Office for National Statistics (ONS) Longitudinal Study

Downloadable Dataset on Ethnicity

- The ONS Longitudinal Study is a record linkage study including individual-level data from the 1971, 1981, 1991 and 2001 Censuses and Vital Registration events. The study sample is approximately 1% of the population of England and Wales. For more information, see http://www.celsius.census.ac.uk.

- Census datasets are of particular value when a detailed breakdown of ethnic groups is required, because of their size. The ONS Longitudinal Study offers the additional strength of a longitudinal design. A question on ethnicity was included in the 1991 and 2001 Censuses; the 2001 Census also included a question on religion. (Country of birth is asked in every Census 1971-2001.)

The dataset

- The population included in this dataset is all ONS Longitudinal Study sample members who were enumerated at all three of the most recent Censuses (1981, 1991 and 2001). Sample members who were not present at all three of these Censuses have been excluded.

- The minimum age (in 2001) is therefore 20 years.

- This teaching dataset is based on records from 333,015 sample members, extracted from the ONS Longitudinal Study in October 2008.

Exercises (with hints using SPSS)

NB If you are not familiar with SPSS some notes are offered at the end of this exercise.

1. Download the dataset from http://www.celsius.census.ac.uk/aggregat/index.html. Save it, noticing where it is saved!

2. Start SPSS.

3. Open the dataset (File, Open, Data).

4. To analyse this aggregated dataset, run the command WEIGHT BY CASCOUNT. before you begin. The results will be identical to those obtained from a non-aggregated dataset.

5. Look at the aggregated dataset. In an aggregated dataset, instead of each row representing one case, each row represents one combination of characteristics, and includes a count of the cases which share this particular combination. The variable showing the count is CASCOUNT.

Instruction: Look at the Data View tab in the Data Editor window. (Turn View, Value labels on if necessary, to see what the values represent.) Examine the variable CASCOUNT –
variable names are in the column heading boxes.

Answer/comment: Taking the first row as an example, it shows 157 people with this combination of characteristics:

White British women, aged 60 or over in 2001, born in the UK (or not stated), not known to be economically active in 1981 or 1991, but economically active in 2001.

6. Run a frequency for each variable except CASCOUNT. For each variable, note from the value labels how missing responses have been dealt with.

*Instruction:* WEIGHT by CASCOUNT.
*FREQUENCIES VARIABLES=sex ethgrp0 agegrp0 bornexuk econac8 econac9 econac0.

Answer/comment: Missing responses have generally been united with a non-missing response, in order to reduce the number of possible categories in the dataset (see 'Why an Aggregated Dataset?' for the reason at www.celsius.census.ac.uk/aggregat/index.html).

7. What is the gender distribution, by age group in 2001?

*Instruction:* CROSSTABS /TABLES=sex BY agegrp0 /CELLS= COUNT COLUMN.

Answer/comment: Women are overrepresented but especially in age group 60 years or over. This is likely to be due in part to longer life expectancy of women compared with men. Another likely reason is that young men are known to be under-enumerated in each Census, and may therefore be less likely than other groups to be enumerated at three successive Censuses.

8. Look at the gender distribution by ethnicity. Which ethnic groups have the highest proportion of men, and which have the highest proportion of women?

*Instruction:* CROSSTABS /TABLES=ethgrp0 BY sex /CELLS=ROW.

Answer/comment: Bangladeshis have the highest proportion of men at 60%, followed by Other Asian at 58%; Pakistanis are also high at 55%. With regard to women, the highest proportion (59%) is found among those whose ethnicity was not recorded. The next highest (58%) was in the Black Caribbean group.

9. Create a less detailed variable for ethnic group using the headings in the 2001 Census form (plus a 'not stated' category): (1) White (2) Mixed (3) Asian or Asian British (4) Black or Black British (5) Chinese or other ethnic group (6) Not stated

*Instruction:*
RECODE ethgrp0 (17=6) (1 thru 3=1) (4 thru 7=2) (8 thru 11=3) (12 thru 14=4) (15 thru 16=5) INTO ethnic6 .
VARIABLE LABELS ethnic6 'Ethnicity in 2001, 6 codes'.
VALUE LABELS ethnic6 1 "White" 2 "Mixed" 3 "Asian or Asian British" 4 "Black or Black British" 5 "Chinese or other ethnic group" 6 "Not stated".
EXECUTE.
CROSSTABS /TABLES=ethgrp0 BY ethnic6.

Answer/comment: Check the tabulation to make sure the new variable has been created
10. Were there differences by (six-group) ethnicity in the probability of being economically active in 1981?

*Instruction: CROSSTABS /TABLES=econac8 BY ethnic6 /CELLS=COLUMN.*

*Answer/comment:* People of Mixed ethnicity had a lower rate of economic activity in 1981 (at 37%) than the sample as a whole (55%). People with a missing response on ethnicity also had a lower rate at 44%.

11. How far may the differences found in exercise 6 be accounted for by age or sex distribution or whether the respondent was born in the UK?

*Instruction: LOGISTIC REGRESSION econac8 WITH sex agegrp0 bornexuk ethnic6 /CATEGORICAL sex agegrp0 bornexuk ethnic6 /CONTRAST (sex)=Indicator(1) /CONTRAST (agegrp0)=Indicator(1) /CONTRAST (bornexuk)=Indicator(1) /CONTRAST (ethnic6)=Indicator(1) /ENTER sex agegrp0 bornexuk ethnic6.*

*Answer/comment:* The model suggests that age, sex and whether born in the UK are all very important in predicting economic activity in 1981, but differences between ethnic groups appear to persist even when these characteristics are accounted for. However the economic activity rate of people of Mixed ethnicity no longer appears noticeably lower than that of other minority ethnic groups (by comparison with the White majority). A more striking finding from the model is that people of Black or Black British ethnicity were considerably more likely than the White group to be economically active in 1981, even when the other factors have been taken into consideration.


*Instruction: COMPUTE allecon=0+econac8+econac9+econac0. VARIABLE LABELS allecon 'How many Censuses economically active'. EXECUTE.*

13. Does the count of economic activity at three Censuses vary by (six-group) ethnicity? It might be safest to limit this table to the middle age group (40-59 in 2001) who would have been of working age at all three Censuses.

*Instruction: COMPUTE filter_$=(agegrp0 = 2). FILTER BY filter_$.*

*CROSSTABS /TABLES=allecon BY ethnic6 /CELLS=COLUMN. USE ALL.*

*Answer/comment:* Considerable variation! Now continue to explore the data on your own.
Notes on SPSS

To load a dataset (having loaded SPSS) simply choose File, Open, Data. Notice the two tabs at the bottom left of the screen: Data View and Variable View – try them. In Data View you can swap between values (usually numbers) and value labels by choosing View, Value Labels.

SPSS offers two ways of running commands. The first is the menu system; this is mainly used by beginners. The second is more efficient: it is an editor especially for working on commands in text form (known in SPSS as 'syntax'). To open the Syntax Editor choose File, New, Syntax and notice the Run command in the menu.

Commands in this exercise are given in text form; you can paste them into a syntax window, amend them if you wish and run them.

• Do not include the word 'Instruction:' at the beginning
• DO include the full stop at the end
• Don't forget to start with WEIGHT BY CASCOUNT.

It is also possible to build up commands through the menu system and then choose Paste instead of OK which will insert the command in text form into the syntax window.