



Teaching students quantitative methods using resources from the British Birth Cohorts

Measuring behaviour difficulties of children and teenagers

RutterBehaviourExercises.doc: The SPSS exercises detailed here use data from the 1958 National Child Development Study (NCDS). Students could replicate these exercises using data from the 1970 British Cohort Study (BCS70) for cross-cohort comparisons. After a short summary of the two data files, *b016rutter.sav* and *n016rutter.sav* the document provides a few step-by-step exercises for students to work through. Additional exercises to complete are suggested. The exercises use the drop down menus but the SPSS syntax to run the commands also included. The **MalaiseExercises.doc** contains a complete set of step-by-step exercises ranging from introductory descriptive statistics to multiple and logistic regression. By replacing the Malaise score variables with the Rutter behaviour score variables these exercises could be easily adapted for students using either the *b016rutter.sav* or *n016rutter.sav* data files. The techniques covered in this document are

- 1) Frequencies
- 2) Cross-tabulations
- 3) T-tests
- 4) Factor analyses

The exercises in **MalaiseExercises.doc** include

- 1) Frequencies
- 2) Means
- 3) Cross-tabulations
- 4) Constructing summary scores
- 5) Recoding continuous into dichotomous variables

- 6) Correlations
- 7) Linear regression
- 8) Logistic regression.

SPSS Data Files: NCDS data 'n016rutter.sav', BCS70 data 'b016rutter.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 is provided by the cohort members' mother, data collected at age 5, age 10 and age 16 from the mother (or father) and the cohort member (child). Data from age 26, 30 and 34 is 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

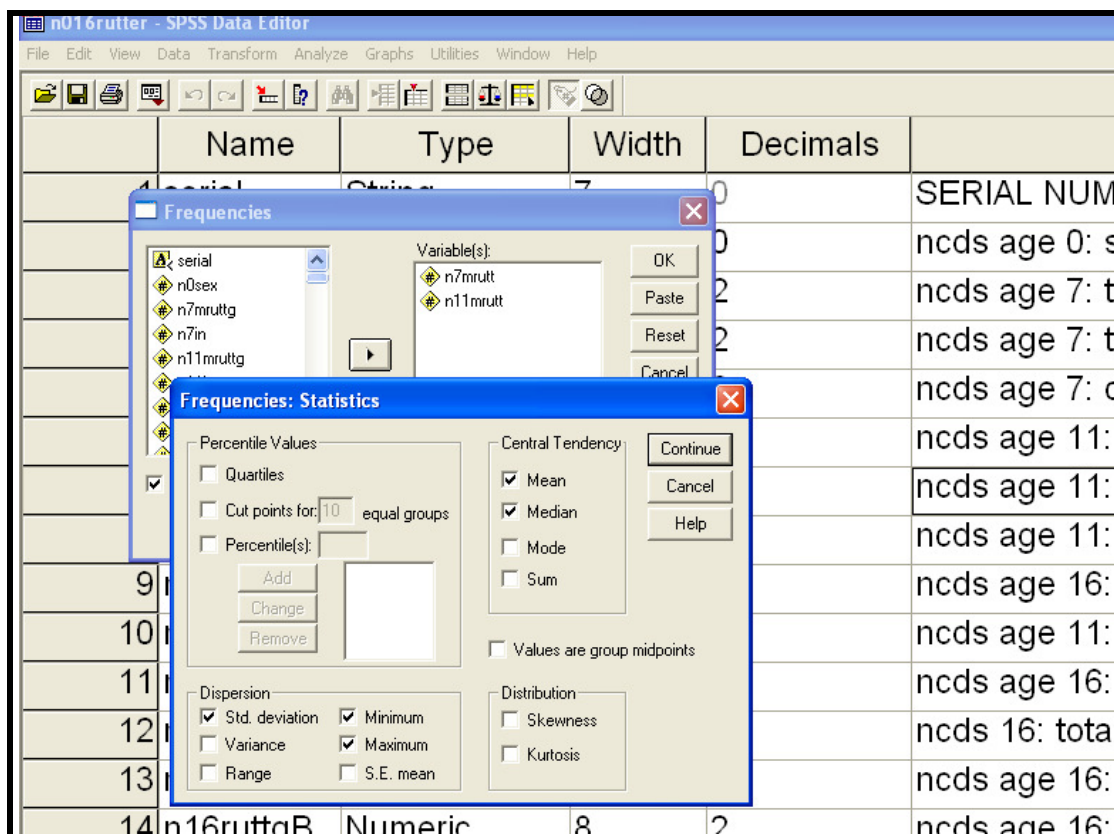
This longitudinal dataset includes information from the 1958 National Child Development Study (NCDS) for looking at the relationship between socio-economic circumstances in childhood and mental well-being in adulthood over time. Information included in this file from birth is provided by the cohort members' mother. Information collected at age 7, age 11 and age 16 from the mother (or father). 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.

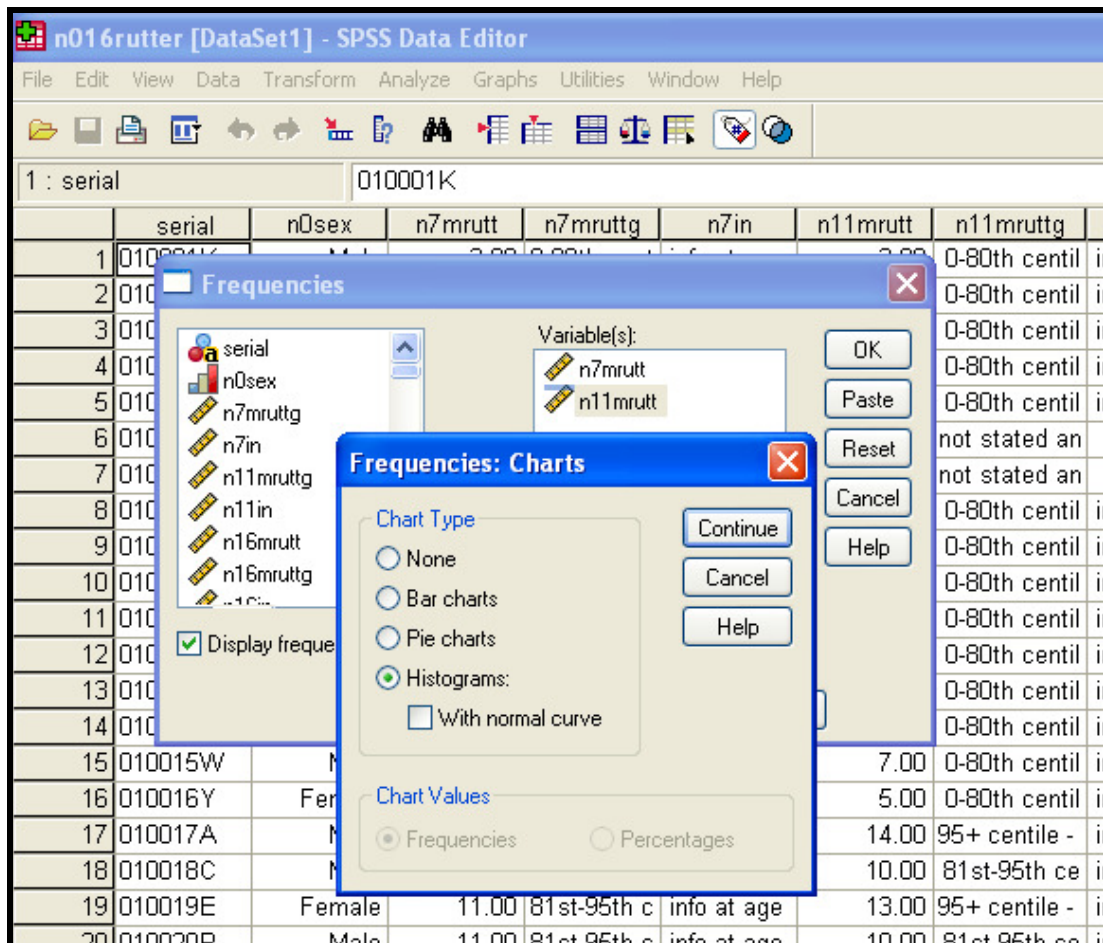
Step-by-Step Exercises using SPSS

Exploring the data

1. Look at the distribution of the Rutter behaviour scores at age 7 and age 11. How many cohort members have a valid score at each age? Were the average (mean) scores the same or were they higher at age 7 or at age 11? At what age was the highest individual score recorded? (Hint: frequency of n7mrutt n11mrutt variables and select additional summary statistics and charts).

From the *Analyze* drop-down menu select *Descriptive Statistics* and then *Frequencies*. Select n7mrutt and n11mrutt and click on the ► button to move the two variables into the Variable(s) box. Click on the Statistics button at the bottom of the Frequencies dialogue box. A new dialogue box Frequencies: Statistics will appear. Under Central Tendency select Mean and Median. Under Dispersion, select Std deviation, Minimum and Maximum, to explore the spread and range of the values in the behaviour score scale. Click on Continue. Click on the Charts button and select Histogram. Click on Continue, then click on OK.





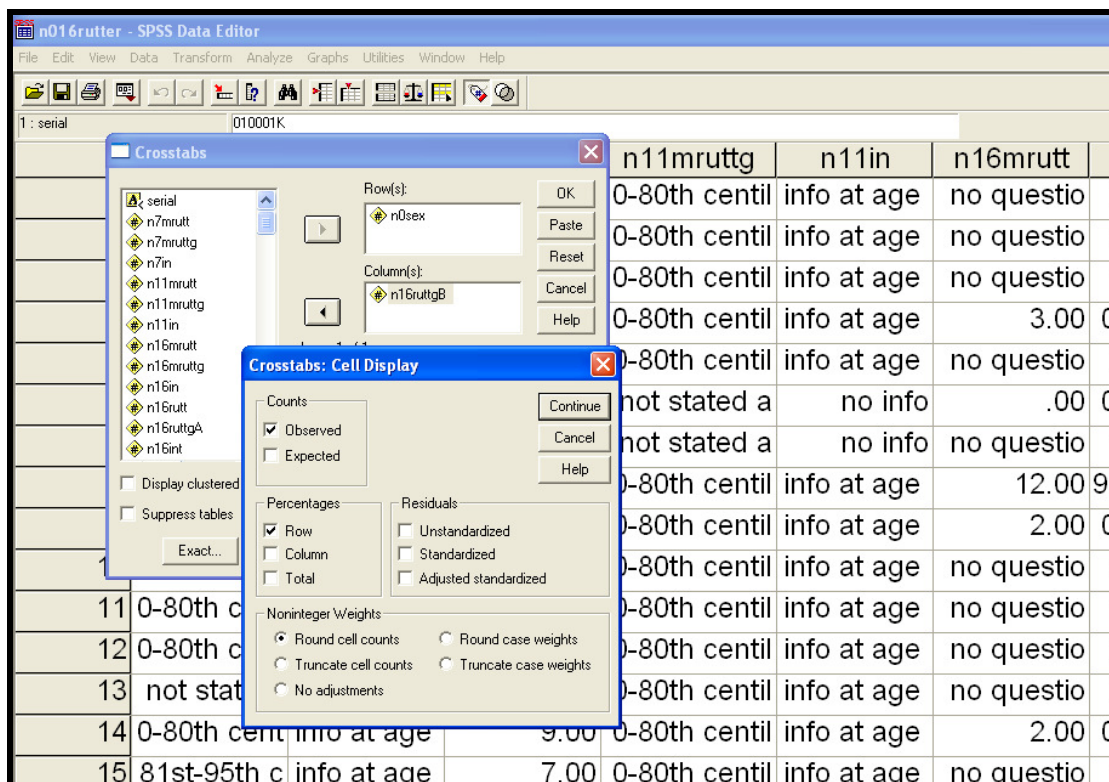
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
FREQUENCIES
  VARIABLES=n7mrutt n11mrutt
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
  /BARCHART FREQ
  /ORDER= ANALYSIS .
```

2. Were boys or girls more likely to show signs of behaviour problems at age 16 according to their teachers? Is the difference statistically significant? (Hint: cross-tab of n16ruttgB and n0sex).

From the *Analyse* drop-down menu select *Descriptive Statistics* and then *Crosstabs*. Select n0sex and click the ► button to place the variable into the Row(s) box. Click on n16ruttgB 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. To see if there is a significant difference between the percentage of boys and girls who were thought to show more signs of behaviour problems at age 16, Click on the 'Statistics' button and select Chi-square at the top of the dialogue box. Click on Continue, then click on OK.

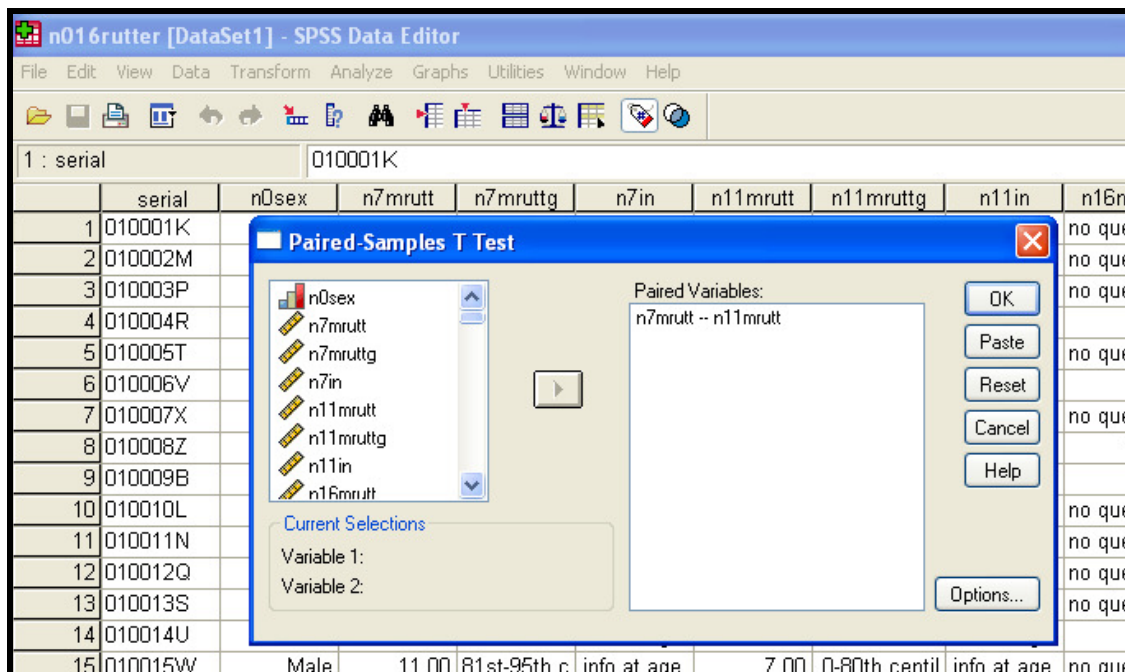
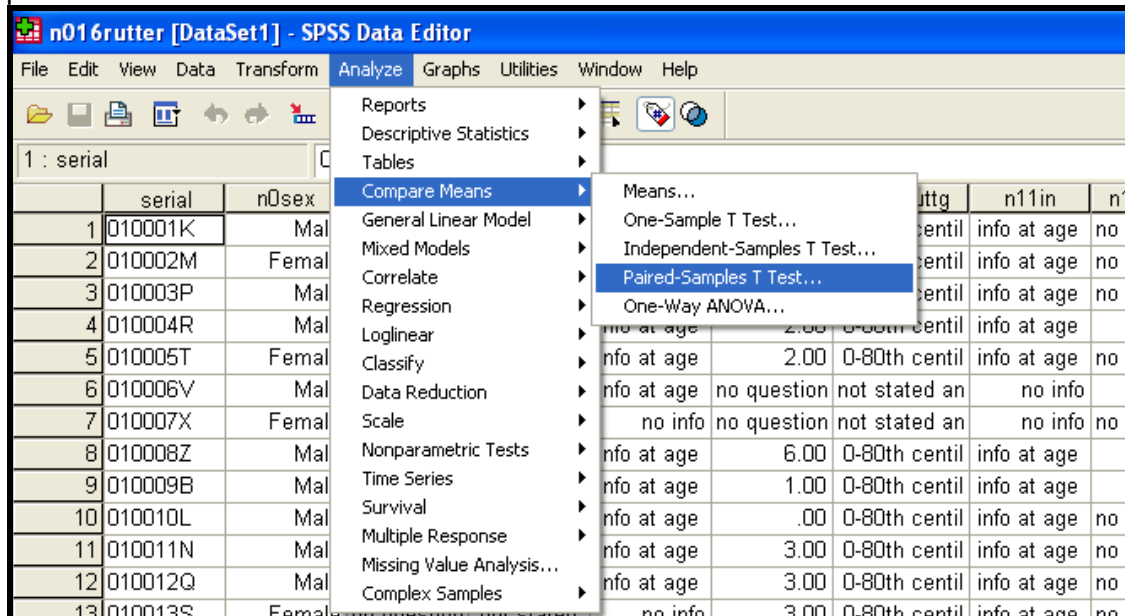


If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ► button on the toolbar to run the command.

```
CROSSTABS
/TABLES=n0sex BY n16ruttgB
/FORMAT= AVALUE TABLES
/STATISTIC=CHISQ
/CELLS= COUNT ROW
/COUNT ROUND CELL.
```

3. What about earlier on in life? Were children more likely to show signs of behaviour problems at age 7 or age 11 according to their mother? Is the difference statistically significant? (Hint: paired samples T-test with variables n7mrutt and n11mrutt).

From the *Analyze* drop-down menu select *Compare Means* and then *Paired-Samples T Test*. Select n7mrutt and then n11mrutt and click the ► button to place the variables into the Paired Variables: 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.

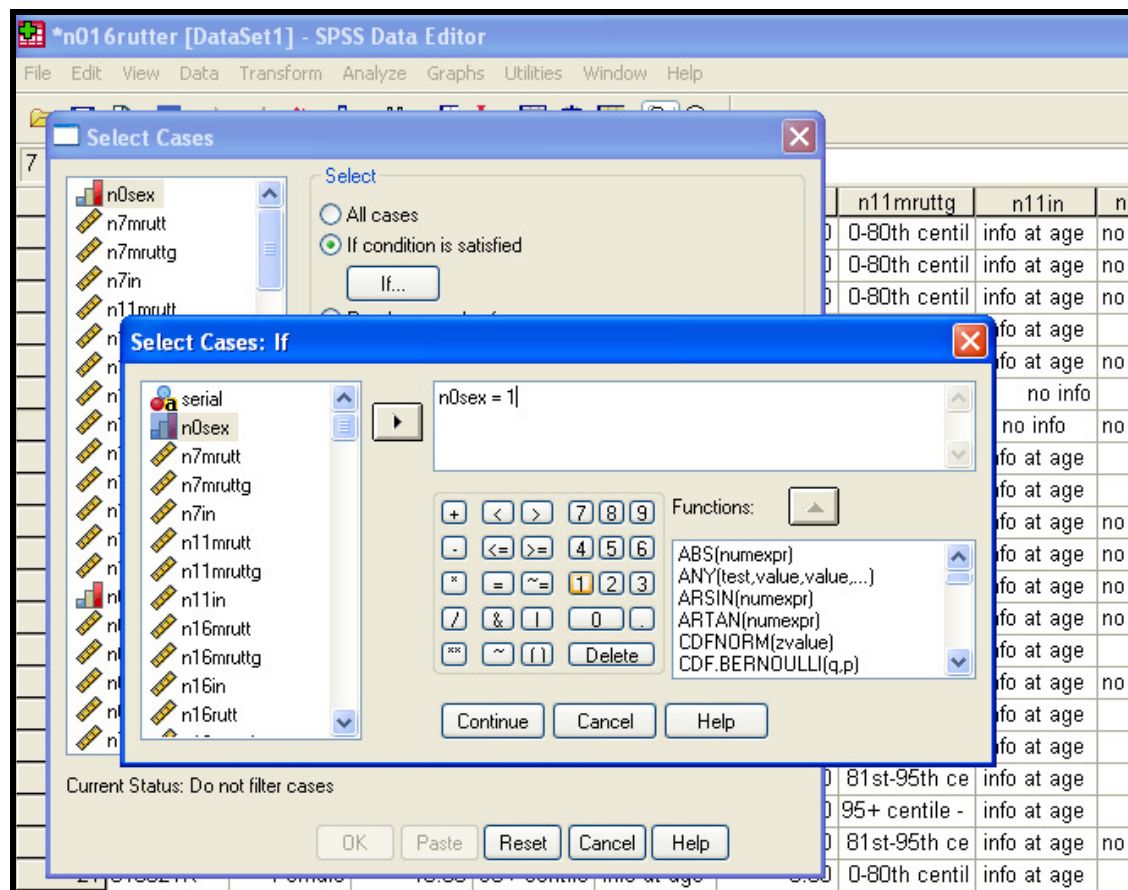
T-TEST

PAIRS = n7mrutt WITH n11mrutt (PAIRED)
/CRITERIA = CI(.95)
/MISSING = ANALYSIS.

You could repeat this analyses for boys and girls separately. Is the result the same?

How to only select boys for the analyses.

From the *Data* drop-down menu select *Select Cases*. In the new window select the *If condition is satisfied* button and click on the *If...* Button. In the new window select n0sex from the list of variables and click on the ► button to move the variable into the empty box. To select boys only, click on = and then 1. These will automatically appear in the box next to n0sex. Click on Continue, then OK.



You can check who has been included in your sample by running a quick frequency of n0sex or look at the Data View window. An angled line through the case numbers of all female cohort members indicates that girls have not been selected for this analyses.

	serial	n0sex	n7mrutt	n7mruttg	n7in	n11mrutt	n11mruttg	n11in
1	010001K	Male	3.00	0-80th cent	info at age	3.00	0-80th centil	info at age
2	010002M	Female	some info	incomplete	info at age	3.00	0-80th centil	info at age
3	010003P	Male	5.00	0-80th cent	info at age	8.00	0-80th centil	info at age
4	010004R	Male	5.00	0-80th cent	info at age	2.00	0-80th centil	info at age
5	010005T	Female	2.00	0-80th cent	info at age	2.00	0-80th centil	info at age
6	010006V	Male	some info	incomplete	info at age	no question	not stated an	no info
7	010007X	Female	no question	not stated	no info	no question	not stated an	no info
8	010008Z	Male	12.00	81st-95th c	info at age	6.00	0-80th centil	info at age
9	010009B	Male	3.00	0-80th cent	info at age	1.00	0-80th centil	info at age
10	010010L	Male	3.00	0-80th cent	info at age	.00	0-80th centil	info at age
11	010011N	Male	4.00	0-80th cent	info at age	3.00	0-80th centil	info at age
12	010012Q	Male	3.00	0-80th cent	info at age	3.00	0-80th centil	info at age
13	010013S	Female	no question	not stated	no info	3.00	0-80th centil	info at age

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_$=(n0sex = 1 ).
VARIABLE LABEL filter_$ 'n0sex = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

Another way of selecting a subset of the datafile, in this case boys or girls, is to use the *Temporary* command. By adding the two lines of syntax to the start of the T-Test command, SPSS will only select boys. To only select girls, replace 'n0sex = 1' with 'n0sex = 2'.

Temporary.

Select if (n0sex = 1).

T-TEST

PAIRS = n7mrutt WITH n11mrutt (PAIRED)

/CRITERIA = CI(.95)

/MISSING = ANALYSIS.

Differentiating between different types of behaviour

Factor Analyses

The individual Rutter behaviour questions cover a broad range of 'behaviours', from fighting, being restless, worries, being disobedient, etc. In the overall summary score variables these behaviours are all summed together. However, 'neurotic' and 'antisocial' behaviour sub-scales can be constructed when all 31 questions in all three components of the Rutter scale are administered (Rutter et al, 1970, p412-413)¹. Using BCS70 data at age 5 and age 10, Thompson et al (2003)² used *Factor Analyses* and identified three subscales from the majority of the questions included in these sweeps. These were conduct problems, hyperactivity and emotional problems. Further details on the overall Rutter scale are provided in **RutterBehaviourQuestions.doc**.

4. Using all 14 individual items that made up the overall Rutter behaviour score at age 7, how many subscales of behaviour can be identified? (Hint: use Factor Analysis)

What is Factor analysis?

Factor analysis is used to reduce a (relatively) large number of correlated variables to a more manageable number of independent factors – or components – that can be used in subsequent analyses.

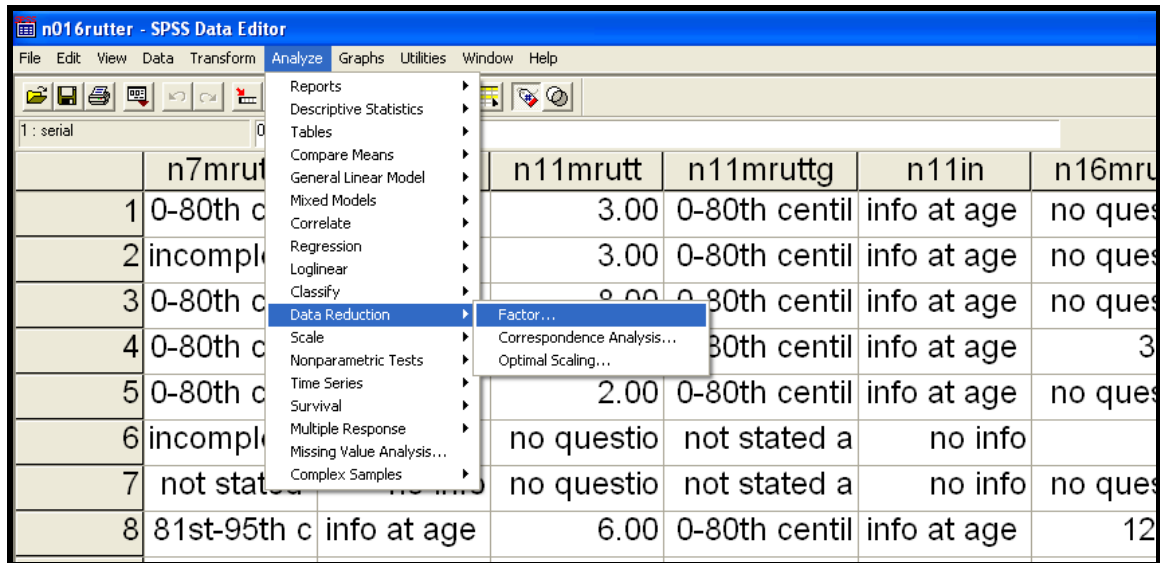
The simplest way of extracting a factor is by Principle Components Analyses (PCA). In PCA, linear combinations of observed variables are formed. The first PCA is the combination that accounts for the largest amount of variance in the sample, the second the next largest, etc. The total variance that can be explained is equal to the number of variables in the analyses. In our example, this is 14.

Factor scores can be used in subsequent analyses. Using our example, we could test the null hypothesis that men and women have the same average antisocial behaviour factor score.

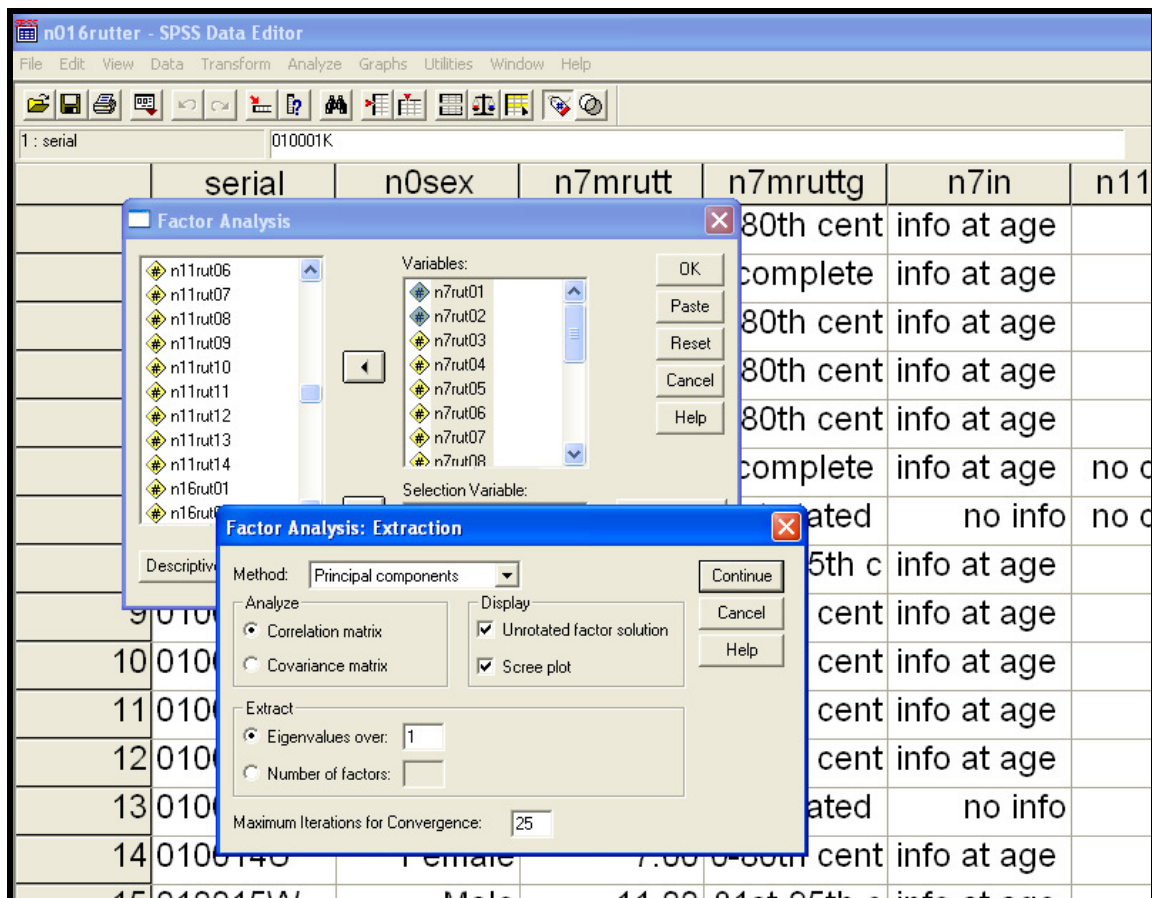
From the *Analyse* drop-down menu select *Data Reduction* and then *Factor*. Select variables n7mrut01 to n7mrut14 and click the ► button to place them into the Variables box.

¹ RUTTER, M., TIZARD, J., & WHITMORE, K. (1970). Education, health and behaviour. London: Longmans.

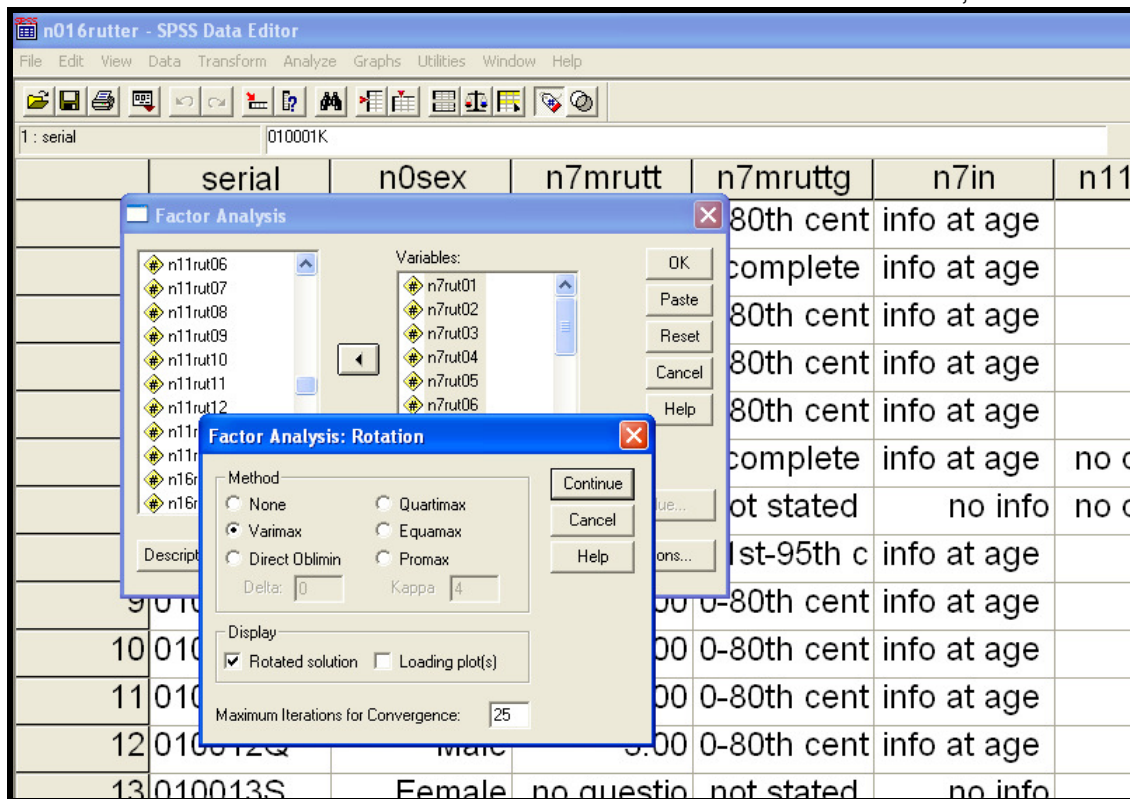
² THOMPSON, A., HOLLIS, C. & RICHARDS, D. (2003). Authoritarian parenting attitudes as a risk for conduct problems – results from a British national cohort study. *European Child and Adolescent Psychiatry*, 12(2), 84-91.



Open the *Extraction* dialogue box. The SPSS default selects extraction by *Principle Components Method*, factors with *Eigenvalues* greater than 1 and displays the *Unrotated Factor Solution*. Also select to display the *Scree plot*.



Then click on the *Rotation* button and select *Varimax* method of rotation. The *Rotated solution* will automatically be displayed. Click on Continue. Now click on *Options* and select *Sorted by Size* under *Coefficient Display Format*. Finally, click on the *Scores* button and select *Save as variables*. 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.

FACTOR

/VARIABLES n7rut01 n7rut02 n7rut03 n7rut04 n7rut05 n7rut06 n7rut07 n7rut08
n7rut09 n7rut10 n7rut11 n7rut12 n7rut13 n7rut14

/MISSING LISTWISE /ANALYSIS n7rut01 n7rut02 n7rut03 n7rut04 n7rut05
n7rut06 n7rut07 n7rut08 n7rut09 n7rut10 n7rut11 n7rut12
n7rut13 n7rut14

/PRINT INITIAL CORRELATION EXTRACTION ROTATION

/FORMAT SORT

/PLOT EIGEN

/CRITERIA MINEIGEN(1) ITERATE(25)

/EXTRACTION PC

/CRITERIA ITERATE(25)

/ROTATION VARIMAX

/SAVE REG(ALL)

/METHOD=CORRELATION .

The SPSS output (Total Variance Explained or Scree Plot) shows that three factors have been extracted. Look at the Data View window and you will see that the three factor scores have been saved as three variables.

	n16rutt24	n16rutt25	n16rutt26	FAC1_1	FAC2_1	FAC3_1
1	does not a	applies so	does not a	-.68316	-.70795	-1.03687
2	does not a	does not a	does not a	.	.	.
3	info missin	info missin	info missin	-.12407	.14051	-.90858
4	does not a	does not a	does not a	-.10765	-.86209	-.33204
5	.	.	.	-.46008	-.77518	-.37353
6	does not a	applies so	applies so	.	.	.
7	.	.	.	.	.	.
8	info missin	info missin	info missin	2.90623	-.48692	-.54117
9	does not a	does not a	does not a	-1.64624	.56439	-.36003
10	.	.	.	-.30202	-.91182	-.86312
11	.	.	.	-.00853	-.67909	-.27779
12	.	.	.	-.57151	-.41590	.00613
13	does not a	applies so	does not a	.	.	.
14	does not a	does not a	does not a	.03750	.64011	-.31769
15	.	.	.	.86118	1.46303	-1.10299
16	.	.	.	.76433	1.27756	1.89875

5. Use these factor scores to see if men or women have higher antisocial behaviour scores? (HINT: independent samples t-test of Fac1_1 by n0sex)

From the *Analyze* drop-down menu select *Compare Means* and then *Independent Samples T-test*. Select Fac1_1 and click the ► button to place the variable into the *Test Variable(s)* box. Click on n0sex and click the ► button to place into *Grouping Variable* box. Click on *Define groups*. This instructs SPSS which groups are being compared. Place a '1' in group 1 and '2' in group 2 as these are the values in then0sex variable (1=boys, 2=girls). Click on Continue, 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.

T-TEST

```
GROUPS = n0sex(1 2)
/MISSING = ANALYSIS
/VARIABLES = Fac1_1
/CRITERIA = CI(.95) .
```

6. Do children who live in overcrowded conditions show more signs of antisocial behaviour than children who do not experience overcrowding? (HINT: independent samples t-test of Fac1_1 by n7crowdg.) Is it the same for both boys and girls?

A simple way to ask SPSS to select only boys or girls is to use the temporary command. By adding the two lines of syntax to the start of the T-Test command, SPSS will only select boys. To only select girls, replace 'n0sex = 1' with 'n0sex = 2'.

Temporary.

Select if (n0sex = 1).

T-TEST

```
GROUPS = n7crowdg(0 1)
/MISSING = ANALYSIS
/VARIABLES = Fac1_1
/CRITERIA = CI(.95) .
```