



Counterfactual-Analyse

*Masterseminar WT 2020 an der Professur für
Wandel und Nachhaltigkeit*



Vorläufiger Zeitplan

Sitzung Nr.	Datum	Uhrzeit	Thema	Raum
1	14.01.2020	09:00-10:30	Theorie	36 / 1114
2	21.01.2020	09:00-10:30	Praxis	104 / 0117 PC-Pool 4
3	28.01.2020	08:30-11:00	Praxis	104 / 0117 PC-Pool 4
4	03.03.2020	09:00-10:30	Zwischenstand Probleme, Einzeltermine?	36 / 1117
5	31.03.2020	-	Abgabe Seminararbeit	36 / 1117



Exkurs zur Seminararbeit

Aufbau

1. Titel
2. Motivation (Herausarbeiten der Forschungslücke aus der Literatur)
3. Forschungsfragen- und ziele
4. **Methodologie und Methodik**
5. **Ergebnisse**
6. Fazit
7. Referenzen

Methode

Beschreibung des Forschungsprozesses:

- Forschungsdesign (Erläuterung: Kontext Motivation → Forschungsfrage → Beantwortung via empirischer Forschungsphilosophie)
- Techniken zur Datenerhebung, bzw. -beschaffung
- Struktur der Daten, Umfang, deskriptive Statistiken
- Methode zur Datenanalyse (DEA, FDH, Orderalpha)
- Detaillierte Ausführung zu der Anwendung der Methode (Begründung der Variablen, Optionen, Spezifikationen für die gewählte Forschungsfrage)

Einführung STATA – Quantitative Forschung

Man unterscheidet in der empirischen Forschung: (1) qualitative und (2) quantitative Forschungsmethoden

Qualitativ

- Explorativ und hypothesengenerierend
- Schrittweise Theoriebildung
- Ziel: Wirklichkeit anhand der subjektiven Sicht relevanter Gesprächspartner abzubilden.
- Ursachen für Verhalten und das Verhalten verstehen

Quantitativ

- Prüfen vorher festgelegter Hypothesen
- Beobachtbares Verhalten wird beschrieben und vorhersagbar gemacht
- Durch Modelle, Zusammenhänge und zahlenbasierte Ausprägungen
- Vollstandardisiert und strukturiert um Vergleichbarkeit bei erhobenen Messwerten zu erreichen

In der VWL sind beide Ansätze von Nöten, z.B. Makroökonomische Modelle lassen sich quantitativ stützen (Modellierung); In der Mikroökonomie lässt sich das Verhalten der Konsumenten qualitativ erklären (Ernst Fehr – Behavioral Economics)

Einführung STATA – Quantitative Forschung

1. Welche Arten von Daten gibt es?

Nach Art der Erhebung:

- Primärdaten (Erhebung neuer Daten z.B. durch Befragung)
- Sekundärdaten (Bereits vorhanden, z.B. ‚Eurostat‘)

Nach Art der Beschaffenheit:

- Querschnittsdaten (fixer Zeitpunkt, verschiedene Auspr.)
- Längsschnittdaten (Zeitreihen einer fixen Ausprägung)
- Paneldaten (Zeitreihe x Ausprägungen)

2. Typische Fehler in Datensätzen

Selection Bias:

- Verzerrung bei der Auswahl der Merkmalsträger

Sampling Bias:

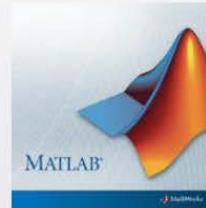
- Nicht-zufällige Auswahl der Merkmalsträger

Measurement Bias:

- Verzerrungen aufgrund fehlerhafter Messungen

Einführung STATA – Quantitative Forschung

Welche Statistiktools eignen sich für quantitative Forschung?



Mächtiger / Open Source

Benutzerfreundliche Bedienung (Programmiergrad)

Eine grundlegende Anwendung der vorgestellten Methode ist mit STATA möglich. Bei komplexeren Vorhaben (z.B. Simulation und Kalibrierung von Modellen) muss auf Matlab umgestiegen werden.

Einführung STATA - Anwendung

Benutzer-
oberfläche:

The screenshot shows the STATA 15.1 user interface. The main window displays the STATA logo and version information. The interface is divided into several panes:

- Review** (left): A pane for reviewing commands, currently empty.
- Command History** (bottom left): A blue callout box indicating the location of the command history pane.
- Result Window** (center): A blue callout box indicating the location of the main output window, which displays the STATA logo and version information.
- Variables** (right): A blue callout box indicating the location of the variables pane, which is currently empty.
- Command Window** (bottom): A blue callout box indicating the location of the command input area.

The main window displays the following text:

```
(R)
-----
Statistics/Data Analysis

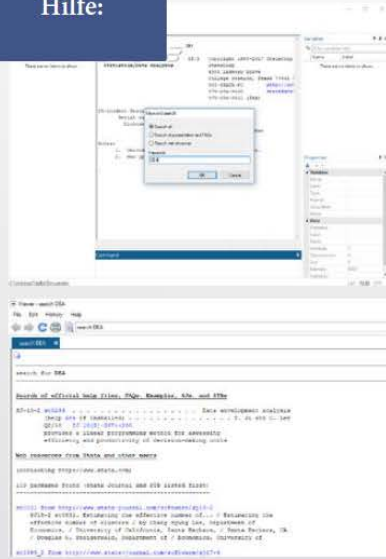
15.1 Copyright 1985-2017 StataCorp
      StataCorp
      4905 Lakeway Drive
      College Station, Texas 77845
      800-STATA-PC http://www
      979-696-4600 stata@stata
      979-696-4601 (fax)

28-student Stata lab perpetual license:
      Serial number: 301506304526
      Licensed to: WGV lab
                  Universität der Bundeswehr München

Notes:
  1. Unicode is supported; see help unicode_advice.
  2. New update available: type -update all-
```

Einführung STATA - Anwendung

Hilfe:



Im Menüreiter ,help‘ können Suchbegriffe, z.B. ein Befehl wie > dea < eingegeben werden



dann werden u.a. ,help files‘, ,packages‘ und ,paper‘ zu dem Suchwort angezeigt.



Zudem finden sich zu den meisten gängigen Problemen Lösungen im ,The Stata Blog‘.



Genauso können Paper im ,The Stata Journal‘ den Umgang mit externen packages erleichtern. (<https://www.stata-journal.com/article.html?article=st0193>)

Kein Tutorial und kein Seminar kann vorbeugen, dass bei der Arbeit mit Statistiksoftware plötzlich ein Fehler auftritt für den es scheinbar keine Lösung gibt. In diesem Fall hilft es, sich per Suchfunktion, Stata-Blog und Journal zu behelfen.

Einführung STATA - Import von Daten

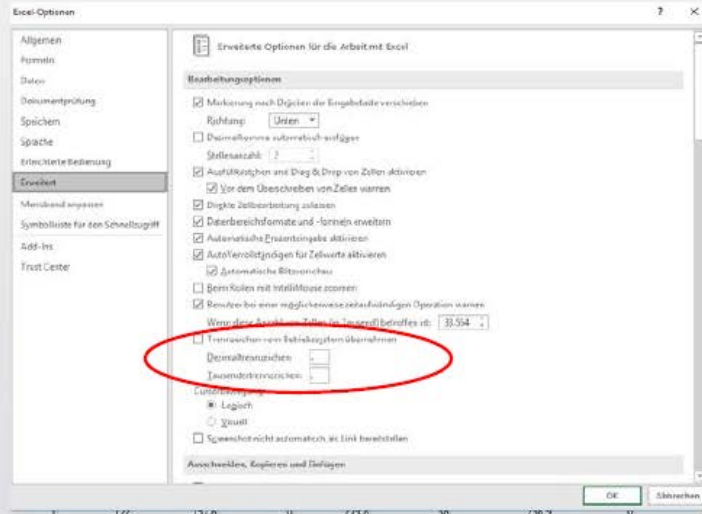
Der Import folgender Dateiformate ist möglich:

1. .dta (Stata Standard)
2. .xls & .xlsx (Excel)
3. .csv
4. .txt

Wenn nicht .dta genutzt wird müssen die Datensätze vorher aufbereitet werden

Bsp: Import von Daten aus Excel-Arbeitsmappe:

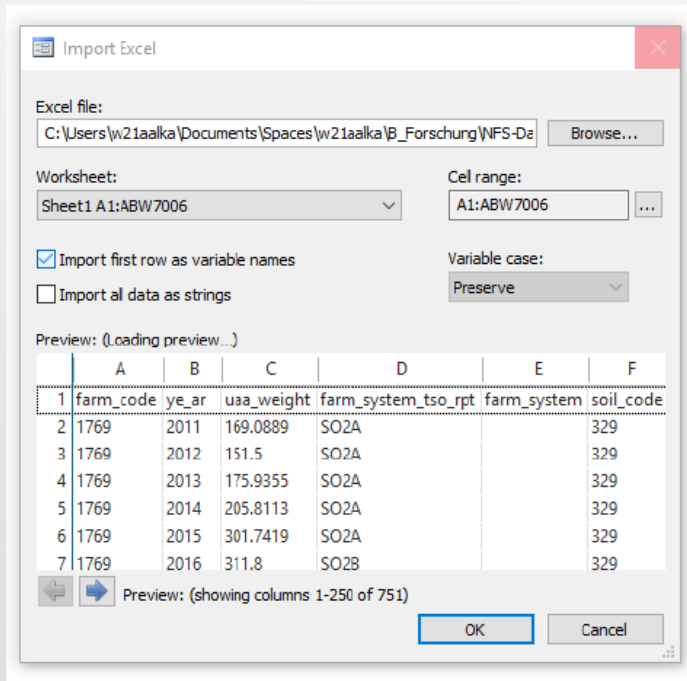
Stata kann nur Zahlen als solche interpretieren, wenn sie mit dem Dezimaltrennzeichen >.< und dem Tausender-trennzeichen >,< gegeben sind.



Datei > Optionen > Erweitert

Einführung STATA - Import von Daten

In STATA: File > Import > Excel Spreadsheet > Browse > Öffnen

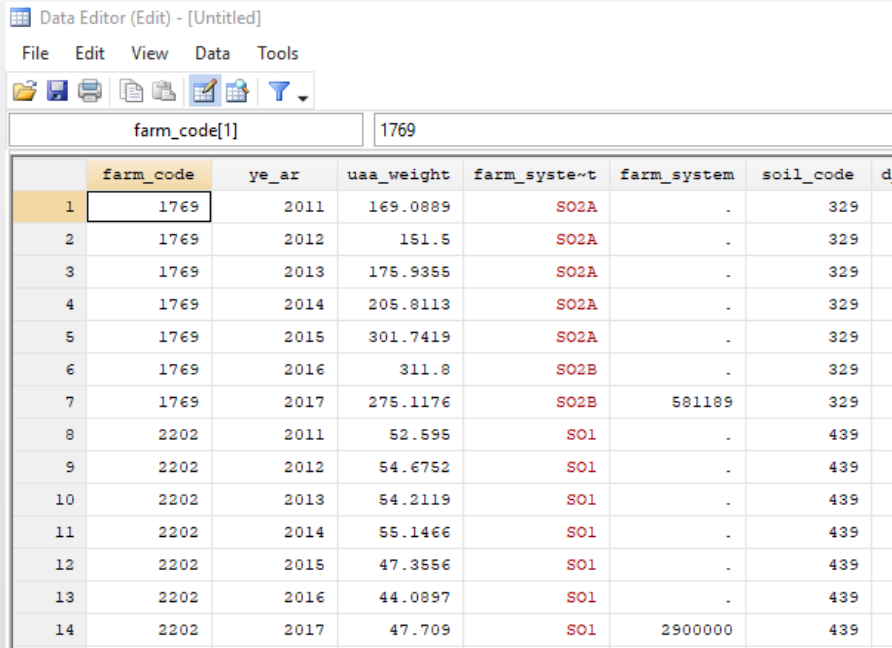


Als 1. macht es Sinn das richtige Worksheet zu bestimmen. Möglicherweise habt ihr in Excel die Daten nach Jahren auf mehrere „Arbeitsblätter“ aufgeteilt.

Wenn ihr euren Variablen bereits im Excel Sheet einen Namen zugewiesen habt, könnt ihr mit **> Import first row as variable names <** etwas Arbeit sparen und habt zudem keine ungewünschten ‚string‘ Variablen (nicht als Zahlen erkannte), welche beim Ausführen eines Befehls eine Fehlermeldung generieren würden

Einführung STATA - Import von Daten

In STATA: Data > Data Editor > Data Editor (Edit)



	farm_code	ye_ar	uaa_weight	farm_syste~t	farm_system	soil_code	d_
1	1769	2011	169.0889	SO2A	.	329	
2	1769	2012	151.5	SO2A	.	329	
3	1769	2013	175.9355	SO2A	.	329	
4	1769	2014	205.8113	SO2A	.	329	
5	1769	2015	301.7419	SO2A	.	329	
6	1769	2016	311.8	SO2B	.	329	
7	1769	2017	275.1176	SO2B	581189	329	
8	2202	2011	52.595	SO1	.	439	
9	2202	2012	54.6752	SO1	.	439	
10	2202	2013	54.2119	SO1	.	439	
11	2202	2014	55.1466	SO1	.	439	
12	2202	2015	47.3556	SO1	.	439	
13	2202	2016	44.0897	SO1	.	439	
14	2202	2017	47.709	SO1	2900000	439	

Im Data Editor könnt ihr eure erfolgreich importierten Daten noch einmal auf Korrektheit überprüfen.

Im Bild links sieht alles korrekt aus.

Die einzigen ‚Strings‘ finden wir unter ‚farm syste~t‘. Das ist auch gut so, denn es handelt sich hier um einen Code aus Buchstaben.

Für die leeren Beobachtungen unter ‚farm_system‘ setzt STATA einen >.< Diese Zellen werden bei Berechnungen, z.B. des Mittelwertes dann auch als leere Zellen behandelt.

Einführung STATA - Erste Befehle

Befehle zum Umgang mit Datenreihen

Alle der folgenden Befehle beziehen sich auf die Variable `,farm_system'`, welche durch jede andere syntaktisch korrekte `,Variablenbezeichnung'` substituiert werden kann

- Löschen einer Datenreihe:
> drop `farm_system` <
- Datenreihe umbenennen:
> rename `farm_system` `betriebsart`
- Datenreihen aufsteigend nach Namen sortieren:
> aorder_all <
- Datenreihen nach einer Datenreihe sortieren
> sort `farm_system`
- Alle Datenreihen des Datensatzes löschen:
> clear < oder > drop_all <
- Einzelne Variablen behalten:
> keep `farm_system`

Befehle zu arithmetischen Manipulationen

Erstellen einer neuen Variable:

- > gen `neu` = <
- Addition > = `alt1` + `alt2`
 - Subtraktion > = `alt1` - `alt2`
 - Multiplikation > = `alt1` * 100
 - Division > = `alt1` / `alt2`
 - Potenzrechnung > = `alt1` ^ 2

Ausserdem:

Mit dem Befehl > sysuse auto.dta < kann ein Datensatz geladen werden, welcher häufig in Foren oder Papern als Beispiel dient



Einführung STATA - Erste Befehle

Aufgabe

1. Löscht euren Datensatz
2. Ladet die ‚auto.dta‘ Daten
3. Wandelt mpg (Miles per Gallon) in Liter je 100 km um
4. ..und generiert die entsprechende Variable
5. Sortiert den Datensatz nach der neuen Variable
6. Welcher Wagen hat die größte Reichweite und wie viele Liter benötigt dieser für 100 km?

Einführung STATA - Deskriptive Statistiken

Deskriptive Statistiken

Instrumente:

- Anzahl der Beobachtungen (Prüfen auf Vollständigkeit)
- Mittelwert, Median (Lage)
- Minimum, Maximum (Spannweite)
- Quantile, Varianz, Standardabweichung zur Bestimmung der Streuung

Diese „einfachen“ deskriptiven Statistiken können für eine Variable mit dem Befehl `> summarize price <` bzw. `> summarize price, detail <` generiert werden (für alle Variablen: einfach Variablenbezeichnung weglassen)

Grafische Analysen

Instrumente:

- Box-Whisker-Plot zur Bestimmung von Lage, Spannweite und Streuung der Beobachtungen
- Histogramm (Symmetrie und Form der Verteilung)
- Scatterplots zur Bestimmung von Zusammenhängen (Korrelationen)
- Liniendiagramme zum Abbilden von Trends bei Zeitreihen

Einführung STATA - Deskriptive Statistiken

Deskriptive Statistiken

Instrumente:

Anzahl der Beobachtungen (Prüfen auf Vollständigkeit)

Mittelwert, Median (Lage)

Minimum, Maximum (Spannweite)


Quantile, Varianz, Standardabweichung zur Bestimmung der Streuung

Diese „einfachen“ deskriptiven Statistiken können für eine Variable mit dem Befehl `> summarize farm_system <` bzw. `> summarize farm_system, detail <` generiert werden (für alle Variablen: einfach Variablenbezeichnung weglassen)

Grafische Analysen

Instrumente:

- Box-Whisker-Plot zur Bestimmung von Lage, Spannweite und Streuung der Beobachtungen
`> graph box price <`
- Histogramm (Symmetrie und Form der Verteilung)
`> histogram price <`
- Scatterplots zur Bestimmung von Zusammenhängen (Korrelationen) `> scatter var_y var_x <`
- Liniendiagramme zum Abbilden von Trends bei Zeitreihen `> graph twoway line var1 var2 <`



Einführung STATA - Counterfactual Schätzverfahren

Aufgabe

- Ladet folgenden Datensatz


use <http://www.stata-press.com/data/r13/cattaneo2>

- Erstellt deskriptive Statistiken zu diesem Datensatz
- Wie sind die Daten beschaffen (normalverteilt, gestreut, etc.)?



Einführung STATA - Counterfactual Schätzverfahren

- Anwendung folgender Schätzer
 - Regression Adjustment (RA)
 - Inverse-Probability-Weighting (IPW)
 - Kombination RA und IPW
 - Matching



Einführung STATA - Counterfactual Schätzverfahren

- Anwendung folgender Schätzer
 - Regression Adjustment (RA)

ATE:

teffects ra (bweight prenatal mmarried mage fbaby) (mbsmoke)

POM:

teffects ra (bweight prenatal mmarried mage fbaby) (mbsmoke), pomeans aequations

Nonnegative Outcomes:

teffects ra (bweight prenatal mmarried mage fbaby, poisson) (mbsmoke)



Einführung STATA - Counterfactual Schätzverfahren

- Anwendung folgender Schätzer
 - Inverse-Probability-Weighting (IPW)

ATE:

`teffects ipw (bweight) (mbsmoke mmarried mage fbaby medu, probit)`



Einführung STATA - Counterfactual Schätzverfahren

- Anwendung folgender Schätzer
 - Kombination RA und IPW (z.B. Augmented Inverse Probability Weighting)

ATE:

```
teffects aipw (bweight prenatal1 mmarried mage fbaby) (mbsmoke mmarried mage  
fbaby medu, probit)
```



Einführung STATA - Counterfactual Schätzverfahren

- Anwendung folgender Schätzer
 - Matching

Nearest-Neighbor-Estimator ATE:

```
teffects nnmatch (bweight mage prenatal1 mmarried fbaby) (mbsmoke)
```

Propensity Score ATE:

```
teffects psmatch (bweight) (mbsmoke mmarried mage fbaby medu)
```



Aufgaben bis zum nächsten Mal

- STATA auf Laptop installieren

<https://td-host.rz.unibw-muenchen.de/primespace/vol01/public/1118/34S0JKZ0003YM005JGYEDBZU9145CJE2/SetupStata16.exe>

→ E-Mail mit Lizenzdatei nach diesem Kurs

- Probedatensatz (Excel) zusammensuchen (z.B. worldbank, eurostat, Penn World Tables) und auf STATA laden (deskriptive Statistiken durchspielen)