

Hinweise zur formalen Dokumentation empirischer Analysen am Lehrstuhl für Volkswirtschaftslehre, insb. Sozialpolitik (Seminar-, Bachelor-, Masterarbeiten)

1. Vorbemerkungen

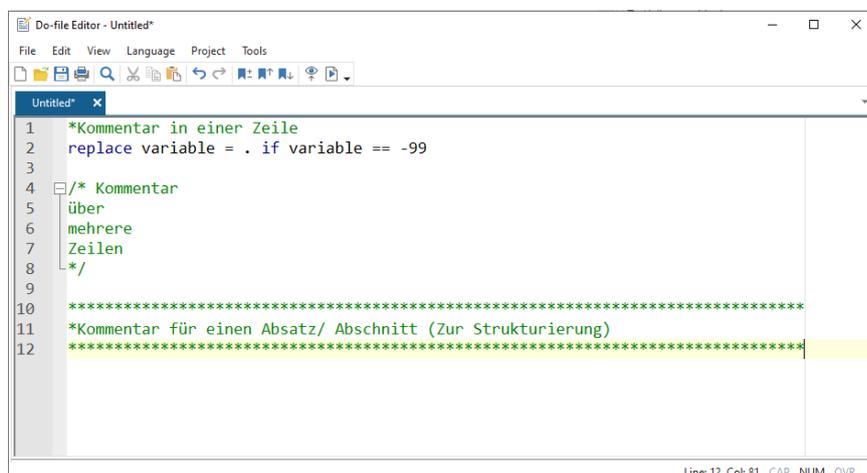
Bei der Abgabe einer empirischen (Abschluss-)Arbeit ist es zwingend erforderlich, die verwendeten Daten, sowie die Syntaxen, welche alle Aufbereitungsschritte und Analysen enthalten, bereitzustellen. Sämtliche in der Arbeit enthaltenen Ergebnisse müssen mit den abgegebenen Dateien reproduzierbar sein. Sofern möglich ist auch der Originaldatensatz ohne bereits erfolgte Aufbereitungsschritte abzugeben. Im Folgenden werden Vorgaben bzgl. der Dokumentation bei empirischen Arbeiten mittels STATA geliefert. Abweichendes Vorgehen bzw. die Verwendung anderer Analysesoftware sollten vorab mit dem bzw. der Betreuer:in besprochen werden.

2. Work-Flow mit STATA

Beim Arbeiten mit STATA werden sämtliche Befehle über Do-Files ausgeführt. Das direkte Nutzen der Konsole für Befehle ist zu vermeiden. Do-Files bieten die Möglichkeit, Arbeitsschritte zu dokumentieren, reproduzieren, strukturieren und zu kommentieren. Für eine bessere Übersicht werden Do-Files nach Arbeitsschritten getrennt. Notwendig ist hierbei die Trennung zwischen Aufbereitungs-Do-File und Analyse-Do-File. Optional können aber noch weitere Do-Files (z.B. fürs Mergen) erstellt werden. Do-Files müssen robust sein, d.h. sie müssen ohne Fehlermeldungen (auch auf anderen Rechnern) durchlaufen.

3. Kommentieren von Do-Files

Do-Files sollten selbsterklärend sein, sodass jeder einzelne Arbeitsschritt in seiner Funktion klar ersichtlich ist. Um dies zu gewährleisten müssen Do-Files so kommentiert werden, dass auch Außenstehende den Befehlen folgen können. In STATA können folgende Formen von Kommentaren verwendet werden (Beachte: Grüne Schriftfarbe bedeutet, dass es sich um einen Kommentar handelt, welchen STATA nicht als Befehl ausführt):

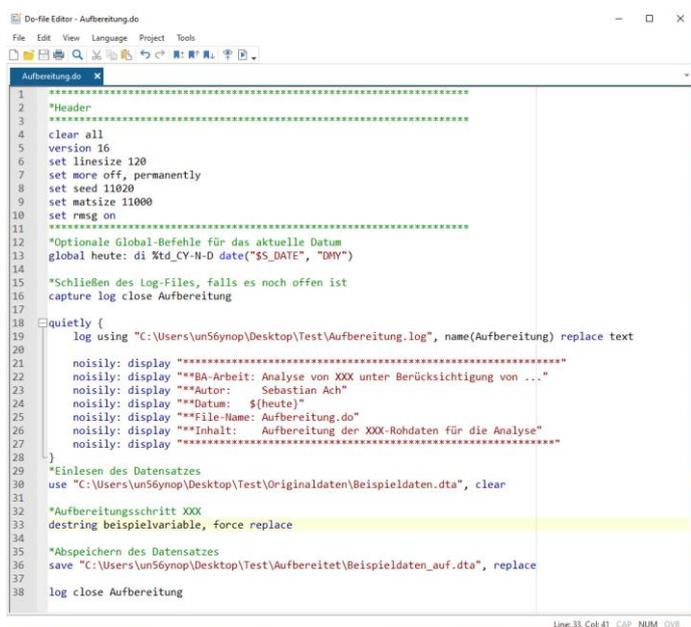


```
1 *Kommentar in einer Zeile
2 replace variable = . if variable == -99
3
4 /* Kommentar
5 über
6 mehrere
7 Zeilen
8 */
9
10 *****
11 *Kommentar für einen Absatz/ Abschnitt (Zur Strukturierung)
12 *****
```

Line: 12, Col: 81 CAP NUM OVR

4. Log-Files

Sämtliche Ergebnisse und Kommentare, welche STATA beim Durchlaufen der Do-Files im Fenster anzeigt, können in einem Textdokument (Log-File) aufgezeichnet werden. Log-Files bieten die Möglichkeit, Ergebnisse wie z.B. Tabellen einzusehen, ohne STATA zu starten und sind insb. nützlich, wenn Originaldatensätze aus Gründen der Größe bzw. des Datenschutzes nicht abgegeben werden können. Aus diesem Grund sind Log-Files für jedes Do-File separat anzufertigen und mit der Arbeit abzugeben. Hier ein kurzes Beispiel für das Erstellen eines Log-Files bzw. die Befehle, welche am Anfang eines Do-Files stehen können:



```

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("%S_DATE", "%mY")
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: %s{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

```

Zu beachten: Das Beispiel enthält eine automatisierte Ausgabe des aktuellen Datums. Es ist jedoch ebenso möglich, das Datum händisch einzutragen und anstelle der Schleife bzw. Display-Befehle einfache Kommentare zu verwenden. Umgekehrt können natürlich auch Makros für die Datenpfade verwendet werden (empfehlenswert aber nicht zwingend notwendig). Das entsprechende Log-File sieht wie folgt aus:

```

Aufbereitung.log - Editor
Datei Bearbeiten Format Ansicht Hilfe
*****
**g: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
. log name: Aufbereitung
. log log: C:\Users\un56ynop\Desktop\Test\Aufbereitung.log
. log type: text
. log closed on: 5 Oct 2022, 11:46:30
*****

```

5. Abgabe

Abzugeben sind sämtliche Do-Files sowie die jeweiligen Log-Files und die Originaldaten, welche für die Aufbereitung genutzt wurden. Falls zusätzliche Ado-Dateien (Nutzergeschrieben STATA-Codes) benutzt werden, sind diese zusätzlich anzugeben und in ihrer Funktion zu beschreiben.

Empfehlung: Nutzung einer klaren Datenstruktur für das Speichern von Do-Files/ Originaldaten/ aufbereiteten Daten/ Log-Files/ Tabellen ...

