Weighting the UK Omnibus data

This document describes how to analyse the Omnibus data. The data have not been weighted so users are urged to read this document.

Summary

UK-level analysis (Section 1 and 3)

For any type of analysis make sure you weight the data. The key variable to look for is called PSEWEIGHT, and the SPSS syntax command "*weight by PSEWEIGHT*" should precede any analysis. If, however, you need to report measures of variance (e.g. confidence intervals, p-values), you need to use the SPSS Complex Samples[1] facility as explained below. The SPSS syntax command '*weight by PSEWEIGHT*' will give p-values and confidence intervals which are almost correct but they will be over-optimistic in that they may incorrectly show that a result is statistically significant when in fact it is not.

GB-level, NI-level or GB vs NI comparison (Section 2)

This is slightly more complicated. Make sure you read all sections.

SECTION 1 - Why complex sample procedures need to be used on the Omnibus

The UK Omnibus is a stratified clustered sample[2], also known as a multistage sample. The calculation of the variance and therefore standard deviation, standard error, p-values and confidence intervals are affected in a way which cannot be fully tackled by using normal weighting procedures (i.e. the "weight by …" command in SPSS)[3]. It is therefore necessary to describe all the sampling characteristics (e.g. strata, clusters) to the software to get the correct variance (and related p-values and confidence interval) estimates. In SPSS this is done with CS (complex sample) commands; the CSPLAN command describes the sample to SPSS while all others (e.g. CSTABULATE) estimate frequencies, crosstabs, regressions etc. using the information declared in CSPLAN. If all you need are simple means or percentages, then normal weighting procedures can be used; however, if variance, standard error or significance tests for these estimates are required or reported, then the CS commands must be used.

The SPSS complex sample plan for the UK 2012 PSE Omnibus is created as follows:

Syntax 1.1 *To produce estimates for the UK as a whole.*

[1] Other software such as STATA, SAS and R have similar procedures.
[2] Northern Ireland is a simple random sample but this has been integrated in the UK sample design for the purposes of both mean and variance estimation.
[3] For more information see http://www.restore.ac.uk/PEAS/exemplar1.php#answers

*Open the UK Omnibus data.

*The syntax below creates a Complex Survey design[4] plan and saves it on your C drive as UKOmnibus.csaplan .

```
CSPLAN ANALYSIS
 /PLAN FILE='C:\Folder\Subfolder\UKOmnibus.csaplan'
 /PLANVARS ANALYSISWEIGHT=PSEWEIGHT
 /SRSESTIMATOR TYPE=WR/PRINT PLAN
 /DESIGN STRATA=Stratum CLUSTER=PSU_UK
 /ESTIMATOR TYPE=WR.
```

If you're interested in population totals (i.e. the number of people thinking an item/activity is a necessity instead of a proportion) see Section 3 below. Once you run the above syntax, all previous "weight by" commands are ignored and overridden when you use other CS commands (e.g. CSTABLE). This is because the CSPLAN saves all required weights and information in a separate plan file, so there is no need for any other weighting procedure. It is best to switch off all weights before you run the CSPLAN above. This is done by running the syntax command "weight off." If you don't, you will get a warning from SPSS[5] when you run CS commands. This warning can be ignored, as it does not affect your results.

The CSPLAN enables the appropriate computation of variance and standard errors (and related p-values, t-tests, confidence intervals etc…) for the whole of the UK. To do this in SPSS, run the following syntax:

```
CSTABULATE
  /PLAN FILE='C:\Folder\Subfolder\UKOmnibus.csaplan'
/TABLES VARIABLES=decorate
 /CELLS TABLEPCT  /STATISTICS SE CIN(95)
 /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

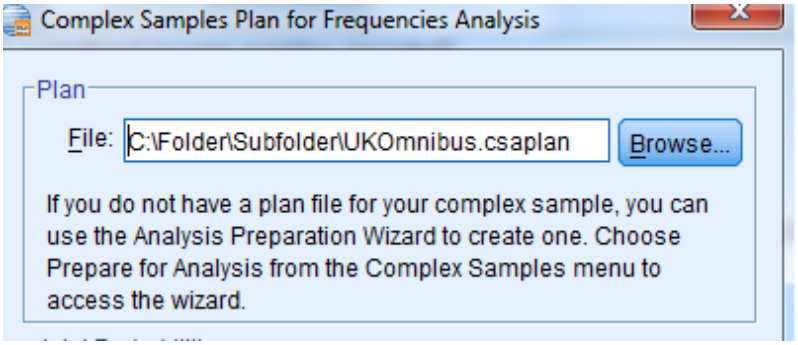
This command tells SPSS where the survey information is stored and which statistical procedure you require (in this case a frequency for the variable named '*decorate*').

You can find other CS commands (regression etc..) under Analyse→ Complex Samples. As syntax 1.1. (above) has already described the sample and created and stored a CS plan file at 'C:\Folder\Subfolder\UKOmnibus.csaplan', you can ignore the first two fields (Select and Prepare) from the CS Plan menu. Whenever you select any of the other procedures using the drop-down

[4] For more details about WR (with replacement) see the Appendix.

[5] The warning will say "This procedure ignores the weight variable." All it means is that it will ignore your previous weight by command when running CS commands. You can ignore this warning or if you find it annoying switch off the weights with weight off.

menu SPSS will ask you to specify a CSPLAN. In the top field paste in C:\UKOmnibus.csaplan or the path where your CSPLAN (see syntax 1.1) is stored



and then press Continue.

If you want to compare Northern Ireland and Great Britain or obtain non-UK-level estimates see Section 2.

Estimation among subgroups

There is a further issue when using a complex sample design plan. From previous experience you may be used to selecting a certain group (e.g. lone mother households) by using, for example, the syntax "*select if lonemother=1*" or using any other kind of filter which either temporarily or permanently deletes those cases not identified as "lonemother". This should not be done at any time when using CS commands (e.g. CSTABULATE) or when carrying out any procedure that uses the UKOmnibus.csaplan, as this will return the wrong standard errors, statistical test results and confidence intervals. Instead you should use the SUBPOP table command and include the relevant group variable (/SUBPOP TABLE=lonemother).

```
CSTABULATE
  /PLAN FILE='C:\Folder\Subfolder\UKOmnibus.csaplan'
 /TABLES VARIABLES=decorate
 /SUBPOP TABLE=lonemother DISPLAY=LAYERED
 /CELLS TABLEPCT
/STATISTICS SE CIN(95)
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Only the /SUBPOP procedure will return the correct standard errors[6]. So, whenever you're looking at a specific subset of the data and want to obtain the appropriate standard errors, use the

[6] Another, more complicated example is:

```
CSTABULATE
  /PLAN FILE='C:\Folder\Subfolder\UKOmnibus.csaplan'
 /TABLES VARIABLES=decorate
 /SUBPOP TABLE=NSSEC3 BY RespSex DISPLAY=LAYERED
 /CELLS TABLEPCT
/STATISTICS SE CIN(95) /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

/SUBPOP procedure. If you are simply deleting a few cases (i.e. outliers) this will not make any difference, but otherwise always use /SUBPOP. As a rule of thumb, think the smaller the group the larger the mistake will be when using ordinary (non CS) selection procedures.

SECTION 2 – Obtaining Estimates (means or percentages) for Northern Ireland or GB only and comparing NI with GB

The sampling method for the 2012 PSE Omnibus for Northern Ireland (NI) was different to that of the GB sample. For NI, the sample is a simple random sample, not a stratified clustered sample (as in GB). The harmonised UK 2012 PSE Omnibus dataset has integrated the NI data into the UK CSPLAN, to make UK-level estimates possible and easy.[7] The focus has been on developing a dataset to produce UK level estimates. To look at NI or GB on their own, or comparing NI to GB, it is very important to do the following.

Northern Ireland (Syntax 2.1)

*Get the UK PSE Omnibus Data and select cases for NI only (SPSS syntax: "select if GBNI=2.").
(Note: this deletes all non-NI cases on the dataset, so do not save your data file, or you will lose all data from GB respondents!).
*Apply the NI sample weight: SPSS syntax: "Weight by PSE**NI**WEIGHT."
*Run your usual analyses (not the CS ones). This is because the NI sample is a simple random sample, so the standard analyses/tests will be appropriate.

Great Britain (Syntax 2.2)

*Get the UK PSE Omnibus Data and select cases for GB only (SPSS syntax: "select if GBNI=1.").
*Provide information about your sample and strata (create the CS plan).

```
CSPLAN ANALYSIS
  /PLAN FILE='C:\Folder\Subfolder\GBOmnibus.csaplan'
  /PLANVARS ANALYSISWEIGHT=PSEGBWEIGHT
  /SRSESTIMATOR TYPE=WR/PRINT PLAN
  /DESIGN STRATA=Stratum CLUSTER=PSU_UK
  /ESTIMATOR TYPE=WR.

*Run the CS procedure (e.g. CSTABULATE with the CSPLAN above).
CSTABULATE
   /PLAN FILE='C:\Folder\Subfolder\ GBOmnibus.csaplan'
/TABLES VARIABLES=decorate
  /CELLS TABLEPCT  /STATISTICS SE CIN(95)
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

[7] For more details on the theory behind this contact Dave Gordon (Dave.Gordon@bristol.ac.uk).

Compare GB and NI

If you want to compare NI to GB simply obtain confidence intervals for NI (without CS commands as described above in 2.1) and then confidence intervals for GB (2.2).

SECTION 3- Estimating population totals

If you want/need population totals for the UK (e.g. the number of UK adults thinking item/activity X is necessary), then apply the variable POPWEIGHT [SPSS syntax command: "*Weight by POPWEIGHT.*"]. When reporting population figures remember to round it and remember that this is an approximation. Most importantly, when using POPWEIGHT **DO NOT** report the standard error or carry out statistical tests, as these will be INCORRECT. The only way to get appropriate variance estimates and statistical test results is to use the CS commands explained above.

APPENDIX

Sampling with or without replacement?

The CSPLAN syntax states that the sample was drawn With Replacement (WR) (as opposed to Without Replacement (WOR)). This distinction is important, since specifying with (WR) or without replacement (WOR) affects the variance and therefore the standard deviations and standard errors of the estimates (e.g. mean). For most nationally representative social surveys, the difference between describing the sample as WR or WOR is minimal as long as the population is much larger than the sample. For example, after estimating the variance of a survey with a complex design, a correction is generally applied for surveys that have used sampling with replacement: the estimated variance is multiplied by $\frac{N-n}{N}$, where N is the population size and n is the sample size. For example, in the case of a population of around 50,000,000 and a sample size of 2000, this correction is around 0.99. This correction does not change significantly until the sample size drops to around 20 million, when it falls to 0.67. Hence even with a sample size of 20,000 the correction remains virtually unchanged, and (more importantly) thus has virtually no effect (it multiplies the variance by 1).

USEFUL LINKS

SPSS Complex Samples Manual
If you are a Stata user look at the svyset command.

CHANGES FROM VERSION 3 TO VERSION 4 (OMNIBUS DATAFILE)

PSPW1 (the normalised weight) became PSEWEIGHT in Version 4.
PSPW2 (the population weight) became POPWEIGHT in Version 4.
Wt (the original normalised weights provided by NISRA and NatCen separetly for NI and GB) was labelled DONOTUSEwt in Version 4.

A new variable PSENIWEIGHT was created. This is the same as Wt for NI cases and is set as missing for all GB cases. A new variable PSEGBWEIGHT was created. This is the same as Wt for GB cases and is set as missing for all GB cases.

For questions on the CSPLAN and this document contact Professor David Gordon, School for Policy Studies, University of Bristol, or on email at: Dave.Gordon@bristol.ac.uk