Centre for Longitudinal Studies, Institute of Education

Teaching students quantitative methods using resources from the British Birth Cohorts

Assessment of Cognitive Development through Childhood

CognitiveExercises.doc: The SPSS exercises detailed here use data from the 1970 British Cohort Study (BCS70). These exercises can be replicated using data from the 1958 National Child Development Study (NCDS) for cross-cohort comparisons. After a short summary of the three data files, *b516reading.sav, n716reading.sav* and *n716maths.sav* the document provides a number of step-by-step exercises to work through and additional exercises to complete. The exercises use the drop down menus but the SPSS syntax to run the commands is also included. The techniques covered in this document range from introductory descriptive statistics to multiple and logistic regression and include:

1 Descriptive statistics
2 Means
3 Correlations
4 Linear regression including interaction terms

SPSS Data Files: BCS70 data *'b516reading.sav'*, NCDS data '*n716reading.sav'* and '*n716maths.sav*'

BCS70 data

This longitudinal dataset *b516reading.sav* includes information from the 1970 British Cohort Study (BCS70). It can be used to examine cognitive development, or more specifically reading, over childhood. The data file contains 14 variables, based around summary reading scores derived from performance of BCS70 cohort members in assessments at age 5, 10 and 16 (1975, 1980 and 1986). This data file only includes data for children who were included in the survey at their birth and had 'age at test' information at age 5. It does not include data on children who were 'recruited' to the survey later on between age 5 and age 16. There are 12,818 cohort members (individuals) in the dataset, 6,631 (51.7%) are men and 6,187 (48.3%) are women.

NCDS data

The two longitudinal datasets, *n716reading.sav* and *n716maths.sav*, include information from the 1958 National Child Development Study (NCDS) and can be used to examine cognitive development, or more specifically, reading and mathematics, over childhood. Each of the data files contains 14 variables, based around summary reading or arithmetic/maths scores derived from performance of NCDS cohort members in assessments at age 7, 11 and 16 (1965, 1969 and 1974). These data files only include data for children who were included in the survey at their birth and had 'age at test' information at age 7. They do not include data on children who were 'recruited' to the survey later on between age 7 and age 16. There are 14,983 cohort members (individuals) in each dataset, 7,689 (51.3%) are men and 7,294 (48.7%) are women.

Step-by-Step Exercises using SPSS

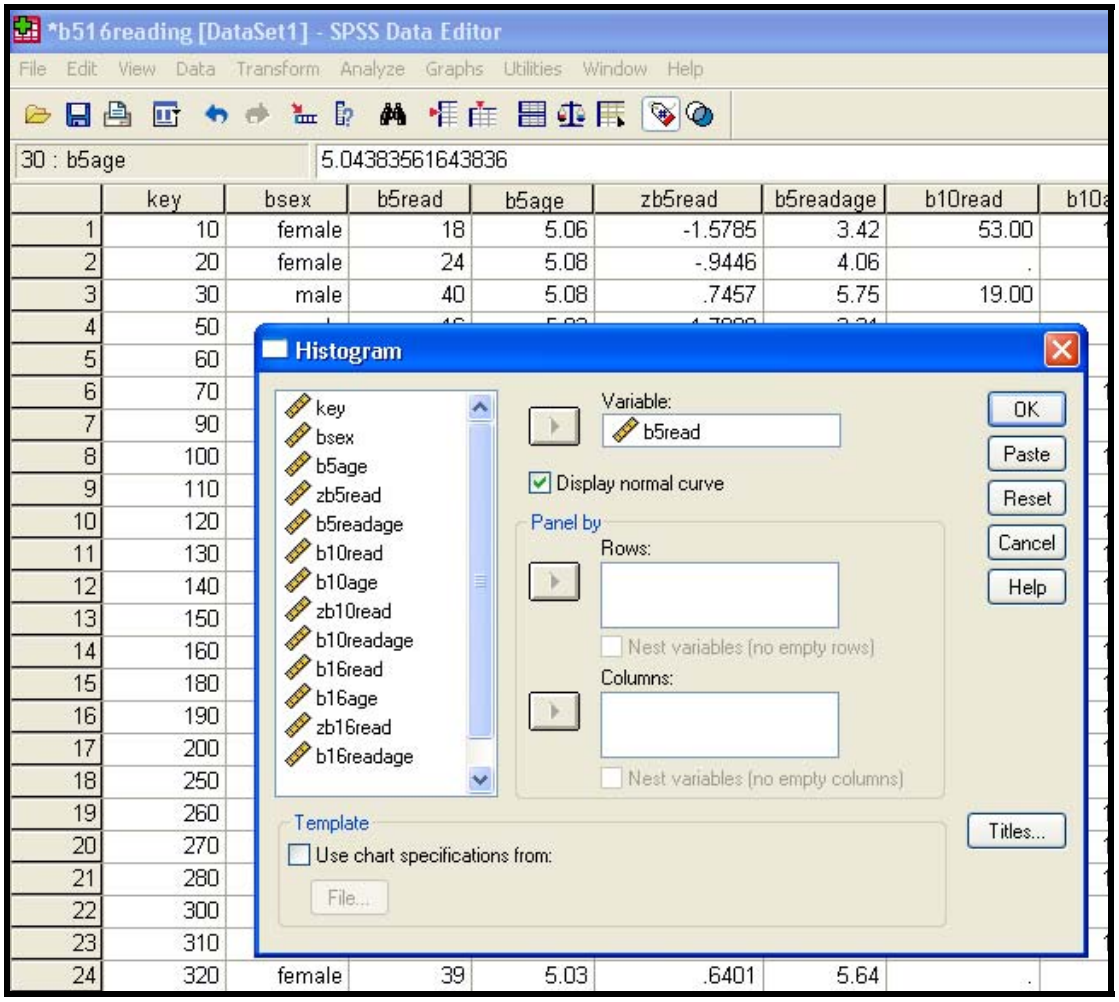
Exploring the b516reading.sav data

a) Descriptive statistics

Comparing the performance scores in the 'reading' tests at age 5, 10 and 16.

i) Produce a histogram for the reading test score at age 5. How would you describe the shape of the distribution? (Hint: use the Graphs menu).

From the Graphs drop-down menu select Histogram. Select b5read and click on the ► button to move the variable into the Variable box. Select to display the 'normal curve' and then click on OK.



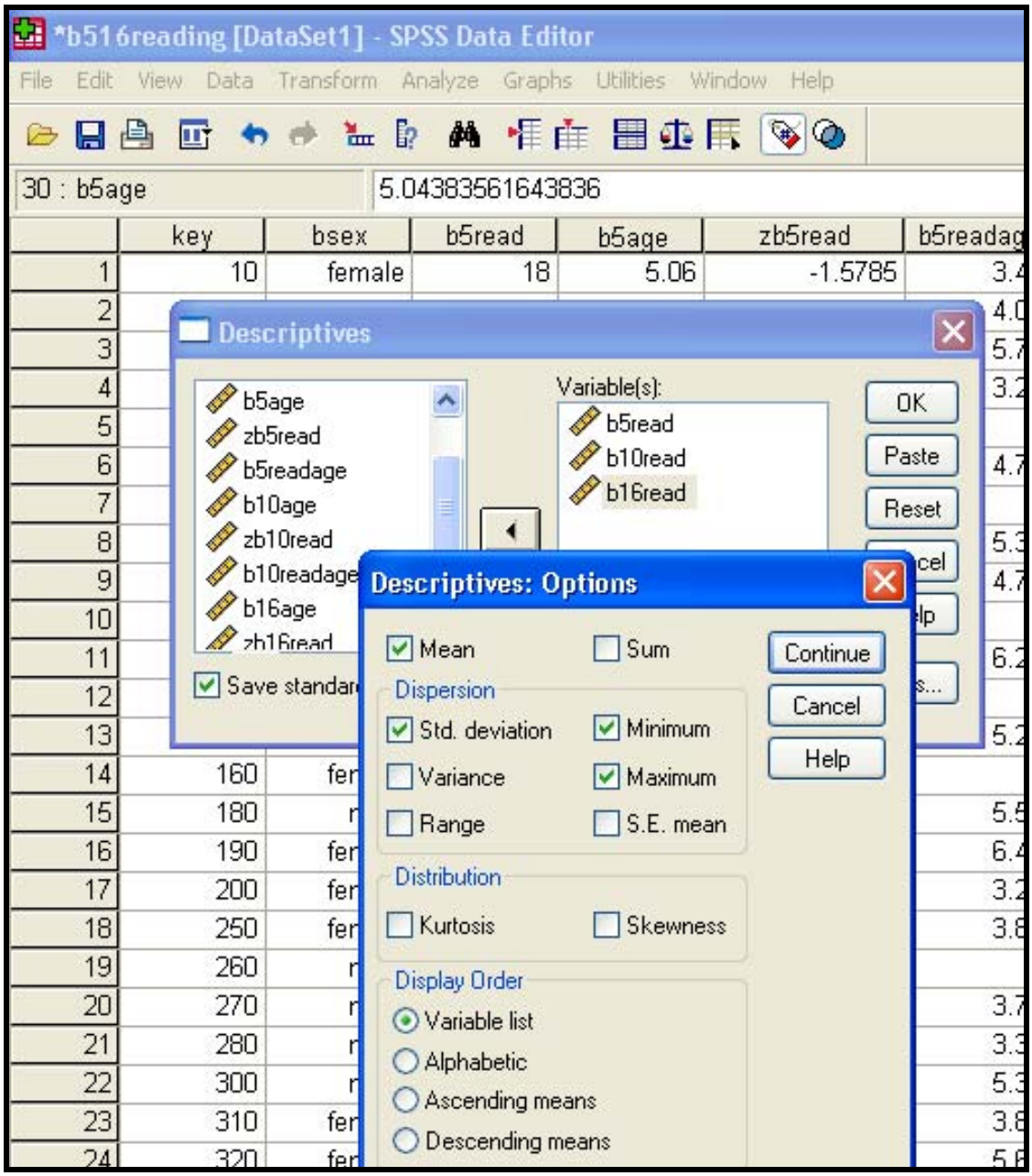
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ▶ button on the toolbar to run the command.

```
GRAPH
  /HISTOGRAM(NORMAL)=b5read .
```

ii) Repeat this for the reading score at age 10 and then at age 16.

iii) At which age did the scores vary the most? (Hint: look at the standard deviations)
iv) Create and save standardised reading scores at age 5, 10 and 16. (Hint: descriptives of b5read b10read b16read variables)

From the Analyse drop-down menu select Descriptive Statistics and then Descriptives. Select b5read b10read b16read and click on the ► button to move the three variables into the Variable(s) box. Select the 'Save standardized values as variables'. Click on Options and select various descriptive measures to further explore the three variables. Click on Continue and then OK.



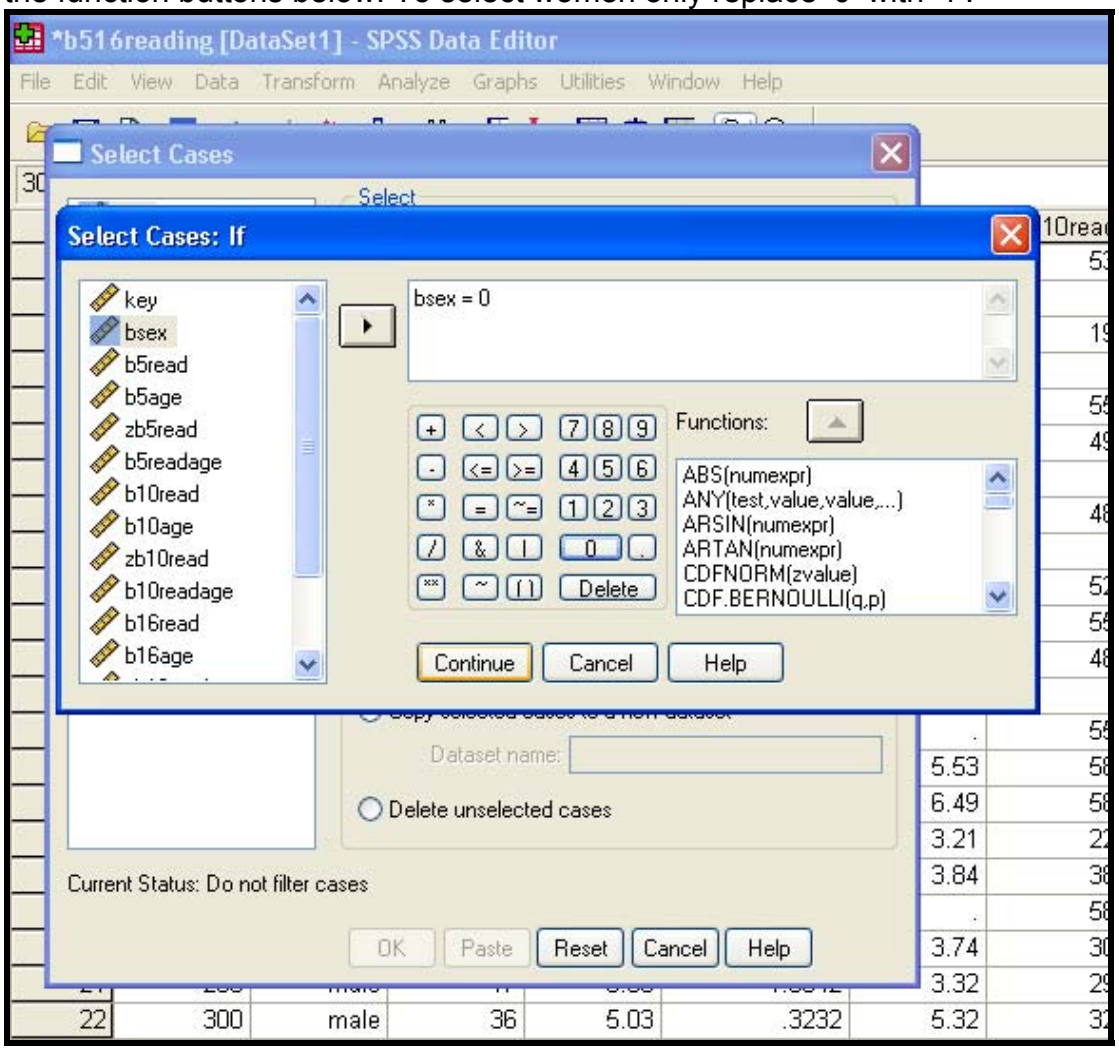
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ▶ button on the toolbar to run the command.

```
DESCRIPTIVES
  VARIABLES=b5read b10read b16read / SAVE
  /STATISTICS=MEAN STDDEV RANGE MIN MAX .
```

Mean scores

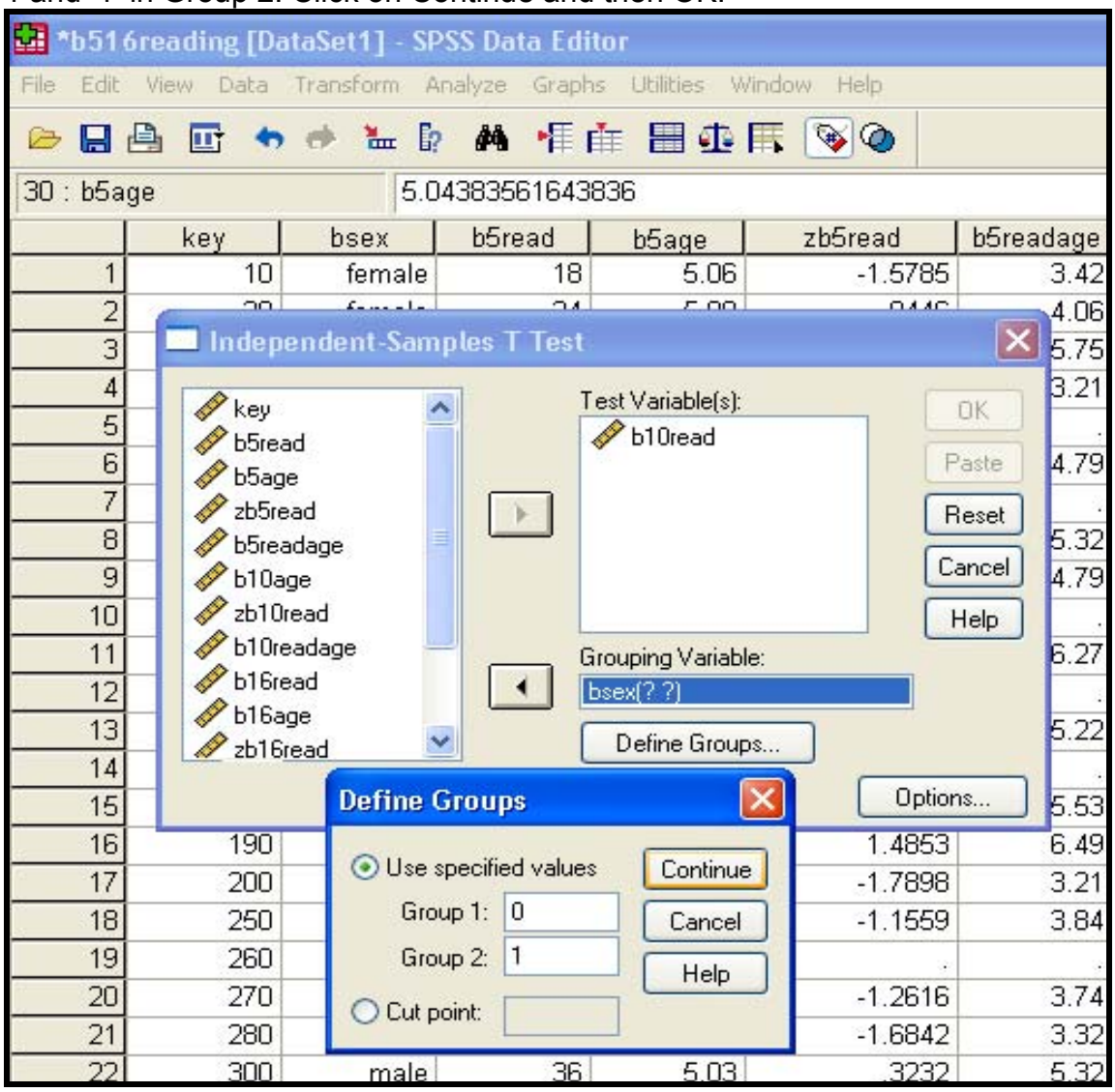
2. Comparing average scores for men and women
   i) Produce a histogram of the reading score at age 10 for men and women separately. (Hint: select men first and then repeat for women).

From the Data drop-down menu choose Select cases. Select If condition is satisfied and click on the If button. Select bsex variable and click on the ▶ button to move the variable across. To select men only choose '=' and '0' from the function buttons below. To select women only replace '0' with '1'.



ii) Was the average (mean) score in the reading test at age 10 higher for men or women? Was the difference significant? (Hint: an independent samples t-test of b10read by bsex).

From the Analyse drop-down menu select Compare Means and then Independent-Samples T Test… Select b10read and click on the ► button to move the variable into the Test Variable(s) box. Similarly, place bsex into the Grouping Variable box. Click on the Define Groups button. Place '0' in Group 1 and '1' in Group 2. Click on Continue and then OK.



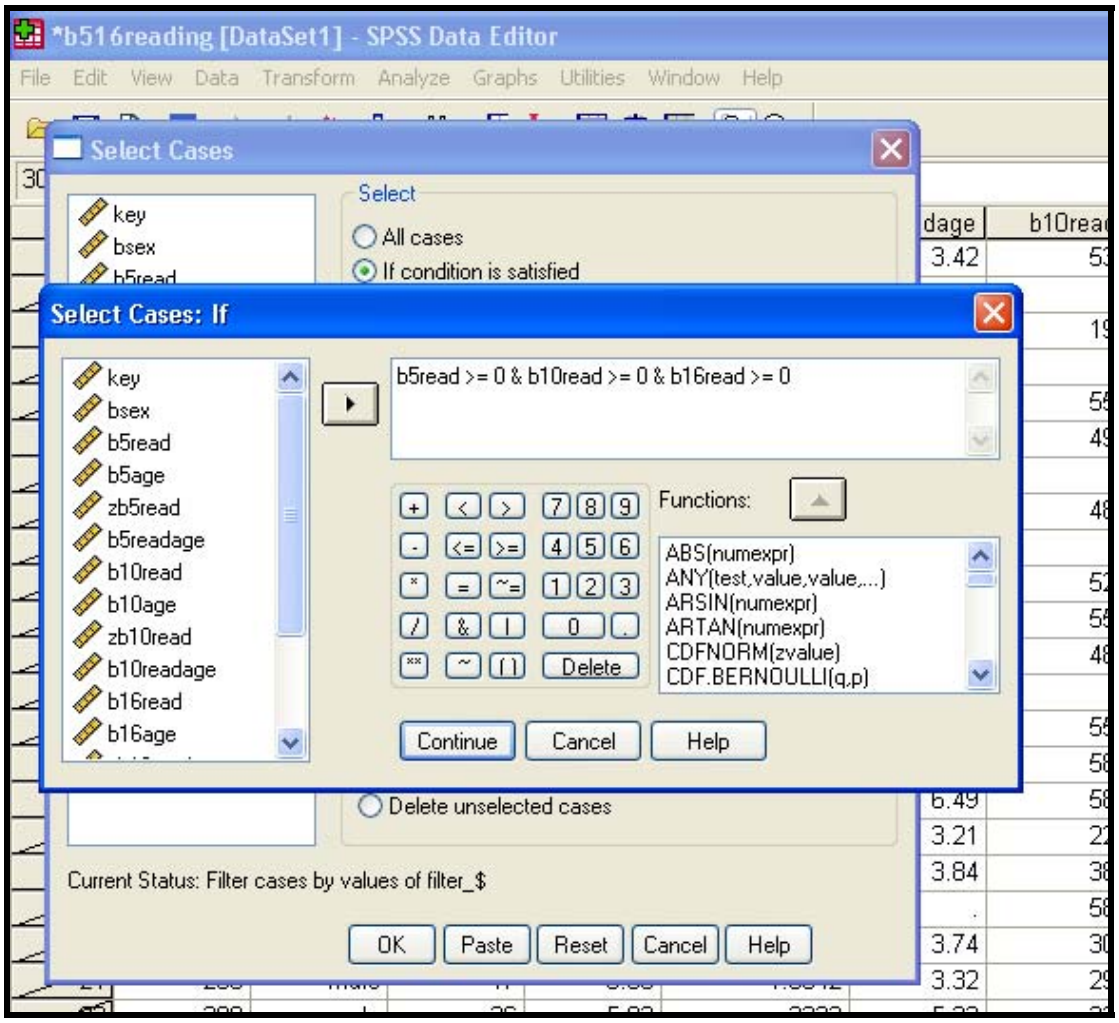
If PASTE is selected instead of OK, a syntax window will open and the following syntax command will appear in it. Highlight the syntax and then click on the ▶ button on the toolbar to run the command.

```
T-TEST
  GROUPS = bsex(0 1)
  /MISSING = ANALYSIS
  /VARIABLES = b10read
  /CRITERIA = CI(.95) .
```

iii) Were the average scores higher or lower for the reduced number of men and women who had a valid score in all three reading tests? (Hint: repeat the t-test after selecting those with a valid score at age 5, 10 and 16).

How to select the 2899 with a valid reading score at age 5, 10 and 16

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 b5read from the list of variables and click on the ▸ button to move the variable into the empty box. To select those with a valid score only, click on >= and then *0*. These will automatically appear in the box next to b5read. Then click on *&* and repeat for b10read and b16read. 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.

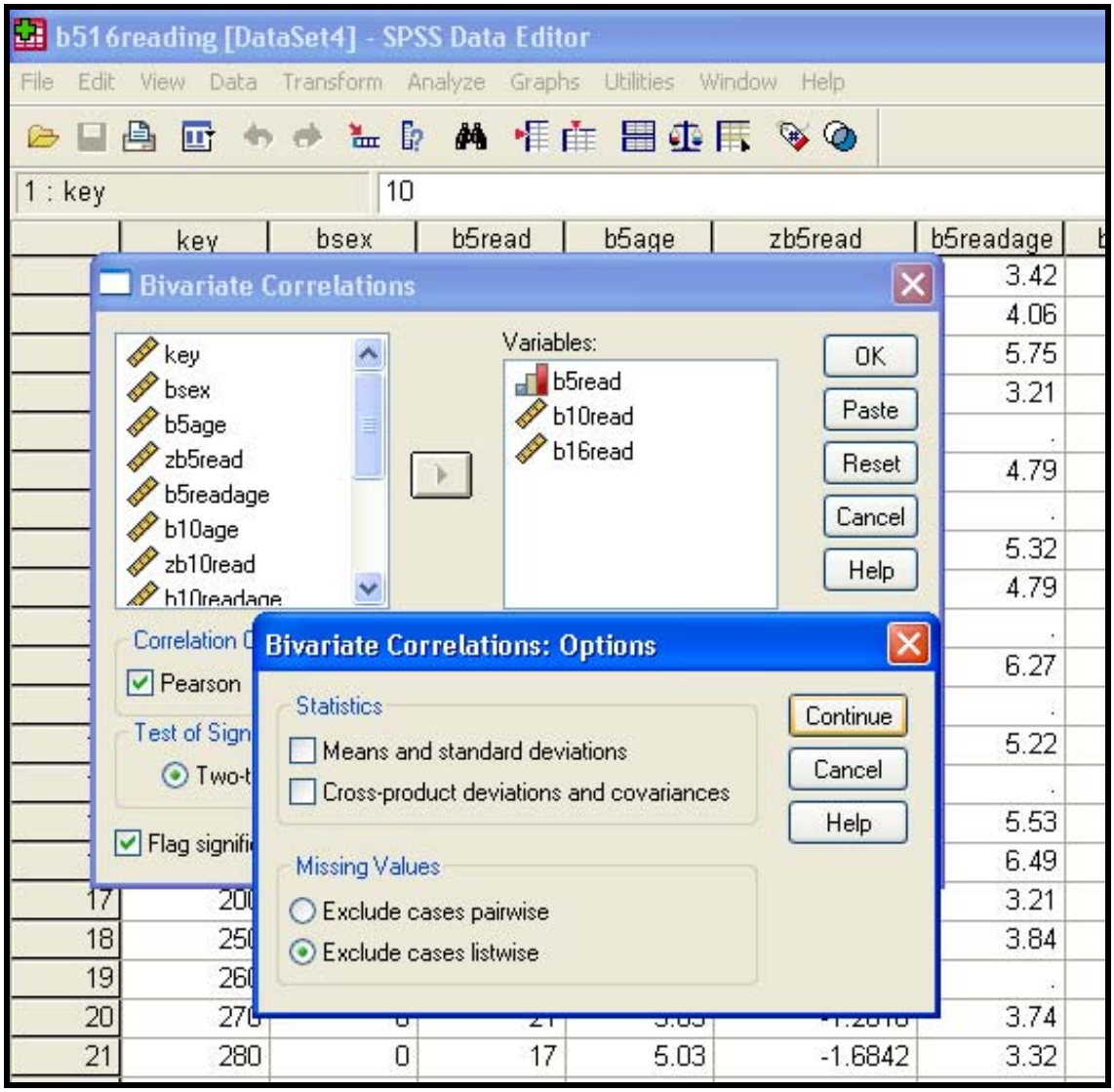
```
USE ALL.
COMPUTE filter_$=(b5read >= 0 & b10read >= 0 & b16read >= 0).
VARIABLE LABEL filter_$ 'b5read >= 0 & b10read >= 0 & b16read >= 0
(FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

Correlations between scores

3. Compare correlations between the three reading scores.
   i) Is the correlation stronger between reading scores at age 5 and age 10, or between reading scores at age 10 and age 16?

   ii) Repeat this exercise for the reduced number of cohort members who have a valid reading score in all three tests. Are the correlations between reading scores now stronger or weaker? What does this suggest? (Hint: correlate the three reading scores b5read b10read b16read. The default position is that cases are excluded *pairwise*. Repeat the exercise and exclude cases *listwise*).

From the *Analyse* drop-down menu select *Correlate* and then *Bivariate*. Select b5read and click the ► button to place it into the Variables box. Repeat for variables b10read b16read. Click OK. After running the analyses, repeat it for the 2899 with valid performance scores at each age. Before clicking on OK, select the options button and under *Missing Values*, select *Exclude cases listwise.*



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=b5read b10read b16read
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .
```

```
CORRELATIONS
  /VARIABLES=b5read b10read b16read
  /PRINT=TWOTAIL NOSIG
  /MISSING=LISTWISE .
```

Linear Regression including interaction terms

4. Before running a linear regression analyses, it is important to get a good feel for the data you will be analysing.

   i) First produce a scatterplot of reading score at age 16 by actual age at time of test. (Hint: select *Simple scatterplot* from the Graphs drop-down menu and select b16read as the Y-axis variable and b16age as the X-

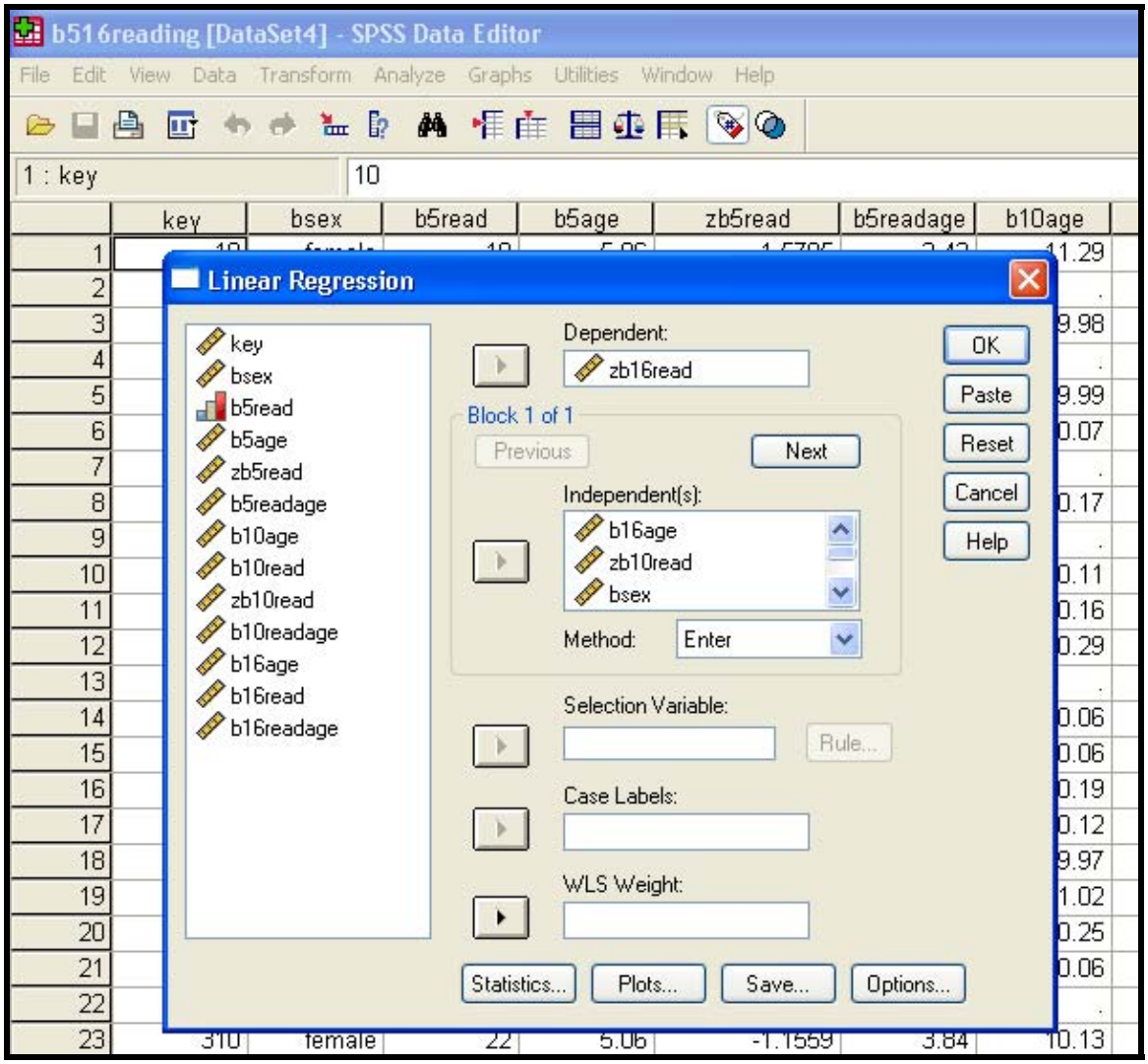
axis variable). How would you define the relationship between age and reading score?

ii) Linear Regression estimates the coefficients by a linear equation, involving one or more explanatory (independent) variables, which best predict the value of the dependent variable, in this example reading score at age 16. There are a number of different stages to this exercise
   a) We first predict reading score at age 16 from the exact age of the child when they sat the test. Do older children have higher scores?
   b) What happens when we add in reading score at age 10? Are children who performed better at age 10 still performing better at age 16?
   c) What about gender? Do boys or girls make better progress between age 10 and 16?
   d) Does the relationship between reading scores at age 10 and age 16 vary by gender, i.e. is there a group covariate interaction?
   e) Does reading score at age 5 add anything to the model? Is gender still significant

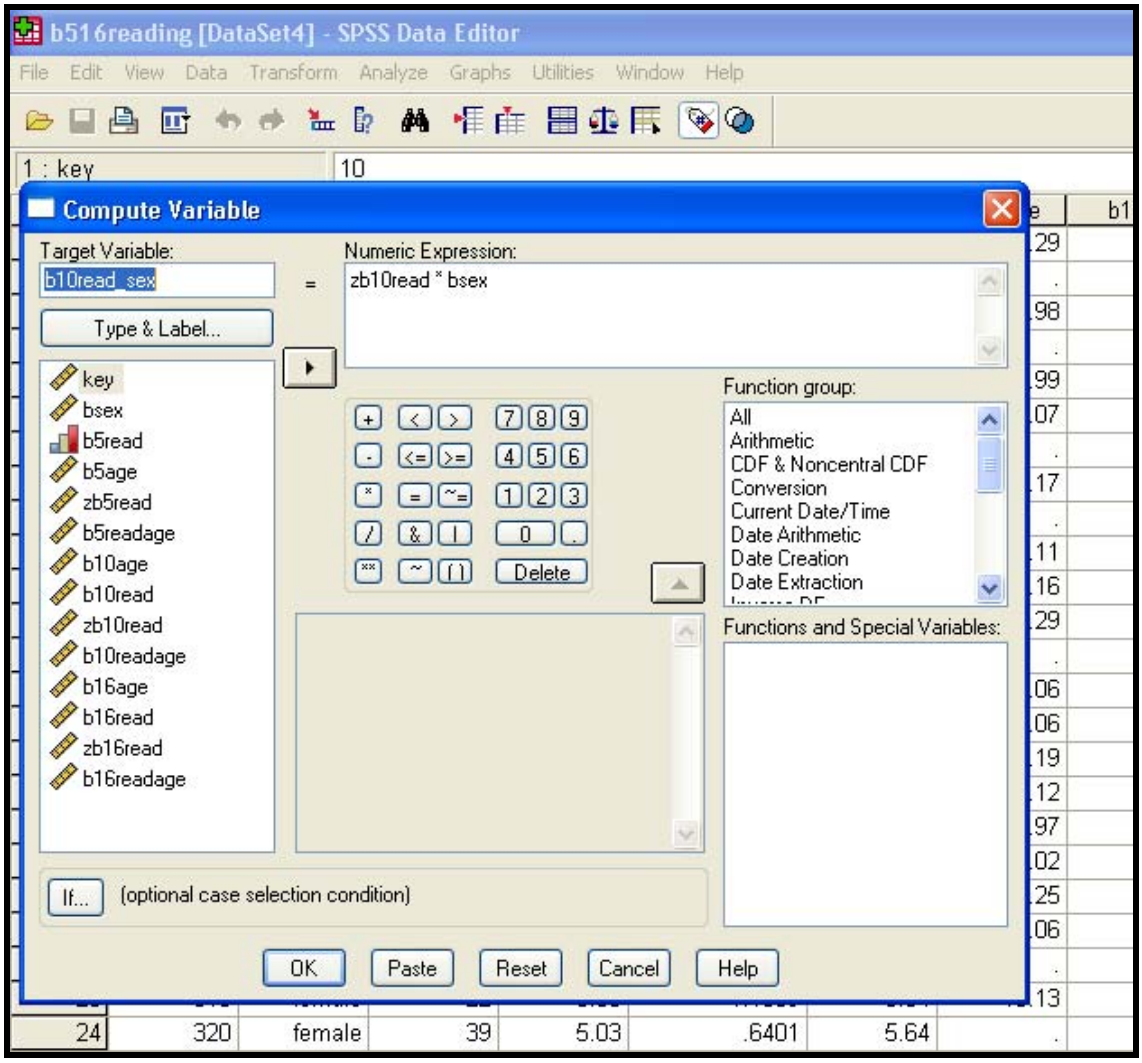
For the first analyses, from the *Analyse* drop-down menu select *Regression* and then *Linear*. Select the standardised reading score variable zb16read and click the ► button to place it into the Dependent box. Select b16age and click the ► button to place it into the Independent(s) box. Click OK. For the second regression, add in zb10read to the Independent(s) box, for the third regression, 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.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT zb16read
  /METHOD=ENTER b16age zb10read bsex  .
```



Before running the fourth regression, create a new variable (b10read_sex) showing the interaction between reading score at age 10 and gender. From the *Transform* drop-down menu select *Compute*. Type in the name of the new variable in the *Target Variable* box. Select zb10read and click the ► button to place it into the *Numeric Expression* box. Click on * (multiply) and then move bsex into the *Numeric Expression* box. Click on OK. After this, include the new variable in the regression model. Is there evidence of an interaction, or do girls fall behind by age 16 regardless of their reading score at age 10?



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 b10read_sex = zb10read * bsex .
EXECUTE .
```

f) What effect does adding in reading performance at age 5 have? Do boys still make more progress than girls between age 10 and age 16?