

**Notes on the formal documentation of empirical analyses
at the Chair of Social Policy (Lehrstuhl für Sozialpolitik)
(Seminarpapers, Bachelor/ Master Thesis)**

1. Introductory remarks

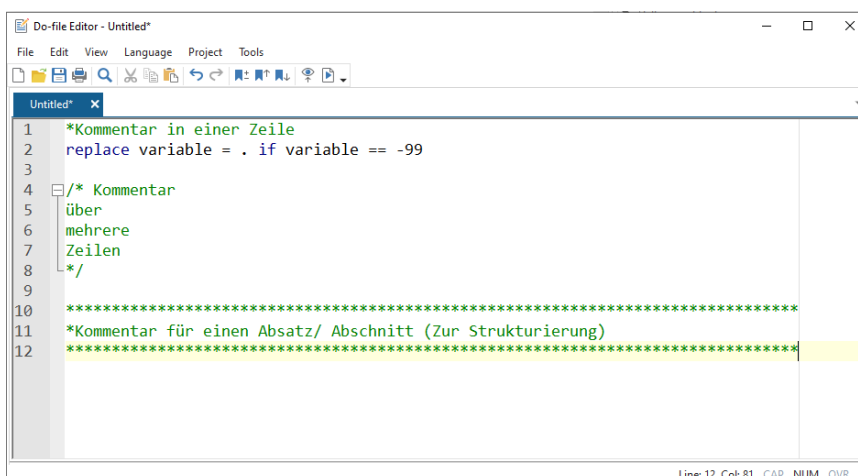
When submitting an empirical thesis, it is imperative to provide the data used and the syntaxes, which contain all processing steps and analyses. All results achieved in the work must be reproducible with the submitted files. If possible, the original data should also be submitted without any processing steps that have already been carried out. In the following, specifications regarding the documentation of empirical work using STATA are provided. Different procedures or the use of other analysis software should be discussed with the supervisor in advance.

2. Workflow with STATA

When working with STATA, all commands are executed via do-files. Avoid using the console for commands. Do-files offer the possibility of documenting, reproducing, structuring and commenting on work steps. For a better overview, do-files are separated according to work steps. As a minimum requirement, it should be distinguished between the preparation do-file and the analysis do-file. If appropriate, further do-files (e.g. for merging) can be created. Do-files must also be robust, i.e. they must run without error messages (also on other computers).

3. Commenting do-files

Do-Files should be self-explanatory so that each individual work step is clearly visible in its function. To ensure this, do-files must be commented on in a way, in which outsiders can understand the underlying function of the commands. As can be seen from the example below, there are multiple ways to comment in STATA (note: green font color indicates a comment, which STATA does not execute as a command):



```
1  *Kommentar in einer Zeile
2  replace variable = . if variable == -99
3
4  /* Kommentar
5  über
6  mehrere
7  Zeilen
8  */
9
10 *****
11 *Kommentar für einen Absatz/ Abschnitt (Zur Strukturierung)
12 *****
```

4. Log-files

All results and comments that STATA displays in the window when running through the do-files can be documented in a text document (log-file). Log-files offer the possibility of viewing results such as tables without STATA. As a result, they are particularly useful if original data cannot be submitted for reasons of size or data protection. Log-files must be prepared separately for each do-file and need to be submitted with the work. Here is a short example of creating a log-file and the respective commands used at the beginning of a do-file:

```

1 *****
2 *Header
3 *****
4 clear all
5 version 16
6 set linesize 120
7 set more off, permanently
8 set seed 11020
9 set matsize 11000
10 set rmsg on
11 *****
12 *Optionale Global-Befehle für das aktuelle Datum
13 global heute: di %td_CV-N-D date("%$DATE", "%dmy")
14
15 *Schließen des log-Files, falls es noch offen ist
16 capture log close Aufbereitung
17
18 quietly {
19     log using "C:\Users\un56ynop\Desktop\Test\Aufbereitung.log", name(Aufbereitung) replace text
20
21     noisily: display "*****"
22     noisily: display "**BA-Arbeit: Analyse von XXX unter Berücksichtigung von ..."
23     noisily: display "**Autor: Sebastian Ach"
24     noisily: display "**Datum: $(heute)"
25     noisily: display "**file-Name: Aufbereitung.do"
26     noisily: display "**Inhalt: Aufbereitung der XXX-Rohdaten für die Analyse"
27     noisily: display "*****"
28 }
29
30 *Einlesen des Datensatzes
31 use "C:\Users\un56ynop\Desktop\Test\Originaldaten\Beispieldaten.dta", clear
32
33 *Aufbereitungsschritt XXX
34 destring beispielvariable, force replace
35
36 *Abspeichern des Datensatzes
37 save "C:\Users\un56ynop\Desktop\Test\Aufbereitet\Beispieldaten_auf.dta", replace
38 log close Aufbereitung
  
```

Please note: The example contains an automated output of the current date. However, it is also possible to enter the date manually and to use simple comments instead of the loop. Conversely, macros can be used for the data paths (recommended but not necessary). The corresponding log-file looks like this:

```

*****
**BA-Arbeit: Analyse von XXX unter Berücksichtigung von ...
**Autor: Sebastian Ach
**datum: 2022-10-05
**file-Name: Aufbereitung.do
**Inhalt: Aufbereitung der XXX-Rohdaten für die Analyse
*****
r; t=0.00 11:46:30

. *Einlesen des Datensatzes
. use "C:\Users\un56ynop\Desktop\Test\Originaldaten\Beispieldaten.dta", clear
r; t=0.00 11:46:30

. *Aufbereitungsschritt XXX
. destring beispielvariable, force replace
beispielvariable: contains nonnumeric characters; replaced as double
(87 missing values generated)
r; t=0.04 11:46:30

. *Abspeichern des Datensatzes
. save "C:\Users\un56ynop\Desktop\Test\Aufbereitet\Beispieldaten_auf.dta", replace
(note: file C:\Users\un56ynop\Desktop\Test\Aufbereitet\Beispieldaten_auf.dta not found)
file C:\Users\un56ynop\Desktop\Test\Aufbereitet\Beispieldaten_auf.dta saved
r; t=0.00 11:46:30

. log close Aufbereitung
name: Aufbereitung
log: C:\Users\un56ynop\Desktop\Test\Aufbereitung.log
log type: text
closed on: 5 Oct 2022, 11:46:30
  
```

5. Submission

All do-files as well as the respective log-files and the original data used for the processing must be submitted. If additional ado-files (user-written STATA codes) are used, these must also be submitted and their function must be described.

Recommendation: Usage of a clear data structure for saving do-files/ original data/ prepared data/ log-files/ tables...

