

LUDWIG-BOLTZMANN-INSTITUT FÜR HISTORISCHE SOZIALWISSENSCHAFT
MATERIALIEN ZUR HISTORISCHEN SOZIALWISSENSCHAFT

1

MANFRED THALLER

Numerische Datenverarbeitung für Historiker

EINE PRAXISORIENTIERTE EINFÜHRUNG
IN DIE QUANTITATIVE ARBEITSMETHODE UND IN SPSS
(STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES)



1982

HERMANN BÖHLAUS NACHF. · WIEN · KÖLN

Inhaltsverzeichnis

Vorwort		1
1.	Einleitung: "Die Quantifizierung" und "der Computer"	5
1.1.	Quantifizierung und quantitative Quelle	5
1.2.	Quantifizierung und abstrakte Sachverhalte	6
1.3.	Quantifizierung als Methode - "Wahrscheinlichkeiten"	7
1.4.	Die Analyse von Zusammenhängen	8
1.5.	Computer und Quantifizierung	11
1.6.	Voraussetzungen des Computereinsatzes	12
1.7.	Programmpakete - der einfachste Zugang zum Rechner	13
1.8.	Ein Beispiel: SPSS	14
2.	Daten und Programme	17
2.1.	Ein erstes Programm	19
2.2.	Die zur Verfügung stehenden Daten	21
2.3.	Die Systemfilebeschreibung	22
2.4.	Der Berufskode	24
2.5.	Der Ortskode	27
2.6.	Die Grobgliederung der Berufe	30
3.	Einfache statistische Auswertungen	31
3.1.	Linearauszahlungen	31
3.2.	Das Meßniveau einer Variablen	31
3.3.	Befehle zur Berechnung deskriptiver Statistiken	34
3.3.1.	Der CONDESCRIPTIVE Befehl	35
3.3.1.1.	Die wichtigsten statistischen Maßzahlen	36
3.3.1.2.	Verteilungen und ihre Parameter	37
3.3.2.	Der FREQUENCIES Befehl	39
3.3.3.	Befehlserweiterungen: OPTIONS und STATISTICS	42
3.3.4.	Weitere Parameter von Verteilungen	43
3.4.	Zusammenhänge zwischen zwei Variablen	44
3.4.1.	Zusammenhänge zwischen einer diskreten und einer kontinuierlichen Variablen	45
3.4.1.1.	Der Schein trügt: die Drittvariablenkontrolle	47
3.4.1.2.	Wider die Papierverschwendung - oder: erst denken, dann rechnen.	49
3.4.2.	Zusammenhänge zwischen zwei diskreten Variablen	50
3.4.2.1.	Lesen und Interpretieren von Tabellen	51
3.4.3.	Der Zusammenhang zweier kontinuierlicher Variabler: die Visualisierung	54
3.4.4.	Zusammenhänge zwischen kontinuierlichen Variablen: die statistische Berechnung	55
3.4.4.1.	Die Bedeutung des Korrelationskoeffizienten	58
3.4.4.2.	Die Signifikanz von Maßzahlen	59
3.5.	Wo bleibt die schließende Statistik?	60
4.	Die Umwandlung von Eingabedaten in eine Systemdatei	61
4.1.	Die Systemfilegenerierung	62
4.1.1.	Wo stehen die Daten? - Das FORTRAN Format	64
4.1.2.	Die Übereinstimmung von Format und Variablenliste	68
4.2.	Die physikalische Lesbarmachung der Daten	68
4.2.1.	Weitere <u>nötige</u> Befehle zur Datenbeschreibung	70
4.2.2.	<u>Empfehlenswerte</u> Befehle zur Datenbeschreibung	72

5.	Veränderung von Variablen und Fileumfang	78
5.1.	Die gezielte Änderung von Einzelwerten	79
5.2.	Die Umrechnung einer Variablen: COMPUTE	82
5.3.	Datenmodifikationen als Hilfsmittel bei der Eingabe schwer meßbarer Variabler	85
5.3.1.	Vorüberlegungen	85
5.3.2.	Die praktische Durchführung	88
5.4.	Bedingte systematische Umformungen - IF	90
5.4.1.	Der IF Befehl als Eingabehilfe - ein Beispiel	92
5.4.2.	Gefahren bei der Konstruktion von IF's	94
5.4.3.	Abkürzungsregeln für IF's	95
5.5.	Der Zugriff auf Teile einer Systemdatei: SELECT IF	96
5.5.1.	Typologien mit IF und SELECT IF	98
5.6.	<u>Permanente</u> und <u>temporäre</u> Modifikationen	101
5.7.	Die Verwendung von SUBFILE Strukturen	103
6.	Stichproben und Gesamterhebungen	107
6.1.	Motive des Computereinsatzes	107
6.1.1.	Wieviele Daten brauchen wir?	109
6.1.2.	Stichproben verfälschen Ergebnisse <u>nicht</u> - ein Beispiel	111
6.2.	Der Nutzen von Pilotstudien	119
7.	Die Kodierung von Quellenmaterial	122
7.1.	Notwendigkeit der Kodierung	123
7.2.	Ablauf und Organisation des Kodierens.	126
8.	"Data Clearing" - die Suche nach Datenfehlern	133
8.1.	Eine Prüfliste für die Fehlersuche	134
8.2.	Fehlerquellen	139
8.2.1.	Das Validitätsproblem	139
8.2.2.	Das Reliabilitätsproblem	142
8.2.3.	Die technischen Fehler	144
9.	Übungsbeispiele	146
10.	Lösungen zu den Übungsbeispielen	151
11.	Literaturvorschläge	155
11.1.	Lehrbücher	155
11.2.	Sammelbände und Kongreßpublikationen	157
11.3.	Überblicke über Datenverarbeitung und Geschichtswissenschaft	158
11.4.	Exemplarische Einzelstudien	159
11.5.	Zeitschriften und Newsletter	161
11.6.	Kommentare zu Einzelproblemen	162
11.6.1.	Kodierungsprobleme	162
11.6.2.	Probleme der Arbeitsorganisation	162
11.6.3.	Stichproben und ihre Tücken	163
11.6.4.	Datenfehler	163
11.6.5.	Nichtstatistische Ansätze in der Datenverarbeitung	164
	Tabellenverzeichnis	165