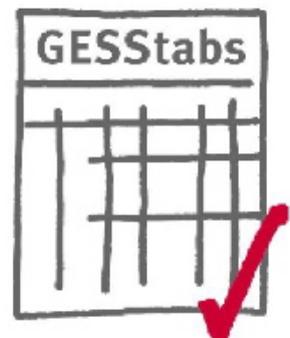


GESStabs

Einführung in die Tabellierung



Gesellschaft für Software
in der Sozialforschung mbH

Waterloohain 6 – 8
22769 Hamburg
Tel.: 040 - 853 753 - 0
Fax: 040 - 853 753 - 33
www.gessgroup.de

GESStabs

Einführung in die Tabellierung

Andreas Rechenberg

Seline Günther

Judith Krüger

Stand: April 2018

GESS Gesellschaft für Software in der Sozialforschung mbH

Waterloohain 6-8

22769 Hamburg

Tel. 040 - 853 753 0

Fax 040 - 853 753 433

software@gessgroup.de

www.gessgroup.de

Inhaltsverzeichnis

1	Einführung in die Tabellierung	1
2	Einlesen von Datensätzen	4
2.1	ASCII	4
2.1.1	Manuelles Erstellen einer var.inc	4
2.1.2	Automatisch generierte var.inc	5
2.2	SPSS	7
2.3	CSV	9
3	Einfache Kreuztabelle aus bestehenden Variablen	10
4	Kreuztabelle mit mehreren Variablen im Kopf	13
5	Tabellen bearbeiten	15
5.1	Groups	15
5.2	Overcodes	15
5.3	Top(2) und Bottom(2) Boxes	17
5.4	Vargroup	18
5.5	Mittelwerte	19
5.6	Sortieren nach Häufigkeiten	21
5.7	Recode	21
6	Tabellenlayout	24
6.1	Tabellen-übergreifende Formatierung	24
6.2	Tabellengestaltung	25
	Stichwortverzeichnis	27

Abbildungsverzeichnis

1.1	Das main.tab	2
2.1	Variablenbezeichnungen aus dem SPSS Datensatz	8
2.2	Vartitle aus dem SPSS Datensatz	8
2.3	Häufigkeitsauszählung	9
3.1	Die GESStabs-Oberfläche	10
3.2	Erstellen von Tabellen	11
3.3	einfache Kreuztabelle	11
4.1	Tabellenkopf mit Textersatz	14
4.2	Kreuztabelle mit mehreren Köpfen	14
5.1	Anpassen der Kopf-Variablen	15
5.2	Overcodes	17
5.3	Tabelle mit Overcodes	18
5.4	Top2-/ Bottom2-Boxes	19
5.5	Mittelwerte	20
5.6	Mittelwerte	20
5.7	Mittelwerte-Übersicht	21
5.8	Absteigend sortierte Tabelle	21
5.9	vor dem Recode	22
5.10	Mit Recode	23
6.1	Übersicht Tabellengestaltung	25

1 Einführung in die Tabellierung

In diesem Tutorial wird die Tabellierung mit GESStabs anhand eines einfachen Beispielprojekts demonstriert. Dafür nutzen wir eine kleine Befragung zum Thema Internetnutzung, welche im Wintersemester 2010/ 2011 an der Universität Hamburg bearbeitet und für diese Einführung mit Zufallszahlen gefüllt wurde.

Zu Übungszwecken empfehlen wir, die hier beschriebenen Beispiele selbst auszuführen. Sie müssen aber nicht alles abtippen, die Dateien stellen wir Ihnen im Archiv EXAMPLE.zip bereit.

Erste Schritte

Installieren Sie GESStabs bitte zunächst an Ihrem Arbeitsplatz.

1. Führen Sie das von uns zur Verfügung gestellte Programm aus und folgen Sie dem Installationsmenü. Standardmäßig wird GESStabs auf Ihrem Computer unter `c://GESS` installiert. Von uns erhalten Sie einen (Test-)Lizenzkey. Zu dessen einmaliger Gültigkeits-Prüfung ist eine Internetverbindung nötig.
2. Für die Arbeit mit den GESStabs-Skripten können Sie jeden beliebigen Texteditor nutzen. Wir empfehlen den Notepad++ Editor: Dieser ist effizient in der Anwendung und wird von uns mit einem Syntax-Highlighting unterstützt (.zip-Datei auf unserer Homepage).
3. Das Lesen, Verarbeiten und Schreiben von SPSS-Dateien erfordert eine SPSS-Input/Output-Schnittstelle. Diese sog. DLL-Dateien stellen wir Ihnen bereit. Speichern Sie diese bitte ebenfalls im `c://GESS`-Verzeichnis ab.
4. Sie können erstellte Tabellen (u.a.) als PostScript-Datei betrachten und als PDF-Dokument abspeichern. Dazu benötigen Sie einen PostScript-Viewer sowie PDF-Destiller. Wir empfehlen GhostView und GPL GhostScript, beides können Sie sich kostenlos im Internet herunterladen.
5. Teilen Sie als letzten Schritt der GESStabs-Software unter **Optionen** die Verzeichnisse mit, in denen auf die notwendigen Arbeitsprogramme (Texteditor, PostScript-Viewer und PDF-Destiller) zugegriffen werden kann.

Im Folgenden werden Sie bei Ihren ersten Schritten mit GESStabs begleitet.

Nachdem Sie dieses Tutorial zusammen mit der beiliegenden Beispieldatei durchgearbeitet haben, können Sie...

- ein Tabellierskript sinnvoll anlegen,
- mit einem ASCII-, SPSS- oder CSV-Datensatz arbeiten,
- Kreuztabellen (mit mehreren Köpfen) erstellen,
- Top(2)- und Bottom(2)-Boxes generieren,
- nach Häufigkeiten sortieren,
- recodieren,
- Mittelwerte und Mittelwertstabellen ausgeben und
- das Aussehen der Tabellen variieren.

Legen wir also los!

GESStabs ist ein Skript-orientiertes Tabellierprogramm. Das bedeutet: Alle Anweisungen an die Software werden als Sammlung von Befehlen in Form eines Skripts festgehalten. Dieses Skript wird anschließend von GESStabs abgearbeitet und erstellt dann die dort angewiesenen Tabellen.

Es ist guter GESS-Programmierstil, ein Skript in mehrere Dateien aufzuteilen. Unter unseren Nutzern hat sich in den letzten 25 Jahren folgendes Vorgehen als effizient erwiesen:

Eine Datei mit der Extension *.tab wird als Hauptprogramm genutzt. Diese Datei bindet mit dem INCLUDE-Statement weitere, thematisch definierte Dateien (Unterprogramme) ein. Dieses Hauptprogramm sieht z. B. so aus:

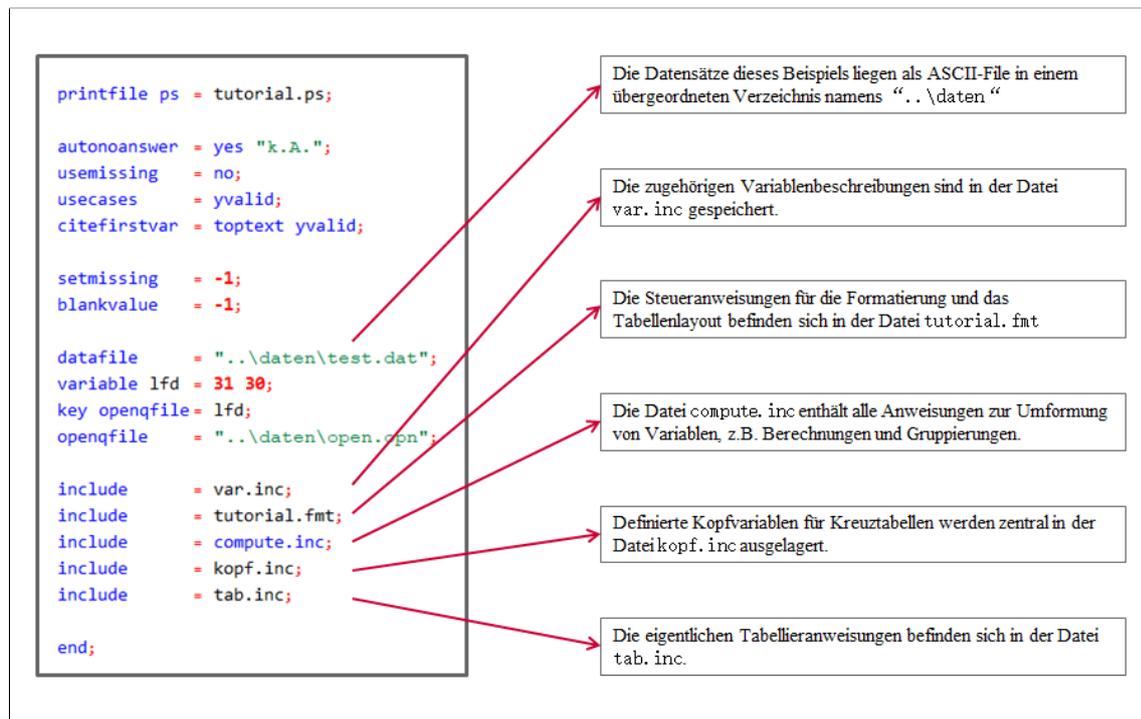


Abb. 1.1: Das main.tab

In diesem Tutorial heißt das Hauptprogramm main.tab. Eingebunden werden in der gelisteten Reihenfolge:

- die Datenquelle (... \data \test.dat),
- die Variablenbeschreibung des eingelesenen Datensatzes,
- das Tabellenlayout in der Datei tutorial.fmt,
- Variablenumformungen in der Datei compute.inc
- die Definition der Tabellenköpfe¹ in der Datei kopf.inc
- und die eigentlichen Tabellieranweisungen mit dem Namen tab.inc.

GESStabs liest diese in unterschiedliche Dateien aufgeteilten Steueranweisungen von oben nach unten, sodass letztendlich ein einziger Stapel als Abfolge aller einzelnen Anweisungen abgearbeitet wird.

¹In der Markt- und Meinungsforschung werden Kreuztabellen oft so dargestellt, dass oben in horizontaler Richtung die sogenannten Aufbruchgruppen aufgeführt werden. Die interessierenden Fragen werden dann senkrecht, meist in Form von Spaltenprozenten, dargestellt. Diese oben in der Tabelle stehenden Aufbruchgruppen werden in GESS-Anleitungen als „Kopf“ bezeichnet.

Die Steueranweisungen im `main.tab` enthalten neben dem Verweis auf die `INCLUDE`-Dateien folgende Befehle:

```
PRINTFILE PS = tutorial.ps;

AUTONOANSWER = YES "k.A.";
USEMISSING   = NO;
USECASES     = YVALID;
CITEFIRSTVAR = TOPTXT YVALID;

SETMISSING   = -1;
BLANKVALUE   = -1;

DATAFILE     = "..\daten\test.dat";
VARIABLE lfd = 31 30;
KEY OPENQFILE = lfd;
OPENQFILE    = "..\daten\open.opn";

END;
```

Mit ein wenig Übung sind viele der Steueranweisungen selbsterklärend. Gehen wir die einzelnen Zeilen einmal durch.

Ein Semikolon beendet jede einzelne Steueranweisung.

Mit `PRINTFILE ps = tutorial.ps;` wird eine Ausgabedatei benannt. Auf das Schlüsselwort `PRINTFILE` folgen die Bezeichnung des Ausgabeformats, in diesem Fall PostScript (`ps`), und die Dateibenennung, hier 'tutorial.ps'.

`AUTONOANSWER = YES;` bewirkt, dass fehlende Antworten standardmäßig und für jede Frage als "k.A." gesetzt werden (Filter werden berücksichtigt)². `USEMISSING` veranlasst, dass fehlende Werte nicht in Berechnungen (z.B. Mittelwerte) eingehen. Mit der Anweisung `USECASES = yvalid;` wird bestimmt, dass in der Tabellierung nur die gültigen Fälle des Seitenaufrisses, also der Y-Variablen, berücksichtigt werden. `CITEFIRSTVAR = toptxt yvalid;` heißt: Drucke den Toptext (üblicherweise ist dies der Fragetext) der Vorspaltenvariable über der Tabelle ab.

Mit `SETMISSING` wird der interne Missing-Wert bestimmt, hier -1. `BLANKVALUE` gibt den Wert an, der bei leeren Datenzellen gesetzt werden soll. Hier entspricht dieser dem Missing-Wert, d.h. leere Datenfelder werden als Missing interpretiert.

Den Schlusspunkt des (aktiven) Skriptes setzt das Schlüsselwort `end;`. Alle danach folgenden Texte werden zwar gelesen, jedoch nicht mehr interpretiert.

² `AUTONOANSWER = NO;` wirkte basisreduzierend.

2 Einlesen von Datensätzen

Nachdem die generelle Systematik eines GESStabs-Skripts erklärt wurde, widmen wir uns nun der nächsten grundsätzlichen Frage: Wie bindet man einen Datensatz ein?

In der Regel wird heutzutage mit ASCII-, SPSS- oder CSV-Datensätzen gearbeitet.

Der einzige Unterschied in der Handhabung der Datenformate liegt in der jeweiligen Variablendefinition. Während in SPSS feste Variablenbeschreibungen gespeichert werden können, müssen die Variablen aus ASCII- und CSV-Datensätzen erst noch definiert werden. SPSS- und CSV-Datensätze ähneln sich strukturell, in ASCII wird die Variablen-Zuordnung über Spaltenangaben gesteuert.

Es folgt eine kurze Einführung in die drei Varianten.

2.1 ASCII

ASCII-Datensätze werden wie folgt in das Skript eingebunden¹:

```
DATAFILE = "Dateiname.dat";
```

In ASCII-Datensätzen entspricht eine Zeile einem Fall. Variablen werden anhand ihrer Spaltenzahl beschrieben.

2.1.1 Manuelles Erstellen einer var.inc

Liegt noch keine vorbereitete `var.inc` vor², muss diese selbst geschrieben werden.

Grundsätzlich braucht man für eine Variablendefinition nur den Variablentypen, den Namen und die Spaltenangabe.

Singlenennungen

Die Definition einer Frage mit Singlenennung sieht beispielsweise so aus:

```
SingleQ f0 = 30 1;
```

Hiermit ist folgendes definiert:

Die Variable enthält maximal eine Antwort pro Interview, ist also eine Variable vom Typ `SingleQ`. Die Variablenbezeichnung lautet `f0`, hiermit kann die Variable später aufgerufen und weiter verarbeitet werden. Die Ziffern nach dem Gleichheitszeichen geben die Spaltenposition im Datensatz und die Anzahl der ab dort genutzten Spalten an. Die Variable `f0` befindet sich demnach in der 30. Spalte des Datensatzes und ist eine Spalte breit.

Anschließend kann man der Frage einen `TEXT` und einen `TITLE` zuweisen.

```
TEXT = "Seit wann nutzen Sie das Internet?";  
TITLE = "Seit...";
```

Liegt keine Zuweisung zu einer bestimmten Frage vor, werden diese der zuletzt definierten Frage zugeordnet. Für eine direkte Zuweisung kann der Variablenname hinter das Befehlswort geschrieben werden³.

¹Siehe [Erste Schritte](#) für das Vorgehen bei zusätzlichem Einbezug eines OpenQ-Datenfiles.

²Für Umfragen, die mit GESS Q. oder GESS Cati durchgeführt wurden, lässt sich diese automatisch generieren, siehe [2.1.2](#)

³Mit diesem Vorgehen sind Sie auf der sicheren Seite, falls sich einmal die Reihenfolge des Skriptinhaltes ändert.

Der LABELS-Befehl bezeichnet die Inhalte der einzelnen Variablencodes.

```
LABELS f0 =
1 "weniger als 6 Monaten"
2 "6 - 12 Monaten"
3 "ein bis zwei Jahren"
4 "zwei bis drei Jahren"
5 "mehr als drei Jahren"
;
```

Multinennungen, Dichotome

Die Beschreibung von Variablen mit mehreren Antwortmöglichkeiten ähnelt dem Vorgehen für SingleQ-Variablen.

```
MultiQ f9 = 392 22 2;
TEXT = "Für welche Zwecke nutzen Sie das Internet?";
TITLE = "Bitte klicken Sie alles zutreffende an.";
LABELS =
1 "Recherche/Ratgeber"
2 "Videos/ Musik"
3 "E-Mail"
4 "Online Gaming"
5 "Shopping / Booking"
6 "Ämter / Verwaltung (Uni, Steuer)"
7 "Chats & Foren"
8 "Online Banking"
9 "Online Medien (Zeitschriften, Magazine, Blogs)"
10 "Soziale Netzwerke (Facebook, Parship, XING)"
11 "Sonstiges"
;
```

Die Definition der Spaltenangabe ist hier ein wenig anders: Insgesamt benötigt eine MultiQ genauso viele Spalten, wie Angaben gemacht werden können (Variable f9 hat 11 Codes, kann daher maximal 11 Antworten erhalten, daher hier: 11 Spalten). Weil die Labelcodes 10 und 11 zweistellig sind, muss jede mögliche Antwort Platz für zwei Ziffern bereitstellen. Dementsprechend werden höchstens $11 \times 2 = 22$ Spalten benötigt. Oben in der Spaltenangabe ist spezifiziert, dass die MultiQ f9 in Spalte 392 des Datensatzes beginnt, insgesamt 22 Spalten benötigt und jede Angabe maximal zwei Stellen besitzt.

2.1.2 Automatisch generierte var.inc

Wurde eine Umfrage in GESS Q. oder GESS Cati durchgeführt, kann eine zugehörige `var.inc` einfach per Mausklick erstellt werden. Hierin sind alle Variablen mit Spaltendefinition und -breite enthalten. Die Vertextung der Variablen entspricht der in der Umfrage verwendeten Texte.

Für unseren Beispiel-Datensatz konnten wir eine automatisch generierte `var.inc` erstellen.

Die Variablendefinition von f1 sieht beispielsweise so aus:

```
SingleQ f1 = 478 1;
TEXT = "An welchem Ort nutzen Sie jetzt gerade das Internet?";
LABELS =
1 "zu Hause"
2 "Am Arbeits- / Ausbildungsplatz"
3 "In der Schule / an der Universität"
4 "An einem öffentlichen Ort (Internet-Cafe, Multimedia-Terminal)"
5 "Von unterwegs (über WLAN Hotspot, Handynetz, ...)"
6 "Woanders außer Haus (bei Freunden, Verwandten)"
;
```

In der zugehörigen Befragung war kein Title vorgesehen. Wird dieser im Zuge der Auswertung gewünscht, kann er mit dem TITLE-Befehlsschlagwort ergänzt werden. Zum Beispiel:

```
TITLE f1 = "Ort der Internetnutzung";
```

Die f3, eine Frage mit Mehrfachantwortmöglichkeit, wurde analog beschrieben. Mit sieben Labels, von denen eines zweistellig ist, benötigt die Variable maximal $7 \times 2 = 14$ Spalten je Fall. Text, Title und Labels sind hier ebenfalls automatisch definiert.

```
MultiQ f3 = 479 14 2;
TEXT = "An welchen der folgenden Orte nutzen Sie zumindest gelegentlich das
Internet?";
TITLE = "Bitte kreuzen Sie alles an, was auf Sie zutrifft!";
LABELS =
1 "Zu Hause"
2 "Am Arbeits- / Ausbildungsplatz"
3 "In der Schule / an der Universität"
4 "An einem öffentlichen Ort (Internet-Cafe, Multimedia-Terminal)"
5 "Von unterwegs (über WLAN Hotspot, Handynetz, ...)"
6 "Woanders außer Haus (bei Freunden, Verwandten)"
99 "Keinen davon"
;
```

Die Beschreibung einer Matrix-Frage kann nicht einfach 1:1 aus dem Fragebogenskript übernommen. Hier wird stattdessen ein Makro erstellt und automatisch in die var.inc integriert.

```
#macro #f10.1(&gItemVal &title &col)
SingleQ f10.&gItemVal.1 = &col 1;
TEXT = "Wie häufig nutzen Sie diese Angebote im Internet?";
TITLE = "&title";
LABELS =
1 "sehr häufig"
2 "häufig"
3 "gelegentlich"
4 "selten"
5 "sehr selten"
;
#endmacro

#f10.1(1 "Recherche/Ratgeber" 520)
#f10.1(2 "Videos/Musik" *)
```

```
#f10.1(3 "E-Mail" *)
#f10.1(4 "Online Gaming" *)
#f10.1(5 "Shopping/Booking" *)
#f10.1(6 "Ämter/Verwaltung (Uni, Steuer)" *)
#f10.1(7 "Chats & Foren" *)
#f10.1(8 "Online Banking" *)
#f10.1(9 "Online Medien (Zeitschriften, Magazine, Blogs)" *)
#f10.1(10 "soziale Netzwerke (Facebook, Parship, XING)" *)
```

Im ersten Abschnitt ist die Fragestruktur definiert. Für die Itemwerte, Title und Angabe der Startspalte werden Platzhalter gesetzt. Im zweiten Abschnitt wird das Makro für jedes Frageitem aufgerufen, das jeweils als eigene Variable abgelegt ist. Werte für Itemwert, Title und Startspalte werden eingesetzt. Ab dem zweiten Makroaufruf ist ein Sternchen anstelle der Startspalte gesetzt. Dies bewirkt ein einfaches Hochzählen der Spaltenangabe: Item 1 liegt auf Spalte 520, für Item 2 wird mit 521 hochgezählt, Item 3 liegt in Spalte 522, usw..

2.2 SPSS

SPSS-Datensätze werden mit dem Schlagwort `SPSSINFILE` in das Skript einbezogen:

```
SPSSINFILE = "Dateiname.sav";
```

Auch SPSS-Datensätze führen einen Fall je Zeile auf. Das Programm kann Variablentexte und -labels speichern. Im Idealfall arbeiten Sie mit einem SPSS-Datensatz, der schön mit Text- und Labeltexten bestückt ist, und können direkt mit der Tabellierung beginnen. Dies ist leider selten der Fall.

Überblick über einen SPSS-Datensatz

GESStabs stellt verschiedene Möglichkeiten bereit, um sich einen Überblick über einen gegebenen Datensatz zu machen.

Mit dem Schlagwort `SYNTAX` können verschiedene Variableninformationen aus dem Datensatz in Textdateien herausgeschrieben werden, nämlich die Variablentypen, -texte, -title und -labels. Zur Extraktion der Eigenschaften aus dem Datensatz wird das jeweilige Stichwort dem `SYNTAX`-Befehl nachgestellt.

Es ist Geschmackssache, ob alle Informationen in eine einzige Textdatei geschrieben werden:

```
SYNTAX VARIABLES TITLE TEXT LABELS = alles_zusammen.inc;
```

oder thematisch aufgeteilt werden:

```
SYNTAX VARIABLES = variablen.inc;
SYNTAX TITLE     = title.inc;
SYNTAX TEXT      = texte.inc;
SYNTAX LABELS    = labels.inc;
```

An dieser Stelle sei erwähnt: SPSS unterscheidet nicht zwischen Texten und Titeln, führt stattdessen einfache Variablenlabels. GESStabs weist diese standardmäßig als Variablentitle aus.

Die erstellten Dateien enthalten nun die Informationen über Typen, Title und Labels der im Datensatz enthaltenen Variablen:

```
#&1 ( f1          "" singl "" "" "" "" ) { An welchem Ort nu
#&1 ( f2          "" singl "" "" "" "" ) { Seit... }
#&1 ( f3_1        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3_2        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3_3        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3_4        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3_5        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3_6        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3_7        "" singl "" "" "" "" ) { Bitte kreuzen Sie
#&1 ( f3.spss.1   "" numer "" "" "" "" ) { Zu Hause }
#&1 ( f3.spss.2   "" numer "" "" "" "" ) { Am Arbeits- / Aus
#&1 ( f3.spss.3   "" numer "" "" "" "" ) { In der Schule / a
#&1 ( f3.spss.4   "" numer "" "" "" "" ) { An einem öffentli
#&1 ( f3.spss.5   "" numer "" "" "" "" ) { Von unterwegs (üb
#&1 ( f3.spss.6   "" numer "" "" "" "" ) { Woanders außer Ha
#&1 ( f3.spss.99  "" numer "" "" "" "" ) { Keinen davon }
#&1 ( f6opn      "" alpha "" "" "" "" ) { Bitte alles zutre
#&1 ( f7.1        "" numer "" "" "" "" ) { Stunden: }
```

Abb. 2.1: Variablenbezeichnungen aus dem SPSS Datensatz

```
VARTITLE f1 ="An welchem Ort nutzen Sie jetzt gerade das Internet?"
VARTITLE f2 ="Seit...";
VARTITLE f3_1 f3_2 f3_3 f3_4 f3_5 f3_6 f3_7 ="Bitte kreuzen Sie al
VARTITLE f3.spss.1 ="Zu Hause";
VARTITLE f3.spss.2 ="Am Arbeits- / Ausbildungsplatz";
VARTITLE f3.spss.3 ="In der Schule / an der Universität";
VARTITLE f3.spss.4 ="An einem öffentlichen Ort (Internet-Café, Mul
VARTITLE f3.spss.5 ="Von unterwegs (über WLAN Hotspot, Handynetz,
VARTITLE f3.spss.6 ="Woanders außer Haus (bei Freunden, Verwandten
VARTITLE f3.spss.99 ="Keinen davon";
VARTITLE f6opn ="Bitte alles zutreffende notieren.";
```

Abb. 2.2: Vartitle aus dem SPSS Datensatz

Eine andere Möglichkeit, sich einen Überblick über einen gegebenen SPSS-Datensatz zu machen, bietet das CODEBOOK-Statement .

Dies bewirkt eine einfache Häufigkeitsauszählung über alle

```
CODEBOOK;
```

oder ausgewählte Variablen

```
CODEBOOK f1;
```

und sieht z.B. so aus:

Table 12:			
An welchem Ort nutzen Sie jetzt gerade das Internet?	Abs.	Col %	Cum. %
Zu Hause	67	16	16
Am Arbeits- / Ausbildungsplatz	75	17	33
In der Schule / an der Universität	73	17	50
An einem öffentlichen Ort (Internet-Café, Multimedia-Terminal, ...)	69	16	66
Von unterwegs (über WLAN Hotspot, Handynet, ...)	70	16	82
Woanders außer Haus (bei Freunden, Verwandten)	78	18	100
N =	432	432	432
GESS mbH			

Abb. 2.3: Häufigkeitsauszählung

Tipp: Codebooks eignen sich auch wunderbar, um Ihre anschließend erstellten Tabellen auf Datenrichtigkeit zu kontrollieren.

2.3 CSV

Mit CSVINFILE werden CSV-Datensätze in ein GESStabs-Projekt eingebunden:

```
CSVINFILE = "Dateiname.csv";
```

Strukturell ähneln CSV-Datensätze SPSS-Files: Jede Zeile enthält einen Interviewfall.

CSV-Datensätze speichern - neben den Rohdaten und Variablennamen - allerdings keine Variableninformationen. Diese können Sie nachträglich definieren, siehe dafür [Manuelles Erstellen einer var.inc](#).

Um die korrekte Darstellung offener Texte (und insbesondere enthaltener Umlaute) im Tabellen-Output sicherzustellen, empfehlen wir, die Kodierung der CSV-Datendatei standardmäßig auf UTF-8 einzustellen:

```
ENCODING CSV = UTF8;
```

3 Einfache Kreuztabelle aus bestehenden Variablen

Das Einfachste ist, zwei bestehende Variablen in einer Kreuztabelle gegeneinander zu tabellieren. Hierfür definieren wir eine Tabelle, die eine Variable im Kopf und eine im Seitenaufriß enthält. Wir möchten die Verteilung der Frage f1 (Ort der Internetnutzung) nach Geschlecht (Variable s3) betrachten und schreiben dazu folgendes Statement in die `tab.inc`:

```
TABLE = s3 BY f1;
```

Das ist die einfachste Form eines GESStabs-Tabellen-Statements. Auf das Schlüsselwort `TABLE` und ein Gleichheitszeichen folgt die Kopfvariable. Hinter `BY` wird die Variable geschrieben, die in der Vorspalte¹ stehen soll. Letztere ist die eigentlich interessierende Variable, deren Verteilung betrachtet wird. Abgeschlossen wird die Anweisung, wie immer, durch das Semikolon.

Zum Tabellieren rufen wir das GESStabs-Menü auf, es öffnet sich folgende minimalistische Programm-Oberfläche:

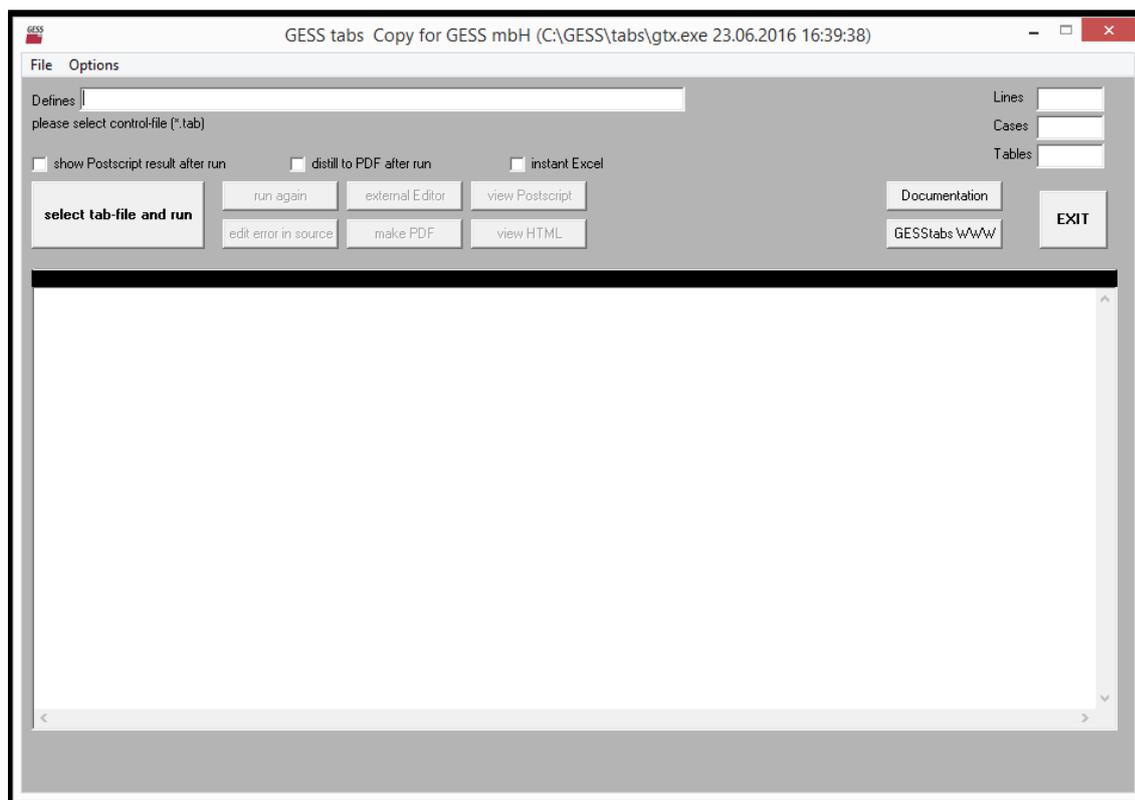


Abb. 3.1: Die GESStabs-Oberfläche

¹Die Bezeichnungen hierfür variieren: Vorspalte, Aufriss, Y-Variable...

Über den Schalter `select tab-file and run` rufen wir unser `main.tab` auf. GESStabs führt das Skript aus und zeigt folgende Meldungen an:

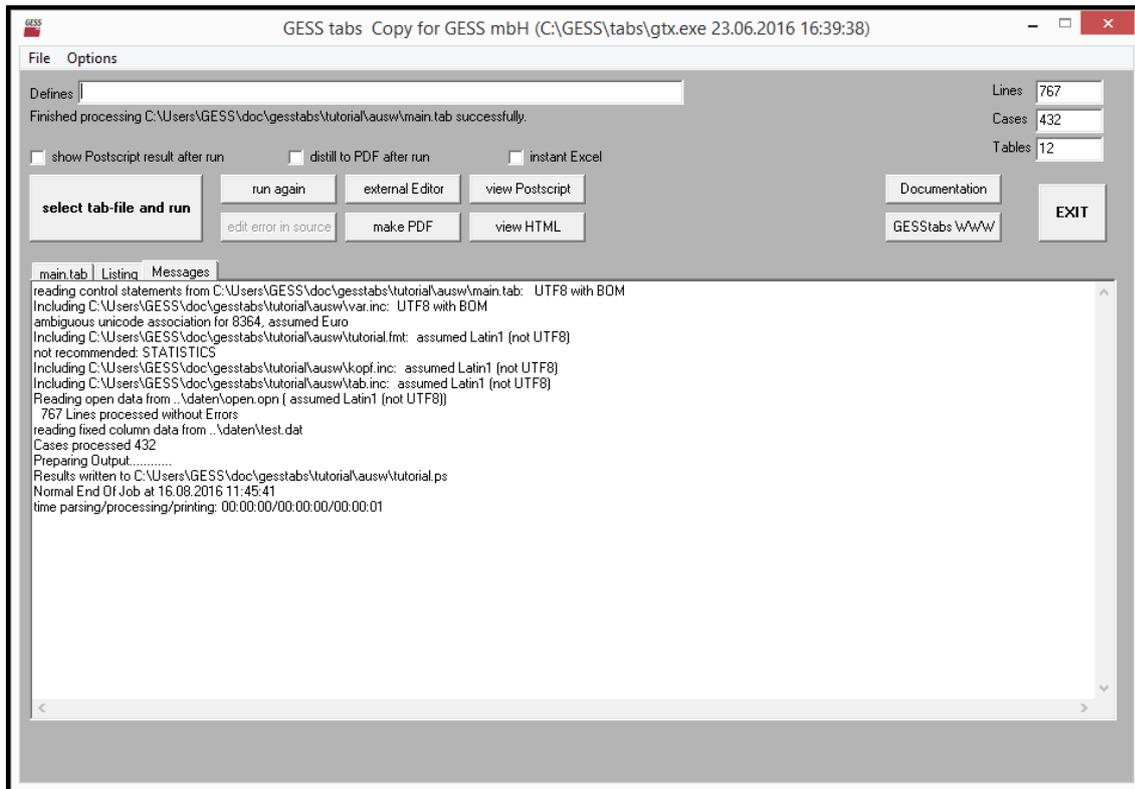


Abb. 3.2: Erstellen von Tabellen

Unser Beispiel-Programm wurde erfolgreich abgearbeitet, rechts oben kann die Anzahl verarbeiteter Skriptzeilen, einbezogener Datenfälle und erstellter Tabellen nachvollzogen werden.

Ein Klick auf den Button `view PostScript` öffnet folgende Tabelle im benutzten PostScript-Viewer:

Table 1			
An welchem Ort nutzen Sie jetzt gerade das Internet?			
Abs.	Total	Geschlecht	
		Männlich	Weiblich
Total	432	223	209
Ort der Internetnutzung			
Zu Hause	67	38	29
Am Arbeits- / Ausbildungsplatz	75	33	42
In der Schule / an der Universität	73	36	37
An einem öffentlichen Ort (Internet-Café, Multimedia-Terminal, ...)	69	35	34
Von unterwegs (über WLAN Hotspot, Handynetz, ...)	70	37	33
Woanders außer Haus (bei Freunden, Verwandten)	78	44	34

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 3.3: einfache Kreuztabelle

Das Ergebnis ist eine einfache Kreuztabelle. Über der Tabelle wird der Text der Vorspalten-Variable abgebildet, über den Labels der beiden Variablen steht der jeweilige Title. Die Tabelle zeigt absolute Werte (siehe Beschriftung 'Abs.') und enthält eine Totalspalte sowie -zeile.

Ein Klick auf `make PDF` würde eine zusätzliche, zur PostScript-Ansicht identische PDF-Datei erzeugen. Für diesen Moment belassen wir es aber bei der PostScript-Datei und lassen diese

geöffnet. Nach erneutem Start des GESStabs-Laufs (Klick auf `run again`) aktualisiert sich diese automatisch, wodurch ein schnelles betrachten vorgenommener Änderungen und Ergänzungen möglich ist.

4 Kreuztabelle mit mehreren Variablen im Kopf

Eine Tabelle mit lediglich zwei Spalten im Kopf sieht nun etwas langweilig aus. Wir fügen dem Tabellenkopf daher weitere Variablen hinzu. Dafür ergänzen wir einfach die entsprechenden Variablennamen im `TABLE`-Statement:

```
TABLE = s1 s2 s3 BY f1;
```

Diese Variablen (Gerätenutzung, Hauptwohnsitze & Geschlecht) sollen also unseren neuen Kopf bilden.

Für Tabellenköpfe, die häufig benutzt werden, ist es unpraktisch, diese in jeder neuen Tabellendefinition mit allen enthaltenen Variablennamen erneut aufzuführen. Zu diesem Zwecke bietet sich ein universaler Mechanismus der GESStabs-Skriptsprache an: das `#EXPAND`. Dabei handelt es sich um einen einfachen Textersatz.

Wir nennen unseren Tabellenkopf `#K1`¹ und definieren mit folgendem Statement alle Variablen, die enthalten sein sollen:

```
#EXPAND #K1 s1 s2 s3
```

Als einfacher Textersatz wird anschließend alles übertragen, was in der Zeile auf das Schlüsselwort `#EXPAND` folgt, inkl. Kommentarzeichen, Semikolon etc.. Hier darf also kein abschließendes Semikolon stehen, da dieses den `TABLE`-Befehl hinter `'#K1'` beenden würde.

Die Tabellenanforderung für die identische Tabelle verkürzt sich dann auf:

```
TABLE = #K1 BY f1;
```

Jetzt kann man fortfahren und die gleiche Frage mit anderen Köpfen tabellieren. Oft möchte man mehr Variablen gegeneinander betrachten, als auf eine Tabellenseite passt. Daher erstellen wir einen zweiten Tabellenkopf, der Fragen nach Einkommen und Bildung enthält.

```
#EXPAND #K2 s8 s9
```

Effektives Skripting erfordert Übersichtlichkeit. Deshalb werden Tabellenköpfe von uns in einer eigenen Datei angelegt, in der eingangs erwähnten `kopf.inc`. In unserem Beispiel enthält diese Datei nur diese zwei Zeilen:

```
#EXPAND #K1 s1 s3 s2
#EXPAND #K2 s8 s9
```

Da unser Beispiel noch recht übersichtlich (weil kurz) ist, mag diese Aufteilung zu streng erscheinen. In der Praxis wird jedoch häufig mit zahlreichen und komplexen Tabellenköpfen gearbeitet. Andere Programmierer (oder man selbst, ein halbes Jahr später) finden sich leichter zurecht, wenn man sich an solche Formalismen hält.

Anschließend tabellieren wir unsere Frage `f1` mit beiden Köpfen:

```
TABLE = #K1 BY f1;
TABLE = #K2 BY f1;
```

¹Namen von `EXPANDs` müssen stets mit einem Lattenkreuz beginnen.

Als Ergebnis erhalten wir die folgenden beiden Tabellen:

Table 2											
An welchem Ort nutzen Sie jetzt gerade das Internet?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Ort der Internetnutzung											
Zu Hause	67	32	19	16	8	18	13	16	12	38	29
Am Arbeits- / Ausbildungsplatz	75	35	24	16	17	12	16	14	16	33	42
In der Schule / an der Universität	73	43	17	13	13	17	13	18	12	36	37
An einem öffentlichen Ort (Internet-Café, Multimedia-Terminal, ...)	69	35	18	16	14	10	19	14	12	35	34
Von unterwegs (über WLAN Hotspot, Handynet, ...)	70	38	14	18	19	9	13	14	15	37	33
Woanders außer Haus (bei Freunden, Verwandten)	78	36	28	14	19	12	13	18	16	44	34

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 4.1: Tabellenkopf mit Textersatz

Table 3															
An welchem Ort nutzen Sie jetzt gerade das Internet?															
Abs.	Total	Höchster Schulabschluss					Netto-Einkommen								
		Kein allgemeiner Schulabschluss	Haupt- / Volksschulabschluss	Realschule / Mittlere Reife	Fachhochschulreife	Hochschulabschluss	weniger als 500 EUR	500 EUR bis unter 1.000 EUR	1.000 EUR bis unter 1.500 EUR	1.500 EUR bis unter 2.000 EUR	2.500 EUR bis unter 3.000 EUR	3.000 EUR bis unter 3.500 EUR	3.500 EUR bis unter 4.000 EUR	4.000 EUR und mehr	kein eigenes Einkommen
Total	432	86	88	95	86	77	65	54	50	53	50	36	35	51	38
Ort der Internetnutzung															
Zu Hause	67	14	17	14	8	14	11	10	3	10	7	5	6	8	7
Am Arbeits- / Ausbildungsplatz	75	18	9	16	19	13	14	8	10	8	7	5	2	15	6
In der Schule / an der Universität	73	14	18	22	12	7	10	8	12	5	6	9	9	8	6
An einem öffentlichen Ort (Internet-Café, Multimedia-Terminal, ...)	69	11	12	13	16	17	8	7	10	8	10	10	6	5	5
Von unterwegs (über WLAN Hotspot, Handynet, ...)	70	10	18	14	14	14	14	11	6	10	10	2	3	8	6
Woanders außer Haus (bei Freunden, Verwandten)	78	19	14	16	17	12	8	10	9	12	10	5	9	7	8

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 4.2: Kreuztabelle mit mehreren Köpfen

5 Tabellen bearbeiten

5.1 Groups

Der zweite Kopf ist nun etwas lang und die Tabelle dadurch nicht sehr leserlich, was hauptsächlich an den zahlreichen Labels aus Variable s9 liegt. Es bietet sich an, die Einkommensklassen im Nachhinein weiter zusammenzufassen¹. Dafür generieren wir eine neue Variable, genauer gesagt eine GROUPS. Mit einer GROUPS können Labels zusammengefasst, neue Labeltexte und -title vergeben und/ oder Variablen gefiltert werden.

```
GROUPS s9neu =
| "Kein eigenes Einkommen"      : 10 in s9
| "1 bis unter 1.000 "          : [1:2] in s9
| "1.000 bis unter 2.500 "      : [3:5] in s9
| "2.500 bis unter 3.500 "      : [6:7] in s9
| "3.500 und mehr"             : [8:9] in s9
;
TITLE s9neu = "Einkommen";
```

Die Labeldefinitionen einer GROUPS beginnen immer mit einem |, gefolgt vom Labeltext. Nach dem Doppelpunkt wird definiert, welche Werte aus s9 in das jeweilige Label einfließen sollen². Unter Einbezug der neu generierten GROUPS wird nun ein neuer Tabellenkopf generiert:

```
#EXPAND #K3 s8 s9neu
```

Tabelliert sieht dies so aus:

```
TABLE = #K3 BY f1;
```

Table 4											
An welchem Ort nutzen Sie jetzt gerade das Internet?											
Abs.	Total	Höchster Schulabschluss					Netto-Einkommen (neu)				
		Kein allgemeiner Schulabschluss	Haupt- / Volksschulabschluss	Realschule / Mittlere Reife	Fachhochschulreife	Hochschulabschluss	Kein eigenes Einkommen	1 € bis unter 1.000 €	1.000 € bis unter 2.500 €	2.500 € bis unter 3.500 €	3.500 € und mehr
Total	432	86	88	95	86	77	38	119	103	86	86
Ort der Internetnutzung											
Zu Hause	67	14	17	14	9	14	7	21	13	12	14
Am Arbeits- / Ausbildungsplatz	75	18	9	16	19	13	6	22	18	12	17
In der Schule / an der Universität	73	14	18	22	12	7	6	18	17	15	17
An einem öffentlichen Ort (Internet-Café, Multimedia-Terminal, ...)	69	11	12	13	16	17	5	15	18	20	11
Von unterwegs (über WLAN Hotspot, Handynetz, ...)	70	10	18	14	14	14	6	25	16	12	11
Woanders außer Haus (bei Freunden, Verwandten)	78	19	14	16	17	12	8	18	21	15	16

Abb. 5.1: Anpassen der Kopf-Variablen

5.2 Overcodes

Mithilfe des LABELS-Statements können einer bestehenden Variable spielend leicht Labels hinzugefügt, diese sortiert und bearbeitet werden.

Schauen wir uns einmal die Variable f11_1 an:

¹Unserer Skript-Systematik folgend nehmen wir Variablenmanipulationen in der compute.inc vor.

²Achten Sie darauf, die Labels sauber zu definieren. Überschneiden sich die Bedingungen, können Werte eines Falls in mehrere Labels der Groups einfließen.

```

SingleQ f11_1 = 713 2;
TEXT f11_1 = "Und zu welchem Thema recherchieren Sie am häufigsten?";
TITLE = "Recherche";
LABELS f11_1 =
1 "Nachrichten"
2 "Familie und Kinder"
3 "Wirtschaft/ Finanzen"
4 "Politik"
5 "Sport"
6 "Essen und Trinken"
7 "Erotik"
8 "Kino- und Fernsehprogramm"
9 "Wetter"
10 "Beauty und Fashion"
11 "Freunde/ Partnerschaft"
12 "Kunst/ Kultur/ Musik"
;

```

Die Themen sind hier ziemlich durcheinandergewürfelt und nicht nach Überpunkten geordnet. Mit GESStabs können wir die Reihenfolge der Labels einfach selbst bestimmen.

```

LABELS f11_1 =
1 "Nachrichten"
4 "Politik"
9 "Wetter"
3 "Wirtschaft/ Finanzen"

10 "Beauty und Fashion"
6 "Essen und Trinken"
8 "Kino- und Fernsehprogramm"
12 "Kunst/ Kultur/ Musik"
5 "Sport"

7 "Erotik"
2 "Familie und Kinder"
11 "Freunde/ Partnerschaft"
;

```

Im Ausdruck einer Tabelle würden jetzt die Themen in dieser Reihenfolge erscheinen. Nun ist es häufig wünschenswert, Oberthemen zu bilden und diese auszählen zu lassen. Hierfür muss man den Labels lediglich die übergeordneten Begriffe als **OVERCODE** mitteilen.

```

LABELS f11_1 =
OVERCODE 1 3 4 9 "Öffentlich (Nettozählung)"
1 "Nachrichten"
4 "Politik"
9 "Wetter"
3 "Wirtschaft/ Finanzen"
OVERCODE 5 6 8 10 12 "\Lifestyle (Nettozählung)"
10 "Beauty und Fashion"
6 "Essen und Trinken"
8 "Kino- und Fernsehprogramm"
12 "Kunst/ Kultur/ Musik"

```

```
5 "Sport"
OVERCODE 2 7 11 "\Privat (Nettozählung)"
7 "Erotik"
2 "Familie und Kinder"
11 "Freunde/ Partnerschaft"
;
```

Würde das Schlüsselwort OVERCODE um das Schlüsselwort SUM ergänzt, würde eine Brutto-Zählung vorgenommen.

Mit #K1 im Kopf sieht die Tabelle der neuen f11_1 so aus:

Und zu welchem Thema recherchieren Sie am häufigsten?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Recherche											
Öffentlich (Nettozählung)	135	79	32	24	30	27	22	25	31	68	67
Nachrichten	58	38	8	12	14	10	11	13	10	27	31
Politik	44	22	16	6	10	10	7	5	12	23	21
Wetter	33	19	8	6	6	7	4	7	9	18	15
Lifestyle (Nettozählung)	181	79	62	40	38	29	42	43	29	89	92
Beauty und Fashion	44	19	10	15	8	4	14	11	7	18	26
Essen und Trinken	37	20	10	7	7	5	8	9	8	19	18
Kino- und Fernsehprogramm	30	12	13	5	7	7	7	5	4	20	10
Kunst/ Kultur/ Musik	34	18	11	5	7	9	5	5	8	15	19
Sport	36	10	18	8	9	4	8	13	2	17	19
Privat (Nettozählung)	116	61	26	29	22	22	23	26	23	66	50
Erotik	30	17	6	7	7	5	4	5	9	19	11
Familie und Kinder	52	26	11	15	8	11	13	11	9	29	23
Freunde/ Partnerschaft	34	18	9	7	7	6	6	10	5	18	16

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 5.2: Overcodes

5.3 Top(2) und Bottom(2) Boxes

Oft sollen Skalen neben der Häufigkeitsverteilung auch in Form von Top- oder Bottom-Boxes abgebildet werden. Auch dies kann einfach mithilfe von OVERCODEs umgesetzt werden.

Hier ein Beispiel für die Formulierung von Overcodes in Frage 10.1.1:

```
SingleQ f10.1.1 = 520 1;
TEXT f10.1.1 = "Wie häufig nutzen Sie diese Angebote im Internet?";
TITLE f10.1.1 = "Recherche/Ratgeber";
LABELS f10.1.1 =
1 "sehr häufig"
2 "häufig"
3 "gelegentlich"
4 "selten"
5 "sehr selten"
Overcode 1 2 "Top-2-Box (1+2)"
Overcode 4 5 "Bottom-2-Box (4+5)"
;
```

... und dessen tabellarische Darstellung:

Table 6											
Wie häufig nutzen Sie diese Angebote im Internet?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Recherche/Ratgeber											
sehr häufig	91	45	27	19	21	17	16	18	19	43	48
häufig	79	39	20	20	14	15	18	16	16	37	42
gelegentlich	92	51	25	16	19	16	24	21	12	52	40
selten	86	43	24	19	20	18	13	21	14	46	40
sehr selten	84	41	24	19	16	12	16	18	22	45	39
Top-2-Box (1+2)	170	84	47	39	35	32	34	34	35	80	90
Bottom-2-Box (4+5)	170	84	48	38	36	30	29	39	36	91	79

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 5.3: Tabelle mit Overcodes

5.4 Vargroup

Datensätze enthalten immer wieder Itembatterien, deren Top- und Bottom-Boxes im Tabellenband zusammengefasst dargestellt werden sollen. In GESStabs gibt es den Variablentyp `VARGROUP`, der hierfür gut geeignet ist.

Eine `VARGROUP` setzt sich aus mehreren einfachen Variablen zusammen. Eine `VARGROUP` aus `n` Variablen hat automatisch `n` Label, die nach definierten Bedingungen berechnet werden.

Das ist in der Praxis einfacher als in der Theorie. Nehmen wir wieder eine Matrix-Variable aus unserem Beispieldatensatz. Diese besteht aus zehn Items, die die Nutzungshäufigkeit bestimmter Internetangebote erfragen (`f10`). Definiert ist diese Itembatterie wie folgt:

```
#macro #f10.1(&gItemVal &title &col)
SingleQ f10.&gItemVal.1 = &col 1;
TEXT = "Wie häufig nutzen Sie diese Angebote im Internet?";
TITLE = "&title";
LABELS =
1 "sehr häufig"
2 "häufig"
3 "gelegentlich"
4 "selten"
5 "sehr selten"
;
#endmacro

#f10.1(1 "Recherche/Ratgeber" 520)
#f10.1(2 "Videos/Musik" *)
#f10.1(3 "E-Mail" *)
#f10.1(4 "Online Gaming" *)
#f10.1(5 "Shopping/Booking" *)
#f10.1(6 "Ämter/Verwaltung (Uni, Steuer)" *)
#f10.1(7 "Chats & Foren" *)
#f10.1(8 "Online Banking" *)
#f10.1(9 "Online Medien (Zeitschriften, Magazine, Blogs)" *)
#f10.1(10 "soziale Netzwerke (Facebook, Parship, XING)" *)
```

Wir möchten nun eine Übersicht der Top-2- und Bottom-2-Boxes aller Items erstellen. Dazu zählen wir alle Einzelvariablen (diese entsprechen den einzelnen Items des Fragekomplexes f10) auf, die einbezogen werden sollen, und definieren anschließend die Berechnung der Top-2- und Bottom-2-Boxes:

```
VARGROUP top2 = (f10.1.1 f10.2.1 f10.3.1 f10.4.1 f10.5.1 f10.6.1 f10.7.1 f10.8.1 f10.9.1 f10.10.1) eq 1 2;
VARTITLE top2 = "Top-2-Box \sehr häufig + häufig";

VARGROUP bot2 = (f10.1.1 f10.2.1 f10.3.1 f10.4.1 f10.5.1 f10.6.1 f10.7.1 f10.8.1 f10.9.1 f10.10.1) eq 4 5;
VARTITLE bot2 = "Bottom-2-Box \sehr selten + selten";
```

Eine VARGROUP setzt automatisch die ursprünglichen TITLE der einfachen Variablen als LABELS der neu generierten Vargroup um.

Tabelliert sieht das Ganze schließlich so aus:

```
TABLE = #K1 BY top2;
TABLE = #K1 BY bot2;
```

Table 7											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Top-2-Box sehr häufig + häufig											
Recherche/Ratgeber	170	84	47	39	35	32	34	34	35	80	90
Videos/ Musik	182	91	45	46	37	37	35	37	36	94	88
E-mail	278	145	82	51	55	51	56	62	54	144	134
Shopping / Booking	231	110	68	53	52	33	49	50	47	121	110
Online Banking	213	105	62	46	41	37	44	52	39	109	104
k.A.	12	10	2	0	2	3	4	1	2	9	3

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Table 8											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Bottom-2-Box sehr selten + selten											
Recherche/Ratgeber	170	84	48	38	36	30	29	39	36	91	79
Videos/ Musik	171	84	54	33	34	30	36	39	32	87	84
E-mail	154	74	38	42	35	27	31	32	29	79	75
Shopping / Booking	89	48	25	16	15	22	19	20	13	47	42
Online Banking	114	49	33	32	30	24	19	19	22	57	57
k.A.	53	29	17	7	9	9	11	9	15	23	30

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 5.4: Top2-/ Bottom2-Boxes

5.5 Mittelwerte

Sobald eine Frage eine Skala enthält, möchte man in der Regel auch einen Mittelwert ausweisen. Das kann in GESStabs denkbar einfach, direkt im Zuge einer Tabellieranweisung erledigt werden.

Als Beispiel ziehen wir noch einmal die Variable f10.1.1 heran³.

Einen Mittelwert fordern wir im TABLE-Statement mit dem Schlüsselwort MEAN und der Spezifikation der gewünschten Variable an:

³Siehe [Top\(2\) und Bottom\(2\) Boxes](#).

```
TABLE = #K1 BY f10.1.1 MEAN(f10.1.1);
```

Hinweis: Es empfiehlt sich, Mittelwerte nur auf Basis der gültigen Antworten zu berechnen und (häufig mit -1 oder 99 vercodete) fehlende Antworten aus der Berechnung auszuschließen. Letztere würden den Mittelwert irrtümlich verfälschen. Dies haben wir in den ersten Zeilen unseres `main.tab` festgelegt. Alle Befragte, welche keine Antwort gegeben haben, sollen automatisch "k.A." erhalten und Missings Values nicht in die Berechnung einfließen: `AUTONOANSWER = YES "{k.A."`; `USEMISSING = no`;

Table 9											
Wie häufig nutzen Sie diese Angebote im Internet?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Recherche/Ratgeber											
sehr häufig	91	45	27	19	21	17	16	18	19	43	48
häufig	79	39	20	20	14	15	18	16	16	37	42
gelegentlich	92	51	25	16	19	16	24	21	12	52	40
selten	86	43	24	19	20	18	13	21	14	46	40
sehr selten	84	41	24	19	16	12	16	18	22	45	39
Top-2-Box (1+2)	170	84	47	39	35	32	34	34	35	80	90
Bottom-2-Box (4+5)	170	84	48	38	36	30	29	39	36	91	79
Recherche/Ratgeber	3,0	3,0	3,0	3,0	3,0	2,9	2,9	3,1	3,0	3,1	2,9

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 5.5: Mittelwerte

In der Zeile, die die Mittelwerte beschreibt, steht nun noch der Variablentitel von f10.1.1. Damit dort korrekterweise "Mittelwert" als Zeilenbeschreibung erscheint, erweitern wir die Tabellier-Anweisung um das `DESCRIPTION`-Stichwort:

```
TABLE = #K1 BY f10.1.1 MEAN :DESCRIPTION "Mittelwert"(f10.1.1);
```

Table 10											
Wie häufig nutzen Sie diese Angebote im Internet?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Recherche/Ratgeber											
sehr häufig	91	45	27	19	21	17	16	18	19	43	48
häufig	79	39	20	20	14	15	18	16	16	37	42
gelegentlich	92	51	25	16	19	16	24	21	12	52	40
selten	86	43	24	19	20	18	13	21	14	46	40
sehr selten	84	41	24	19	16	12	16	18	22	45	39
Top-2-Box (1+2)	170	84	47	39	35	32	34	34	35	80	90
Bottom-2-Box (4+5)	170	84	48	38	36	30	29	39	36	91	79
Mittelwert	3,0	3,0	3,0	3,0	3,0	2,9	2,9	3,1	3,0	3,1	2,9

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 5.6: Mittelwerte

Dies könnte nun für alle folgenden Items des Fragekomplexes fortgeführt werden. Als Ergebnis erhielten wir zehn Tabellenseiten, auf denen jeweils die Skalen-Verteilung plus Mittelwert der einzelnen Items aufgeführt wären. Nicht sehr übersichtlich, oder?

Um die Items auf einen Blick vergleichen zu können, bietet es sich stattdessen an, eine Mittelwerte-Übersicht zu erstellen. Das ist ganz einfach:

```
TABLE = #K1 BY MEAN(f10.1.1) MEAN(f10.2.1) MEAN(f10.3.1) MEAN(f10.4.1) MEAN(f10.5.1) MEAN(f10.6.1) MEAN(f10.7.1) MEAN(f10.8.1) MEAN(f10.9.1) MEAN(f10.10.1);
```

Hier haben wir die :DESCRIPTION-Anweisung wieder herausgenommen und nutzen so den bereits vergebenen Title der einzelnen Items. Hier die Tabelle:

Table 11											
Wie häufig nutzen Sie diese Angebote im Internet?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	76	87	94	83	223	209
Recherche/Ratgeber	3,0	3,0	3,0	3,0	3,0	2,9	2,9	3,1	3,0	3,1	2,9
Videos/ Musik	2,8	2,8	2,9	2,5	2,7	2,7	2,9	2,9	2,7	2,8	2,8
E-mail	2,6	2,5	2,5	2,8	2,7	2,6	2,7	2,5	2,6	2,6	2,6
Shopping / Booking	2,6	2,7	2,6	2,4	2,4	3,0	2,7	2,6	2,4	2,6	2,6
Online Banking	2,5	2,5	2,5	2,6	2,7	2,6	2,5	2,4	2,5	2,5	2,5

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 05.04.2018

Abb. 5.7: Mittelwerte-Übersicht

5.6 Sortieren nach Häufigkeiten

Ebenfalls der Übersichtlichkeit halber bietet es sich häufig an, eine tabellierte Variable nach Häufigkeit ihrer Nennungen zu sortieren. Beispielhaft skripten wir dies für Frage f9 aus unserem Datensatz.

Für eine absteigende Sortierung der absoluten Werte (des Totals) muss die bekannte Tabellierungsanweisung nur um drei Schlüsselwörter ergänzt werden:

```
TABLE = #K1 BY f9 SORT ABSOLUTE DESCEND;
```

Table 12											
Für welche Zwecke nutzen Sie das Internet?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	76	87	94	83	223	209
Bitte klicken Sie alles zutreffende an.											
Recherche/Ratgeber	210	102	61	47	42	37	44	44	43	116	94
Soziale Netzwerke (Facebook, Parship, XING)	173	77	55	41	32	35	37	38	31	96	77
Chats & Foren	173	92	43	38	34	30	35	36	38	90	83
Online Medien (Zeitschriften, Magazine, Blogs)	173	82	51	40	35	36	32	39	31	88	85
Online Gaming	164	84	43	37	36	26	34	38	30	92	72
Sonstiges	159	80	45	34	32	29	28	33	37	79	80
Videos/ Musik	159	84	41	34	35	31	29	40	24	76	83
E-mail	158	80	50	28	34	37	30	39	18	74	84
Ämter / Verwaltung (Uni, Steuer)	148	87	41	20	33	24	31	34	26	75	73
Online Banking	146	75	34	37	31	23	32	27	33	72	74
Shopping / Booking	135	67	40	28	33	25	29	25	23	75	60

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 10.04.2018

Abb. 5.8: Absteigend sortierte Tabelle

5.7 Recode

Häufig enthalten Umfragen auch numerische Fragen. In der Tabellierung macht es meist keinen Sinn, sämtliche Einzelnennungen aufzuführen. Stattdessen können die Einzeleingaben der Befragten für die Auswertung in Kategorien zusammengefasst werden⁴. Zu diesem Zwecke steht in GESStabs der Befehl RECODE bereit. Ein RECODE erlaubt die Umkodierung einzelner Variablenausprägungen der zuletzt definierten Variable oder einer Liste explizit benannter Variablen.

⁴Hierbei muss natürlich beachtet werden, dass eine Gruppierung die ursprüngliche Verteilung der Daten nicht verzerrt.

Als Beispiel nehmen wir die Frage f7.1. Diese ist wie folgt in unser automatisch generierten var.inc beschrieben:

```
#macro #f7(&p1, &title, &col, &width)
SingleQ f7&p1 = &col &width;
TEXT = "Wie viel Zeit haben Sie gestern aktiv im Internet verbracht?";
TITLE = "&title";
#endmacro

#f7(1 "Stunden:" 493 4)
```

Als Eingabe sind ganze Zahlen zwischen 0 und 24 (Stunden) erlaubt gewesen. Ohne einen RECODE wird jede gemachte Eingabe abgebildet:

Table 13											
Wie viel Zeit haben Sie gestern aktiv im Internet verbracht?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Stunden:											
1	19	11	5	3	4	2	4	4	5	10	9
2	37	17	13	7	8	6	8	7	8	24	13
3	40	23	12	5	10	7	10	6	7	13	27
4	99	47	25	27	24	19	23	21	12	56	43
5	57	29	18	10	11	13	8	18	7	24	33
6	14	4	6	4	2	4	2	4	2	9	5
7	48	29	10	9	9	10	8	7	14	23	25
8	21	13	2	6	7	5	2	4	3	11	10
9	13	6	6	1	3	3	3	1	3	5	8
10	23	14	4	5	5	2	5	6	5	13	10
11	17	9	2	6	2	2	3	4	6	12	5
12	18	7	8	3	3	2	5	4	4	9	9
13	4	1	2	1	0	0	1	1	2	3	1
16	12	6	4	2	0	1	2	5	4	5	7
20	10	3	3	4	2	2	3	2	1	6	4

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 10.04.2018

Abb. 5.9: vor dem Recode

Um die Einzel-Werte zu gruppieren, erstellen wir erst einmal eine neue Variable, genauer gesagt eine identische Kopie der f7.1. Natürlich könnten wir auch direkt die ursprüngliche Variable bearbeiten, jedoch gingen dann die Originalwerte verloren⁵. Es ist sicherer – und nicht aufwändiger –, eine neue Variable zu generieren.

```
COMPUTE f7r = f7.1;
COPYTEXT f7r = f7.1;
COPYTITLE f7r = f7.1;
```

Mit dem Befehl COMPUTE wird eine neue Variable erstellt (f7r) in welche die ursprünglichen Werte (aus f7.1) kopiert werden. COPYTEXT und COPYTITLE kopieren die Texte der Ausgangsvariable und weisen sie der neuen Variable zu.

Um nun Gruppen zusammenzufassen, weisen wir die Ausgangswerte (0 bis 24) neuen Labels zu:

```
RECODE f7r
0:5 = 1/
6:10 = 2/
11:15 = 3/
16:20 = 4/
21:24 = 5
;
```

⁵Außerdem werden wir die Originalwerte noch benötigen.

Anschließend vertexten wir die neuen Labelcodes:

```

LABELS f7r =
1 "0 bis 5 Stunden"
2 "6 bis 10 Stunden"
3 "11 bis 15 Stunden"
4 "16 bis 20 Stunden"
5 "21 bis 24 Stunden"
;
    
```

Tabelliert sieht die Verteilung der neuen Variable so aus:

```

TABLE = #K1 BY f7r MEAN(f7.1);
    
```

Table 14											
Wie viel Zeit haben Sie gestern aktiv im Internet verbracht?											
Abs.	Total	Gerätnutzung			Wohnsitz					Geschlecht	
		nur ich alleine	zwei weitere Personen	drei weitere Personen oder mehr	Deutschland	Österreich	Schweiz	Übriges europäisches Ausland	In einem außer-europäischen Land	Männlich	Weiblich
Total	432	219	120	93	90	78	87	94	83	223	209
Stunden:											
1 bis 3 Stunden	96	51	30	15	22	15	22	17	20	47	49
4 bis 6 Stunden	170	80	49	41	37	36	33	43	21	89	81
7 bis 9 Stunden	82	48	18	16	19	18	13	12	20	39	43
10 bis 15 Stunden	62	31	16	15	10	6	14	15	17	37	25
15 bis 24 Stunden	22	9	7	6	2	3	5	7	5	11	11

GESStabs Gesellschaft für Software in der Sozialforschung mbH * Zufallszahlen 10.04.2018

Abb. 5.10: Mit Recode

Noch einen Kommentar zum TABLE-Statement: Den Mittelwert berechnen wir natürlich auf Basis aller Einzelnennungen aus der Ursprungsvariable. Ein Mittelwerte auf Grundlage der gruppierten f7r würde nur die Labelcodes zwischen 1 und 5 zur Mittelwerteberechnungen heranziehen – mathematisch richtig, inhaltlich falsch.

6 Tabellenlayout

Nachdem wir uns den Grundlagen des Tabelleninhaltes gewidmet haben, kommen wir nun zu einem ebenso wichtigen Teil der Auswertung, nämlich der Gestaltung des Tabellenbandes und der einzelnen Tabellen.

Von der Schriftfarbe und -größe über die Farbgestaltung sowie das Einfügen von Logos und Bildern, der Positionierung der Tabelle auf dem Blatt sowie Einstellung einzelner Abstände kann mit der GESStabs-Syntax das gesamte Tabellenlayout individuell eingestellt werden. An dieser Stelle können wir lediglich auf die Grundlagen des Tabellendesigns eingehen. Möchten Sie das Layout Ihres Tabellenbandes neu aufsetzen, stehen wir Ihnen natürlich gern unterstützend zur Seite!

6.1 Tabellen-übergreifende Formatierung

Grundsätzlich bündeln wir tabellenübergreifende Formatierungsanweisungen in einer separaten Format-Datei, hier der `tutorial.fmt`. Alle Anweisungen gelten so lange, bis sie neu definiert werden¹.

Zu den ganz grundlegenden Formatierungsanweisungen zählen die folgenden:

```
NUMBERCHAR    = #;
SPLITCHAR     = ~;
SPLITCHARSTAY = -;
LINEFEEDCHAR  = \;
```

`NUMBERCHAR` gilt als numerischer Platzhalter (s.u.). Wird dieser z.B. in Seitenzahlen verwendet, wird automatisch hochgezählt.

`SPLITCHAR` und `SPLITCHARSTAY` legen Trennzeichen fest: Während letzteres einen dauerhaften Umbruch markiert, stellt `SPLITCHAR` ein bedingtes Trennzeichen dar und markiert die Stelle, an der umgebrochen wird, wenn es notwendig ist².

`LINEFEEDCHAR` erzwingt einen Zeilenumbruch.

Fundamental ist auch `USEFONT`: Hiermit werden Schriftfarbe und -größe festgelegt. Für einen reibungslosen GESStabs-Lauf müssen wir dies einmal generell festlegen:

```
USEFONT = "Helvetica" SIZE 9;
```

Zudem kann `USEFONT` für einzelne Tabellenelemente definiert werden, z.B. `USEFONT ABSOLUTE = "Helvetica" SIZE 8;`, `USEFONT MEAN = "Helvetica-Bold" SIZE 10;` usw..

Mit `FORMAT` werden die Zahlenformate der unterschiedlichen Maßzahlen bestimmt. Wünscht man etwa generell Mittelwerte mit zwei Nachkommastellen, schreibt man:

```
FORMAT MEAN = '#,##';
```

`ALIGN`, `DISTANCE` oder `FRAMECOLOR` sind weitere Schlagworte, mit denen beispielsweise Boxenhalte ausgerichtet, Abstände definiert oder Rahmenfarben festgelegt werden. Eine vollständige Formatierungsdatei steht in unserem Beispielprojekt zur Verfügung.

¹An dieser Stelle sei noch einmal auf die lineare Abarbeitung des Skiptes durch die Software hingewiesen: Formatierungsbefehle gelten für alle nachfolgenden Skriptzeilen und so lange, bis der Formatierungsbefehl neu definiert wird. Dies gilt dann wieder für alle nachfolgenden Skriptteile usw..

²An dieser Stelle wird dann das unter `SPLITCHARSTAY` definierte Zeichen eingesetzt.

6.2 Tabellengestaltung

Eine Tabelle setzt sich aus verschiedenen Bereichen zusammen. Hier eine Übersicht über die Struktur einer typischen GESStabs-Tabelle:

Das Diagramm zeigt die hierarchische Struktur einer Tabelle. Die Hauptebenen sind:

- Header**: Umfasst die gesamte Tabelle.
- Tabletitle**: Ein zentraler Titelbereich.
- Text Vorspalte**: Ein Bereich für Text vor den Spalten.
- Titel Kopfzeile**: Eine Zeile mit Spaltenüberschriften.
- Col % Abs.**: Spaltenüberschriften mit Prozentangaben und Absolutwerten.
- Total**: Eine Zeile mit Gesamtwerten.
- Titel Vorspalte**: Überschriften für die Zeilen.
- Label 1-6**: Zeilenüberschriften für die Hauptdaten.
- Institution**: Ein Bereich für den Namen der Institution.
- Document**: Ein Bereich für eine Anmerkung oder Dokumentnummer.
- Footer**: Ein Bereich am unteren Rand der Tabelle.

Tabletitle		Titel Kopfzeile	
Col % Abs.	Total	Label Kopf 1	Label Kopf 2
Total	100 % 432	100 % 223	100 % 209
Titel Vorspalte			
Label 1	16 % 67	17 % 38	14 % 29
Label 2	17 % 75	15 % 33	20 % 42
Label 3	17 % 73	16 % 36	18 % 37
Label 4	16 % 69	16 % 35	16 % 34
Label 5	16 % 70	17 % 37	16 % 33
Label 6	18 % 78	20 % 44	16 % 34

Abb. 6.1: Übersicht Tabellengestaltung

HEADER und FOOTER werden üblicherweise für den ganzen Tabellenband festgelegt, in letzterem steht üblicherweise die Seitennummer:

```
FOOTER = "-#-";
```

Auch die Positionen INSTITUTION und DOCUMENT ändern sich in der Regel nicht innerhalb eines Tabellenbandes. Hier können beispielsweise der Institutsname und eine beliebige weitere Anmerkung platziert werden:

```
INSTITUTION = "GESS Gesellschaft für Software in der Sozialforschung mbH";
DOCUMENT = "based on randomized numbers \ "DATE;
```

Letzteres haben wir mit DATE ergänzt – hier wird das jeweils aktuelle Datum der Tabellengenerierung eingesetzt.

Im TABLETITLE stehen meist der Name der Studie und/ oder eine thematische Zuordnung. Etwa:

```
TABLETITLE = "Demostudie: GESStabs-Tutorial";
```

Darunter steht der TOPTXT der Vorspalten-Variable³

Linksstehend findet man eine Beschreibung der Aufrissvariable: gegebenfalls ist hier der TITLE angezeigt und die Labelcodes- oder (wenn vergeben) -texte werden aufgelistet. Nach demselben Schema wird in horizontaler Richtung die Kopfvariable aufgeführt.

³Theoretisch kann hier auch etwas anderes stehen, siehe die entsprechende Einstellung in [Das main.tab](#).

In welcher Form die Daten innerhalb der Tabelle präsentiert werden, kann individuell gesteuert werden.

Den Rahmen bilden die sogenannten `FRAMEELEMENTS`. In unserem Schema gibt es eine Totalzeile und -spalte. Diese Elemente sind nicht automatisch vorhanden, sondern werden mit folgendem Befehl hinzugefügt:

```
FRAMEELEMENTS = TOTALROW TOTALCOLUMN;
```

Was innerhalb der einzelnen Datenzeile stehen soll, welche Maßzahlen also für die abgebildeten Daten ausgewiesen werden sollen, steuert das Befehlsstichwort `CELLELEMENTS`. Spaltenprozent und Absolutwerte werden etwa wie folgt angefordert:

```
CELLELEMENTS = COLUMNPERCENT ABSOLUTE;
```

Wie im vorherigen Abschnitt [6.1](#) beschrieben, wird die Formatierung der einzelnen Maßzahlen mit dem `FORMAT`-Schlagwort bestimmt, zum Beispiel:

```
FORMAT COLUMNPERCENT = "`#,# \%"';
```

Schließlich können wir noch steuern, wie die Zelleninhalte beschrieben werden. Dies erscheint dann im oberen linken Tabellenbereich, unter dem Fragetext. Zum Beispiel:

```
DESCRIPTION COLUMNPERCENT = "Col \%";  
DESCRIPTION ABSOLUTE      = "Abs.";
```

Stichwortverzeichnis

- Align, 24
- ASCII, 2, 4
 - Multinennungen, 5
 - Singlenennungen, 4
 - var.inc, 4, 5, 22
- Autonoanswer, 3
- Blankvalue, 3
- By, 10
- Cellelements, 26
- Citefirstvar, 3
- Codebook, 8
- Compute, 22
- compute.inc, 2, 15
- Copytext, 22
- Copytitle, 22
- CSV, 4, 9
- CSVINFILE, 9
- Date, 25
- Description, 20, 21
- Distance, 24
- Document, 25
- Encoding, 9
- end;, 3
- Erste Schritte, 1
- Expand, 13
- Footer, 25
- Format, 24, 26
- Formatierung, 24
- Framecolor, 24
- Frameelements, 26
- GESStabs-Oberfläche, 10
- Groups, 15
- Häufigkeitssortierung, 21
- Header, 25
- Include, 2
- Institution, 25
- kopf.inc, 2, 13
- Kopfvariable, 13
- Kreuztabelle, 10
- Labels, 5, 8, 15
- Layout, 24
- Linefeedchar, 24
- main.tab, 2, 11, 20
- Make PDF, 11
- Makro, 6
- Mean, 19
- Missing, 3, 20
- Mittelwerte, 19
- Mittelwerte-Übersicht, 20
- MultiQ, 5
- Notepad++, 1
- Numberchar, 24
- Overcode, 15, 17
- PDF, 1, 11
- PostScript, 1, 3, 11
- Printfile, 3
- Recode, 21
- run again, 12
- Select tab-file and run, 11
- Setmissing, 3
- SingleQ, 4
- Sort, 21
- Splitchar, 24
- Splitcharstay, 24
- SPSS, 1, 4, 7
- SPSSINFILE, 7
- Syntax, 7
- tab.inc, 2, 10
- Tabellengestaltung, 25
- Table, 10, 13, 19
- Tabletitle, 25
- Text, 4
- Title, 4, 6, 8
- Top-, Bottom-Boxes, 17, 18
- Toptext, 25
- tutorial.fmt, 24
- Usecases, 3
- Usefont, 24
- Usemissing, 3
- UTF-8, 9
- Vargroup, 18