

Organisation der Auswertungsprogramme

für eine Datenfernverarbeitungsanfrage zur Nutzung der Daten des IAB-Betriebspanels im Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB

Die übersendeten Auswertungsprogramme müssen zur Minimierung des Bearbeitungsaufwandes und damit zur möglichst raschen Übersendung der Ergebnisse eine bestimmte Ablaufstruktur aufweisen. Dem Lesen und Schreiben von Datensätzen kommt dabei elementare Bedeutung zu.

Für jede Anfrage einer bestimmten Person oder Organisation wird auf einem FDZ-internen Server ein Ordner vergeben. Die Nomenklatur ist immer Nachname, Vorname (Kürzel der Organisation) wie beispielsweise in Musterfrau, Martha (ABC). Zur Vermeidung datenschutzrechtlicher Komplikationen ist als Pfad lediglich X:\Nutzer anzugeben, die tatsächliche Pfadangabe wird intern im FDZ in das jeweilige Auswertungsprogramm eingetragen.

Die laut Anfrageformular genutzten Originaldatensätze werden nach der Nomenklatur der Testdaten aufgerufen. Dafür werden die Originaldaten schreibgeschützt in das jeweilige Nutzerverzeichnis eingestellt.

Beispiele:

Für eine Anfrage in SPSS wäre zum Einlesen des 2002er-Datensatzes folgende Befehlszeile anzusetzen:

```
GET FILE='X:\Nutzer\Musterfrau, Martha (ABC)\test02.sav'.
```

In STATA wird die Verwendung des cd-Befehls empfohlen, also

```
cd "X:\Nutzer\Musterfrau, Martha (ABC)"  
use test02.dta, clear
```

Dann kann auch das Anlegen des log-files einfach durch den Befehl

```
log using ergebnisse.log, replace
```

 erfolgen.

Zusätzlich sollten in STATA einige Parameter bevorzugt wie folgt an den Kopf eines *.do-files (optional des Mutter-Do-Files) gesetzt werden:

```
clear  
capture log close  
set mem 200m  
set matsize 800  
set more off  
set logtype text
```

An diese Befehlsabfolge schließt sich meist der cd-Befehl an.

(Ende Beispiele)

Es ist möglich, innerhalb eines Nutzerverzeichnis weitere Unterverzeichnisse anzulegen. Diese und das Lesen und Schreiben von Datensätzen in solche Unterordner sollten in einer begleitenden E-Mail kurz beschrieben werden.

Nach dem Durchlauf des Auswertungsprogramms werden die Originaldaten aus dem Nutzerverzeichnis gelöscht. Es besteht jederzeit die Möglichkeit, Datensätze im Nutzerverzeichnis abzuspeichern und bei Bedarf wieder auf die eigens generierten Datensätze zurückzugreifen. Obligatorisch werden solche Datensätze drei Monate ab Erstellungsdatum im Nutzerverzeichnis aufbewahrt. Auf Antrag kann auch eine längere Lagerung (maximal ein Jahr) vorgenommen werden. Die datenschutzrechtlichen Bestimmungen verlangen dann aber eine *eindeutige* Benennung der Speicherfrist. Unbestimmte Speicherfristen wie beispielsweise „bis auf Weiteres“ oder „bis Ablauf des Projekts“ können nicht berücksichtigt werden. Nach drei Monaten werden solche Datensätze zur Freigabe von Ressourcen aus den Nutzerverzeichnissen gelöscht.

Ausgabe- bzw. Ergebnisdateien (*.log in STATA und *.spo in SPSS) werden normalerweise im Arbeitsverzeichnis der Nutzer abgelegt. Vor der Übersendung der Ergebnisse erfolgt eine Datenschutzprüfung. Oberstes Gebot ist dabei die Vermeidung von Reidentifikationsrisiken für die befragten Betriebe. Ungewichtete Fallzahlen kleiner vier in einzelnen Auswertungszellen - auch in hochgerechneten Ergebnissen - werden nicht ausgewiesen und durch einen "/" gekennzeichnet. Es ist in Einzelfällen möglich, dass dann auch die Gesamtzahl der analysierten Betriebe nicht ausgewiesen werden kann, nämlich, wenn durch diese Zahl auf die nicht ausgewiesenen Betriebe in einer bestimmten Auswertungszelle geschlossen werden kann.

Die Genauigkeit von Hochrechnungen hängt von der Anzahl der zugrunde liegenden Fälle ab. Daher werden Ergebnisse, auch hochgerechnete, die auf einer ungewichteten Fallzahl von mehr als drei und weniger als 21 Betrieben beruhen, gesondert eingeklammert. Solche Ergebnisse sollten vorsichtig interpretiert werden. Beim Auftauchen kritischer Fallzahlen erfolgt eine Konvertierung der Ausgabedateien in das Excel-Format. In diesem erfolgt dann auch die Übersendung der Ergebnisse. Abweichungen sind nur nach Rücksprache mit und Prüfung durch den jeweiligen Ansprechpartner im FDZ möglich. (Zu) Kleine Fallzahlen können beispielsweise durch die Verwendung weniger feiner Einteilungen vermieden werden. In der Rubrik [Hinweise zur inhaltlichen Arbeit mit den Daten](#) gibt es beispielsweise für Branchen und Betriebsgrößen verschiedene praktikable Vorschläge (Programmierbeispiele in SPSS).