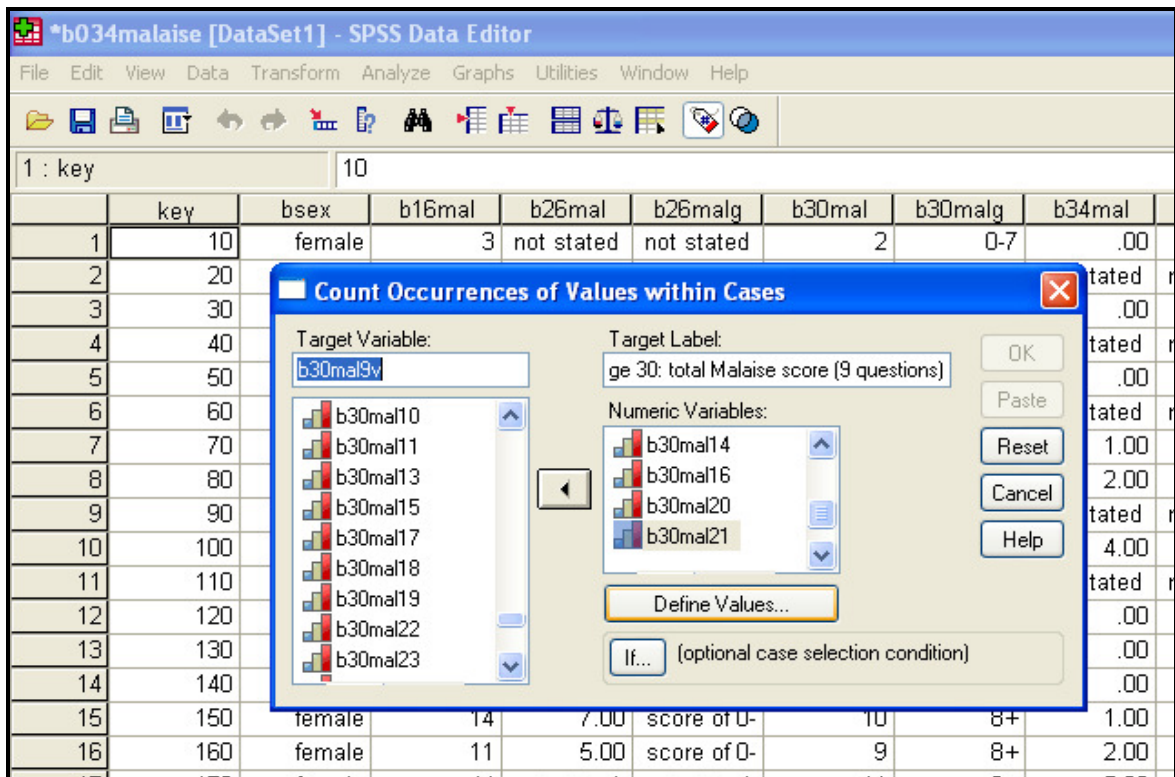
```
count b34miss = b34mal02 b34mal03 b34mal05 b34mal09 b34mal12 b34mal14
b34mal16 b34mal20 b34mal21 (missing).
compute malmiss =b34mal + b34miss.
if (b34miss > 0 and b34mal <= 3 and malmiss >= 4) b34mal = -1.
If (b34miss = 9) b34mal = -2.
recode b34mal (0 thru 3=1) (4 thru highest = 2) (-1=-1) (-2=-2) into b34malg.
missing values b34mal b34malg (-1,-2).
variable labels b34mal 'bcs70 age 34: total Malaise score (9 questions)'.
variable labels b34malg 'bcs70 age 34: total Malaise score - grouped'.
value labels b34mal -1'incomplete info' -2'not stated any questions'.
value labels b34malg 1'0-3' 2'4+' -1'incomplete info' -2'not stated any questions'.

freq b34mal b34malg.
```

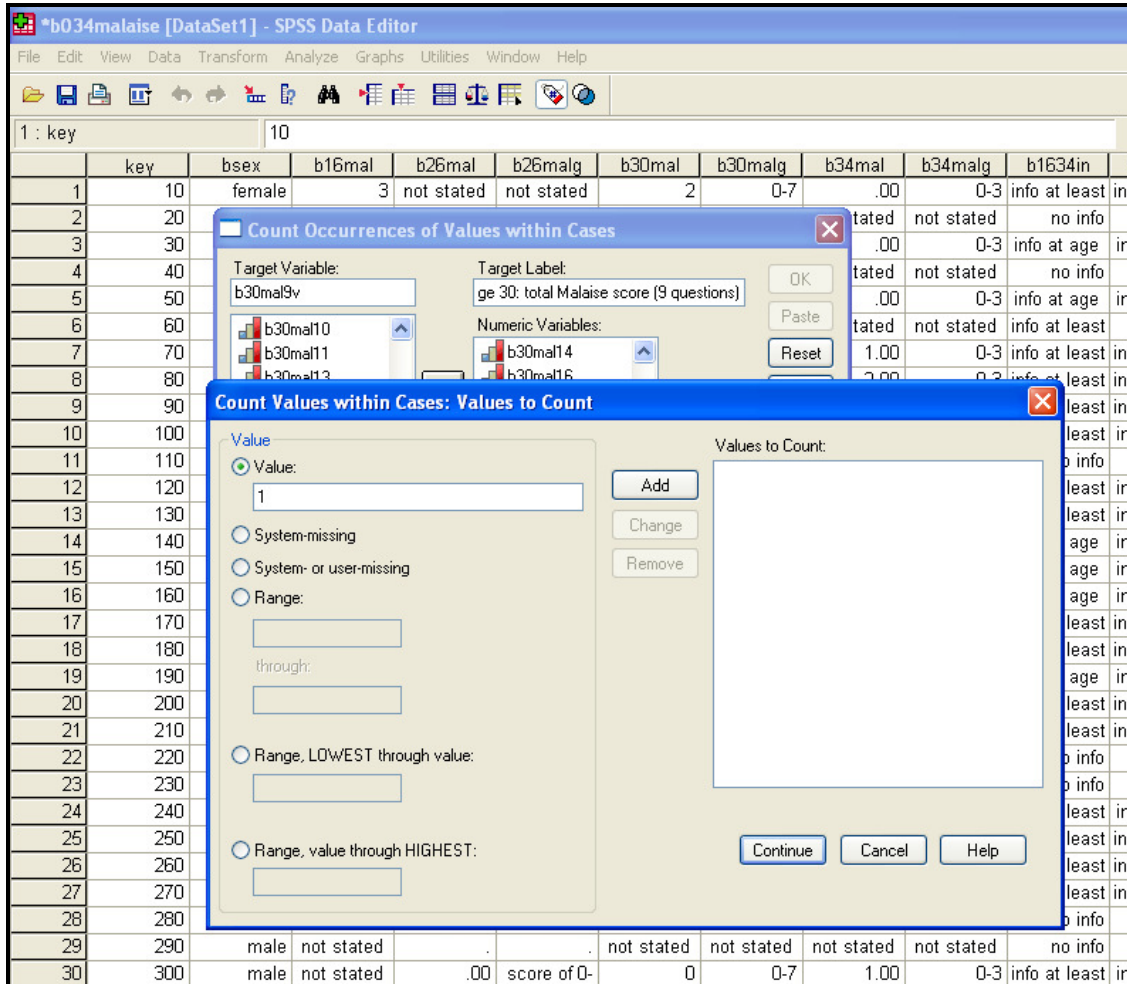
Constructing the score from the drop down menus

STEP 1 Under the *Transform* drop down menu, select *Count*. In the *Target Variable* box type in the name of the new summary variable (b30mal9v). In the *Target Label* box type in an appropriate label to help identify what the variable is. For example 'BCS70 age 30: total Malaise score (9 questions)'. Scroll down the list of variables and select the 9 malaise variables that are the same as those asked at age 34 (b34mal02, b34mal03, b34mal05, b34mal09, b34mal12, b34mal14, b34mal16, b34mal20, b34mal21) and click on the ► button to place

them in the *Variables:* box. Once this has been done, click on the *Define Values* button.



On the left hand side of the new screen, click on the *Value* button and enter 1 into the empty box. Click on the *Add* button on the right hand side of the screen to move the information into the *Values to Count* box. Click on *Continue*, then OK.



If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

COUNT

b30mal9v = b30mal02 b30mal03 b30mal05 b30mal09 b30mal12 b30mal14
b30mal16 b30mal20 b30mal21 (1) .

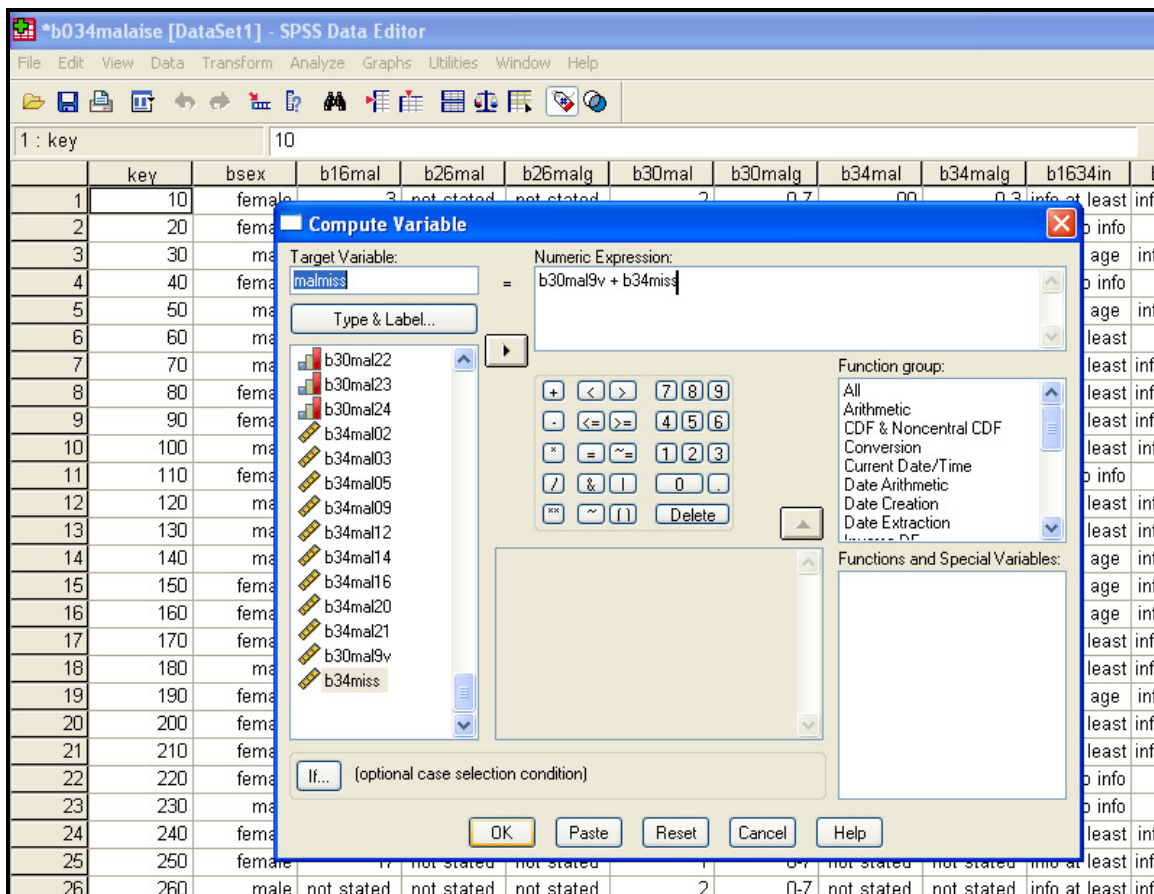
VARIABLE LABELS b30mal9v 'BCS70 age 30: total Malaise score (9 questions)'.
EXECUTE .

STEP 2 Repeat this process to count the number of missing responses for each cohort member (variable name b30miss). Instead of counting the number of yes

(value 1) responses, select *System- or user-missing* in the Values to count window.

STEP 3 The next stage is to maximise the number of cohort members we can include in the new dichotomous variable we will construct (b30mal9vg).

Under the *Transform* drop down menu, select *Compute*. In the *Target Variable* box type in the name of the new variable that will add together the number of 'yes' responses and the number of 'missing' responses (e.g. malmiss). Scroll down the list of variables to the two new variables you have just created b30mal9v and b30miss. Select b30mal9v and click on the ► button to move it into the *Numeric Expression:* box. Click on the + symbol. Then select b30miss and click on the ► button to now move this variable into the *Numeric Expression:* box. Once this has been done, click on OK.

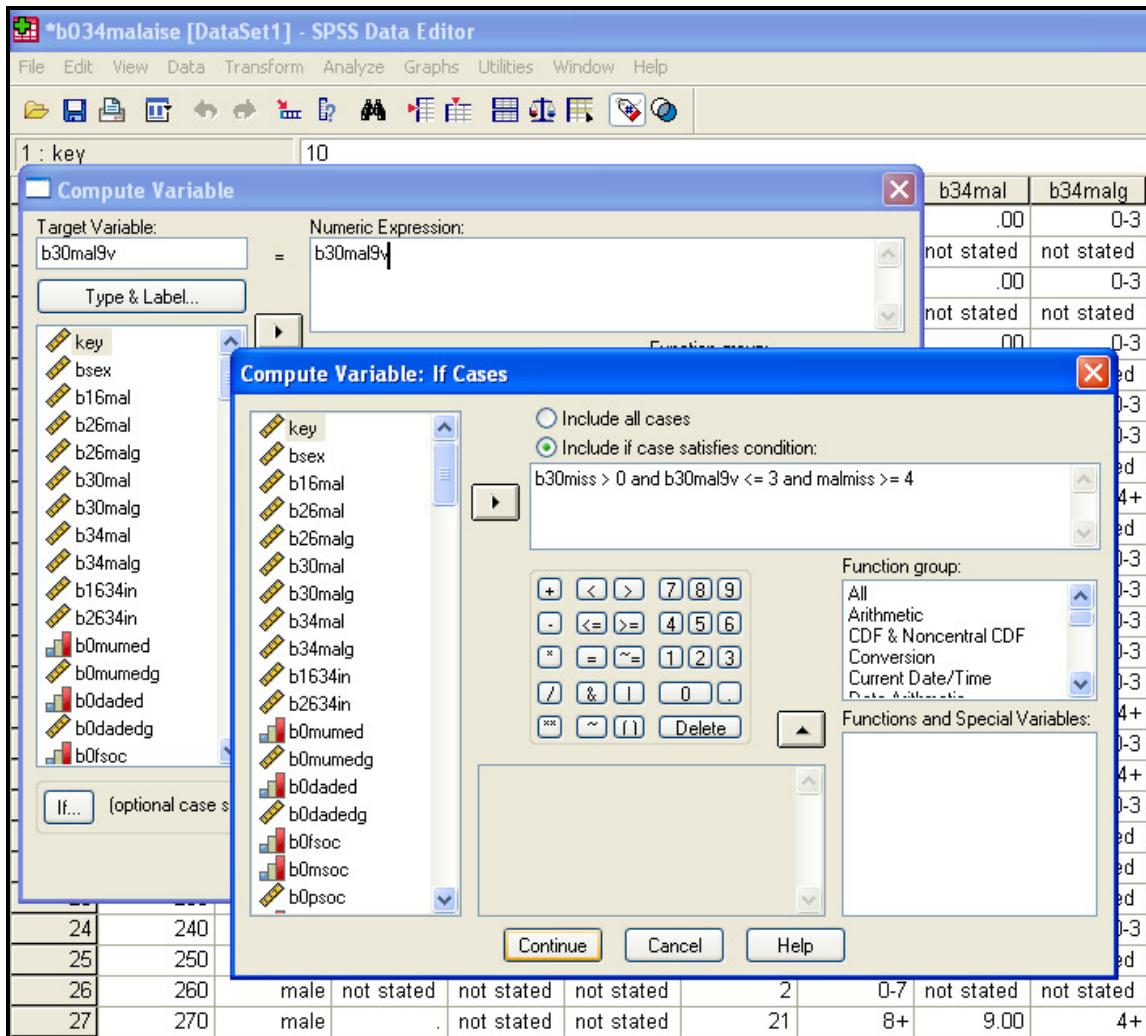


If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
COMPUTE malmiss = b30mal9v + b30miss .
EXECUTE .
```

STEP 4 Recoding the cohort members who have a valid score in b30mal9v to 'missing' if they did not answer all 9 questions and the number of missing answers could effect whether they were assigned a depressed label or not in the new dichotomous variable b30mal9vg.

Under the *Transform* drop down menu, again select *Compute*. In the *Target Variable* box type in the name of the existing variable b30mal9v. In the *Numeric Expression* box type in -1. (We do not want to make a new variable; we are just altering the existing one.) Click on the *If* button at the bottom of the screen. A new screen will open. Click on the button *Include if case satisfies condition*: Either type in the syntax rules (shown below) or select the variables from the variable list and click on the ► button to move them into the box. Likewise click on the mathematical symbols and numbers. Click on Continue and then OK. SPSS will ask if you want to change the existing variable. Click Yes.



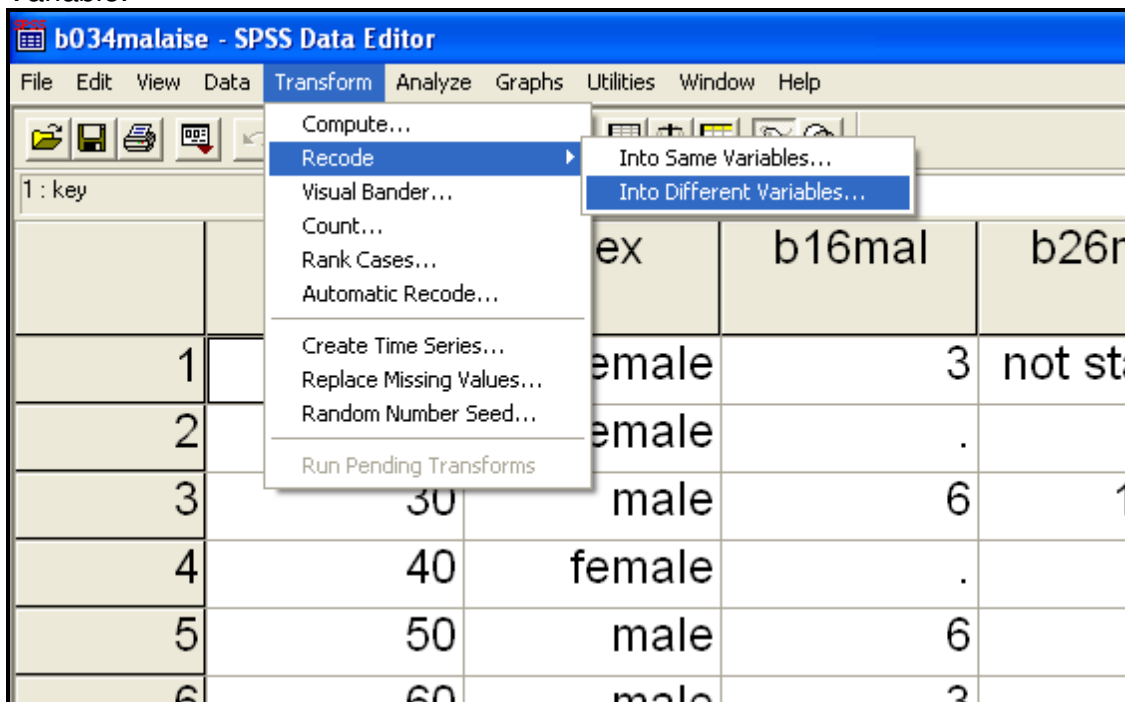
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

IF (b34miss > 0 and b30mal9v <= 3 and malmiss >= 4) b30mal9v = -1.
EXECUTE .

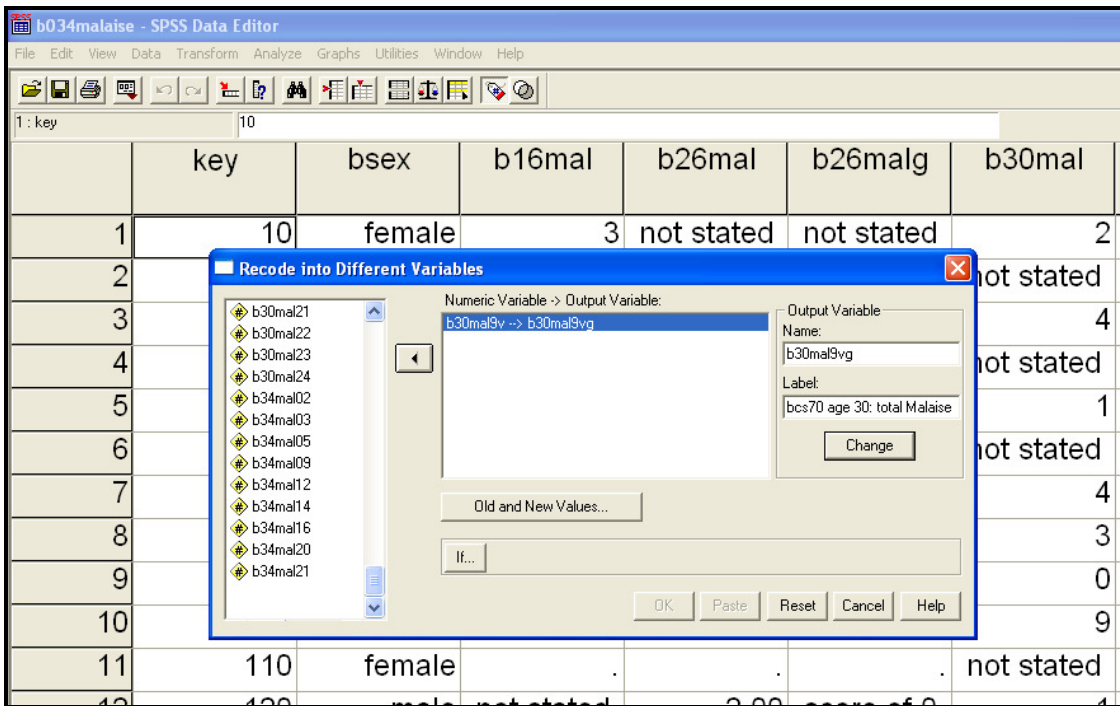
Repeat this process to assign a value of -2 to those who did not answer any of the 9 Malaise questions. (HINT: if b30miss = 9, b30mal9v = -2).

STEP 5: Recode the continuous variable into a dichotomous variable.

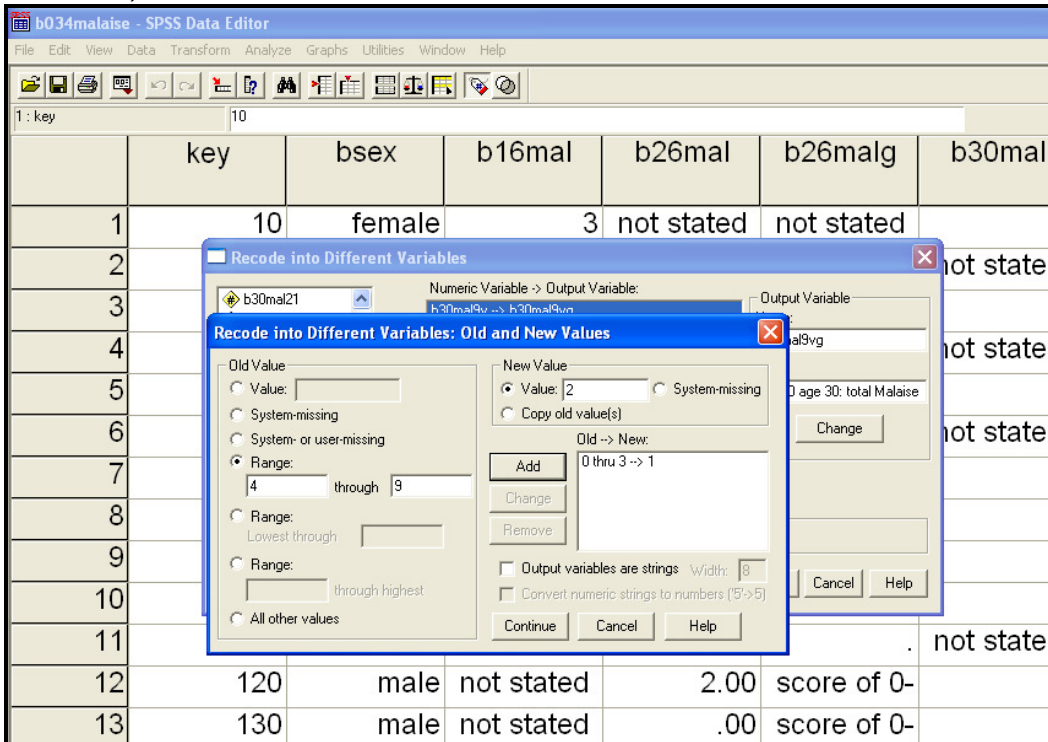
Under the Transform drop down menu select Recode and then Into Different Variable.



Scroll down the list of variables and select b30mal9v, click on the ► button to move the variable into the Numeric variable → Output variable box. Type the name of the new variable in the Name box, and label in the Label box. Click on the Change button to move the information into the Numeric variable → Output variable box. Now click on the Old and New Values button.



Under Old value select Range and type 0 in the first box and 3 in the second box. Under New value select Value and type in 1. Click on the Add button to move the information into the Old → New box. Repeat this for the values in the range 4 – 9 in b30mal9v and assign 2 as the new value in b30mal9vg. Under Old Value select Value and type in -1. Under New Value select Copy old value(s), then the Add button. Repeat this to copy the value -2 to the new variable. Click on Continue, then OK.



If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
RECODE
  b30mal9v
  (-1=Copy) (-2=Copy) (0 thru 3=1) (4 thru 9=2) INTO b30mal9vg .
VARIABLE LABELS b30mal9vg 'bcs70 age 30: total Malaise score (9 questions -
grouped'.
EXECUTE .
```

STEP 6: Assigning missing values and value labels to b30mal and b34mal9vg. In the Variable View window, scroll down the variables to b30mal9v and b30mal9vg. Move across to the Values column and click on the box for b30mal9vg. A new screen will open. Type in 1 in the Value box and 0-3 in the Value Label box. Click on Add. Repeat this for the following 3 values: 2= 4+, -1= incomplete info, -2=not stated any questions. Click on OK.

Assign the same value labels to -1 and -2 in b30mal9v.

The screenshot shows the SPSS Data Editor Variable View window. The table below represents the data shown in the window:

	Name	Type	Width	Decimals	Label	Values	Missing
104	b30mal23	Numeric	1	0	bcs70 age 30: {*, dont know/	-1	
105	b30mal24	Numeric	1	0	bcs70 age 30: {*, dont know/	-1	
106	b34mal02	Numeric	2	0	bcs70 age 30: {-1, dont know	-9 - -1	
107	b34mal03	Numeric	2	0	bcs70 age 30: {-1, dont know	-9 - -1	
108	b34mal05	Numeric	2	0	bcs70 age 30: {-1, dont know	-9 - -1	
109	b34mal09	Numeric	2	0	bcs70 age 30: {-1, dont know	-9 - -1	
110	b34mal12	Numeric	2	0	bcs70 age 30: {-1, dont know	-9 - -1	
111	b34mal14	Numeric	2	0			--1
112	b34mal16	Numeric	2	0			--1
113	b34mal20	Numeric	2	0			--1
114	b34mal21	Numeric	2	0			--1
115	b30mal9v	Numeric	8	2			one
116	b30miss	Numeric	8	2			one
117	malmiss	Numeric	8	2		None	None
118	b30mal9vg	Numeric	8	2	bcs70 age 30:	None	None
119							

The 'Value Labels' dialog box for variable b30mal9v is open, showing the following content:

- Value: -2
- Value Label: not stated any questions
- Buttons: Add, Change, Remove, OK, Cancel, Help
- Current list of labels:
 - 1.00 = "incomplete info"
 - 1.00 = "0-3"
 - 2.00 = "4+"

Click on the cell in the Missing column for variable b30mal9vg, and then select Discrete missing values. Type in -2 and -1 into the separate cells. Repeat this for variables b30mal9v.

The screenshot shows the SPSS Data Editor window with a list of variables and a 'Missing Values' dialog box open. The variable list has columns for Name, Type, Width, Decimals, Label, Values, and Missing. The dialog box is for variable 'b30mal9v' and shows 'Discrete missing values' selected with values '-2' and '-1'.

	Name	Type	Width	Decimals	Label	Values	Missing
104	b30mal23	Numeric	1	0	bcs70 age 30:	{*, dont know/	-1
105	b30mal24	Numeric	1	0	bcs70 age 30:	{*, dont know/	-1
106	b34mal02	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
107	b34mal03	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
108	b34mal05	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
109	b34mal09	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
110	b34mal12	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
111	b34mal14	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
112	b34mal16	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
113	b34mal20	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
114	b34mal21	Numeric	2	0	bcs70 age 30:	{-1, dont know	-9 - -1
115	b30mal9v	Numeric	8	2			None
116	b30miss	Numeric	8	2			None
117	malmiss	Numeric	8	2			None
118	b30mal9vg	Numeric	8	2			-2.00 - -1.00
119							

As you can see, once you know the syntax to write, it is a much faster process than using the drop down menus. If you save the syntax file you will also retain a record of how you made the variables.

Now (finally!) compare the frequencies of the original Malaise dichotomous variable and the new dichotomous Malaise variable from the 9 questions. Are the same percentage of cohort members identified as depressed by the reduced number of questions? (Hint: Frequencies of b30malg b30mal9vg). Is this the same for men and women? (Hint: run a cross tab of b30malg b30mal9vg by bsex)

Correlations between scores

- Depression (malaise) is measured at age 16, 26, 30 and 34. At which two time points, ages, are the scores most strongly associated? Is this the same for men and women? (Hint: correlate the four continuous variables b16mal b26mal b30mal b34mal).

From the Analyse drop-down menu select Correlate and then Bivariate. Select b16mal and click the ► button to place it into the Variables box. Repeat for variables b26mal, b30mal, b34mal. Click OK.

	key	bsex	b16mal	b26mal	b26malg	b30mal
372	3680	male	not stated	11.00	score of 8+	not stated
373	3690	female	7	6.00	score of 0-	5
374	3700	female				0
375	3710	female				1
376	3720	male				t stated
377	3730	female				1
378	3740	female				2
379	3750	female				t stated
380	3760	male				t stated
381	3770	female				t stated
382	3780	male				t stated
383	3790	female	17	11.00	score of 8+	4
384	3800	male	not stated	not stated	not stated	not stated

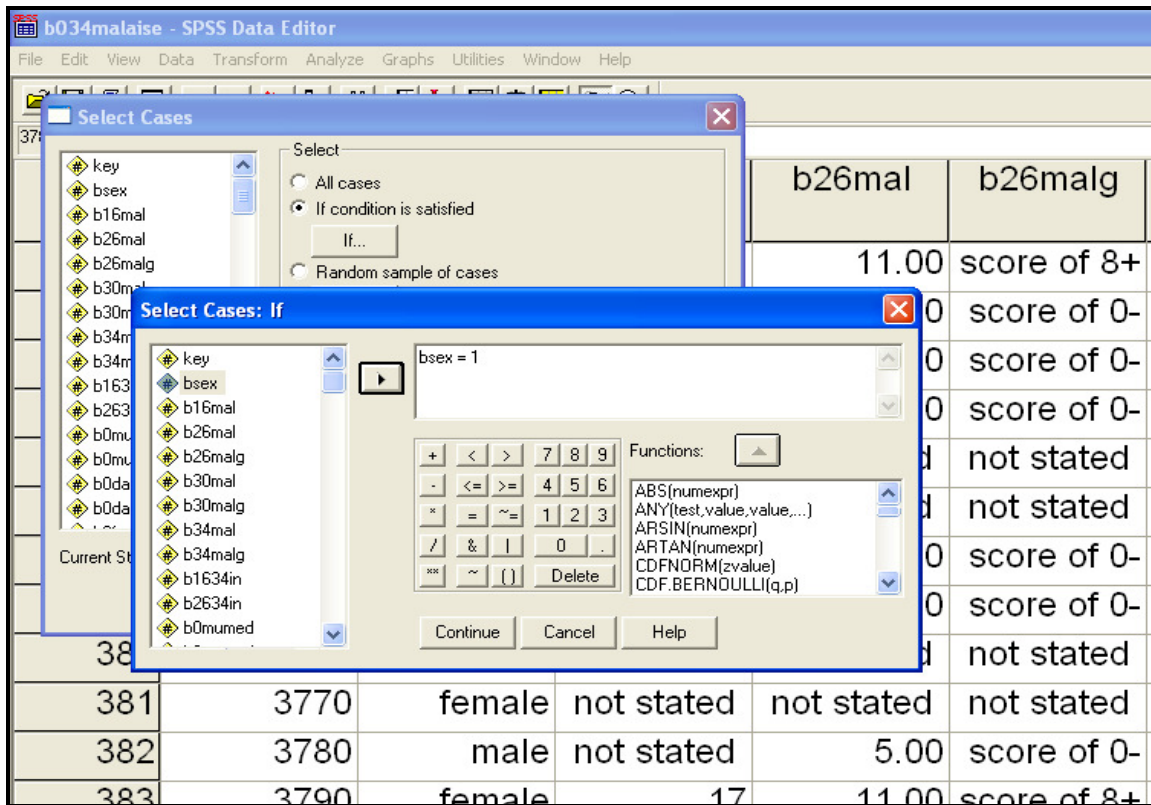
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```

CORRELATIONS
/VARIABLES=b16mal b26mal b30mal b34mal
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE .
    
```

Re-run the analyses for men and women separately. (Hint: use Select cases).

From the Data drop down menu, select Select Cases. Click on If condition is satisfied and then the If button. Select the bsex variable and click the ► button to place it into the white box. To select men, add = 1 after bsex. Click on Continue and then OK. Now re-run the Correlate command, first for men then for women. To select women only, add = 2 after bsex.



If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
USE ALL.
COMPUTE filter_$=(bsex = 1).
VARIABLE LABEL filter_$ 'bsex = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

NB: if you are using syntax, a simpler way of selecting men only would be to add the following instruction before an analyses.

```
Temporary.
Select if (bsex = 1).
CORRELATIONS
/VARIABLES=b16mal b26mal b30mal b34mal
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE .
```

Linear Regression

Linear Regression estimates the coefficients by a linear equation, involving one or more independent variables, which best predict the value of the dependent variable, in this example malaise score at age 34.

- Although the correlation between malaise scores at age 16 and 34 was the weakest among the four scores, the positive correlation was highly significant. Using simple linear and then multiple regression we first predict a high malaise score at age 34 from malaise score at age 16, and then control for birthweight, gender, mother's age of leaving full-time education and own highest qualification by age 34.

The screenshot shows the SPSS Data Editor window with a data table. Overlaid on this is the 'Linear Regression' dialog box. The 'Dependent:' field contains 'b34mal'. The 'Independent(s):' field contains 'b16mal', 'bsex', and 'b0bwght'. The 'Method:' dropdown is set to 'Enter'. At the bottom of the dialog box, there are four buttons: 'Statistics...', 'Plots...', 'Save...', and 'Options...'. Two arrows point from these buttons to callout boxes below the image.

Click here to select *Descriptive Statistics* and *Casewise Diagnostics* – this will identify any outliers in the data

Click here to request a plot of Standardised Predicted Residuals (Y) against Standardised Predicted Values (X)

The following syntax is produced when the paste command is selected

If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT b34mal
/METHOD=ENTER b16mal /METHOD=ENTER b16mal bsex b0bwght
b0mumed b34lefted
/SCATTERPLOT=(*ZPRED ,*ZRESID )
/CASEWISE PLOT(ZRESID) OUTLIERS(3) .
```

Fitting real data into the Linear Regression equation

The Linear Regression equation is

$$\text{(Predicted) } \hat{Y} = b_0 = b_1(x_1) + b_2(x_2) + \dots + b_p(x_p)$$

Using the regression equation (above), use the statistics produced in the *B* column in the *Coefficients* table in the SPSS output file and the information below to calculate the estimated Malaise score for a cohort member at age 34.

This particular cohort member

- is female
- had a low birth weight of 2410 grams
- had a very high malaise score at age 16 of 44
- had a mother who left education at age 15,
- had an A/S level qualification (highest academic qualification achieved)

SPSS syntax to identify this cohort member is

temporary.

select if (bsex = 2 and b16mal = 44 and b0mumed = 15 and b0bwght = 2410 and b34hq13 = 4).

SUMMARIZE

/TABLES=key bsex b16mal b0mumed b0bwght b34hq13.

/FORMAT=VALIDLIST NOCASENUM TOTAL LIMIT=100

/TITLE='Case Summaries'

/MISSING=VARIABLE

/CELLS=COUNT .

Compare this with their actual Malaise score at age 34 (Y). How accurate was the predicted \hat{Y} ? (Hint: to find the individual cohort member in the data file, go to the *Data View* window and place the cursor on the *Key* column, select *Find* from the *Edit* drop down menu. The Key number to find is 170. Look under b34mal to see the actual score obtained at age 34.

	key	bsex	b16mal	b26mal	b26malg	b30mal	b30malg	b34mal	b34malg	
2	20	female	not stated	not stated	not stated	not stated	not stated	not stated	not stated	
3	30	male	6	10.00	score of 8+	4	0-7	.00	0-3	
4	40	female	not stated	not stated	not stated	not stated	not stated	not stated	not stated	
5	50	male	6	1.00	score of 0-	1	0-7	.00	0-3	
6					not stated	not stated	not stated	not stated	not stated	
7					not stated	not stated	4	0-7	1.00	0-3
8					not stated	not stated	3	0-7	2.00	0-3
9					not stated	not stated	0	0-7	not stated	not stated
10					8.00	score of 8+	9	8+	4.00	4+
11					not stated	not stated	not stated	not stated	not stated	not stated
12	120	male	not stated	2.00	score of 0-	1	0-7	.00	0-3	
13	130	male	not stated	.00	score of 0-	0	0-7	.00	0-3	
14	140	male	5	2.00	score of 0-	3	0-7	.00	0-3	
15	150	female	14	7.00	score of 0-	10	8+	1.00	0-3	
16	160	female	11	5.00	score of 0-	9	8+	2.00	0-3	
17	170	female	44	not stated	not stated	11	8+	5.00	4+	
18	180	male	3	.00	score of 0-	not stated	not stated	.00	0-3	

You could repeat the above analysis using different but similar variables in the data file. For example, replace highest qualification at age 34 with the age the cohort member left full-time education (b34lefted), mothers age of leaving full-time education with fathers age of leaving full-time education (b0daded) or family social class at some point in childhood (b0psoc, b5psoc, etc).

Logistic Regression

We can run a similar analysis using logistic regression to predict depression at age 34 from signs of depression at age 16, birth weight, gender, mother's age of leaving full-time education and own highest qualification by age 34.

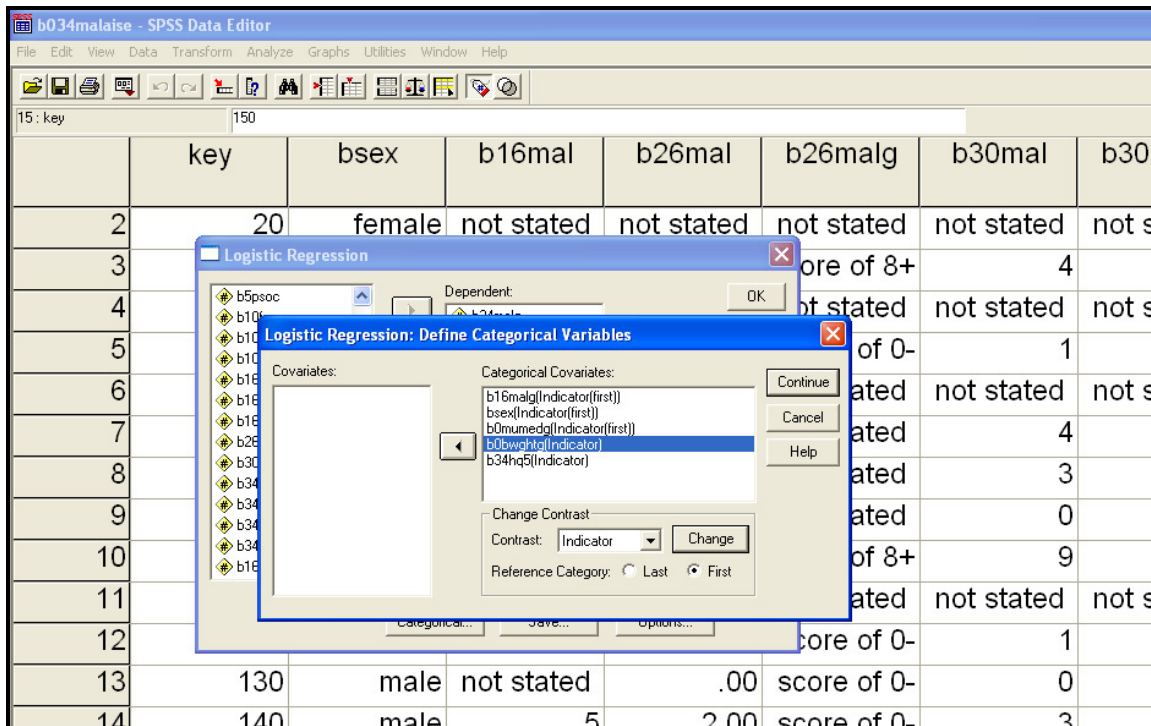
From the *Analyse* drop down menu select *Regression* and then *Binary Logistic* to open the Logistic Regression dialog box. Transfer the categorical Malaise score variable *b34malg* into the *Dependent* box, and the (predictor) variable into the *Covariates* box. In the first analysis just move the grouped malaise score at age 16 (*b16malg*). In the second stage of the analysis include the other measures from birth using the categorical and not continuous measures for age mother left full-time education (*b0mumedg*), birth weight (*b0bwghtg*), gender (*bsex*) and highest qualification at age 34 (*b34hq5*).

Reference Category

Categorical variables require a reference category to be set. There are different methods of contrasting category membership in a variable. The default method in SPSS is *Indicator*, with the last (highest value) category being set as the reference category. (In the SPSS output, the reference category is represented in the contrast matrix as a row of zeros.)

In our model, this would mean that the reference category in the variable *b34hq5* would be set as category 5 'higher degree / PGCE'. The relative importance of membership to each of the other qualification groups (values 0 'none' up to 4 'degree....nursing qual') will be compared against membership to category 5 'higher degree / PGCE'.

Click on the *Categorical* button to open the *Define Categorical Variables* box. Move the covariates into the *Categorical Covariates* box. The default makes the last (highest value) category of a variable the reference category. To change the reference category to the first (lowest value) category, click on *First* and then *Change*. Do this for each of the following variables *b16malg*, *bsex*, *b0bwghtg*, *b0meumedg*. Click on *Continue* and then *OK*.



If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

Initial analyses

```
LOGISTIC REGRESSION b34malg
/METHOD = ENTER b16malg
/CONTRAST (b16malg)=Indicator(1)
/CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
```

Second analyses

```
LOGISTIC REGRESSION b34malg
/METHOD = ENTER b16malg bsex b0mumedg b0bwghtg b34hq5
/CONTRAST (b16malg)=Indicator(1) /CONTRAST (bsex)=Indicator(1)
/CONTRAST (b0mumedg)=Indicator(1) /CONTRAST
(b0bwghtg)=Indicator(1) /CONTRAST (b34hq5)=Indicator
/CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
```

Looking at the Variables in the Equation table

- How much more likely were depressed teenagers to grow up to be a depressed 34 year old?
- After taking into account gender, birth weight, age mother left education and own highest qualification, how much more likely were depressed teenagers to grow up to be depressed at age 34?

- After malaise score at age 16, which variable makes the biggest contribution to the model?

Further stages of analysis

Repeat the above analysis adding in new variables or replacing variables with similar measures. For example, fathers age of leaving full-time education (b0dadgedg). Remember to think about whether it makes more sense to set the first or last category as the reference category. If a variable has many categories (e.g. family social class), recode to have fewer categories.

- Is mothers or fathers age of leaving full-time education the stronger measure (bigger coefficient) for predicting depression at age 34?
- Which variable makes a significant contribution to the model when mothers age of leaving full-time education is included, but not when fathers age of leaving full-time education is included?

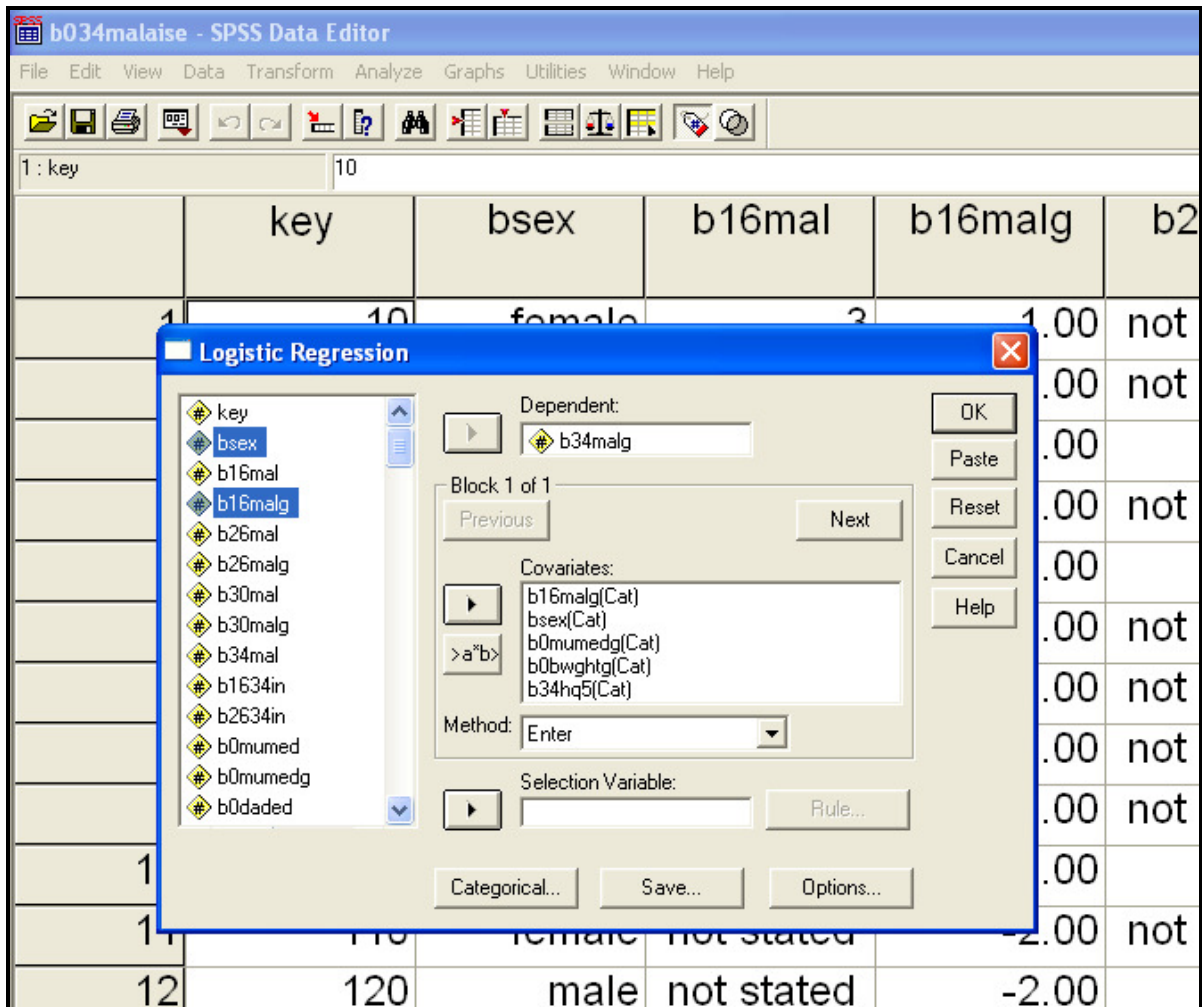
If you have the inclination, you could also try repeating the above analyses for men and women separately.

- Are the results the same for men and women separately?

Interaction Terms

For each variable in the logistic model, it is assumed that the effect is the same for all values of other variables in the model. For example, the effect of level of a mother's education is the same for men as it is for women. If this is not the case, there is an interaction. You can include a term for the *level of mother's education* (b0mumedg) by gender (bsex) interaction in your model.

To include an interaction term between depression at age 16 and gender in your model click on bsex and then b16malg whilst simultaneously pressing the *Ctrl* key on your PC keyboard. This will mean both variables are highlighted. Click on the *>a*b>* button to place an interaction term in the Covariates dialog box. Continue as before.



If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
LOGISTIC REGRESSION b34malg
/METHOD = ENTER b16malg bsex b0mumedg b0bwghtg b34hq5
b16malg*bsex
/CONTRAST (b16malg)=Indicator(1) /CONTRAST (bsex)=Indicator(1)
/CONTRAST (b0mumedg)=Indicator(1) /CONTRAST
(b0bwghtg)=Indicator(1) /CONTRAST (b34hq5)=Indicator
/CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
```

- Is the interaction significant? Try repeating this using the continuous malaise score at age 16 variables (b16mal). Are the same results obtained?

Remember all of these exercises can be repeated for cross-cohort comparisons using similar variables in the NCDS data file *n042malaise.sav*.