Scottish Schools Adolescent Lifestyle and Substance Use Survey

Introduction to analysing the SALSUS trend dataset

Chris Martin, April 2016
Contents

1. Overview of creation of the trend dataset
2. Introduction to the questionnaire coverage (and how to find variables)
3. Practical session
4. Tips for planning analysis and avoiding errors
5. Practical session (continued)
The 12 waves of data

1982
- First in the survey series measuring smoking prevalence

1984
- 

1986
- 

1990
- Drinking prevalence introduced to the survey

1992
- 

1994
- 

1996
- Prevalence of drug use introduced to the survey

2000
- 

2002
- Separate Scottish survey introduced. Local level data for the first time

2004
- 

2006
- 

2008
- Local boosts available for the first time

2010
- 

2013
- 

SALSUS TIME SERIES DATASET
Stage 1 was to see if it was feasible

‘Is it feasible?’ comes down to answering the question, is the data consistent enough between years to allow meaningful comparison?

(And if not, could the data be amended to allow comparisons to be made in a meaningful way?)
“Consistent enough to allow meaningful analysis”

- Always used **paper self-completion surveys** administered in school under ‘exam’ conditions.
- **Similar sampling strategies** (designed to provide nationally representative estimates at a minimum and clustered around schools or classes or both).
- **State and independent schools** included and special schools excluded.
- **Weighting** has been used in most waves.
- Schools have been encouraged to follow up on **pupils who were absent** on the day of completion.
From the 1990 User Guide...

Survey of Smoking
1990

Most of the questions can be answered by putting a tick in the box next to the answer that applies to you - like this:

Yes 1
No 2

or sometimes you have to write a number in the box, for example:

5

Some questions don't apply to everybody, so sometimes it tells you by the box which question you should answer next. If there is nothing by the box, you should answer the next question.
From the 2002 User Guide...

**Question 33**

How much do you weigh without clothes?

*Answer either in kilograms OR stones and pounds*

- I weigh ____________ kilograms
- OR
- I weigh ____________ stone ____________ pounds
- OR
- I don't know what I weigh ☐

**HBSC - q33a, q33b.1, q33b.2**
From the 2006 User Guide...

10) Do you smoke cigarettes at all nowadays?

Yes
No

Go to Q11
From the 2013 User Guide...

**These next questions are about smoking tobacco**

7. Do you smoke cigarettes at all nowadays?
   - [ ] Yes
   - [ ] No

8. Now read the following statements carefully and cross the box next to the one which best describes you
   - [ ] I have never smoked
   - [ ] I have only ever tried smoking once
   - [ ] I used to smoke sometimes but I never smoke a cigarette now
   - [ ] I sometimes smoke cigarettes now but I don’t smoke as many as one a week
   - [ ] I usually smoke between one and six cigarettes a week
   - [ ] I usually smoke more than six cigarettes a week
### Mapping the questionnaire coverage over time

<table>
<thead>
<tr>
<th>Drinking</th>
<th>Variable name</th>
<th>90</th>
<th>92</th>
<th>94</th>
<th>96</th>
<th>98</th>
<th>00</th>
<th>02</th>
<th>04</th>
<th>06</th>
<th>08</th>
<th>10</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behaviour - ever had an alcoholic drink</td>
<td>a_everdrink00</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - usual frequency of having a drink</td>
<td>a_drinkfreq00</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - DERIVED Frequency of drinking – main variables</td>
<td>a_drinkfreq00</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - when last had a drink</td>
<td>a_lastdrin90</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - days of the week they had a drink</td>
<td>a_sun2drink00</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - ever been drunk</td>
<td>a_everdrink04</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - been drunk in last 7 days</td>
<td>a_drinkweek06</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - deliberately tried to get drunk in last week</td>
<td>a_deliberate06</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - vomited in last week</td>
<td>a_vomited06</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - 5 or more drinks on the same occasion in last 30 days</td>
<td>a_fivedrinks02</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - money usually spent each week on alcohol</td>
<td>a_spentweek02</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - who usually with when drinking</td>
<td>a_withnum04 to a_alone04 &amp; a_withf06 a_withgt690</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - usual location when drinking</td>
<td>A_drinkpub06</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - whether experienced various things while drinking (injured, seen a doctor, vomited, etc.)</td>
<td>A_argument04</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour - DERIVED No. of effects experienced while drinking in last year</td>
<td>a_eleffect02</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour (units) - amount of Beer/Lager/Cider in last seven days</td>
<td>a_drinkbeer90</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour (units) - whether drink strong or normal beer.</td>
<td>a_brstrength98</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Putting the data together – part 1

- For each separate wave of data,
  - **rename variables** to agreed names for the combined dataset
  - **recode any variables** to ensure that values are consistent across years
  - potentially **derive new variables/re-banded versions**
  - **drop variables** that are considered completely redundant
  - create a 'year' variable
Putting the data together – part 2

- Combine all the revised individual datasets into a single dataset.
  - Add variable labels and value labels to the combined dataset.
  - Where necessary, amend data to make base definitions consistent across years.
  - Ensure that missing and not answered responses are coded as consistently as possible.

- Create additional derived variables (such as revised weights).

- Re-order variables to aid analysis.
### Key Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Year</td>
</tr>
<tr>
<td>Weight</td>
<td>Weight</td>
</tr>
<tr>
<td>LA</td>
<td>c_la02</td>
</tr>
<tr>
<td>Sex</td>
<td>c_sex90</td>
</tr>
<tr>
<td>Year group</td>
<td>c_yeargroup90</td>
</tr>
</tbody>
</table>

### Naming Conventions

#### Smoking variables
- **S__XX** (s_smokefreq90)

#### Alcohol variables
- **A__XX** (a_firstdrink96)

#### Drugs
- **D__XX** (D_drugever98)

#### Sport, leisure, social contact, mental health
- **P__XX** (P_truant02)

#### Contextual/other
- **C__XX** (C_numbro06)
Finding information on variables

From the User Guide

“Smoking - General attitudes to smoking:
General attitudes toward smoking (13 variables): True or False responses were changed to agree or disagree in 2006. These have been collated into the same variables. However, a new set (16 variables) were introduced in 2013, when distinction between 'strongly' and 'tend to' agree/disagree was added.”
Practical session... Over to you

You should have...
- The data
- Workbook
- Syntax files
- Variable reference file

Work at your own pace...
- Get a feel for the data
- Get to know how to use of the user guide and variable reference file

In the workbook...
- Chapter 3 – Data management in SPSS.
- Chapter 4 – Six examples fully worked
  - Syntax file for examples 1 to 6.
- Chapter 5 – Six further examples
  - Syntax file for examples 7 to 12.

Don’t hesitate to ask questions
Key Variables

Year
Weight
LA
Sex
Year group

Year
Weight
c_la02
c_sex90
c_yeargroup90

Naming Conventions

Smoking variables
S_....XX (s_smokefreq90)

Alcohol variables
A_.....XX (a_firstdrink96)

Drugs
D_...XX (D_drugever98)

Sport, leisure, social contact, mental health
P_...XX (P_truant02)

Contextual/other
C_...XX (C_numbro06).
Its easy to make mistakes...

- Mechanical mistakes...
  - Easy to delete cases, variables, overtype values
  - Shared files can be altered

- Procedural mistakes...
  - Forget to apply weights, use the wrong weight, forget to change weights
  - Making mistakes when changing a variable

- Analytical mistakes...
  - Misunderstand the question: think it’s asked of someone it isn’t, think it represents something it doesn’t
  - Misinterpret the output: think something’s important when it isn’t, think something shows something it doesn’t
Pre-flight checks

- Data file is clean original
  - How do you know if someone has changed the file?
- Cases and variables are intact
  - How many cases should you have?
  - How many variables?
- All variables are visible
- No filters or weights active
  - Check status bar at the bottom

Visible: 786 of 786 Variables

SPSS Processor is ready  Filter On  Weight On  Split File On
Preparing for analysis

- Questions to be answered
  - Make the terms of the question specific
  - If possible, reformulate the question as a testable hypothesis

- Link analysis to survey questions and variables
  - Identify which SHS variables will address the question
  - Inspect the variables individually and ask
    - Are there any problems with the variables?
    - Are they in the right format for the analysis?
    - Do they need to be merged / combined to make the analysis possible or easier?

- The limits of analysis
  - Does the analysis actually answer the question?
  - Is the analysis sufficiently robust to support the conclusions?
The five stages

- Formulate your question/hypothesis
- Decide who you are interested in
- Decide if the analysis should be weighted
- Work out what variables you need (and do they need to be changed in any way)
- Decide on the type of analysis needed
Ways to avoid mistakes

- Don’t touch the data!
- Write down what you are going to do
- Work from a clean file – an unaltered and up-to-date version
- Make sure analysis questions are specific and testable
- Build in simple checks on data and analysis
- Avoid trawling for results
- Be clear about significance – chance and importance
- Talk through analysis with someone else
- (If you can) have someone else check your analysis
- (If not) stand back from analysis and ask if it makes sense
Practical session continued...

Key Variables

- Year
- Weight
- LA
- Sex
- Year group

- Year
- Weight
- c_la02
- c_sex90
- c_yeargroup90
For more information

Chris Martin
Research Director
0131 220 5699
chris.martin@ipsos.com
@_chris_d_martin

www.ipsos-mori.com/scotland
Additional if needed
Hypothesis driven analysis

- A hypothesis a statement of some testable outcome
  - More than a question that you want to answer
  - A question with rules about what will determine the answer
- A hypothesis helps you construct analysis to rigorously test the hypothesis and decide whether to accept or reject it
- The value of hypothesis driven analysis
  - provides focus to the analysis
  - provides – before the analysis is undertaken – a basis for deciding how to evaluate the outcome
  - Provides transparency about the thinking that has structured the analysis
- Hypotheses may be derived from some broader theory
- The alternative is a fishing expedition – scanning data looking for something interesting and reporting that
Formulating hypothesis

- The null hypothesis
  - A statement of no difference or no effect
    - There is no difference in smoking rates between boys and girls
    - Children attending private school are as likely to truant than children attending state schools

- The alternative hypothesis
  - A statement of difference or effect
    - There is a significant difference in smoking rates between boys and girls
    - Children attending private school are less likely to truant than children attending state schools

- Testing is in terms of the null hypothesis – to accept or reject.
  - Unless our analysis can clearly demonstrate otherwise, you accept that there is no difference or effect