

Inhaltsverzeichnis

Vorwort	XI
0 Zu diesem Buch	1
0.1 Aufbau	2
0.2 Material und Hinweise zur Benutzung	3
0.3 Hinweise für Lehrende	5
1 „Das erste Mal“	7
1.1 Aufruf von Stata	7
1.2 Gestalten der Bildschirmansicht	8
1.3 Erste Analysen	9
1.4 Do-Files	25
1.5 Stata verlassen	27
2 Arbeiten mit Do-Files	29
2.1 Von der interaktiven Arbeit zum Do-File	29
2.2 Do-Files sinnvoll gestalten	35
2.2.1 Kommentare	35
2.2.2 Zeilenwechsel	35
2.2.3 Befehle, die in keinem Do-File fehlen sollten	37
2.3 Arbeitsorganisation	40
2.4 Kurzzusammenfassung	45
3 Die Stata-Grammatik	47
3.1 Elemente der Stata-Kommandos	47
3.1.1 Der Befehl	47
3.1.2 Die Variablenliste	48
3.1.2.1 Variablenliste optional oder vorgeschrieben	48
3.1.2.2 Abkürzungen der Variablenliste	49
3.1.2.3 Spezielle Variablenlisten	51
3.1.3 Optionen	51
3.1.4 Die in-Bedingung	53
3.1.5 Die if-Bedingung	54
3.1.6 Ausdrücke	57
3.1.6.1 Operatoren	57
3.1.6.2 Funktionen	59
3.1.7 Die Nummernliste	60

3.1.8	Dateinamen	60
3.2	Wiederholung ähnlicher Befehle	62
3.2.1	Das by-Präfix	62
3.2.2	Die foreach-Schleife	64
3.2.3	Die forvalues-Schleife	67
3.3	Die Gewichtungsanweisung	68
4	Eine allgemeine Bemerkung zu den Statistik-Kommandos	73
5	Erstellen und Verändern von Variablen	77
5.1	Die Befehle generate und replace	77
5.1.1	Variablennamen	78
5.1.2	Einige Beispiele	79
5.1.3	Rekodieren mit by, _n und _N	82
5.1.4	Explizite Subscripte	86
5.2	Spezielle Rekodierungs-Befehle	88
5.2.1	recode	88
5.2.2	egen	89
5.3	Weitere Werkzeuge zum Rekodieren von Daten	90
5.3.1	String-Funktionen	91
5.3.2	Datums- und Zeitfunktionen	95
5.3.2.1	Datumsangaben	95
5.3.2.2	Zeit	98
5.4	Befehle zum Umgang mit Missings	101
5.5	Beschriftung von Variablen	102
5.6	Storage-Types oder: der Geist in der Maschine	105
6	Erstellen und Verändern von Grafiken	107
6.1	Eine Vorbemerkung zur Syntax	107
6.2	Typen von Grafiken	109
6.2.1	Beispiele	109
6.2.2	Spezielle Grafiken	109
6.3	Elemente der Grafiken	113
6.3.1	Erscheinungsbild der Daten	113
6.3.1.1	Auswahl der Marker	115
6.3.1.2	Farbe der Marker	117
6.3.1.3	Größe der Marker	117
6.3.1.4	Linien	118
6.3.2	Grafik- und Plotregion	121
6.3.2.1	Größe der Grafik	121
6.3.2.2	Plotregion	122
6.3.2.3	Skalierung der Achsen	122
6.3.3	Informationen innerhalb der Plotregion	124
6.3.3.1	Referenzlinien	125
6.3.3.2	Beschriftungen innerhalb der Plotregion	125

6.3.4	Informationen außerhalb der Plotregion	129
6.3.4.1	Beschriftung der Achsen	130
6.3.4.2	Tick-Lines	132
6.3.4.3	Achsentitel	133
6.3.4.4	Die Legende	135
6.3.4.5	Grafik-Titel	136
6.4	Multiple Grafiken	137
6.4.1	Überlagerung mehrerer twoway-Grafiken	137
6.4.2	Befehlsoption byC)	138
6.4.3	Zusammenführung von Grafiken	139
6.5	Speichern und Drucken von Grafiken	141
7	Die Beschreibung von Verteilungen	145
7.1	Wenig oder viele Ausprägungen?	146
7.2	Variablen mit wenig Ausprägungen	147
7.2.1	Tabellarische Darstellungen	147
7.2.1.1	Häufigkeitstabellen	147
7.2.1.2	Mehr als eine Häufigkeitstabelle	148
7.2.1.3	Vergleich von Verteilungen	149
7.2.1.4	Zusammenfassende Maßzahlen	151
7.2.1.5	Mehr als eine Kontingenztabelle	151
7.2.2	Grafische Verfahren	152
7.2.2.1	Histogramme	152
7.2.2.2	Balkendiagramme	154
7.2.2.3	Kuchendiagramme	156
7.2.2.4	Dot-Chart	157
7.3	Variablen mit vielen Ausprägungen	158
7.3.1	Häufigkeitsverteilung gruppierter Daten	159
7.3.2	Beschreibung durch Maßzahlen	162
7.3.2.1	Wichtige Maßzahlen	162
7.3.2.2	sumraarize	164
7.3.2.3	tabstat	165
7.3.2.4	Vergleich von Verteilungen mit Maßzahlen	165
7.3.3	Grafische Verfahren	171
7.3.3.1	Box-Plots	171
7.3.3.2	Histogramme	172
7.3.3.3	Kern-Dichte-Schätzer	174
7.3.3.4	Quantil-Plot	179
7.4	Kurzzusammenfassung	183
8	Einführung in die Regressionstechnik	185
8.1	Lineare Einfachregression	188
8.1.1	Das Grundprinzip	188
8.1.2	Lineare Regression mit Stata	192
8.1.2.1	Der Koeffizientenblock	193

	8.1.2.2	Standardfehler	195
	8.1.2.3	Der Anova-Block	197
	8.1.2.4	Der Modellfit-Block	199
8.2		Die multiple Regression	201
	8.2.1	Multiple lineare Regression mit Stata	202
	8.2.2	Spezielle Kennzahlen der multiplen Regression	204
	8.2.3	Was bedeutet eigentlich „unter Kontrolle“?	207
8.3		Regressions-Diagnostik	208
	8.3.1	Die Verletzung von $E\{e_i\} = 0$	209
	8.3.1.1	Linearität	212
	8.3.1.2	Einflussreiche Beobachtungen	215
	8.3.1.3	Übersehene Einflussfaktoren	224
	8.3.2	Die Verletzung von $\text{VAR}(e_i) = \sigma^2$	225
	8.3.3	Die Verletzung von $\text{COV}(e_i, e_j) = 0; i \neq j$	227
8.4		Verfeinerte Modelle	228
	8.4.1	Kategoriale unabhängige Variablen	228
	8.4.2	Interaktionseffekte	231
	8.4.3	Regressionsmodelle mit transformierten Daten	235
	8.4.3.1	Modellierung nichtlinearer Zusammenhänge	235
	8.4.3.2	Transformation zur Beseitigung von Heteroskedastizität	238
8.5		Mehr zu Standardfehlern	239
	8.5.0.3	Bootstrap-Techniken	239
	8.5.0.4	Konfidenzintervalle in Klumpenstichproben	241
8.6		Weiterführende Verfahren	243
	8.6.1	Median-Regression	243
	8.6.2	Regressionsmodelle für Paneldaten	245
	8.6.2.1	Die Stata-Diät: Aus breit wird lang	245
	8.6.2.2	Fixed-Effects-Modell	249
	8.6.2.3	Fehlerkomponenten-Modelle	253
8.7		Zusammenfassung	256
9		Regressionsmodelle für kategoriale abhängige Variablen	257
	9.1	Das lineare Wahrscheinlichkeitsmodell	258
	9.2	Grundkonzepte	262
	9.2.1	Odds, Log-Odds und Odds-Ratios	262
	¹ 9.2.2	Exkurs: Das Maximum-Likelihood-Prinzip	267
	9.3	Logistische Regression mit Stata	271
	9.3.1	Der Koeffizientenblock	273
	9.3.1.1	Vorzeicheninterpretation	274
	9.3.1.2	Interpretation mit Odds-Ratios	274
	9.3.1.3	Wahrscheinlichkeitsinterpretation	275
	9.3.2	Der Iterationsblock	277
	9.3.3	Der Modellfit-Block	278
	9.3.3.1	Klassifikationstabellen	279

9.3.3.2	Pearson-Chi-Quadrat	281
9.4	Diagnostik der logistischen Regression	283
9.4.1	Linearität	283
9.4.2	Einflussreiche Fälle	287
9.5	Likelihood-Ratio-Test	291
9.6	Verfeinerte Modelle	293
9.7	Weiterführende Verfahren	297
9.7.1	Probit-Modelle	297
9.7.2	Multinomiale logistische Regression	300
9.7.3	Ordinale Logit-Modelle	304
9.8	Kurzzusammenfassung	307
10	Daten lesen und schreiben	309
10.1	Das Ziel: Die Datenmatrix	309
10.2	Import maschinenlesbarer Daten	311
10.2.1	Einlesen von System-Files anderer Programme	312
10.2.2	Einlesen von ASCII- bzw. Textdateien	312
10.2.2.1	Einlesen von Daten im Spreadsheet-Format	313
10.2.2.2	Einlesen von Daten im freien Format	315
10.2.2.3	Einlesen von Daten im festen Format	317
10.3	Dateneingabe	320
10.3.1	Dateneingabe über den Editor	320
10.3.2	Der input-Befehl	322
10.4	Zusammenführung von Datensätzen	325
10.4.1	Die Datenstruktur des GSOEP	326
10.4.2	Der Befehl merge	328
10.4.2.1	Grundlegendes zur merge-Prozedur	329
10.4.2.2	Kontrolle der Beobachtungen	332
10.4.2.3	Zusammenführen von mehr als zwei Dateien	333
10.4.2.4	Datenbankspezifische merge-Werkzeuge	334
10.4.2.5	Zusammenführen hierarchischer Daten	335
10.4.3	Der Befehl append	338
10.5	Datensätze speichern und exportieren	340
10.6	Große Datensätze, große Probleme	342
10.6.1	Regeln zum Umgang mit dem Arbeitsspeicher	342
10.6.2	Die Verwendung zu großer Datensätze	344
10.7	Kurzzusammenfassung	345
11	Do-Files für Fortgeschrittene und eigene Programme	347
11.1	Zwei Anwendungsbeispiele	347
11.2	Vier Programmierwerkzeuge	349
11.2.1	Makros	349
11.2.2	Do-Files	353
11.2.3	Programme	353
11.2.4	Ado-Files	356

11.3	Selbst programmierte Stata-Befehle	360
11.3.1	Weitergabe von Variablenlisten	363
11.3.2	Weitergabe von Optionen	365
11.3.3	Weitergabe von if und in	366
11.3.4	Bilden von Variablen unbekannter Anzahl	367
11.3.5	Voreinstellungen	370
11.3.6	Erweiterte Makrofunktionen	372
11.3.7	Veränderungen am Datensatz vermeiden	373
11.3.8	Help-Files	374
11.4	Kurzzusammenfassung	376
12	Rund um Stata	377
12.1	Ressourcen mit Informationen	377
12.2	Pflege von Stata	37g
12.3	Zusätzliche Prozeduren	3§Q
12.3.1	SJ- und STB-Ados	330
12.3.2	SSC-Ados	382
12.3.3	Andere Ados	382
12.4	Bezugsquellen	334
12.5	Kurzzusammenfassung	385
	Literaturverzeichnis	337
	Index	391