

Weights and survey design variables and the calculation of standard errors for the LFS and FRS modules

Both the LFS and FRS use complex sampling and grossing weights (to population counts). These features create challenges for users wanting to calculate standard errors, outlined below.

Context

If we are writing a learning module on these datasets, we should include some advice on the following topics. The advice would be mostly likely connected to a data exercise that looks at calculating whether point estimates vary significantly across years or groups.

Also I have asked the LFS team at ONS for their advice and they point me to this guide:

<https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesno9guidetocalculatingstandarderrorsforonsocialsurveys>. The FRS team have not offered any advice yet.

No survey design variables in the EUL version

One challenge is that the EUL versions of these datasets do not contain the survey design variables. Some users (typically PhD students who want to do it right) contact us to learn what they should do. In general, the options are to ignore it (like many others do), apply for secure versions of the data (I do not know if anyone has ever applied for secure data, just to do that), or to try and use some of the design factors published by the survey producers. In last 12 months of so, we have had users ask why government surveys (Wealth and Assets Survey and Family Resources Survey) do not have anonymised versions of the survey design variables as other surveys do.

Q - What advice should we be providing to users in these situations? Should we make a case to the producers to provide anonymised survey design variables?

How statistics packages deal with grossing weights

The weights with the FRS and LFS gross to population counts. Some statistics software (for some analyses) will use weighted counts when calculating standard errors. As a result, there is a massive effect on the size of the standard error and therefore very problematic if not dealt with. The problem is associated with weights in the standard SPSS but may also apply other e.g., some SAS procedures. Our standard advice has been to rescale the weight to have a mean of 1. We also do this for teaching datasets.

However, this is a rough solution where the calculation is still based on the weighted count, which varies from the sample size. As a result, the standard error will still vary from that provided by software that uses the sample size to calculate the standard error (but they are now much closer).

One work round in SPSS is to use the complex sample add-on. It works on a limited set of functions only. We currently have no advice on how to set it up.

Q-What advice should we be providing to users in these situations?

Commented [VH1]: i would say yes. we may have asked this before but not for a long while so worth asking again. If you have an example of another survey that already does this that would help - partic if ONS. HSE used to provide these but not sure if they still do

Commented [SK2R1]: I did a search on VQB for 'psu' and found variables for USoc, Community Life Survey and others so there are a number of data producers who may provide this info in EUL data. I've not checked that the info is actually available in any of the surveys though and the psu variable may have another name. From the ONS, there's only 2008 and 2009 EU-SILC and an Opinions and Lifestyle Survey - Well-being module from 2015-2016 data that contain a 'psu' variable (I think the ONS says not to use the well-being data any longer as users should use the full APS dataset).

Commented [VH3]: advising to use the complex sample add-on won't meet the needs of all users because there is an extra cost. I would have Pierre look at it and ask Natalie if still unsure.