Practical session

Brian Dodgeon and Peter Shepherd
Centre for Longitudinal Studies
Institute of Education
University of London
Objectives of Practical Session

1. Practise using *Idealist* Data Dictionary (on the complimentary CD)

2. Choose one or more of four model examples

3. Use *Idealist* Data Dictionary to identify relevant variables.

4. Use SPSS model datasets provided on data sticks to carry out analysis, or use corresponding STATA datasets (**NB:** data sticks must be returned at the end of the workshop)

5. **Optional:** Gain experience of using NESSTAR via The Data Archive website.
Using Idealist Data Dictionary

Follow instructions in the handout for:

- opening the NCDS data dictionary
- carrying out a simple search
- narrowing the search criteria
- widening the search criteria
- exporting a hit list
- marking records

Once you have completed this exercise, you are ready to use the data dictionary to identify relevant sets of variables
Using Idealist Data Dictionary

Opening the data dictionary
1. To access the NCDS data dictionary first click on the folder icon “NCDS Data Dictionary” and then click on the folder icon “NCDS 0-50 Data Dictionary”.
2. Clicking on the icon “Iwinread.exe” will open the data dictionary at the first page.

How to carry out a simple search
a) Using the ‘Search’ drop down menu and then selecting ‘Find’ [or by pressing F5], search for all variables in the NCDS datasets on voting. The best way to do this is to search using the ‘contains a word starting with’ option and type in ‘vot’. This will then find the words ‘vote’, ‘voted’, ‘voting’ and should result in a hit list with 20 entries.

b) If you were only interested in the age 42 sweep carried out in 1999/2000 you may wish to ‘narrow’ your hit-list. Either use the Search drop down menu again and select ‘narrow’ or press F7. Now type in 2000 and click on OK. Your hit-list should now be reduced to 6 records and, if you scroll through them, you will see they all relate to the 1999-2000 data collection.

c) It is often helpful to ‘Export’ your hit list so that you can create your own data dictionary. Use the drop-down menu ‘File’ then ‘Export’. In the first box under ‘What?’ select ‘Hit list’. For ‘Method’ select ‘Idealist Natural’, then click OK and the hit list will be placed on the clipboard. Now go into a Word document and select Paste to paste the contents of the clipboard into the file.

d) Marking records: sometimes you may need to create your own customised ‘hit-list’ by marking the records you are most interested in. Search for all records containing words beginning with ‘cook’. You should have 109 in your hit-list. Now scroll through the first 21 of these (i.e. age 0-42 sweeps) and identify those that ask about cooking in relation to the domestic division of labour (i.e. who does most of the cooking a partnership). Mark these records by selecting the ‘Record’ drop-down menu and then ‘Mark’. Finally select ‘Record’ and ‘Drop Unmarked’ to leave you with a hit-list of just two records. These can then be exported to your data dictionary as above.

e) Repeat the search for records including words beginning with ‘vot’ using the BCS70 data dictionary. You should find 9 records. Export these to your data dictionary too.

f) Repeat the search for records including words beginning with ‘cook’ using the BCS70 data dictionary. You should find 32 records. Scroll through these and mark any that look interesting. Create a new hit-list by dropping the unmarked records. Export these to your data dictionary too.

g) TIP: Having now carried out this exercise you are ready to use the data dictionary to identify relevant sets of variables. Bear in mind that the data dictionary includes a number of idiosyncrasies, such as abbreviated or contracted words, that you sometimes need to take into account when you are using it to search.
Choose one or more of four model examples

1. Does eating fruit make you more likely to vote?
2. Cognitive Ability in Middle Age
3. How family life of child impacts on outcomes in adulthood (BCS70)
4. Testing if quality of family life really does predict academic success, by logistic regression
Choose one or more of four model examples

The SPSS datasets for the worked examples are:

- Does eating fruit make you more likely to vote? - **EXAMPLE 1**: Example1.sav
- Cognitive Ability in Middle Age - **EXAMPLE 2**: Example2.sav
- How family life of child impacts on outcomes in adulthood - **EXAMPLE 3**: Ex_3_BCS70 childhood data (ages 5 and 10).sav; Ex_3_BCS70 highest qualification at age 30.sav;
- Testing if quality of family life really does predict academic success, by logistic regression – **EXAMPLE 4**: Ex_4_merged BCS70 longitudinal dataset (with dichotomous variables).sav

The corresponding STATA datasets are:

- **EXAMPLE 1**: Example1.dta
- **EXAMPLE 2**: Example2.dta
- **EXAMPLE 3**: Ex_3_BCS70 childhood data (ages 5 and 10).dta; Ex_3_BCS70 highest qualification at age 30.dta;
- **EXAMPLE 4**: Ex_4_merged BCS70 longitudinal dataset (with dichotomous variables).dta

You should:

- Try to locate the variables in the Data Dictionary without first looking at the data files (which contain exactly the variables you’re looking for!).
- Then try to do the analyses in SPSS or STATA, using either syntax or drop-down menus.
- If you don’t succeed, you can take advantage of the syntax provided on the memory sticks in the directories ‘SPSS Syntax’ and ‘STATA syntax’.
Example 1: *Does eating fruit make you more likely to vote?*

- For NCDS cohort members investigate the relationship between eating fruit and tendency to vote at General Elections.
- Look at a possible ‘mediating’ variable – social class.
Example 1: *Does eating fruit make you more likely to vote?*

- Use the Data Dictionary to find which variables at the NCDS 1999/2000 survey (NCDS6) record frequency of eating fruit, and whether the cohort member voted at the 1997 election.

- Do a cross-tabulation of these two variables, in order to compare the percentage who voted, in each category of fruit-eating frequency. If you’re working with SPSS, use the data set Example1.sav, or if STATA, use Example1.dta.

- Find the variable for social class at the same survey.

- Selecting just those who are unskilled (category 5), do the same cross-tabulation and see if the gradient is still the same.

- Finally, do a cross-tabulation of social class against whether they voted in 1997.
Example 2: *Cognitive Ability in Middle Age*

- For NCDS cohort members investigate how people’s cognitive ability at age 50 varies by gender and health behaviours: drinking, smoking and exercise.
Example 2: Cognitive Ability in Middle Age

- Use the Data Dictionary to find which variables at the NCDS 2008 survey (NCDS8) record the results of the four cognitive tests administered during the survey:
  - *Word Recall* (‘Number of words correctly recalled’)
  - *Delayed Word Recall* (‘Number of words correctly recalled after delay’)
  - *Animal Naming* (‘Number of animals mentioned’)
  - *Letter Cancellation*. You’ll find there are two scores: ‘Speed’ and ‘Accuracy’.

- Using the data set Example2.sav, do a frequency count of these variables (or if using STATA, Example2.dta).

- Use the Data Dictionary to find which variables at the NCDS 2008 survey represent cohort members’ sex and patterns of drinking, smoking and exercise. You’ll find there are a number of these health behaviour variables – the best to use are probably:
  - ‘Smoking frequency’,
  - ‘Frequency of drinking alcohol’
  - ‘Whether exercises regularly’.

- Do a frequency count of these variables, then find the mean value of each of the test scores for each value of the sex and health behaviour variables.
Example 3: How family life of child impacts on outcomes in adulthood (BCS70)

- Data from childhood contains variables looking at quality of family life:
  - Parents reading to child
  - Dad’s role in management of child
  - Doing activities together:
    - Walks
    - Outings
    - Holidays
    - Shopping
    - Meals
    - Chatting
    - Restaurant

- See if you can find these variables in the Data Dictionary, then do some frequency counts

- Merge data subset from childhood with data subset from adulthood

- Do some cross-tabulations of the family-life variables at age 5 and 10 against the variable showing qualifications at age 30 (bhqual00 or its more simple, dichotomous version ‘degree’).

- Check if any associations you see might be due to external variables like social class
Example 3: *How family life of child impacts on outcomes in adulthood (BCS70)*

- Use BCS70 Data Dictionary to find the relevant variables.

- Open data subset ‘Ex_3_BCS70 childhood data (ages 5 and 10).sav’ (or Ex_3_BCS70 childhood data (ages 5 and 10).dta if using STATA).

- Do frequency counts of these variables.

- Merge childhood data subset with the dataset ‘Ex_3_BCS70 highest qualification at age 30.sav’ (or STATA equivalent).

- Use the merged (now longitudinal) dataset to do cross-tabulations of the family-life variables against qualifications at age 30.

- You may feel that some of the associations you see are mediated by external factors such as social class (e.g. professional parents may be more likely to read to their children and take them on holidays, and high academic achievement may be more directly related to the prosperity of the family).

- Try using the social class of parents at age 10 to investigate this (e.g. select only one social class band, and see if the associations are as strong).
Example 4: *Testing if quality of family life really does predict academic success, by logistic regression*

- Use the merged dataset from example 3 to do a logistic regression analysis
- Use dichotomous variable ‘degree’ as the outcome, and the family-life variables as predictors.
- Need to create dichotomous versions of the family-life variables (or use ready-made ones!)
Example 4: Testing if quality of family life really does predict academic success, by logistic regression

- Use the merged dataset from Example 3, containing both the childhood family life variables and the adult data showing highest qualification.

- You’ll need to create new ‘2-value’ versions of the family-life variables (e.g. recode the four values of variable m105 to a new variable with 1=’dad has big or equal role; 2=’dad has lesser role’).

- If you prefer not to do this yourself, you can use the ready-made dataset on the memory stick ‘Ex_4_merged BCS70 longitudinal dataset (with dichotomous variables).sav’ (or the corresponding STATA version).


- Which results are significant at the 95% level?

- As with example 3, try using the social class variable to see whether it is really family life that is predicting academic success, or if it’s mediated by socio-economic factors.
Once you have completed as many of the worked examples as you wish to do, try using NESSTAR to browse the variables for the 1999-2000 survey.
1. Using your web browser open: http://www.data-archive.ac.uk
2. Choose “FIND DATA” then “MAJOR STUDIES – VIEW AT ESDS” then “ESDS Longitudinal datasets”
3. Choose “National Child Development Study” then “National Child Development Study”
4. For Dataset serial number 5578 (1999-2000 data) click on the icon for “Nesstar”.
5. Open the relevant folder for NCDS.
6. Click on the “Variable Description” folder in the left-hand pane.
7. To explore data on, for example, “current accommodation” click on the relevant folder.
8. To run tabulations of, for example, ACCOM (type of accommodation) by HSETYPE (whether house/bungalow etc) use the following instructions.

   a) First click on ‘Type of accom. CM currently occupies” in the left-hand pane and choose ‘Table’ in the right-hand pane, then revisit ‘Type of accom. CM currently occupies” in the left-hand pane and choose ‘Add to row’.

   b) Next click on ‘Whether house/bungalow detached, semi or terraced’ in the left-hand pane and choose ‘Table’ in the right-hand column, then revisit ‘Whether house/bungalow detached, semi or terraced’ in the left-hand pane and choose ‘Add to column’.

   c) During this you will then be asked to provide a userID and password. For this session these are:

   `userID: ioestaffdodgeb`  \(\text{password: hellojan1}\)
Please complete an ‘Evaluation Form’
QUESTIONS