

## Der Studierendensurvey

Daten- und Methodenbericht zu den Erhebungen von 1983 bis 2016

### I Einleitung

Der Studierendensurvey ist eine Dauerbeobachtung zur Studiensituation und den studentischen Orientierungen an deutschen Hochschulen. Der Fragebogen wurde von der AG Hochschulforschung der Universität Konstanz entwickelt und 1982 zum ersten Mal durchgeführt. Alle 2-3 Jahre erfolgte eine weitere Befragung, zuletzt wurde 2015/16 die 13. Welle erhoben. Die breit angelegte Themenvielfalt umfasst den gesamten Hochschulbereich, von der Zeit vor der Hochschule bis hin zu den beruflichen Übergängen und bindet berufsbezogene, gesellschaftliche und politische Haltungen mit ein. In einzelnen Erhebungen wurden spezifische Themen vertieft eingebunden. Die Befragung erfolgte an jeweils ca. 25 nach theoretischen Merkmalen ausgewählten Universitäten und Fachhochschulen, die ein breites Fächerspektrum aufweisen, über das gesamte Bundesgebiet verteilt sind und eine bestimmte Mindestgröße aufweisen. An diesen wurden die Studierenden für die Befragung zufällig ausgewählt, wobei die Größe der Zufallsstichprobe, angepasst an die Hochschulgröße, vorgegeben wurde. Seit Beginn der Erhebungen wird das Projekt vom BMBF gefördert.

Für die Übergabe der Datensätze an das Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) zum Zwecke der Bereitstellung der Daten für wissenschaftliche Analysen, wurden die 13 Datensätze neu aufbereitet, zusammengeführt und dokumentiert. Dort werden sie als *Scientific Use Files* (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung gestellt. Neben den Datensätzen der Erhebungen werden auch Dokumentationsmaterialien zu den Datensätzen und zur Durchführung der Studien bereitgestellt.

### II Übersicht zum 13. Studierendensurvey vom WS 2015/16

<b>Studienreihe</b>	Konstanzer Studierendensurvey; Studiensituation und studentische Orientierungen an Universitäten und Fachhochschulen
<b>Erhebende Institution</b>	AG Hochschulforschung, Universität Konstanz
<b>ProjektmitarbeiterInnen</b>	Werner Georg, Frank Multrus, Tino Bargel, Sandra Majer, Monika Schmidt, Hans Simeaner, Doris Lang
<b>Förderung</b>	Bundesministerium für Bildung und Forschung (BMBF)
<b>Version des Datensatzes</b>	13. Welle
<b>Grundgesamtheit:</b>	322 deutsche Universitäten und Fachhochschulen, die der OECD-Klassifizierung der Stufe 5 ISCED (5A und 5B) im tertiären Sektor des Bildungswesens entsprechen.
<b>Stichprobenverfahren</b>	Geschichtete Zufallsstichprobe, Schichten nach theoretischen Vorgaben. Auswahl Hochschulen: 16 Universitäten und 12 Fachhochschulen nach theoretischen Vorgaben (Größte Lage, Fächerangebot). Auswahl Studierende: einfache geschichtete Zufallsstichprobe, auf Hochschulgröße angepasst.
<b>Erhebungsmethode</b>	standardisierte postalische Erhebung (paper pencil) und standardisierte Online Erhebung.

<b>Erhebungszeitraum</b>	Ende November 2015 – Ende März 2016
<b>Rücklaufquote und auswertbare Fälle</b>	Bruttostichprobe: Papier: 36.902 (bereinigter Versand), Online: 33.369 Nettostichprobe: Papier: 5.620, Online: 2.929 Rücklaufquote: Papier: 15,2%; Online: 8,8%, zusammen: 13,3%
<b>Fragebogenumfang</b>	92 Fragen mit 413 Items, plus 19 weitere Fragen mit 125 Items, verteilt auf in 4 Zusatzmodule
<b>Themen</b>	<ol style="list-style-type: none"> <li>1. Hochschulzugang, Fachwahl, Motive und Erwartungen</li> <li>2. Studienstrategien, Studienverlauf und Qualifizierungen</li> <li>3. Studienintensität, Zeitaufwand und Studiendauer</li> <li>4. Studienanforderungen, Regelungen und Prüfungen</li> <li>5. Kontakte und Kommunikation, soziales Klima, Beratung</li> <li>6. Fachstudium, Situation der Lehre und Studienqualität</li> <li>7. Studentische Lebensform, soziale Situation, Erwerbstätigkeit</li> <li>8. Schwierigkeiten, Beeinträchtigungen und Belastungen</li> <li>9. Hochschulpolitik: Partizipation und studentische Vertretung</li> <li>10. Internationalität, Auslandsstudium und Europäischer Hochschulraum</li> <li>11. Berufswahl, Berufswerte und Tätigkeitsbereiche</li> <li>12. Berufsaussichten und Arbeitsmarktreaktionen</li> <li>13. Demokratische Einstellungen und gesellschaftliche Vorstellungen</li> <li>14. Politische Beteiligung, Standort und Ziele</li> <li>15. Wünsche und Forderungen zur Hochschulentwicklung</li> <li>16. Demographische und bildungsbiographische Daten</li> </ol>
<b>DOI</b>	
<b>Datenprodukte</b>	SUF
<b>Datenstruktur</b>	Personendatensätze
<b>Materialien zum 13. Studierendensurvey</b>	<ul style="list-style-type: none"> <li>• Daten- und Methodenbericht (deutsch)</li> <li>• Studienübersicht (deutsch)</li> <li>• Datensatz 13. Erhebung (STATA)</li> <li>• Codebuch zur 13. Erhebung als codierter Fragebogen</li> <li>• Variablenliste zur 13. Erhebung mit labels und Kennwerten</li> <li>• Originalfragebogen zur 13. Erhebung</li> <li>• Datensatz Mehrwellendatei, Erhebungen 1-13</li> <li>• Codebuch zur Mehrwellendatei, Erhebungen 1-13</li> <li>• Karte mit geografischer Lage aller Samplehochschulen</li> <li>• Anschreiben an Studierende im WS 2015/16 für Papierbefragung</li> <li>• Anschreiben an Studierende im WS 2015/16 für Onlinebefragung</li> <li>• Anschreiben an Studierende im WS 2015/16 für Papier- und Onlinebefragung (Wahlgruppe)</li> <li>• Begrüßungsschreiben-Studierende-Online im WS 2015/16</li> <li>• Informationsschreiben für Studierende im WS 2015/16</li> <li>• Erinnerungsschreiben für Studierende im WS 2015/16</li> <li>• Dokumentation Studienfächer im Studierendensurvey</li> <li>• Publikationsverzeichnis zum Studierendensurvey</li> </ul>

### **Basispublikationen zum 13. Studierendensurvey**

- Multrus F./ S. Majer / T. Bargel / M. Schmidt (2017): Studiensituation und studentische Orientierungen. 13. Studierendensurvey an Universitäten und Fachhochschulen. BMBF.
- Multrus F./ S. Majer / T. Bargel / M. Schmidt (2017): Studiensituation und studentische Orientierungen. Zusammenfassung zum 13. Studierendensurvey an Universitäten und Fachhochschulen. BMBF.
- Simeaner, H., Multrus, F., & Kolbert-Ramm, C. (2017). Datenalmanach. Studierendensurvey 1993 – 2016 nach Geschlecht. Hefte zur Bildungs- und Hochschulforschung (98), AG Hochschulforschung, Universität Konstanz, September 2017.
- Simeaner, H., Multrus, F., & Kolbert-Ramm, C. (2017). Datenalmanach. Studierendensurvey 1993 – 2016. Hefte zur Bildungs- und Hochschulforschung (97), AG Hochschulforschung, Universität Konstanz, September 2017.
- Grundauszählung Universitäten und Fachhochschulen 13. Erhebung
- Multrus F. / S. Majer (2017): Methodenbericht zum 13. Studierendensurvey. Vergleich Papier-Onlinebefragung. Werkstattbericht. Hefte zur Bildungs- und Hochschulforschung (95), AG Hochschulforschung, Universität Konstanz, September 2017.

### **Weiterführende Publikationen**

- Multrus, F. (2018). Bedingungen und Effekte von realisierter und geplanter Auslandsmobilität bei Studierenden. In: Becker, K. & Heißenberg, S. (Hrsg): Dimensionen studentischer Vielfalt. Empirische Befunde zu heterogenen Studien- und Lebensarrangements (S. 196-250). Bielefeld: wbv Publikationen.
- Majer, S. (2018). Studienerfolg von traditionellen und beruflich qualifizierten Studierenden. In Bornkessel, P. (Hg.), Erfolg im Studium. Konzepte, Befunde und Desiderate (S. 175-225). Bielefeld: wbv Publikationen.
- Bargel, T. (2017). Studium im Wandel. Im Zeitraum von 30 Jahren: 1983 – 2013. Studentische Erfahrungen und Evaluation. Hefte zur Bildungs- und Hochschulforschung (92), AG Hochschulforschung, Universität Konstanz, Juli 2017.
- Bargel, T. (2017). Studentische Orientierungen gegenüber Studium, Beruf und Politik im Wandel. Zeitreihe des Studierendensurveys 1983-2013. Hefte zur Bildungs- und Hochschulforschung (91), AG Hochschulforschung, Universität Konstanz, Juli 2017.
- Multrus, F. (2016). Der Studierendensurvey. Konzept, Aufbau und theoretische Grundlagen. Hefte zur Bildungs- und Hochschulforschung (87), AG Hochschulforschung, Universität Konstanz, Juni 2016.
- Schmidt, M., Bargel, T., Georg, W., & Gorbachyk, A. (2016). Studienzwecke und Berufsaussichten - ein Vergleich Studierender in Kiew und München. Hefte zur Bildungs- und Hochschulforschung (88), AG Hochschulforschung, Universität Konstanz, September 2016.
- Georg/Bargel: Parental qualifications as determinants of university entrance and choice of a field of study in Germany. European Journal of Higher education.
- Georg, W. & Bargel, T. (2016). Soziale Determinanten der Studienaufnahme und Fachwahl. In Grossmann, D. & Wolbring, T. (Hg.), Evaluation von Studium und Lehre (S. 93-121). Wiesbaden: Springer.
- Bargel, T., Multrus, F. & Majer, S. (2017). HSW-Gespräch zum Zusammenhang von Veranstaltungsbesuch und Studienqualität mit Tino Bargel, Frank Multrus und Sandra Majer. „Das Hochschulwesen“. HSW 4+5, 154-162.

### III Datennutzungshinweise

**[Voraussetzungen der Datennutzung]** Die Daten des Studierendensurveys werden durch das FDZ des DZHW als Scientific Use Files (SUF) ausschließlich zur wissenschaftlichen Sekundärnutzung freigegeben. Die SUF kann über die Website des FDZ (<https://fdz.dzhw.eu>) beantragt werden. Für die Nutzung wird ein Datennutzungsvertrag abgeschlossen. Studierende oder Promovierende ohne eine Anstellung an einer wissenschaftlichen Einrichtung müssen gemeinsam mit einer betreuenden Person einen Datennutzungsvertrag abschließen. Im Zuge des Vertragsabschlusses wird durch das FDZ auch das Vorliegen eines wissenschaftlichen Nutzungsinteresses geprüft.

**[Datenzugang]** Das SUF wird über zwei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind:

- **Download:** Die Daten werden verschlüsselt auf der Website des FDZ zum Download bereitgestellt. Datennutzer\*innen können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.
- **Remote-Desktop:** Die Daten werden auf einem Terminal-Server des FDZ bereitgestellt. Über eine besonders gesicherte Verbindung zwischen dem lokalen Computer der nutzenden Person und dem Terminal-Server des FDZ können die Daten mit der auf dem Terminal-Server vorhandenen Software analysiert werden. Das Übertragen der Daten auf den lokalen Computer ist nicht möglich. Analyseergebnisse werden erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

**[Gebühren der Datenbereitstellung]** SUF werden derzeit (Stand: Oktober 2021) kostenfrei zur Verfügung gestellt. Änderungen bzw. das aktuelle Kostenmodell können auf der Web-site des FDZ-DZHW (<https://fdz.dzhw.eu>) eingesehen werden.

**[Pflichten der Datennutzer\*innen]** Die Datennutzer\*innen sind verpflichtet, folgende Regeln<sup>1</sup> einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer\*innen Kenntnis von Sicherheitslücken hinsichtlich des Datenschutzes bzw. der Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** Die Daten dürfen nur durch die Personen genutzt werden, die einen Datennutzungsvertrag abgeschlossen haben.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von wissenschaftlichen Arbeiten (z.B. Publikationen, Abschlussarbeiten, Dissertationen), die aus der Arbeit mit Daten des FDZ hervorgehen, ist dem FDZ anzuzeigen. Dabei ist dem FDZ eine elektronische Version der Druckfassung zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z.B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden.

---

<sup>1</sup> Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

## 1 Inhalt und Anlage des Studierendensurveys (alle Erhebungen)

### [Studienreihe]

Die Entwicklung des Studierendensurveys geht auf Arbeiten aus dem Projekt „Hochschulsozialisation“ im Sonderforschungsbereich „Bildungsforschung“ (SFB 23) an der Universität Konstanz zurück, in dessen Rahmen Abiturienten, Studierende und Akademiker untersucht wurden. Im Konzept versteht sich der Studierendensurvey als gesellschaftliche Dauerbeobachtung, wobei der Kern seit Beginn der Erhebungen die Leistungsmessung im Hochschulbereich darstellt, um Grundlagen für die Hochschulpolitik und deren öffentliche Debatte bereit zu stellen.

Ausführliche Informationen zu Entwicklung, Aufbau und theoretischer Rahmung sind in Heft 87 der Reihe „Hefte zur Bildungs- und Hochschulforschung“ zusammengetragen: Multrus, F. (2016). Der Studierendensurvey. Konzept, Aufbau und theoretische Grundlagen. AG Hochschulforschung, Universität Konstanz.

Die Erhebungen des Studierendensurvey umfassen einen Zeitraum von 35 Jahren, in denen die Studiensituation und die studentischen Orientierungen an deutschen Hochschulen (Universitäten und Fachhochschulen) erhoben und analysiert wurden.

1. Erhebung: WS 1982/83
2. Erhebung: WS 1984/85
3. Erhebung: WS 1986/87
4. Erhebung: WS 1989/90
5. Erhebung: WS 1992/93
6. Erhebung: WS 1994/95
7. Erhebung: WS 1997/98
8. Erhebung: WS 2000/01
9. Erhebung: WS 2003/04
10. Erhebung: WS 2006/07
11. Erhebung: WS 2009/10
12. Erhebung: WS 2012/13
13. Erhebung: WS 2015/16

Bei den einzelnen Erhebungen handelt es sich um Querschnitterhebungen. Allerdings wurde darauf geachtet, dass möglichst immer die gleichen Hochschulen befragt wurden, sodass vergleichbare Umgebungsbedingungen vorhanden sind. Dies konnte für die Mehrzahl der beteiligten Hochschulen über lange Zeiträume auch realisiert werden.

### [Hochschulsample]

Die 25-30 ausgewählten Samplehochschulen wurden jeweils im Frühjahr des Jahres der Befragung angeschrieben und zur Beteiligung eingeladen. Der Erstkontakt ging über das jeweilige Rektorat, die weiterführenden Kontakte über die zuständigen Mitarbeiter, mit denen telefonisch die Vorgehensweise abgestimmt wurde.

Die Auswahl der Hochschulen erfolgte dabei nach bestimmten Kriterien. Es sollten Universitäten, Technische Universitäten, Gesamthochschulen (in den frühen Erhebungsphasen) und Fachhochschulen beteiligt sein. An den Hochschulen sollte ein breites Fächerspektrum angeboten sein, sodass möglichst alle Fächer vertreten sind. Daher wurden keine Spezialhochschulen aufgenommen. Für die Auswahl der Fachhochschulen bedeutete dies, dass sie ein Fächerangebot aus mindestens drei Fä-

chergruppen ausweisen müssen (Sozialwissenschaften, Wirtschaftswissenschaften, Ingenieurwissenschaften), damit differenzierte Vergleiche zu den Universitäten möglich werden. Des Weiteren sollten Hochschulen aus allen Bundesländer beteiligt sein, damit Aussagen für ganz Deutschland möglich werden. In einigen definierten Fällen sollten die beteiligten Universitäten und Fachhochschulen möglichst aus der gleichen Stadt stammen, sodass Vergleiche mit kontrollierten Regionalbedingungen ermöglicht werden.

Im Laufe der Erhebungen wurde das Hochschulsample ausgeweitet und es wurden neue Hochschulen nach den gleichen Bedingungen rekrutiert. Nach der Wende wurden Anfang der 90er Jahre Hochschulen aus den neuen Bundesländern in das Hochschulsample aufgenommen. Bei den Neurekrutierungen zur Jahrtausendwende wurden gezielt Neugründungen aus den 50iger Jahren ausgewählt, um Vergleiche zu traditionsreichen Hochschulen zu ermöglichen. All diese Hochschulen wurden ab Erstkontakt in allen darauffolgenden Erhebungen wieder eingeladen.

Samplehochschulen im Studierendensurvey 1982-2015



### **[Studierendensample]**

Die Auswahl der Studierenden erfolgte an den jeweiligen Hochschulen. Als Vorgabe erhielten die dortigen zuständigen Mitarbeitenden von der Arbeitsgruppe Hochschulforschung (AGH) die Anweisung, dass nur deutsche Studierende ausgewählt werden sollen, die sich nicht in einem Urlaubssemester befinden. Dabei sollen möglichst alle Fächern und allen Semester vertreten sein. Aus diesem Pool sollten die Hochschulen dann eine bestimmte Anzahl an Studierenden zufällig aus ihrer Studierendenkartei auswählen. Diese Anzahl war abhängig von der Größe der Hochschule und wurde von der AGH vordefiniert.

### **[Durchführung der Erhebungen]**

Die postalischen Unterlagen, die zum einen aus Fragebogen, Anschreiben und Rückkuverts, zum anderen aus einem Briefkuvert mit Erinnerungsschreiben bestanden, wurden ab Ende November verpackt und an die beteiligten Hochschulen per Post versandt. Dort mussten für die ausgewählten Studierenden zwei Adress-Etikettensätze ausgedruckt werden. Diese wurden von den Hochschulen auf die Fragebogenunterlagen und die Erinnerungsschreiben geklebt. Sodann wurden zuerst nur die Fragebogenunterlagen an die Studierenden versendet. Dies markierte den Start der Feldzeit. Die Anschreiben an die Studierenden, die Hintergrundinformationen sowie die Erinnerungsschreiben sind in den Dokumentationen einsehbar, ebenso die Verteilung der Hochschulen pro Erhebung.

Die Erhebungen wurden jeweils im Wintersemester Ende November / Anfang Dezember durchgeführt, sodass die Studienanfänger zumindest 6-8 Wochen Erfahrungen im Studium machen konnten. Die Feldzeit wurde zumeist bis Ende März / Anfang April ausgeweitet, wobei im Januar das Erinnerungsschreiben an alle angeschriebenen Studierenden von den Hochschulen versendet wurde.

Während der Feldzeit wurden die per Rückumschlag direkt an die AG zurückgeschickten Fragebogen ausgepackt, einer ersten Prüfung unterzogen und mit Eingangsstempel und Laufnummer versehen. Die gesammelten ausgefüllten Fragebogen wurden in Tranchen einer Erfassungsfirma übersandt, die die schriftlichen Angaben in eine digitale Form übertrug. Dieser Vorgang wurde mit mehreren Kontrollprüfungen begleitet.

Ab Mai / Juni lagen die digitalisierten Rohdaten der Befragung vor und konnten in der AGH aufbereitet werden. Dann erfolgten die ersten Analysen und die Vorbereitungen der Publikationen.

### **[Fragebogen]**

Die Fragebogen bestanden aus doppelseitig bedruckten Din A3 Seiten, die in der Mitte geklammert wurden, sodass ein Din A4 Format resultierte. Zusätzlich hatten sie eine farbige Umschlagsseite. Auf der vorderen Umschlagsseite stand der Name des Projektes (Forschungsprojekt Studiensituation) und Name und Adresse der durchführenden Institution (Projektleiter und Universität Konstanz). Zusätzlich wurden auf dem Deckblatt kurze Hinweise zur Freiwilligkeit und zur Beantwortung des Fragebogens angegeben. Weitere Informationen wurden bis 1992 mit einem Begleitschreiben versendet.

Ab der sechsten Erhebung im WS 1994/95 wurde die Vorder- und Rückseite des Fragebogens zusätzlich mit den Informationen zur Studie bedruckt und auf ein eigenes Begleitschreiben verzichtet. Ab dann wurden die Studierenden auch darauf hingewiesen, dass sie eine Kurzfassung der Ergebnisse erhalten können, wenn sie eine Postkarte mit ihrer Adresse beilegen. Diese Maßnahme sollte einen Anreiz darstellen und wurden von vielen Studierenden auch genutzt. Außerdem wurden vier verschiedene Farben für die Umschlagsgestaltung gewählt, die nach Hochschulart und alte bzw. neue Bundesländer unterschieden. Dies diente dazu, dass bei fehlender Hochschulnennung wenigstens eine grobe Zuordnung zum Aggregat möglich war (z.B. Universitäten neue Länder).

Zur 12. Erhebung im WS 2012/13 wurde der gesamte Fragebogen in der Gestaltung überarbeitet und nach Vorschlägen von GESIS neu angeordnet. Dabei wurde auch die Umschlagsgestaltung neu erstellt und die Informationen zur Studie wieder in ein Begleitschreiben übernommen. In der letzten Erhebung vom WS 2015/16 wurde das Deckblatt nochmals umgestaltet, diesmal nach Vorschlägen des Grafikbüros der Universität Konstanz.

Der innere Teil des Fragebogens umfasste in den ersten sechs Erhebungen 20 Din A4 Seiten. Im neuen Jahrtausend wurde die Befragung auf 16 Seiten gekürzt. Zur 13. Erhebung wurde nochmals gekürzt und auf 13 Seiten begrenzt. Bis zur 11. Erhebung war das Layout des inneren Fragebogens tabellenartig angelegt, sodass die Fragen und die Antworten jeweils in einer eigenen Spalte standen. Diese Darstellung wurde für die 12. Erhebung nach Vorschlägen von GESIS aufgebrochen. Zu 13. Erhebung wurde allerdings wieder die Tabellenstruktur favorisiert.

Alle Fragebogen mit Deckblättern und Begleitschreiben sind in den Dokumentationen einzusehen.

### **[Themen]**

In jeder Erhebung wurde eine breite Themenvielfalt erhoben. Bestimmte Themen kamen in allen Erhebungen vor, dazu gehören die Sozialdaten und die Angaben zum Studium (Fach, Semester, Abschluss, etc.), aber auch Fragen zu der Zeit vor dem Studium, dem Erleben des Studiums, den Anforderungen, den Schwierigkeiten und Belastungen, den Wünschen, dem Ertrag oder den Bewertungen der Lehre, sowie der Zeit nach dem Studium mit Berufsvorstellungen oder gesellschaftliche und politische Einstellungen.

Daneben wurden in manchen Erhebungen auch Schwerpunktthemen gesetzt, die teilweise auch über mehrere Erhebungen aufrechterhalten wurden. So z.B. Fragen zur Wiedervereinigung, zur Existenzgründung, zur Computernutzung oder zum europäischen Hochschulraum.

Insgesamt lassen sich alle Themenbereich in folgende Liste zusammenstellen:

- Demographische und bildungsbiographische Daten
- Hochschulzugang, Fachwahl, Motive und Erwartungen
- Studienstrategien, Studienverlauf und Qualifizierungen
- Studienintensität, Zeitaufwand und Studiendauer
- Studienanforderungen, Regelungen und Prüfungen
- Kontakte und Kommunikation, soziales Klima, Beratung
- Fachstudium, Situation der Lehre und Studienqualität
- Studentische Lebensform, soziale Situation, Erwerbstätigkeit
- Schwierigkeiten, Beeinträchtigungen und Belastungen
- Hochschulpolitik: Partizipation und studentische Vertretung
- Internationalität, Auslandsstudium und Europäischer Hochschulraum
- Berufswahl, Berufswerte und Tätigkeitsbereiche
- Berufsaussichten und Arbeitsmarktreaktionen
- Demokratische Einstellungen und gesellschaftliche Vorstellungen
- Politische Beteiligung, Standort und Ziele
- Wünsche und Forderungen zur Hochschulentwicklung
- Neue Medien in Studium und Lehre
- Computer, Internet und neue Medien
- Akzeptanz neuer Studienstrukturen und Abschlüsse
- Vorstellungen über Hochschule und Wissenschaft
- Rolle und Chance von Hochschulabsolventen
- Austausch zwischen alten und neuen Bundesländer

## **[Publikationen zu den Erhebungen]**

Es wurden in jeder Erhebung folgende Publikationsarten erstellt:

- Grundauszählungen aller Variablen, nach Universitäten und Fachhochschulen.
- Indikatorenbände für die beteiligten Hochschulen., Grundauszählungen ausgewiesener Merkmale nach den dortigen Fächergruppen. Diese Bände wurden ausschließlich den jeweils beteiligten Hochschulen zugesandt und sind hier nicht als Dokumentationen einzusehen.
- Datenalmanach, eine Grundauszählung aller Merkmale in Zeitvergleich, nach Hochschulart und Fächergruppen.
- Datenalmanach, zusätzlich getrennt nach Geschlecht.
- Hauptbericht „Studiensituation und studentische Orientierungen“. Publikation der wichtigsten Befunde zu einer breiten Auswahl an Themenbereichen.
- Kurzfassung des Hauptberichtes.
- Schwerpunktberichte. Vertiefende Analysen und Befunde zu ausgewählten Bereichen. Darunter fallen die Fachmonografien, die für einzelne Fächer (Medizin, Jura) oder Fächergruppen (Geisteswissenschaften, Ingenieurwissenschaften, Naturwissenschaften) erstellt wurden, sowie für Abschlussarten (Bachelor, Master). Andere Schwerpunktthemen waren: Methodenvergleich Paper – Online, Studieneingangsphase, Forschung und Praxis, Frauen im Studium, Wandel politischer Orientierungen, wissenschaftlicher Nachwuchs unter Studierenden.
- Ausgaben der Reihe „Hefte zur Bildungs- und Hochschulforschung“, die unterschiedliche Themen beinhalten, worunter Dokumentationen ebenso fallen wie Fachberichte oder methodischen Untersuchungen.
- Hinzu kommen fachwissenschaftliche Buchbeiträge und Artikel sowie sonstige Transferveröffentlichung.

Sämtliche Publikationen sind auf der Website der AG Hochschulforschung zusammengestellt (<https://www.soziologie.uni-konstanz.de/ag-hochschulforschung/publikationen/>).

## **[Besonderheiten der 13. Erhebung]**

Im WS 2015/16 wurden neben dem Papierfragebogen auch ein Onlinefragebogen eingesetzt. Ziel war es zu prüfen, ob ein Umstieg auf zukünftige Onlinebefragungen Auswirkungen auf das Antwortverhalten der Studierenden haben wird. Daher wurde der Fragebogen in eine Onlineversion umgestaltet, inhaltlich aber identisch belassen. Damit ein Vergleich zwischen den Antworten der Papier- und der Onlinebefragten methodisch abgesichert ist, wurde das Studierendensample verdoppelt, sodass gleich viele Studierende einen Papierfragebogen erhielten wie eine Einladung zur Onlinebefragung mit Link und Zugangscode.

Außer dem direkten Vergleich der Papier- und der Onlineversion wurde die 13. Erhebung des Studierendensurveys auch dazu genutzt, weitere Möglichkeiten von Onlinebedingungen zu testen. Dazu wurde ein komplexes Design entwickelt, dass das Onlinesample in sieben Untergruppen unterteilt.

- Eine Gruppe erhielt sowohl den Papierfragebogen als auch en Onlinezugang und konnte auswählen, welches Format sie beantworten wollen.
- Zwei Gruppen erhielten den vollständigen, sehr langen Onlinefragebogen. Für eine Gruppe wurden Motivationsseiten in den Fragebogen eingefügt, um zu prüfen, ob dieser Ansatz das Durchhaltevermögen stärken kann.
- Vier Gruppen erhielten eine um 40 % gekürzte Fassung. Für diese Gruppen wurden zusätzlich extra Module entwickelt, die am Ende des Fragebogens eingefügt wurden und sehr spezielle Zusatzfragen enthielten.

Unabhängig davon wurde an den Samplehochschulen nachgefragt, ob erstens die Möglichkeit einer zweiten Erinnerung über E-Mail besteht und zweitens, ob eine Bewerbung der Befragung auf den Websites und in den social media Kanälen möglich wäre. Zu jeder dieser Maßnahmen waren jeweils rund die Hälfte der Samplehochschulen bereit, sodass Vergleiche zwischen den Maßnahmen möglich waren.

Eine Hochschule hat nur einer Onlinebefragung zugestimmt, sodass dort keine Post verschickt wurde, sondern die Einladung per E-Mail erfolgte.

Die Umsetzung und die Datenerhebung der Onlinebefragung wurden von DZHW durchgeführt. Nach Abschluss der Feldzeit wurden die Rohdaten an die AGH geschickt.

Die Analysen zum Vergleich der Antworten der Studierenden aus der Papier- und der Onlinegruppe konnte nachweisen, dass ein Umstieg möglich ist, ohne dass das ein signifikant anderes Ergebnis im Antwortverhalten zu beobachten wäre. Die unterschiedlichen Maßnahmen zeigten unterschiedliche Effekte. Eine ausführliche Dokumentation zu diesen Vergleichen wurde in einem Methodenbericht publiziert und ist nachzulesen in Heft 95 der Reihe „Hefte zur Bildungs- und Hochschulforschung“: Multrus, F. & S. Majer (2017). Methodenbericht zum 13. Studierendensurvey. Vergleich Papier-Onlinebefragung. Arbeitsgruppe Hochschulforschung, Universität Konstanz.

#### [Rücklauftabelle]

Der Rücklauf der einzelnen Erhebungen lag bis Mitte der 90er Jahre bei über 40 %. Danach hat die Beteiligung der Studierenden systematisch nachgelassen. Die einzelnen Rückläufe und bereinigte Stichprobengröße sind in der Tabelle aufgelistet.

<b>Versandte Fragebogen und Beteiligung an den dreizehn Erhebungen des Studierendensurveys nach Hochschulart (1983 - 2016)</b> (Angaben in Absolut und Prozent)			
<b>Früheres Bundesgebiet</b>	<b>Insgesamt</b>	<b>Universitäten</b>	<b>Fachhochschulen</b>
<b>WS 1982/83</b>			
Versand (bereinigt)	18.940	16.586	2.354
Rücklauf: absolut	7.817	6.607	1.059
in Prozent	41,3	39,8	45,0
<b>WS 1984/85</b>			
Versand (bereinigt)	22.470	17.500	4.970
Rücklauf: absolut	10.038	7.663	2.324
in Prozent	44,7	43,8	46,8
<b>WS 1986/87</b>			
Versand (bereinigt)	22.400	17.400	5.000
Rücklauf: absolut	9.852	7.532	2.279
in Prozent	44,0	43,3	45,6
<b>WS 1989/90</b>			
Versand (bereinigt)	19.730	15.380	4.350
Rücklauf: absolut	8.812	6.999	1.813
in Prozent	44,7	45,5	41,7
<b>Bundesrepublik Deutschland (Alte und Neue Länder)</b>			
<b>WS 1992/93</b>			
Versand (bereinigt)	20.175	15.519	4.656
Rücklauf: absolut	9.240	7.192	2.048
in Prozent	45,8	46,3	44,0
<b>WS 1994/95</b>			
Versand (bereinigt)	19.644	15.051	4.593
Rücklauf: absolut	8.461	6.582	1.879
in Prozent	43,1	43,7	40,9
<b>WS 1997/98</b>			

Versand (bereinigt)	19.641	15.118	4.523
Rücklauf: absolut	7.271	5.799	1.472
in Prozent	37,0	38,4	32,5
<b>WS 2000/01</b>			
Versand (bereinigt)	23.391	18.064	5.327
Rücklauf: absolut	8.130	6.385	1.745
in Prozent	34,8	35,4	32,8
<b>WS 2003/04</b>			
Versand (bereinigt)	27.423	21.498	5.475
Rücklauf: absolut	9.975	8.307	1.668
in Prozent	36,4	37,8	30,5
<b>WS 2006/07</b>			
Versand (bereinigt)	26.064	20.844	5.220
Rücklauf: absolut	8.350	6.894	1.456
in Prozent	32,0	33,1	27,9
<b>WS 2009/10</b>			
Versand (bereinigt)	27.307	21.570	5.737
Rücklauf: absolut	7.590	6.117	1.473
in Prozent	27,8	28,4	25,7
<b>WS 2012/13</b>			
Versand (bereinigt)	26.274	19.927	6.347
Rücklauf: absolut	4.884	3.792	1.092
in Prozent	18,6	19,0	17,2
<b>WS 2015/16 Papier</b>			
Versand (bereinigt)	30.789	23.479	7.312
Rücklauf: absolut	4.959	3.885	1.074
in Prozent	16,1	16,5	14,7
<b>Wahlgruppe<sup>2</sup> Papier</b>			
Versand (bereinigt)	6.122	4.763	1.359
Rücklauf: absolut	661	508	153
in Prozent	10,8	10,7	11,3
<b>WS 2015/16 Papier gesamt</b>			
Versand (bereinigt)	36.902	28.235	8.669
Rücklauf: absolut	5.620	4.393	1.227
in Prozent	15,2	15,6	14,2
<b>WS 2015/16 online</b>			
Versand (bereinigt)	27.244	21.296	5.948
Rücklauf: absolut	2.656	2.071	581
in Prozent	9,7	9,7	9,8
<b>WS 2015/16 Wahlgruppe Online (Lang- und Kurzfassung)</b>			
Versand (bereinigt)	6.122	4.763	1.359
Rücklauf: absolut	273	201	70
in Prozent	4,5	4,2	5,1
<b>WS 2015/16 Online gesamt</b>			
Versand (bereinigt)	33.369	26.062	7.307
Rücklauf: absolut	2.929	2.272	651
in Prozent	8,8	8,7	8,9
<b>Wahlgruppe gesamt (Papier + online)</b>			
Versand (bereinigt)	6122	4.763	1.359
Rücklauf: absolut	934	709	223
in Prozent	15,3	14,9	16,4

<sup>2</sup> Wahlgruppe konnte auswählen, ob sie Papier- oder online ausfüllen will. Daher bilden Rücklaufberechnungen für die Wahlgruppe, unterschieden nach Papier oder online, nur in der Summe korrekte Werte ab.

<b>WS 2015/16 gesamt (Papier + online)</b>			
Versand (bereinigt)	64.286	49.642	14.646
Rücklauf: absolut	8.549	6.665	1.878
in Prozent	13,3	13,4	12,8

Quelle: Studierendensurvey 1983-2016, AG Hochschulforschung, Universität Konstanz

### [Datensätze und Codebücher]

In jeder der 13 Erhebungen wurde ein großer Anteil an gleichbleibenden Variablen erhoben. Ein kleinerer Teil wurde variiert, einerseits aufgrund methodischer Weiterentwicklungen, andererseits aufgrund aktueller Entwicklungen oder neuer Schwerpunktsetzungen. Insofern existiert ein großer Datenfundus, der über die Zeit hinweg stabil geblieben ist und mehrere bis alle Wellen umfasst. Gleichzeitig existieren Daten, die nur in einer einzigen Erhebung vorzufinden sind. Sämtliche Variablen, die in mehr als einer Welle erfragt wurden, sind Teil des Mehrwellendatensatzes. Daten, die nur ein einziges Mal erhoben wurden sind auch nur in dem entsprechenden Einzeldatensatz zu finden.

Aufgrund der Überarbeitung der Datensätze und deren Dokumentationen wurden die Variablen neu codiert. Jede Variable hat nun in jeder Welle und in der Mehrwellendatei den immer gleichen Namen bzw. die gleiche Nummer. So wurde z.B. die Variable „Alter“ in allen Datendateien als „v4“ codiert.

In der Reihenfolge der Variablenbenennungen wurde mit den Variablen der 13. Welle begonnen, dann systematisch die späteren Wellen angehängt. Variablen, die nur einmal vorkommen erhielten ihre Nummer erst nachdem alle Mehrwellenvariablen aus allen Erhebungen benannt wurden. Daher ist nicht in jeder Welle eine vollständige Reihenfolge an Variablennummern enthalten. In den Datensätzen erhielten alle Variablen und alle Antworten Labels.

Zu jedem Datensatz existiert ein Codebuch für die Zuordnung der Daten zu den Variablen. Für die Einzelwellen sind dies die Fragebogen selbst, in denen die Variablennummer bei den entsprechenden Items eingefügt wurden. Dazu wurde für jede Welle eine Variablenliste erstellt, die alle Variablen auflistet mit Zusatzangaben wie Anzahl der Beobachtungen, Anzahl differenter Ausprägungen, Wertebereich, Mittelwert, Minimum, Maximum sowie die zugehörigen Labels für die Variablen. Die Dateien sind jeweils mit der Erhebungswelle und der Jahreszahl der Erhebung gekennzeichnet.

In der 13. Erhebung wurde die Gesamtstichprobe nach neun Gruppen unterschieden, die jeweils unterschiedliche Designs aufweisen (s.o.). Dabei sind nicht alle Variablen in allen Gruppen gestellt worden. Daher wurden im Fragebogen-Codebuch drei verschiedene Farben für die Variablennummern gewählt, um anzuzeigen, welche Variablen zu welchen Gruppen gehören. Die blauen Variablennummern kennzeichnen die Variablen, die in den (um 40% gekürzten) Kurzversionen (Gruppen 5-8) vorhanden sind. Diese Variablen sind natürlich auch in den Langversionen und somit in allen Gruppen vorhanden. Die roten Variablennummern bezeichnen die Variablen, die nur in den Langversionen gestellt wurden (Gruppen 0-4 und 9). Die grünen Variablennummern kennzeichnen jene Variablen, die nur in den vier Modulen der Kurzversionen zusätzlich gestellt wurden, also nur in jeweils einer der vier Gruppen (5-8) vorkommen.

Für die Mehrwellendatei wurde ein gesondertes Codebuch erstellt, das in Tabellenform die Variablennummern den Fragen und Items zuordnet, mit Angabe der Ausprägungen und Wellenzugehörigkeit sowie dem Bezug zu den Fragennummern der Einzelwellen. Dazu gibt es eine Extraspalte für Anmerkungen, in der wichtige Angaben zu Besonderheiten und zur Nutzung der Daten stehen. Die Abfolge der Variablen wurde dabei einerseits nach inhaltlichen Themen bestimmt und andererseits

nach ihrem Auftreten im Zeitverlauf. D.h., es wurde mit jenen Variablen begonnen, die in der 13. Erhebung vorhanden waren und auch in früheren Wellen erfragt wurden. Dabei wurden allerdings auch ältere Daten aufgenommen, wenn sie thematisch dazu gehören. Danach folgen die Variablen, die nur bis zur 12. Erhebung erfragt wurden und danach die Variablen, die bis zur 11. erfragt wurden, usw. Die letzten Einträge sind dann jene Variablen, die nur in den ersten beiden Wellen vorhanden waren. Insgesamt enthält der SUF-Remote-Datensatz der Mehrwellendatei 1.196 Variablen bei einem Umfang von 108.969 Fällen. Der stärker anonymisierte SUF-Download-Datensatz enthält eine Variable weniger.

Für bestimmte Variablengruppen wurden neue Variablen kreiert, vorrangig um Aggregate herzustellen. Das betrifft z.B. die Unterscheidung nach Hochschulart (Universitäten vs. Fachhochschulen) oder die Bundesländer (alte vs. neue Bundesländer) der Hochschulen. Darüber hinaus wurde die Vielzahl an Studienfächer in Fächergruppen und Studienbereiche kategorisiert. Ebenso die Angaben zu den Fremdsprachen. Für einige wenige Variablen wurden Summenscores kreiert. Diese neuen Variablen sind alle im Codebuch zur Mehrwellendatei aufgelistet und werden dort erklärt. In den Codebüchern der Einzelwellen sind sie in den Variablenlisten aufgenommen und mit labels versehen, in den codierten Fragebogen werden sie vor den Fragebogenseiten aufgelistet. Die meisten dieser Variablen wurden bereits bei der Erstellung der Dateien eingeplant, weshalb sie in die laufenden Variablennummern eingebunden sind. Einige wenige wurden später erstellt, weshalb sie Zusatzkennungen zu den bereits bestehenden Variablen erhielten. Dazu wurde ein Buchstabe an die Variablennummer angehängt (z.B. 75b, v85a, v274k oder v352i), die für Bereinigung (s.u.), Neukreierung, Kategorisierung oder Indikatorerstellung zu der betreffenden Variable steht.

### **[Dokumentationen]**

Neben den Datensätzen und Codebücher sind weitere Dokumentationen erhältlich. Das betrifft zum einen Originalmaterialien, wie Originalfragebogen, Einladungsschreiben, Erinnerungsschreiben und Informationsmaterial. Zum anderen wurde eine spezielle Dokumentation zu den Studienfächern des Studierendensurveys und deren Codierung im Zeitverlauf erstellt. Dort sind auch die Zusammenstellungen der aggregierten Variablen beschrieben, die Fächergruppen und Studienbereiche enthalten. Dabei wird auch ausführlich auf die Vereinbarkeit mit den Codierungen des Statistischen Bundesamtes verwiesen.

### **[Wertebereiche]**

Die Antworten der Studierenden sind alle als Zahlenwerte codiert. Am häufigsten sind siebenstufige Skalen, die zumeist einen Wertebereich von 0-6 aufweisen. Bewertungsskalen, die von -3 bis +3 rangieren, sind mit 1-7 codiert, in den Labels aber benannt. Diese Codierung stammt noch aus den frühen Erhebungsjahren und wurde hier aus traditionellen Gründen beibehalten. Alle anderen Skalen (z.B. vier- oder fünfstufig) sowie auch nominale Auswahlkategorien erhielten Wertebereiche von 1 bis x (=höchste Zahl der Stufen oder Auswahlen).

Für alle Werte wurden die entsprechenden Labels in die Datensätze aufgenommen. Zudem ist in den Fragebogen-Codebüchern sowie dem Codebuch der Mehrwellendatei einsehbar, welche Zahlen-codes welche Antwortvorgaben bzw. Werte abbilden. Veränderungen von Wertebereichen über die Zeit sind in der Dokumentation der Mehrwellendatei angegeben.

Nur ganz wenige Fragen in den Erhebungen hatten offene Antwortmöglichkeiten. Mit Ausnahme der Fremdsprachen sind diese offenen Antworten nicht im Datensatz enthalten.

## **[Missing values]**

In allen Datensätzen sind fehlende Werte vorhanden, da nicht alle Studierenden immer alle Fragen beantwortet haben. Hinzu kommt, dass in jeder Erhebung festgelegt wurde, ab welchem Ausmaß von Abbrüchen die Antworten noch einbezogen werden. Daher sind auch Fragebogen vorhanden, bei denen z. B. nur die Hälfte beantwortet wurde.

In der neuerstellten Mehrwellendatei des Studierendensurveys für das FDZ sind alle Variablen enthalten, die in mindestens zwei Erhebungen gestellt wurden. Bei der Zuordnung der Daten wurden fünf verschiedene Missing-Codes eingesetzt.

1. [.] einfaches Missing: Damit werden alle fehlenden Werte bezeichnet, die darauf zurückgehen, dass die Befragten keine Angaben zu dem entsprechenden Item abgegeben haben. Dazu wird der Standardmissing-Code der Statistiksysteme verwendet. In STATA ist dies der Punkt „.“.
2. [-9] Systemmissing für Erhebungswellen. Nicht alle Variablen wurden in allen Wellen wieder abgefragt. Variablen, die in einer Welle nicht abgefragt wurden, erhalten dort den Code: „-9“.
3. [-10] Systemmissing für Online-Kurzbefragung. Dieser Missing-Code gibt es nur in der 13. Erhebung. Dort wurden in den Online-Langversionen alle Variablen verwendet, die auch in den Papierversionen vorhanden waren. Darüber hinaus wurden in vier Zusatzmodulen Kurzversionen erstellt, die nicht mehr alle Variablen enthielten. Die Variablen, die in diesen Onlinekurzversionen der 13. Erhebung nicht abgefragt wurden, erhalten den Code „-10“.
4. [-11] Systemmissing für Filterfragen. Im Studierendensurvey sind einige Filterfragen vorhanden, deren Angaben sich auf eine Vorfrage beziehen. Typische Beispiele für solche Vorfragen sind Fragen zu Tätigkeiten (z.B. Auslandsaufenthalt), die mit „ja“ oder „nein“ beantwortet werden können. In solchen Fällen wird meist zusätzlich ein weiteres Datum abgefragt, wie z.B. die Dauer des Aufenthaltes. Geben in diesen Fällen die Befragten in der Vorfrage ein „Nein“ an und bleibt die Folgefrage unbeantwortet, dann wird der Code „-11“ vergeben, um herauszustellen, dass diese Angaben gar nicht gemacht werden können, da die Bedingung hierfür nicht gegeben ist. In den Fällen, in denen die Befragten in der Vorfrage ein „Ja“ wählen oder keine Angaben dazu machen (= Missing), wird für Nichtbeantwortungen der Folgefrage der einfache Missing-Code verwendet.
5. [-12;] Systemmissing für Bereinigung. Einige Items wurden in ihren Wertebereichen bereinigt. Diese bereinigten Variablen werden im Datensatz mit der Endung b (z.B.: v75b) versehen, während die Originalvariablen weiterhin erhalten bleiben (v75). Die Regeln der Bereinigungen können in einer gesonderten Datei eingesehen werden. Angaben, die aufgrund der Bereinigung herausgenommen wurden, erhalten den Missing-Code „-12“ (siehe dazu unten Punkt „Bereinigung“).

## **[Gewichtung]**

In den Datensätzen sind keine Gewichtungsvariablen vorhanden.

## **[Fachcodierungen]**

Die Studienfächer werden im Studierendensurvey über eine Fächerliste auf der Innenseite des Deckblatts im Fragebogen ausgewählt und erhalten dort eine Zahlencodierung. In den Datensätzen sind die Fächernamen der Liste als Labels gesetzt. Aus den Fächern wurden zusätzlich Variablen für Studienbereiche und für Fächergruppen erstellt, darunter auch Zusammenstellungen, wie sie das Statistische Bundesamt ausweist. Die Codierung der Studienfächer sowie deren Zuweisungen zu den Fächer-

gruppen wie auch zu den statistischen Bundesamtkennungen ist in einer gesonderten Dokumentation beschrieben (siehe Datei: „Studienfächer“). Dort wird auch auf Besonderheiten und Probleme der Codierungen über den Zeitverlauf hinweg eingegangen.

Die Listen der Fragebogen-Codebücher und die der Originalfragebögen unterscheiden sich an einigen Stellen. Dies liegt daran, dass die Codelisten für die Mehrwellendatei neu erstellt werden musste. Diese Codierung wurde dann auch auf die Einzelwellen übertragen, damit in allen Dateien die immer gleichen Codes vorhanden sind. Die Originalfragebogen weisen dagegen die ursprüngliche Codierung auf, die über die Zeit hinweg nicht einheitlich war.

### **[Anonymisierung]**

Aufgrund der Datenschutzbestimmungen dürfen von den Daten ausgehend keine Rückschlüsse auf einzelne Personen möglich sein. Personen könnten in den Datensätzen des Studierenden surveys nur dann identifiziert werden, wenn Zugang zu den jeweils hochschuleigenen Studierendenkarteien und eine Zuordnung der eingeladenen Studierenden zu den Fragebögen bestehen würde. Eine solche Zuordnung der beteiligten Studierenden konnte bei den Papierbefragungen weder von den Hochschulen noch von der AGH erstellt werden, da die Fragebögen bereits vorverpackt waren und die zufällig ausgewählten Adressen von den Hochschulen auf diese aufgeklebt wurden. Da die ausgefüllten Fragebogen mittels mitgesandten Rückkuverts direkt an die AGH zurückgeschickt wurden, hatten auch die Hochschulen keinen Zugriff mehr auf die Unterlagen.

Bei der Onlineerhebung 2015 wurden Zugänge mittels Tokens verwendet, die mit den E-Mail-Einladungen zufällig verknüpft wurden. Die Tokenliste wurde vom DZHW verwaltet, die E-Mailadressliste von den Hochschulen. Beide Listen blieben getrennt und es wurde nur für die Verknüpfung der E-Mail-Einladung auf beide Listen zugegriffen, wobei zu keinem Zeitpunkt eine gemeinsame Datei erstellt wurde. Die Hochschulen hatten sich vorab bereit erklärt die E-Maillisten nach Ausgabe der Einladungen sofort wieder zu löschen. Die AGH erhielt nur die Daten der Onlinefragebögen mit Tokens, jedoch keine E-Maildaten.

Eine Zuordnung der Fragebögen zu Personen ist ohne Studierendendaten der Hochschulen nicht möglich. Mit viel Aufwand könnten vielleicht Bezüge zu Personen hergestellt werden, wenn sehr spezifische Subsample verwendet werden, die an bestimmten Hochschulen nur wenige Personen umfassen.

Gleichzeitig soll auch verhindert bzw. deutlich erschwert werden, dass die Hochschulen bestimmt werden können. Dazu wurden die Variablen, die Angaben zu den Hochschulen beinhalten, ebenfalls anonymisiert.

Die Anonymisierungsvorgaben werden auf zwei Stufen durchgeführt:

A. Für die Download-Dateien als SUF werden höhere Anforderungen an die Anonymisierung vorgegeben. Daher werden einige Variablen ausgeschlossen und andere kategorisiert bzw. einem Topcoding unterzogen. Im Einzelnen sind dies folgende Veränderungen:

1. Die Variable mit der Hochschulkennung wurde aus den Datensätzen entfernt (v8). Dafür wurde eine aggregierte Variable erstellt, die zwischen Universitäten und Fachhochschulen unterscheidet (v9).
2. Die Codierung der Bundesländer, in denen sich die Hochschulen befinden, wurde aggregiert, sodass nur nach „alten“ und „neuen“ Bundesländern unterschieden wird (v10).
3. Die Variable zu Behinderungen und chronischen Krankheiten wurde entfernt (v18, v19).
4. Die Variable Alter wurde kategorisiert (v4).
5. Die Variable Anzahl der Kinder wurde einem Topcoding unterzogen (v7).

6. Die Variable Bundesland der Hochschulreife wurde kategorisiert, so dass nur zwischen alten und neuen Ländern unterschieden wird (v64).
7. Die Variablen für das Zweit- und Drittfach wurden zu Fächergruppen kategorisiert (v41, v42).
8. Die Variable Note im Abschlusszeugnis der Hochschulberechtigung wurde kategorisiert (v74).
9. Die Variablen Hochschul- und Fachsemester wurden einem Topcoding unterzogen (v75, v76).
10. Die Variablen zu Fremdsprachenkursen im Studium wurden kategorisiert (v274-v277).
11. Die Variablen zum Jahr der Hochschulreife wurden kategorisiert (v60-v63, v66).
12. Für manche Variablen wurden kleine Subgruppen in die Kategorie „sonstige“ verschoben. Das betrifft die Variablen zu Leistungskurse in der Oberstufe (v71-v73) sowie die Fächercodierung der Studienfächer (v40) bzw. Studienbereiche (v44).

B) für die SUF-Dateien, die über einen Remote-Zugriff verwendet werden, bleiben viele dieser Variablen im Original erhalten. Hier werden nur folgende Veränderungen durchgeführt:

1. Die Variable mit der Hochschulkennung (v8) wurde beibehalten, jedoch wurden die Codes unsystematisch als Zahlencodes vergeben, sodass über die Erhebungen hinweg Zuordnungen bestehen, aber keine namentliche Zuordnung vorhanden sind.
2. Die Variable zu Behinderungen und chronischen Krankheiten wurde entfernt (v18, v19).
3. Die Variable Alter wurde kategorisiert (v4).
4. Die Variable Anzahl der Kinder wurde einem Topcoding unterzogen (v7).
5. Die Variablen Hochschul- und Fachsemester wurden einem Topcoding unterzogen (v75, v76).

### **[Bereinigung von Daten]**

In den Datendateien wurden bestimmte Variablen bereinigt. Dabei wurden jeweils die Original-Variablen beibehalten (z.B. v74) und die bereinigten Variablen mit dem Zusatz „b“ versehen (z. B. v74b). All diese Variablen werden auch im Codebuch zur Mehrwellendatei beschrieben.

Die Bereinigung betrifft alle offenen Zahlenangaben der Studierenden in den 13 Fragebogen. Darunter fallen Angaben für:

- Anzahl Jahre bzw. Jahreszahlen (Jahr der HS-Reife)
- Anzahl Semester (Fachsemester, Hochschulsemester, Hiwi/Tutor-Zeiten, Regelstudienzeit, geplante Fachsemester bis Abschluss)
- Anzahl Monate (Tätigkeiten vor Studium, Praktikum, Forschungsprojekt, Unterbrechung des Studiums, Auslandsstudium, Sprachkurse im Ausland)
- Anzahl ECTS-Punkte
- Stunden (Timebudget)

Ziel der Bereinigung war, jene Angaben herauszunehmen, die logisch/mathematisch nicht möglich oder real sehr unwahrscheinlich sind. Gleichzeitig sollten dabei aber möglichst viele Informationen beibehalten werden. Daher wurden auch eher extreme Angaben zugelassen, wenn sie theoretisch möglich sind und mit weiteren Merkmalen sinnhaft übereinstimmen. Dadurch verbleiben allerdings auch in den bereinigten Daten immer noch große Spannweiten.

Daten, die durch die Bereinigung wegfallen, erhalten den Systemmissing-Code -12.

Im Einzelnen sind folgende Bereinigungen durchgeführt worden:

Für die Bereinigung des Jahrs der Hochschulreife wurde ein Abgleich mit bereinigtem Alter und Jahr der Erhebung durchgeführt. Die bereinigte Spanne für das HSR-Jahr beträgt 1927-2015.

Für die Angaben zu Fach- und Hochschulsesemester wurden ein Abgleich mit dem Alter durchgeführt, sowie ein Abgleich mit der Zeitspanne zwischen Erhebungsjahr und Jahr der Hochschulreife. Zusätzlich wurde geprüft ob die Angaben zum Fachsemester größer sind als zum Hochschulsesemester. In letzterem Falle werden die Hochschulsesemester missing (-12) gesetzt. Die Fachsemester werden zusätzlich auf missing (-12) gesetzt, wenn Bachelorstudierende ohne ersten Abschluss mehr als 20 Semester angeben.

Für die 12 Variablen mit Monatsangaben zu Tätigkeiten vor dem Studium wurden ein Abgleich mit Alter, der Zeitspanne zwischen Erhebungsjahr und Jahr der Hochschulreife sowie den Hochschulsesemestern durchgeführt.

Für die Zeitdauer für Hiwi- und Tutor-Tätigkeiten (Semester) wurde ein Abgleich mit Alter und Hochschulsesemester durchgeführt. Die bereinigten Spannen betragen nun 0-24.

Für die weiteren Variablen mit Monatsangaben wurden ebenfalls Abgleiche durchgeführt. Für die meisten Angaben wurden Alter und Fach- bzw. Hochschulsesemester einbezogen. Für die Zeit zwischen erstem Abschluss und Masterstudium wurde zusätzlich die Zeit zwischen HSR-Jahr und Erhebungsjahr beachtet. Die Angaben für die Regelstudienzeit wurden auf mindestens 2 gesetzt.

Die Wochenstundenangaben zum Studienbetrieb (Timebudget) wurden nicht bereinigt, da keine der einzelnen Variablen ungültige (zu große) Werte hat. Aus diesen Einzelvariablen wurden vier Summenscores gebildet. Diese Scores erreichen allerdings auch ungültige Werte (= mehr als eine Woche Stunden hat). Da aber nicht überprüfbar ist welche der summierten Einzelvariablen falsch benannt wurde, kann auch keine bereinigt werden. Daher wurde für jede der erstellten vier Summenscores jeweils eine neue Variable definiert, die angibt, ob die resultierende Summe realistisch (Wert=0, wenn Wochenstunden unter 140 liegen), unwahrscheinlich (Wert =1, wenn Wochenstunden zwischen 140 und 168 liegen) oder unmöglich ist (Wert=2, wenn mögliche Wochenstunden überschritten werden).

Für die Angaben zu den zur Verfügung stehenden Geldbeträgen wird ein Abgleich zwischen „Geld gesamt“ und „davon BaFöG“ durchgeführt. Für das BaFöG wird missing (-12) gesetzt, wenn der Wert über 940 liegt.