FRS 2122 Dataset: Measuring Uncertainty in FRS Estimates

As the FRS is a sample from the UK population, any statistics derived from it are only estimates of the true numbers for the overall population.

Quantification of uncertainty in estimates produced by the FRS is performed by deriving a series of alternative samples from the main sample and quantifying the variation observed across these samples.

Each of these samples is a subsample of the main sample, known as a 'resample', and sampling units are selected and weighted using a method that replicates the original sampling methodology.

The datasets:

- saf_resamples_dataset_2122.csv
- eul_resamples_dataset_2122.csv

are datasets containing factors assigned at the household, benefit unit (family), and individual level that describe the 500 different randomly chosen resamples used for uncertainty quantification this year. An estimate based on a given resample can be arrived at by appling the particular set of grossing factors associated with a particular resample to any calculations. Weighting factors will be 0 when the household has not been selected as part of a particular resample.

The availability of these factors should allow users to reproduce all measures of uncertainty published, as well as calculate uncertainty for any analysis they undertake independantly.

For more information on the quantification of uncertainty in FRS estimates, please see the following slides and the FRS Background Information and Methodology document.

This guide provides the following information:

- Slide 2: List of Variables in the FRS Resamples Datasets.
- Slide 3: How to read the dataset.
- Slide 4: Process of Measuring Uncertainty in FRS Estimates
- Slide 5: Flow Diagram Example of Measuring Uncertainty in FRS Estimates

List of Variables in the FRS Resamples Datasets

Each FRS resamples dataset contains 503 variables which are the following (where only 1 version of “SerNum” is present):

Variable name	**Description**
SerNum	[SAF dataset only] Family Resources Survey (FRS) unique serial number of the household [EUL dataset only] ANONYMISED Family Resources Survey (FRS) unique serial number of the household
BenUnit	Identifier for the benefit unit (family) in the household
Person	Identifier for the person in the household
gs_indpp1, gs_indpp2, …, gs_indpp500	Resample population grossing factors for resamples #1 through to #500

How to Read the FRS Resamples Datasets.

Both

- saf_resamples_dataset_2122.csv
- eul_resamples_dataset_2122.csv

are comma separated values (CSV) files, a universal format for supplying structured information. CSV files can be read by most data processing applications, and can be transformed into SAS and SPSS dataset files among others.

To import these files using SAS for example, run code similar to the following:

```
proc import   datafile="<File Location>"
              dbms=csv
              out=<Output Table Name>;
run;
```

Process of Measuring Uncertainty in FRS Estimates

1. Join any FRS data on FRS main tables for which an estimate is desired with the resamples dataset containing all resample grossing factors.
2. Calculate the relevant FRS estimate many times.
   a) Firstly, calculate the 'Central Estimate' using full sample grossing factors (i.e. Gross4)
   b) Then calculate all 'Resample Estimates' using the resample grossing factors: (i.e. gs_indpp1 to gs_indpp500).
3. For 95% confidence intervals, order the resample estimates by ascending values: e.g. from lowest value (rank=1) to highest value (rank=500).
4. For standard error, calculate the standard deviation of the resample estimates.
5. Identify the 2.5th and 97.5th percentiles value in the ranked resample estimates. Calculate confidence limits with the small sample correction:
   a) Calculate Lower Limit: 2*Central Estimate minus 97.5th percentile.
   b) Calculate Upper Limit: 2*Central Estimate minus 2.5th percentile

Flow Diagram Example of Measuring Uncertainty in FRS Estimates

1. Combine:

FRS Main Dataset

FRS Resamples Dataset

2. Calculate Estimates:

For:

a. Central Estimate

Applying original grossing factor.

Gross4
E.g. 16.36

b. Resample Estimates

Applying resample grossing factors.

gs_indpp1 to gs_indpp500
E.g. …

3. Order resample estimates by ascending values:

1. 14.25
2. 14.56
3. 14.66
4. 14.74
5. 14.75
6. 14.76
7. 14.82
8. 14.88
9. 14.89
10. 14.92
11. 14.93
12. **14.94**
13. **14.95**
14. 14.96
15. 14.98
…
485. 17.48
486. 17.49
487. **17.53**
488. **17.55**
489. 17.66
490. 17.69
491. 17.70
492. 17.79
493. 17.80
494. 17.92
495. 17.94
496. 18.04
497. 18.21
498. 18.22
499. 18.25
500. 18.27

4. For standard error, calculate the standard deviation of the resample estimates:

Standard deviation of resample estimates is standard error on central result
E.g. 1.34

5. Identify values for lower and upper limits:

a. Select the 2.5% value
E.g. 2.5% for 500 resamples is 12.5th value so compute mean of 12th and 13th ranked estimate: **14.95**

b. Select the 97.5% value
E.g. 97.5% for 500 resamples is 97.5th value so compute mean of 487th and 488th ranked estimate: **17.54**

and calculate Confidence Intervals:
With small sample correction

a. Calculate Lower Limit:
2*Central Estimate – 97.5% value: **E.g. 2*16.36 – 17.54 = 15.18**

b. Calculate Upper Limit
2*Central Estimate – 2.5% value: **E.g. 2*16.36 – 14.95 = 17.77**