

Datenauswertung

Software

Die Datenauswertung besteht stets aus vielen Einzelschritten. Im Rahmen der Beispielstudie wurde die Datenauswertung mithilfe des Statistikprogramms SPSS[®] durchgeführt. Natürlich wären auch alle anderen kommerziellen Statistikprogramme wie zum Beispiel Stata[®] oder das Open-Source-Statistikprogramm R (das aber keine Menüsteuerung besitzt) zu diesem Zweck geeignet gewesen. Prinzipiell wäre es sogar möglich, seine Daten mithilfe eines Tabellenkalkulationsprogramms auszuwerten (vgl. Duller 2007). Die Durchführung der statistischen Datenanalyse setzt neben grundlegenden Programmkenntnissen auch ein Mindestmaß an statistischer Vorbildung voraus (vgl. Raithel 2008).

Einlesen von Daten

In Kapitel 8.3 des Lehrbuchs wurden als Möglichkeiten der Datenerfassung (1) das Codieren mit Erfassungsbögen aus Papier (mit anschließendem Abtippen der Daten) und (2) das Codieren anhand einer elektronischen Eingabemaske vorgestellt. Überträgt man beide Erfassungsmöglichkeiten auf die Beispielstudie, sind jeweils unterschiedliche Teilschritte durchzuführen, um zu einer sogenannten **Rohdatenmatrix** zu gelangen, die sämtliche Codierentscheidungen quantitativ abbildet:

1. Codieren anhand von Erfassungsbögen aus Papier → Abtippen der Erfassungsbögen → Rohdatenmatrix
2. Erfassen der Daten mithilfe einer elektronischen Eingabemaske → Rohdatenmatrix

Fehlerkontrolle

Die Rohdatenmatrix bildet den datentechnischen Ausgangspunkt jeder statistischen Datenanalyse, wobei die Zeilen der Rohdatenmatrix die Untersuchungseinheiten und die Spalten die Variablen des Codeplans repräsentieren. Allerdings ist auch bei größter Sorgfalt während der Feldphase damit zu rechnen, dass die Rohdatenmatrix fehlerhaft ist. Das bedeutet,

dass eine eingelesene Rohdatenmatrix nicht umgehend ausgewertet werden sollte, sondern vor der eigentlichen Datenanalyse eine Fehlerkontrolle durchgeführt werden muss.

Die Fehlerkontrolle einer Rohdatenmatrix beginnt üblicherweise damit, dass die Rohdaten mithilfe eines Statistikprogramms erst einmal komplett durchnummeriert werden. Dadurch ist es möglich, die Rohdaten zu jeder Zeit in ihre ursprüngliche Reihenfolge zu bringen. In unserem Fall wurde eine zusätzliche Variable im Datensatz mit dem Namen „id“ erzeugt (vgl. Abb. 6). Es empfiehlt sich nach erfolgter Durchnummerierung der Rohdatenmatrix eine Sicherungskopie anzulegen und diese separat abzuspeichern. Auf diese Weise bleibt die ursprüngliche Rohdatenmatrix auf jeden Fall erhalten.

Abbildung 6: Auszug aus der SPSS-Rohdatenmatrix der Beispielstudie

	id	t0	t1	t2	t3a	t3b	t3c	t4	p1	p2	p3
1	1	1	81	9	15	6	7	1	Eine Brücke zum Urknall	621	1
2	2	1	82	9	15	6	7	1	Die Max-Planck-Gesellschaft trif	623	1
3	3	2	82	9	15	6	7	1		.	.
4	4	2	82	9	15	6	7	1		.	.
5	5	2	82	9	15	6	7	1		.	.
6	6	1	83	9	15	6	7	1	Der Mittelmeerraum und Europa Ku	328	1
7	7	2	83	9	15	6	7	1		.	.
8	8	2	83	9	15	6	7	1		.	.

Die eigentliche Fehlerkontrolle beginnt dann mit einer **univariaten Häufigkeitsauszählung** aller numerischen Variablen in der Rohdatenmatrix. Mit dem Begriff „univariat“ ist gemeint, dass die Häufigkeitswerte je Variable einzeln ausgewiesen werden. Diesen Schritt der Fehlerkontrolle bezeichnet man auch als **Grundauszählung**.

Abbildung 7: Auszug aus der Grundauszählung der Variable Z5

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Titelseite	5	,5	,7	,7
	2 Politik	67	6,4	10,0	10,7
	3 Wirtschaft	41	3,9	6,1	16,9
	4 Unternehmen	1	,1	,1	17,0
	5 Lokales	104	10,0	15,5	32,5
	6 Kultur (Feuilleton)	104	10,0	15,5	48,1
	7 Wissenschaft	221	21,2	33,0	81,0
	8 Seite 3	3	,3	,4	81,5
	9 Medien	5	,5	,7	82,2
	10 Vermischtes/Aus aller Welt	24	2,3	3,6	85,8
	11 Anderes Ressort (u.a. Leserbriefe)	46	4,4	6,9	92,7
	12 Campus/Hochschule	26	2,5	3,9	96,6
	13 Sonderbeilage	23	2,2	3,4	100,0
	Total	670	64,2	100,0	
Missing	System	373	35,8		
Total		1043	100,0		

Abbildung 7 zeigt, wie häufig jede Variablenausprägung pro Zeitungsartikel codiert wurde. Würde sich zum Beispiel bei diesem Kontrollschritt zeigen, dass irrtümlicherweise auch der Code 14 vergeben wurde, obwohl dieser Code gar nicht im Codeplan vorgesehen ist, müsste der Forscher herausfinden, welcher Codierer diesen Zeitungsartikel codiert hat und diesen bitten, die Variable Z5 erneut zu codieren. Die Grundauszählung dient somit einerseits der Kontrolle der Variablenausprägungen (Sind nur gültige Werte vorhanden?) und andererseits der Kontrolle, ob das Material komplett codiert wurde (Stimmt die Fallzahl?).

Der zweite Schritt der Fehlerkontrolle besteht typischerweise darin, sämtliche Filterregeln im Datensatz nachzuvollziehen und zu prüfen, ob die Filterregeln von den Codierern eingehalten wurden. Eine Filterregel im Codeplan der Beispielstudie lautet beispielsweise: Wenn sich die Codierer bei Variable T4 für den Wert 2 entscheiden, sollen sie die Codierung der Variablen für die Pressemitteilungen (P1 bis P3) überspringen und bei Variable Z1 weiter codieren (vgl. Codeplan im Anhang des Lehrbuchs). Außerdem empfiehlt es sich, die Rohdatenmatrix auf sachlogische Codierfehler zu überprüfen. Sachlogische Codierfehler können dann auftreten, wenn die Codierungen verschiedener Variablen einer Untersu-

chungseinheit nicht komplett unabhängig sind. Ein Beispiel: Bei der Codierung der 670 Zeitungsartikel haben die Codierer jeden Artikel einer Zeitungsrubrik bzw. einem Zeitungsressort (Variable Z5) sowie einem Genre (Variable Z6) zuordnen müssen. Die Gegenüberstellung dieser beiden Variablen mittels einer Kreuztabelle macht zum Beispiel eine Plausibilitätsprüfung möglich.

Die Ergebnisse der Grundauszählung und der Fehlerkontrolle werden allerdings nie publiziert, sondern dienen dem Forscher einzig und allein der **Plausibilitätskontrolle**.

Datenmanagement

Da man als Forscher schon allein aus Platzgründen nicht alle empirischen Ergebnisse dokumentieren kann, ist es notwendig, bei der Datenanalyse komplexitätsreduzierend zu arbeiten. Damit ist zum Beispiel gemeint, dass in Tabellen oder Abbildungen meistens nicht sämtliche beobachteten Variablenausprägungen ausgewiesen werden. Stattdessen werden selten beobachtete Variablenausprägungen zu einer Restkategorie (z.B. „Sonstiges“) zusammengefasst (vgl. Abb. 9 und Abb. 10). Das nachträgliche Zusammenfassen von Variablenausprägungen bezeichnet man auch als **Recodierung**.

Abbildung 8: Recodierung der Rubrikenvariable (Anzahl)

		ZF5R Rubrik/Ressort recodiert						Total
		1,00 Wissenschaft	2,00 Lokales	3,00 Kultur (Feuilleton)	4,00 Politik	5,00 Wirtschaft	6,00 Sonstiges	
ZF5 Rubrik/Ressort	1 Titelseite	0	0	0	0	0	5	5
	2 Politik	0	0	0	67	0	0	67
	3 Wirtschaft	0	0	0	0	41	0	41
	4 Unternehmen	0	0	0	0	1	0	1
	5 Lokales	0	104	0	0	0	0	104
	6 Kultur (Feuilleton)	0	0	104	0	0	0	104
	7 Wissenschaft	221	0	0	0	0	0	221
	8 Seite 3	0	0	0	0	0	3	3
	9 Medien	0	0	0	0	0	5	5
	10 Vermischtes/Aus aller Welt	0	0	0	0	0	24	24
	11 Anderes Ressort (u.a. Leserbriefe)	0	0	0	0	0	46	46
	12 Campus/Hochschule	0	0	0	0	0	26	26
	13 Sonderbeilage	0	0	0	0	0	23	23
Total		221	104	104	67	42	132	670

Abbildung 8 zeigt, dass die Variablenausprägungen der 670 Zeitungsartikel recodiert wurden: Die ursprünglichen Ausprägungen für die „Titelseite“, „Seite 3“, „Medien“, „Vermischtes“, „Andere Ressorts“, „Campus/

Hochschule“ und „Sonderbeilage“ wurden der Restkategorie „Sonstiges“ zugerechnet.

Der Vorteil dieser komplexitätsreduzierenden Vorgehensweise entfaltet sich jedoch erst im Rahmen weitergehender Analysen, wie der Vergleich von Abbildung 9 und 10 zeigt: Die Ergebnisse in Abbildung 10 sind infolge der Zusammenfassung von selten vorkommenden Ausprägungen sehr viel leichter zu interpretieren als die Ergebnisse in Abbildung 9.

Recodierungsentscheidungen trifft ein Forscher meistens erst nach Beendigung der Feldphase. Aber bereits bei der Entwicklung des Untersuchungsinstruments und den damit vielfach verbundenen Entscheidungen, wieviele und welche Ausprägungen die einzelnen Variablen aufweisen sollen, lohnt es sich, das künftige Aussehen der Ergebnistabellen und Abbildungen im Blick zu haben.

Abbildung 9: Rubrikenverteilungen der MPG-Artikel in Welt, FAZ, FR und SZ auf Basis der Originalvariable (Prozent)

		ZF3 Titel der Zeitung				Total
		1 FAZ	2 SZ	3 WELT	4 FR	
ZF5 Rubrik/Ressort	1 Titelseite			2,8%		,7%
	2 Politik	9,6%	15,1%	10,1%	4,4%	10,0%
	3 Wirtschaft	8,6%	5,0%	6,2%	3,7%	6,1%
	4 Unternehmen	,5%				,1%
	5 Lokales	11,1%	6,3%	19,7%	27,4%	15,5%
	6 Kultur (Feuilleton)	32,3%	14,5%	5,6%	5,2%	15,5%
	7 Wissenschaft	20,7%	33,3%	47,8%	31,1%	33,0%
	8 Seite 3	,5%	,6%	,6%		,4%
	9 Medien		1,9%		1,5%	,7%
	10 Vermischtes/Aus aller Welt	4,5%	5,7%	,6%	3,7%	3,6%
	11 Anderes Ressort (u.a. Leserbriefe)	6,6%	8,8%	2,8%	10,4%	6,9%
	12 Campus/Hochschule	5,1%	5,0%		5,9%	3,9%
	13 Sonderbeilage	,5%	3,8%	3,9%	6,7%	3,4%
Total		100,0%	100,0%	100,0%	100,0%	100,0%

Abbildung 10: Rubrikenverteilungen der MPG-Artikel in Welt, FAZ, FR und SZ auf Basis der recodierten Variable (Prozent)

		zf3 Titel der Zeitung				Gesamt
		1 FAZ	2 SZ	3 WELT	4 FR	
zf5r Rubrik/Ressort recodiert	1 Wissenschaft	20,7%	33,3%	47,8%	31,1%	33,0%
	2 Lokales	11,1%	6,3%	19,7%	27,4%	15,5%
	3 Kultur (Feuilleton)	32,3%	14,5%	5,6%	5,2%	15,5%
	4 Politik	9,6%	15,1%	10,1%	4,4%	10,0%
	5 Wirtschaft	9,1%	5,0%	6,2%	3,7%	6,3%
	6 Sonstiges	17,2%	25,8%	10,7%	28,1%	19,7%
Gesamt		100,0%	100,0%	100,0%	100,0%	100,0%

Literatur

- Duller, Christine (2007): Einführung in die Statistik mit EXCEL und SPSS: Ein anwendungsorientiertes Lehr- und Arbeitsbuch. Heidelberg: Physica-Verlag.
- Raithel, Jürgen (2008): Quantitative Forschung. Ein Praxiskurs. Wiesbaden: VS Verlag.

Medienresonanzanalyse

Eine Einführung in Theorie und Praxis

Raupp, J.; Vogelsang M.A., J.

2009, 210 S. 17 Abb. Mit Online-Extras., Softcover

ISBN: 978-3-531-16000-9