

Hinweise zur inhaltlichen Arbeit mit den Daten des IAB-Betriebspanels

Diese Datei enthält Informationen zur inhaltlichen Arbeit mit den Daten des IAB-Betriebspanels. Dabei wurde versucht, den von Nutzern geäußerten Informationsbedürfnissen entgegen zu kommen. Anregungen, Kritik und Fragen können in einer Mail an das FDZ (iab.fdz@iab.de) unter dem Stichwort User-Support IAB-Betriebspanel formuliert werden.

Die Schalterstelle des IAB-Betriebspanels befindet sich derzeit noch im Probetrieb. Eine umfassendere Übersicht wird erst mit dem Vollbetrieb des FDZ möglich. Wir bitten um Verständnis.

Derzeit sind Informationen zu folgenden Themen verfügbar:

- Erläuterungen zum Gewichtungungsverfahren (Auswahl der Betriebe, Wellenkennzeichen u.ä.)
- SPSS-Beispielsyntax (mit Beispielen zur Branchen- und Betriebsgrößenklasseneinteilung)
- FAQ (anonymisiert)
- Literaturhinweise zum Ansatz, Aufbau und zur Nutzung des IAB-Betriebspanels

Erläuterungen zum Gewichtungsverfahren

Erläuterungen zu den Gruppenkennzeichen

Die Wellenkennzeichen erlauben u.a. die Identifizierung von Erstbefragten und von Panel-Betrieben. Die Variablen befinden sich im Organisationsfile und müssen bei Verwendung zu den Datensätzen hinzugespielt werden. Für die Befragung 1993 ist der Name welle01, für 1994 welle02 usw. . Bitte beachten sie, dass ab dem Jahr 2001 die Variablennamen geändert wurden. Folgende Ausprägungen sind möglich:

| | Kennbuchstaben | |
|---|--|----------------|
| | <i>Mit</i> | <i>ohne</i> |
| | sozialversicherungspflichtig Beschäftigten | |
| 1. Fälle mit Interview in lfd. Welle | | |
| 1.1. Erstmals befragte Betriebe | A | nicht zulässig |
| 1.2. wiederholt befragte Fälle | | |
| 1.2.1. mit Interview im Vorjahr | | |
| gleiche Einheit wie Vorjahr | B | C |
| andere Einheit wie Vorjahr | D | nicht zulässig |
| 1.2.2. ohne Interview im Vorjahr | | |
| selbe Einheit wie Vorjahr | E | nicht zulässig |
| andere Einheit wie Vorjahr | G | nicht zulässig |
| 2. Fälle ohne Interview in lfd. Welle | | |
| 2.1. Ausfall, in Zukunft wieder befragbar | H | |
| 2.2. Fälle aus Aufstockungen, die nicht mehr befragt werden | W | |
| 2.3. Ausfall, nicht mehr befragbar | X | |
| 2.4. Betrieb erloschen | | |
| 2.4.1. in der lfd. Welle | Y | |
| 2.4.2. bereits früher | Z | |

Das Hochrechnungsverfahren

Bei der Gewichtung werden die Sollstrukturen über die Betriebsdatei der Bundesanstalt für Arbeit zum Zeitpunkt der Stichprobenziehung gebildet, also jeweils ein Jahr vor der Befragung. Dies hat für die Hochrechnung insbesondere von Querschnittsauswertungen zur Folge, dass für Gesamtwerte der Beschäftigten (und Betriebe) in der laufenden Welle bei der Hochrechnung jeweils der Wert des Vorjahres (zum Stichtag 30.06.) erreicht wird, so, wie die Amtlichen Nachrichten der Bundesanstalt für Arbeit ihn ausweisen (mit einer Fehlertoleranz von ca. 0,1%, da die Hochrechnungsfaktoren sowohl auf Betriebe wie Beschäftigte angepasst werden). Dies gilt natürlich nur dann, wenn die betriebliche Vorjahresbeschäftigung, die in jedem Datensatz i.d.R. für jeden Betrieb mit gültigem Interview enthalten ist, verwendet wird. Wenn also beispielsweise

die Gesamtbeschäftigung im aktuellen Querschnittsjahr hochgerechnet wird, erhält man *nicht* den exakten Wert, wie er von der Bundesanstalt für Arbeit für das betreffende Jahr zum Stichtag 30.06. ausgewiesen wird. Seit den Aufstockungen durch die Bundesländer werden auch bei Auswertungen auf Bundeslandebene die jeweiligen Landeswerte (des Vorjahres) entsprechend dem vorher Gesagtem erreicht.

Einzelheiten zum Ansatz und Aufbau sowie zur Stichprobenziehung beschreibt:

Bellmann, Lutz: Das IAB-Betriebspanel: Konzeption und Anwendungsbereiche, Allgemeines Statistisches Archiv 86, S. 177-188, Physica-Verlag 2002

Querschnittsgewichtung

In die Querschnittsgewichtung werden alle Fälle einbezogen, die jeweils zum Stichtag 30.06. mindestens einen sozialversicherungspflichtig Beschäftigten hatten. Die zugehörige Definition im Datensatz ergibt sich aus dem Kriterium `quer0X='Q'`, wobei für X die Nummer der betreffenden Welle eingesetzt wird, also z.B. für das erste Befragungsjahr 1993 eine 1 (`quer01='Q'`), für 1996 eine 4 (`quer04='Q'`).

Für Querschnittsauswertungen ist der Querschnittsgewichtungsfaktor `hrf0Xq` einzusetzen, wobei für X analog zur Querschnittsdefinition die jeweilige Wellennummer eingesetzt wird (z.B. für den Datensatz 1996 `hrf04q`).

SPSS-Beispiel für den Datensatz 2000 (=Welle 8):

```
USE ALL.  
COMPUTE filter_$=(quer08 = 'Q').  
VARIABLE LABEL filter_$ "quer08 = 'Q' (FILTER)".  
VALUE LABELS filter_$ 0 'Nicht ausgewählt' 1 'Ausgewählt'.  
FORMAT filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE .
```

```
WEIGHT BY hrf08q .
```

Panelgewichtung

Längsschnittauswertungen ab Welle 1 (1993) sind nur für Westdeutschland möglich, für Ostdeutschland ab Welle 4 (1996). Das Erlöschen von Betrieben gehört mit zu den tatsächlichen Bewegungen und wird durch gesonderte Kennzeichen miterfasst (s. Erläuterungen Gruppenkennzeichen). Für Längsschnittauswertungen muss der Organisationsfile mit den entsprechenden Jahresdatensätzen zusammengefügt werden.

Die zugehörige Definition im Datensatz wird zum einen am Beispiel von 1993 bis 2000 und zum anderen am Beispiel von 1996 bis 2000 vorgestellt. Das Kriterium für 1993 bis 2000 ist `pan01_08='P'`, für 1996 bis 2000 `pan04_08='P'`. Für „Zwischenjahre“ ist der Prozess analog ab 1996 (Beispiel 1996 auf 1998 `pan04_06='P'`), von 1993 bis 1995 wurde eine etwas andere

Nomenklatur verwendet (Beispiel: 1993 auf 1995 panel03='P'). Im Organisationsfile sind nicht alle möglichen Kombinationen enthalten; empfohlen wird, sich vor der Längsschnittbildung die zur Verfügung stehenden Definitionskriterien im Organisationsdatensatz anzuschauen. Verbleiben Fragen, hilft die Schalterstelle gerne weiter.

Die Gewichtung der Längsschnitte berücksichtigt das vorher Gesagte. Um eine optimale Anpassung der Panelstichprobe zu erreichen, wird zweistufig vorgegangen. Im ersten Schritt wird die Hochrechnung durch die Strukturvorgaben der Betriebsdatei der Bundesanstalt für Arbeit vorgenommen, im zweiten Schritt wird mit Hilfe multivariater Analyseverfahren die Ausfallwahrscheinlichkeit zwischen den einzelnen Wellen untersucht. Beide Gewichtungsschritte werden so lange (iterativ) wiederholt, bis eine optimale Anpassung der Werte für Gesamtdeutschland erreicht ist (regionale Merkmale wie West/ Ost oder Bundesländer werden in der multivariaten Analyse berücksichtigt).

Analog zu den Querschnitten wird für Längsschnittauswertung von 1993 bis 2000 ist der Längsschnittgewichtungsfaktor HR01_08P eingesetzt, für Auswertungen von 1996 bis 2000 HR04_08P.

Neubenennung der Variablenamen ab 2002

Von der Umstellung sind folgende Variablen betroffen (alphabetische Sortierung nach Überschriften).

Bisheriger Variablenname (bis 2001)

Neuer Variablenname (ab 2002)

Hochrechnungsfaktoren für Querschnittfälle:

| | |
|--------|---------|
| HRF01Q | HR1993Q |
| HRF02Q | HR1994Q |
| HRF03Q | HR1995Q |
| HRF04Q | HR1996Q |
| HRF05Q | HR1997Q |
| HRF06Q | HR1998Q |
| HRF07Q | HR1999Q |
| HRF08Q | HR2000Q |
| HRF09Q | HR2001Q |

Hochrechnungsfaktoren für Panelfälle ab 1993 (Welle 1):

| | |
|----------|----------|
| HRF02P | HR93_94P |
| HRF03P | HR93_95P |
| HRF04P | HR93_96P |
| HR01_05P | HR93_97P |
| HR01_06P | HR93_98P |
| HR01_07P | HR93_99P |
| HR01_08P | HR93_00P |
| HR01_09P | HR93_01P |

Hochrechnungsfaktoren für Panelfälle ab 1996 (Welle 4):

| | |
|----------|----------|
| HR04_05P | HR96_97P |
| HR04_06P | HR96_98P |
| HR04_07P | HR96_99P |
| HR04_08P | HR96_00P |
| HR04_09P | HR96_01P |

Panelfälle ab 1993 (Welle 1)

| | |
|----------|--|
| PANEL01 | entfällt, da identisch mit Querschnittfällen |
| PANEL02 | PAN93_94 |
| PANEL03 | PAN93_95 |
| PANEL04 | PAN93_96 |
| PAN01_05 | PAN93_97 |
| PAN01_06 | PAN93_98 |
| PAN01_07 | PAN93_99 |
| PAN01_08 | PAN93_00 |
| PAN01_09 | PAN93_01 |

Panelfälle ab 1996 (Welle 4):

| | |
|----------|----------|
| PAN04_05 | PAN96_97 |
| PAN04_06 | PAN96_98 |
| PAN04_07 | PAN96_99 |
| PAN04_08 | PAN96_00 |
| PAN04_09 | PAN96_01 |

Querschnittfälle:

| | |
|--------|----------|
| QUER01 | QUER1993 |
| QUER02 | QUER1994 |
| QUER03 | QUER1995 |
| QUER04 | QUER1996 |
| QUER05 | QUER1997 |
| QUER06 | QUER1998 |
| QUER07 | QUER1999 |
| QUER08 | QUER2000 |
| QUER09 | QUER2001 |

Wellenkennzeichen:

| | |
|---------|----------|
| WELLE01 | WELL1993 |
| WELLE02 | WELL1994 |
| WELLE03 | WELL1995 |
| WELLE04 | WELL1996 |
| WELLE05 | WELL1997 |
| WELLE06 | WELL1998 |
| WELLE07 | WELL1999 |
| WELLE08 | WELL2000 |
| WELLE09 | WELL2001 |

SPSS-Beispielsyntax

Anhand eines fiktiven Auswertungsinteresses wird ein vollständiges und lauffähiges SPSS-Auswertungsprogramm kommentiert und erläutert. Die zu beantwortende Frage lautet:

Wie verteilt sich die Summe der Teilzeitbeschäftigten auf Branchen und Betriebsgrößenklassen im Jahr 2001?

Dazu kann ein SPSS-Auswertungsprogramm beispielsweise wie folgt konzipiert werden. Im Anschluss an den Programmcode wird die Organisation des Beispielprogramms beschrieben.

```
GET FILE='X:\Nutzer\Musterfrau, Martha (ABC)\test01.sav'.
```

```
USE ALL.
```

```
COMPUTE filter_$=(hrf09q > 0).
```

```
VARIABLE LABEL filter_$ 'hrf09q > 0 (FILTER)'.
```

```
VALUE LABELS filter_$ 0 'Nicht ausgewählt' 1 'Ausgewählt'.
```

```
FORMAT filter_$ (f1.0).
```

```
FILTER BY filter_$.
```

```
EXECUTE .
```

*** Vorschläge zur Brancheneinteilung ***

```
RECODE i92b (1,2=1) (3 THRU 18 = 2) (19, 20 = 3) (21 THRU 23 = 4) (24, 25 = 5)  
          (26, 27 = 6) (28 THRU 32 = 7) (33 THRU 38 = 8) (39 THRU HIGHEST = 9)  
          INTO branche.
```

```
VAR LAB branche "9er Branchengliederung".
```

```
VAL LAB branche 1 'Land- und Forstwirtschaft; Bergbau/Energie/Wasser'  
                2 'Verarbeitendes Gewerbe' 3 'Baugewerbe' 4 'Handel und Reparatur'  
                5 'Verkehr/Nachrichten' 6 'Kredit/Versicherungen'  
                7 'Dienstleistungen überw. f. Unternehmen'  
                8 'Sonstige Dienstleistungen' 9 'Öffentlicher Sektor'.
```

*** Wahlweise können Dienstleistungsbetriebe in öffentlicher Hand dem öffentlichen Sektor zugeordnet werden ***

```
IF (branche ge 5 AND i80=5) branche=9.
```

```
EXECUTE.
```

*** Vorschläge zur Betriebsgrößenklasseneinteilung ***

*** ursprüngliche 10er-Einteilung ***

```
RECODE grkl_01 (ELSE=COPY) INTO betrgr.
```

```
VAR LAB betrgr "10 Betriebsgrößenklassen" .
```

*** oder beispielsweise ***

```
RECODE i01ges01 (1 THRU 9 = 1) (10 THRU 19 = 2) (20 THRU 49 = 3) (50 THRU 199 = 4)
(200 THRU 499 = 5) (500 THRU HIGHEST = 6) INTO betrgr1.
```

```
VAR LAB betrgr1 "sechs Betriebsgrößenklassen" .
```

```
VAL LAB betrgr1 1 '1 bis 9 Beschäftigte' 2 '10 bis 19 Beschäftigte' 3 '20 bis 49 Beschäftigte'
4 '51 bis 199 Beschäftigte' 5 '200 bis 499 Beschäftigte' 6 '500 und mehr
Beschäftigte' .
```

*** bzw. ***

```
RECODE i01ges01 (1 THRU 19 = 1) (20 THRU 99 = 2) (100 THRU 499 = 3) (500 THRU
HIGHEST = 4) INTO betrgr2.
```

```
VAR LAB betrgr2 "vier Betriebsgrößenklassen" .
```

```
VAL LAB betrgr2 1 '1 bis 19 Beschäftigte' 2 '20 bis 99 Beschäftigte' 3 '100 bis 499 Beschäftigte'
4 '500 und mehr Beschäftigte' .
```

```
EXECUTE.
```

*** Test: wie viele Betriebe befinden sich in einer Zelle ***

*** gewichtet ***

```
WEIGHT BY hrf09q.
```

```
CROSSTABS
```

```
/TABLES=branche BY betrgr1
/FORMAT= AVALUE TABLES
/CELLS= COUNT .
```

*** ungewichtet ***

```
WEIGHT OFF.
```

```
CROSSTABS
```

```
/TABLES=branche BY betrgr1
/FORMAT= AVALUE TABLES
/CELLS= COUNT .
```

*** Aufbereitung der Teilzeit-Rohdaten ***

*** Für eventuelle spätere Auswertungen wird empfohlen, die Originalwerte der Ursprungsvariable beizubehalten und einen anderen Variablennamen zu verwenden ***

RECODE i53tz (ELSE=COPY) INTO teilzeit.
EXECUTE.

*** Schritt 1a: bei Betrieben ohne Teilzeitbeschäftigte wird die Summe der Teilzeitbeschäftigten
zunächst auf -9, dann auf missing gesetzt ==> Summe der Teilzeitbeschäftigten
nur in Betrieben mit Teilzeitbeschäftigten ***

IF (i53a=2) teilzeit=-9 .
EXECUTE.

*** Schritt 1b: bei Betrieben ohne Teilzeitbeschäftigte wird die Summe der Teilzeitbeschäftigten
auf Null gesetzt ==> Summe der Teilzeitbeschäftigten in allen Betrieben der
Branche/ Betriebsgrößenklasse ***

***IF (i53a=2) teilzeit=0.
***EXECUTE.

*** IN EINER AUSWERTUNG KANN NUR 1a ODER 1b BENUTZT WERDEN !!! ***

*** Schritt 2: Die Angaben "weiss nicht" (= -8) und "keine Angabe" (= -9) werden als missing
values behandelt ***

MIS VAL teilzeit (-8, -9) .

*** Die Auswertung ***

*** 1. Schritt: Querschnittsgewichtung ***

WEIGHT BY hrf09q.

*** 2. Schritt: Summe über Mittelwerte vergleichen ***

MEANS
TABLES=teilzeit BY betrgr1 branche
/CELLS COUNT SUM .

*** Die Ausgabe kann auch mit Ost/West-Trennung, nach Bundesländern und/ oder für Betriebe
mit und ohne Betriebsrat getrennt erfolgen, wenn folgende Zeilen vor den Befehl MEANS
gesetzt werden ***

SORT CASES BY wo_01.


```
SPLIT FILE BY wo_01.
```

```
***oder
```

```
SORT CASES BY wo_01 i82.
```

```
SPLIT FILE BY wo_01 i82 .
```

Die letzten beiden Befehlszeilen erzeugen eine Ausgabe getrennt nach alten und neuen Bundesländern sowie für Betriebe mit und ohne Betriebsrat.

Aus den so produzierten Ausgaben kann auch der Anteil der Betriebe mit Teilzeitbeschäftigten an allen Betrieben der jeweiligen Zelle Größenklasse*Branche bestimmt werden, wenn auch die Alternativdefinition Schritt 1b ausgeführt wird (dazu muss der Variablenname teilzeit umbenannt werden). Selbstverständlich sind für alle hier gezeigten Befehlszeilen und Beispiele auch andere Herangehensweisen denkbar.

Der Beispielcode ist wie folgt organisiert: zuerst wird – entsprechend den Hinweisen zur Organisation der Auswertungsprogramme – der gewünschte Datensatz eingelesen. Anschließend wird die Auswertung auf gültige Querschnittsfälle begrenzt. Danach werden die interessierenden Recodierungen vorgenommen, die Anzahl der gewichteten und ungewichteten Fälle ausgegeben und schließlich die interessierende Auswertung durchgeführt. Das hier verwendete Beispiel produziert ohne Ausnahme sowohl für die Datenschutzprüfung (ungewichtete Fallzahl nie kleiner vier) wie auch für die Hochrechnung (nie größer drei und kleiner 20) einwandfreie Ergebnisse (dies gilt nicht, wenn die SORT CASES Kommandos vorgeschaltet werden). Eine Konvertierung in das Excel-Format ist also nicht notwendig und es erfolgt die Übersendung der Originalausgabedatei im *.spo-Format.

FAQ

Hier werden einige häufig gestellte Fragen diskutiert und nach Möglichkeit beantwortet. Das Angebot wird schrittweise ergänzt und ist derzeit noch nicht vollständig verfügbar. Wir bitten um Verständnis.

In welchen Grenzen sind Auswertungen aus dem IAB-Betriebspanel repräsentativ?

Sofern die Hochrechnungsmatrix (bis 1999 10 Betriebsgrößenklassen * 16 Branchen, ab 2000 10 Betriebsgrößenklassen * 20 Branchen) oder eine Zusammenfassung der einzelnen Zellen verwendet wird, gelten folgende Grenzen:

1993 bis 1995: westdeutsches Bundesgebiet

1996 bis 1999: westdeutsches Bundesgebiet und ostdeutsche Bundesländer

ab 2000 : west- und ostdeutsche Bundesländer

Auf diesen Ebenen sind mindestens repräsentative Querschnittsauswertungen möglich. Bezüglich der Aussagekraft der Ergebnisse für bestimmte weitere Merkmale wird auf ungewichtete Anzahl der zugrunde liegenden Fälle verwiesen. Die Repräsentativität von Längsschnittauswertungen hängt zusätzlich von der zur Verfügung stehenden Anzahl der Betriebe und ihrer Lage bezüglich der Merkmale Branche*Betriebsgröße ab und sollte im Einzelfall überprüft werden.

Sind (hochgerechnete) Ergebnisse nach Branchen differenziert vor und nach der Erhebungswelle 2000 miteinander vergleichbar?

Im Jahr 2000 wurde die Stichprobenziehung und die Wirtschaftszweigklassifikation auf WZ93/NACE umgestellt. Betriebe befinden sich vor und nach dem Jahr 2000 oft in verschiedenen Hochrechnungszellen. Einwandfreie Umschlüsselungen sind wenig erfolgversprechend, da etwa zwei Drittel aller Betriebe umgeordnet wurden. Für die Neueinteilung können keine klaren Grenzlinien bzw. Daumenregeln angegeben werden. Selbst ein Vergleich von Betrieben des Verarbeitendes Gewerbe und der Dienstleistungen enthält vor und nach dem Jahr 2000 keine auch nur annähernd gleichen Grundgesamtheiten. Betriebsgrößenklassen sind nicht betroffen. Sie sind über die Jahre hinweg uneingeschränkt miteinander vergleichbar.

Welche Möglichkeiten bestehen bei einer unbrauchbaren Auswertung aufgrund zu geringer Fallzahlen?

Korrekturmöglichkeiten gibt es im Bereich der Gruppierung, indem Zellen zusammengefasst werden. Manchmal ist die optimale Intervallbreite auch nach mehreren Versuchen noch nicht gefunden. Im Bedarfsfall hilft der/ die Ansprechpartner/-in im FDZ weiter.

Ist die Branchengliederung im IAB-Betriebspanel zwingend anzuwenden?

Bis zum Jahr 1999 bestehen keine zusätzlichen Optionen. Ab dem Jahr 2000 kann über den Organisationsfile die WZ93 (=NACE) zugespielt und die Betriebe danach gruppiert werden. Von einem solchen Vorgehen wird aber abgeraten, da solche Einteilungen i.d.R. von der Hochrechnungsmatrix des IAB-Betriebspanels abweichen und so nicht näher bestimmbare Ungenauigkeiten bei der Hochrechnung der Ergebnisse produzieren.