




Workshop: Assessing Data Quality and Disclosure Risk in Numeric Data

Exercise 1: Identifying issues

In this exercise you will open a data file and evaluate it, manually. Please work in small groups and follow the steps below.

- 1) Open a statistical software of preference SPSS/Stata/PSPP/Gretl
- 2) Open the teaching data file (the Google Drive Teaching Data folder contains 3 versions of the same data **'teaching data%set'** in sav, dat and csv formats)

Name	Date modified	Type	Size
 teaching data%set	18/02/2019 15:05	Microsoft Excel C...	4,052 KB
 teaching data%set	18/02/2019 15:04	DTA File	2,282 KB
 teaching data%set	18/02/2019 13:37	SAV File	2,417 KB

- 3) What issues can you identify in the data?
 - File Issues:
 - Metadata Issues:
 - Data Integrity Issues:
 - Disclosure Control:
- 4) What commands did you use/run to check the data and metadata quality