

# GESStabs

---

## Invertierte Datenspeicherung



Gesellschaft für Software  
in der Sozialforschung mbH

Waterloohain 6 - 8  
22769 Hamburg  
Tel.: +49 40 - 853 753 - 0  
Fax: +49 40 - 853 753 - 33  
[www.gessgroup.de](http://www.gessgroup.de)

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
1.1	Definition und Struktur invertierter Datensätze . . . . .	1
1.2	Basissyntax . . . . .	1
<b>2</b>	<b>Anwendung</b>	<b>2</b>
2.1	Erzeugen invertierter Datensätze . . . . .	2
2.2	Modifikation der gespeicherten Datenmenge . . . . .	2
2.3	Einlesen invertierter Datensätze . . . . .	2
2.4	Speichern und Verwenden von Fallgewichten . . . . .	3
2.5	Besonderheiten . . . . .	4
2.5.1	Alpha-Variablen . . . . .	4
2.5.2	Mehrfachnennungsvariablen . . . . .	4
2.5.3	Web-Anwendungen . . . . .	4
2.5.4	Aktualisiertes INVERTOUT . . . . .	4
<b>3</b>	<b>Syntax</b>	<b>5</b>

# 1 Einführung

Als proprietäres Dateiformat kann GESStabs invertierte Datensätze schreiben und lesen.

Der Vorteil bei der Arbeit mit invertierten Datensätze sind wesentlich verkürzte Verarbeitungszeiten bei Datensätzen mit einer hohen Fallzahl, für deren Auswertung eine vergleichsweise geringe Variablenzahl benötigt wird. Interaktive Anwendungen sind ein typischer Anwendungsfall, bei dem in kurzer Zeit einzelne Auswertungen aus einem begrenzten Variablen-Pool eines großen Datensatzes abgebildet werden sollen.

## 1.1 Definition und Struktur invertierter Datensätze

In den üblichen Dateiformaten, die in GESStabs mittels `DATAFILE`, `CSVINFILE` oder `SPSSINFILE` eingebunden werden, sind die Daten fallweise angeordnet: Die Dateien enthalten sequenziell Interviewfälle und zusammenhängend deren dazugehörige Werte. Bei invertierter Speicherung sind die Daten variablenweise angeordnet: Alle Werte einer einzelnen Variable sind zusammenhängend in einem großen Block gespeichert. Werden nur wenige Variablen benötigt, erlaubt diese Speicherform ein sehr effizientes Verarbeiten des Datensatzes.

Der Inhalt invertierter Datensätze ist auf numerische Werte beschränkt. Dies hat bspw. Auswirkungen auf den Übertrag von Alpha-Variablen, die im Zuge der Invertierung umgeformt werden müssen. Siehe hierzu den Abschnitt über [Alpha-Variablen](#).

## 1.2 Basissyntax

Mit `INVERTOUT` wird eine invertierte Datei herausgeschrieben und benannt, ähnlich wie bei `SPSSOUTFILE`:

```
INVERTOUT = "<filename/-path>";
```

Invertierte Datensätze sind reine Daten-Dateien, sie enthalten weder Variablentitel, -texte noch Variablenlabels. In dieser Hinsicht ist das Dateiformat mit CVS-Datensätzen<sup>1</sup> vergleichbar.

`INVERTIN` liest eine invertierte Datei ein. Im Anschluss an den Befehl sind die Variablennamen und Daten im GESStabs-Skript bekannt.

```
INVERTIN = "<filename/-path>";
```

<sup>1</sup>Mit zwei Unterschieden: Einerseits wird der Zustand fehlender Werte (das globale `SETMISSING` und das variablenbezogene `MISSING`) in invertierte Dateien übertragen. Andererseits werden durch das `CSVINFILE`-Statement die atomaren Bestandteile von Mehrfachnennungsvariablen wieder zusammengesetzt.

## 2 Anwendung

### 2.1 Erzeugen invertierter Datensätze

Syntax:

```
INVERTOUT = "<filename/-path>;
```

Während eines GESStabs-Laufs kann ein invertierter Datensatz erzeugt werden. Unabhängig von der Platzierung des INVERTOUT-Statements im Skript wird die Datei nach dem Einlesen aller Skriptzeilen erstellt. Alle Variablen, die während eines GESStabs-Laufs erzeugt werden – durch Einlesen eines anderen Datensatzes oder mittels `COMPUTE` –, fließen in den invertierten Datensatz ein. Der Datensatz erhält die Extension “inv“, die die Software automatisch setzt.

Unter der Voraussetzung, dass der Schalter `NOINVERTADDON` auf `NO` steht, werden zusätzlich zur binären Datendatei eine Beschreibung der Studie in JSON-Format sowie GESStabs-SYNTAX-Form ausgegeben<sup>1</sup>.

Invertierte Datensätze können zudem komprimiert abgespeichert werden:

```
ZIPINVERTOUT = [YES | NO];
```

### 2.2 Modifikation der gespeicherten Datenmenge

Mit `INVERTOUTVARS` kann der Datensatz auf ausgewählte Variablen beschränkt werden. Mit `DELETEVARS` werden alle aufgezählten Variablen ausgeschlossen, `KEEPVARS` speichert nur die genannten Variablen.

```
INVERTOUTVARS [KEEPVARS | DELETEVARS] = <varlist>;
```

Es besteht zudem die Möglichkeit, die Blockgröße invertierter Dateien anzupassen. Diese beträgt im Standardfall 16.384 Fälle und kann optional geändert werden:

```
INVERTOUTMAX = <number>;
```

### 2.3 Einlesen invertierter Datensätze

Invertierte Datensätze werden mit dem Befehlswort `INVERTIN` eingelesen.

```
INVERTIN = "Datensatz";
```

GESStabs prüft in diesem Zuge, ob sich in dem Dateiverzeichnis ein zugehöriges `INCLUDE`-File (mit Dateiondung “.inc“) mit Variablenbeschreibungen befindet – ist dies der Fall, wird diese Datei automatisch interpretiert und die enthaltenen Variablen entsprechend vertextet. Voraussetzung ist, dass der Schalter `NOINVERTADDON` auf `NO` gesetzt ist<sup>2</sup>.

Das `INVERTIN`-Statement erkennt, ob es sich bei der eingelesenen Datei um komprimierte Daten handelt oder nicht.

<sup>1</sup>Siehe hierzu die Abschnitte [2.5.2](#) und [2.5.3](#).

<sup>2</sup>Siehe hierzu die Abschnitte [2.5.2](#) und [2.5.3](#).

## 2.4 Speichern und Verwenden von Fallgewichten

In Variablen abgelegte Fallgewichte werden beim Export eines invertierten Datensatzes automatisch gespeichert. Für das Abspeichern von Fallgewichten, die mit iterativer Randgewichtung erzeugt wurden<sup>3</sup>, gibt es ein spezielles Statement<sup>4</sup>:

```
INVERTFILEWEIGHTOUT = <varname>;
```

Das systemweite Gewicht wird im Zuge des Datensatz-Exports in die genannte Variable geschrieben. Existiert diese bereits, wird sie überschrieben, ansonsten wird eine neue Variable angelegt.

Sukzessive iterative Gewichtungen würden beispielsweise so ausgeführt:

a.tab:

```
INFILE = <ascii.dat>;
SINGLEQ = TEXT ... LABELS ...
SINGLEQ = TEXT ... LABELS ...
WEIGHTCELLS = ...
NOINVERTADDON = NO;
INVERTFILEWEIGHTOUT = Gewicht1;
INVERTOUT = stufe1;
END;
```

b.tab:

```
NOINVERTADDON = NO;
INVERTIN = stufe1;
INCLUDE = stufe1.inc;
WEIGHT = Gewicht1;
WEIGHTCELLS = ...
INVERTFILEWEIGHTOUT = Gewicht2;
INVERTOUT = stufe2;
END;
```

c.tab:

```
NOINVERTADDON = NO;
INVERTIN = stufe2;
INCLUDE = stufe2.inc;
WEIGHT = Gewicht2;
OPTIMIZE = YES;
TABLE ...
TABLE ...
END;
```

Die invertierte Datei "stufe2.inv" enthält nach Ablauf des a.tab und b.tab beide Gewichte (Gewicht1 und Gewicht2) parallel. Um den Geschwindigkeitsvorteil der invertierten Struktur zu nutzen, sollte im c.tab die Option OPTIMIZE genutzt werden. Damit die Variablentexte, -titel und Wertelabel in der Tabellierung erscheinen, werden die im a.tab definierten Informationen nach Einstellung von NOINVERTADDON = NO; automatisch herausgeschrieben und in den folgenden Schritten wieder eingelesen.

<sup>3</sup>WEIGHTCELLS

<sup>4</sup>analog zur Vorgehensweise von SPSSOUTFILE

## 2.5 Besonderheiten

### 2.5.1 Alpha-Variablen

Alpha-Variablen können nicht 1:1 in die Struktur einer invertierten Datei übertragen werden, weil diese ausschließlich numerische Variablenwerte enthält. Beim Einlesen von Alpha-Variablen aus einem spaltenfixierten oder CSV-Datensatz wird daher jedem Text ein eindeutiger Code zugeordnet – eine Art Wertelabel, das beim Einlesen des Datensatzes aus den Daten generiert wird. Die Zuordnung von Code zu Text bleibt erhalten, indem diese Information via **SYNTAX** in eine **INCLUDE**-Datei übertragen wird. Da diese Vercodung dem System jedoch erst nach dem Verarbeiten des gesamten GESStabs-Skripts und der Invertierung bekannt ist, wird die automatische Generierung der **INCLUDE**-Datei bis zum Ende des Einlesevorgangs verzögert.

### 2.5.2 Mehrfachnennungsvariablen

Auch Mehrfachnennungsvariablen (einschl. dichotomer Variablen) können nicht unmittelbar in das invertierte Format übertragen werden. Sie werden daher im Zuge des Exports zunächst in ihre atomaren Bestandteile geteilt und in das invertierte Dateiformat übertragen. Anstelle der MultiQ “Gründe“ weist der invertierte Datensatz die atomaren Variablen “Gründe\$1“, “Gründe\$2“ und “Gründe\$3“ auf. Ist die automatische Erzeugung eines GESStabs-**SYNTAX**-Skripts im Zuge der Invertierung aktiviert<sup>5</sup>, werden die Einzelvariablen nach dem Einlesen der invertierten Datei wieder zu einer MultiQ zusammengesetzt. Wurde diese Steuerung deaktiviert<sup>6</sup>, muss der Anwender dies händisch erledigen: `VARGROUP Gründe = ("Gründe$1" "Gründe$2" "Gründe$3") EQ 1;.`

### 2.5.3 Web-Anwendungen

Sollen invertierte Datensätze in interaktiven (Web-)Anwendungen eingesetzt werden, können die dafür nötigen Basisfunktionen im Zuge von **INVERTOUT** automatisch in JSON-Format exportiert werden. Voraussetzung hierfür ist die folgende Einstellung: `NOINVERTADDON = NO;`

### 2.5.4 Aktualisiertes **INVERTOUT**

Findet auf Basis eines **INVERTIN**-Files ein Tabellierlauf statt, kann mit **UPDATEINVERT**; gleichzeitig eine neue invertierte Datendatei herausgeschrieben werden. Der aktualisierte Datensatz erhält denselben Namen wie das einbezogene **INVERTIN**-File. Der zur Eingabe benutzte invertierte Datensatz wird archiviert und erhält einen Dateinamen, der sich aus dem ursprünglichen Dateinamen plus fortlaufender Archivnummer (beginnend bei “\_1“, wird diese mit jedem weiteren **UPDATEINVERT** erhöht) zusammensetzt. Parallel dazu werden entsprechende JSON- und **SYNTAX**-Dateien nach demselben Schema geschrieben und archiviert.

GESStabs prüft nicht, ob ein solches Update inhaltlich sinnvoll ist. Es entstehen also möglicherweise lediglich Kopien bestehender Datensätze und Zusatzdateien.

Beachte: Ein gleichzeitiges **INVERTOUT** und **UPDATEINVERT** in einem GESStabs-Lauf sind nicht erlaubt.

---

<sup>5</sup>`NOINVERTADDON=NO;`

<sup>6</sup>`NOINVERTADDON=YES;`

## 3 Syntax

Schreiben eines invertierten Datensatzes:

```
INVERTOUT = "<filename/-path>";
```

Komprimieren eines invertierten Datensatzes:

```
ZIPINVERTOUT = [YES | NO];
```

Auswahl der zu exportierenden Variablen:

```
INVERTOUTVARS [KEEPVARS | DELETEVARS] = <varlist>;
```

Anpassen der Blockgröße eines invertierten Datensatzes:

```
INVERTOUTMAX = <number>;
```

Herausschreiben von durch iterative Randverteilung erzeugten Fallgewichten

```
INVERTFILEWEIGHTOUT = <varname>;
```

Einlesen eines invertierten Datensatzes:

```
INVERTIN = "<filename/-path>";
```

Ein-/ Ausschalten des automatischen Generierens bzw. Einlesens von JSON- und SYNTAX-Dateien:

```
NOINVERTADDON = [NO | YES];
```

Herausschreiben eines neuen INVERTOUT-Files bei gleichzeitiger Archivierung des einbezogenen invertierten Datensatzes:

```
UPDATEINVERT;
```