

Elke Middendorff / Marten Wallis

# 11. Sozialerhebung

Daten- und Methodenbericht zur  
Studierendenbefragung 1985

Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



#### **Projektleitung**

Dr. Elke Middendorff  
Telefon +49 (0)511 450670-432  
E-Mail: middendorff@dzhw.eu

Marten Wallis  
Telefon +49 (0)511 450670-434  
E-Mail: wallis@dzhw.eu

#### **Projektmitarbeiter\*in**

Cagla Belgin Varol

#### **Impressum**

##### **Herausgeber**

Deutsches Zentrum für Hochschul- und  
Wissenschaftsforschung GmbH (DZHW)  
Lange Laube 12 | 30159 Hannover | [www.dzhw.eu](http://www.dzhw.eu)  
Postfach 2920 | 30029 Hannover  
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

##### **Geschäftsführung:**

Prof. Dr. Monika Jungbauer-Gans  
Dr. Thorsten Kowalke  
Vorsitzender des Aufsichtsrats:  
Ministerialdirigent Peter Greisler

##### **Registergericht:**

Amtsgericht Hannover | B 210251  
Umsatzsteuer-Identifikationsnummer:  
DE291239300

Oktober 2022

# Inhaltsverzeichnis

Tabellenverzeichnis	I
<b>1 Einleitung</b>	<b>2</b>
<b>2 Datenaufbereitung</b>	<b>4</b>
2.1 Vergabe von Variablennamen, Variablen- und Wertelabels .....	4
2.1.1 Schema der Variablennamen .....	4
2.1.2 Präfix und Suffix.....	5
2.2 Systematik fehlender Werte .....	5
<b>3 Gewichtung</b>	<b>7</b>
3.1 Vorgehen und Anwendungshinweise .....	7
3.2 Gewichtung des Datensatzes .....	8
<b>4 Anonymisierung</b>	<b>10</b>
Literaturverzeichnis	13

## Tabellenverzeichnis

Tabelle 1: Teilelemente und Zusammensetzung des Variablenstammes .....	4
Tabelle 2: Themengebiete in den Variablennamen .....	5
Tabelle 3: Systematik fehlender Werte im Datensatz des Primärforschungsprojektes und im SUF .....	6
Tabelle 4: Bereitgestellte Gewichte zur 11. Sozialerhebung (1985) .....	8
Tabelle 5: Überblick zur Anonymisierung der 11. Sozialerhebung (1985) .....	12

# 1 Einleitung

Die Sozialerhebung des Deutschen Studentenwerks (DSW) ist eine seit 1951 bestehende Untersuchungsreihe zur wirtschaftlichen und sozialen Lage der Studierenden in Deutschland.<sup>1</sup> Sie wird seit 1982<sup>2</sup> (10. Sozialerhebung) im Auftrag des bzw. seit der 21. Sozialerhebung in Kooperation mit dem DSW durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)<sup>3</sup> durchgeführt. Das Bundesministerium für Bildung und Forschung (BMBF) fördert die Studie seit der 6. Sozialerhebung (1967/1968). Die Sozialerhebung dient – in Ergänzung zur amtlichen Hochschulstatistik – unter anderem dem nationalen und internationalen Bildungsmonitoring. Darüber hinaus liefert sie wichtiges Steuerungswissen für hochschul- und sozialpolitische Fragen sowie belastbare und umfassende Daten für die Forschung.

Die 11. Sozialerhebung wurde im Sommersemester 1985 durchgeführt und bezog Studierende an 53 Hochschulen in der Bundesrepublik Deutschland zuzüglich Berlin-Wests ein: 26 wissenschaftliche Hochschulen, darunter drei Pädagogische Hochschulen und eine Kunsthochschule sowie 27 Fachhochschulen.<sup>4</sup> Bundeswehrhochschulen und Verwaltungsfachhochschulen waren nicht einbezogen.

Befragt wurden deutsche Studierende im Vollzeitstudium. Der Rücklauf von beurlaubten, exmatrikulierten, ausländischen Studierenden oder Gasthörer\*innen blieb unberücksichtigt.<sup>5</sup>

Der Fragekatalog wurde ergänzt durch einen Zusatzbogen „Studium und Freizeit“, mit dem eine sich Teilstichprobe (14 % der Fälle im Datensatz) befragt wurde. Das Primärforschungsprojekt hatte die Daten des Zusatzbogens in einem eigenen Datensatz abgelegt. Für die Erstellung des Scientific Use Files wurden dieser mit den Daten der Hauptbefragung zusammengeführt. Darüber hinaus erfassten zwei Zusatzbogen die lokalen Besonderheiten am Studienort München (6 % der Fälle im Datensatz) bzw. Tübingen (3 % der Fälle im Datensatz). Das Primärforschungsprojekt hatte diese Daten separat

---

<sup>1</sup> Berichte verschiedener Sozialerhebung stehen auf der Website des Nachfolgeprojekts zur Verfügung ([https://www.dzhw.eu/forschung/projekt?pr\\_id=650](https://www.dzhw.eu/forschung/projekt?pr_id=650)). Mit der neuen Studienreihe Die Studierendenbefragung in Deutschland – Eine für alle werden neben der Sozialerhebung der Konstanzer Studierendensurvey und die Studienreihe Beeinträchtigt studieren zusammengeführt.

<sup>2</sup> Die 1. (1951) und 2. Sozialerhebung (1953) wurden vom Studentenwerk Frankfurt am Main im Auftrag des Verbands Deutscher Studentenwerke durchgeführt. Das Studentenwerk führte auch die 3. (1956) bis 9. Sozialerhebung (1979) durch, die vom Deutschen Studentenwerk (DSW) beauftragt wurden. Einen detaillierten Überblick über Akteure, Methoden, Themen und projektbezogene Publikationen der Untersuchungsreihe von ihren Anfängen bis zur 21. Sozialerhebung bietet ein Working Paper von Middendorff (2022), das im MDM des FDZ-DZHW zur Verfügung steht. (<https://t1p.de/se-history>).

<sup>3</sup> Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informationssystem GmbH. Im nachfolgenden Text wird stets der Begriff DZHW verwendet, auch wenn die Studie vor der Ausgründung 2013 durchgeführt wurde. Auf den Originaldokumenten der 11. Sozialerhebung (Fragebogen, Anschreiben) sowie in den Berichten zum Projekt ist das Deutsche Studentenwerk (DSW) als Akteur benannt. Der Hauptbericht wurde vom Bundesminister für Bildung und Wissenschaft herausgegeben. In seinem Impressum ist die HIS GmbH (HIS) als mit der Untersuchung vom DSW beauftragte Institution gekennzeichnet.

<sup>4</sup> Siehe Codierliste [cl-dzhw-64.xlsx](#)

<sup>5</sup> Erst ab der 14. und bis zur 16. Sozialerhebung wurden Studierende mit ausländischer Staatsangehörigkeit innerhalb des gemeinsamen Fragebogens mit einem auf sie zugeschnittenen Frageblock angesprochen und über nicht passende Fragen hinweggefiltert. Im Rahmen der 17. – 21. Sozialerhebung erhielten bildungsausländische Studierende einen eignen, zwei- bzw. mehrsprachigen Fragebogen (vgl. auch Middendorff 2022, S. 6 f.).

vom Datensatz der Hauptbefragung abgelegt und nicht für die bundesweite Auswertung verwendet. Diese Regionaldaten sind demzufolge auch nicht Bestandteil der SUF-Aufbereitung. Darüber hinaus enthielt der Fragebogen einmalig Fragen Zufriedenheit mit dem Studium sowie zu Antizipationen in Bezug auf die Beschäftigungssituation (s. Middendorff 2022, S. 12). Eine Unterstichprobe (ca. 18 %) beantwortete den Zusatzbogen „Studium und Freizeit“. Diese Daten waren Grundlage für den Sonderbericht *Hochschulische und außerhochschule Aktivitäten im Studienalltag* (Weidner-Russell, & Haase, 1987). Ein weiterer Sonderbericht widmete sich dem Weg zur Hochschule (Weidner-Russell & Haase, 1986).

Im Rahmen der Tätigkeit des vom BMBF geförderten Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten einiger jüngerer Erhebungen dieser Reihe nachträglich zum Zweck der Nachnutzung aufbereitet und dokumentiert.<sup>6</sup> Die 11. Sozialerhebung wird als Scientific Use File (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung gestellt. Neben dem Datensatz der Erhebung wird auch Dokumentationsmaterial zum Datensatz und zur Durchführung der Studie bereitgestellt.<sup>7</sup>

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur 11. Sozialerhebung (doi: 10.21249/DZHW:ssy11:1.0.0). Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Kapitel 2. Kapitel 3 beschreibt Aspekte der Datenaufbereitung, Kapitel 4 und 5 enthalten die Beschreibung der vorgenommenen Gewichtung bzw. der Anonymisierung.

Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebogen etc.) können frei im Metadatensystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Es ist geplant, einen kumulierten Datensatz der 10.-21. Sozialerhebung zu erstellen. Dieser Datensatz wird voraussichtlich in der ersten Jahreshälfte 2023 sowohl als Scientific Use File als auch als Campus Use File zur Verfügung stehen.

---

<sup>6</sup> Da zum Erhebungszeitpunkt der Daten keine Datennachnutzung vorgesehen war, sind einige Informationen zur Erhebung nicht mit dem Fokus einer späteren Datennachnutzung dokumentiert worden und teilweise nicht mehr rekonstruierbar. An entsprechenden Stellen ist dies im Text angemerkt.

<sup>7</sup> Informationen zu verfügbaren Datensätzen und Dokumentationen können im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu/#!/de/studies/stu-ssy11?version=1.0.0>) heruntergeladen werden.

## 2 Datenaufbereitung

### 2.1 Vergabe von Variablennamen, Variablen- und Wertelabels

#### 2.1.1 Schema der Variablennamen

Das FDZ-DZHW hat einen Standard zur Variablenbenennung entwickelt, der in den hier aufbereiteten SUF angewendet wird. Es besteht aus einer Präfix-Stamm-Suffix-Systematik: Der Variablenname enthält in Präfix und Suffix zentrale Metadaten, die für die strukturierte Verarbeitung von Variablen nötig sind. Der Stamm enthält zwei hierarchisch zusammenhängende Differenzierungen: Kennzeichnung des Themas sowie eine numerische Ordnung innerhalb des Themas.

Die systematische Vergabe von Stamm und Präfix sind unerlässlich, da sie Metadaten enthalten, die für die weitere (Meta)Datenaufbereitung notwendig sind. Nach der Evaluation der bisherigen Erfahrungen wurde die „thematische Freigabe“ des Stamms als bestes Mittel der Ressourcenverminderung bei gleichzeitig möglichst hohem Informationsgehalt identifiziert.

Tabelle 1: Teilelemente und Zusammensetzung des Variablenstammes

Teilelement	Beschreibung
Themendifferenzierung*	Mit einem (englischen) Kürzel aus drei Buchstaben wird die Variable einem inhaltlichen Themengebiet zugeordnet.
Nummerierung*	Innerhalb der definierten Themenbereiche werden die Variablen auf minimal zwei, maximal drei Stellen durchnummeriert.
Indizierung	Mit Hilfe eines Buchstabens am Ende des Stamms können verschiedene Variablen, die zur gleichen Frage gehören und dadurch die gleiche Themendifferenzierung und Nummerierung aufweisen (z. B. bei Itembatterien, Mehrfachnennungen oder Fragen, in denen geschlossene und offene Fragen kombiniert werden), gekennzeichnet werden (z. B. 01a, 01b, 01c, ...). Falls eine Frage den Umfang von 26 Einzelvariablen (a-z) überschreitet, wird die Itembezeichnung ab dem 27. Item mit zwei Buchstaben fortgesetzt (aa, ab, ac, ...).

\* muss zwingend vergeben werden

Im Folgenden wird das Variablennamenschema dargestellt, welches für den gepoolten Datensatz der 13.-21. bzw. der 17.-21. Sozialerhebung verwendet wurde und das sich am sogenannten Goldstandard des im FDZ-DZHW entwickelten einheitlichen Variablennamenschema orientiert (vgl. ebenda). Die Zusammenfügung von Datensätzen setzt voraus, dass identische Variablen und/oder identische Fälle als solche eindeutig identifizierbar sind. Für die beiden bisher vorliegenden gepoolten Datensätze aus neun (13.-21. Sozialerhebung) bzw. fünf Sozialerhebungen (17.-21. Sozialerhebung) wurde deshalb ein einheitliches Variablennamenschema angewandt, um dieser Anforderung zu entsprechen. Darüber hinaus gibt es kohortenbezogene Variablenspezifika, wie z. B. zusätzliche Items einer Itematterie, Modifikationen in der Formulierung der Frage und/oder Antwort(en). Diese Besonderheiten sollen im Variablennamen systematisch kenntlich gemacht werden, damit Nutzer\*innen sowohl die thematische Zugehörigkeit als auch die Besonderheit einer Variablen erkennen

können. Für das SUF der 11. Sozialerhebung wird dieses Schema insoweit übernommen, wie es für einen Einzeldatensatz erforderlich ist, d. h. die Ausweisung kohortenbezogener Variablenspezifika entfällt. Mit der Übernahme des Variablennamenschemas wird die Voraussetzung dafür geschaffen, die bestehenden gepoolten Datensätze um die Daten der 11. Sozialerhebung zu erweitern. Darüber hinaus wird dadurch den zeitreiheninteressierten Nutzer\*innen des gepoolten Datensatzes die Orientierung im SUF der 11. Sozialerhebung erleichtert.

Tabelle 2: Themengebiete in den Variablennamen

Nr.	Themengebiete-Kürzel (= Stamm)	Themengebiet (englisch)	Themengebiet (deutsch)
1	<b>dem</b>	socio-demographic characteristics	sozio-demographische Merkmale
2	<b>par</b>	characteristics of parents	Merkmale der Eltern
3	<b>stu</b>	characteristics of study	Merkmale des Studiums
4	<b>ped</b>	prior education and entry into HE	Vorbildung und Hochschulzugang
5	<b>fin</b>	financing (of living during studies)	Finanzierung (des Lebensunterhalts während des Studiums)
6	<b>baf</b>	BAföG (German Federal Grant on Training and Education Promotion)	BAföG (Bundesausbildungsförderungsgesetz)
7	<b>tim</b>	time usage (studies/job)	Zeitaufwand für Studium, Erwerbstätigkeit und Freizeit
8	<b>job</b>	job during studies	Erwerbstätigkeit während des Studiums
9	<b>abr</b>	studying abroad	studienbezogener Auslandsaufenthalt
10	<b>liv</b>	living (accommodation)	Wohnsituation
11	<b>nut</b>	mensa and nutrition	Mensa und Ernährung
12	<b>way</b>	way and mode of transportation to university	Weg zur Hochschule und Verkehrsmittelwahl

### 2.1.2 Präfix und Suffix

Das Präfix kennzeichnet die Welle, mit der eine Wiederholungsmessung erfolgt ist. Da die Sozialerhebung eine Untersuchungsreihe im Querschnittsdesign ist, entfällt das Präfix im Variablennamen.

Im Suffix des Variablennamens wird gekennzeichnet, ob eine Variable generiert, versioniert, anonymisiert, plausibilisiert oder harmonisiert wurde bzw. auf welchem Zugangsweg (Download, Remote-Desktop, On-Site) sie bereitgestellt wird. Für das vorliegende SUF wird ausschließlich das Suffix `_g` verwendet um Variablen zu kennzeichnen, die aus einer oder mehreren Variablen des Ursprungsdatensatzes erzeugt wurden (Recodierungen, Indizes, vercodete Variablen, Aggregationen).

## 2.2 Systematik fehlender Werte

Der Datensatz des Primärforschungsprojektes unterschied zwei Missings: „keine Angabe“ und „trifft nicht zu“. Es wurden verschiedene Codes für „keine Angabe“ verwendet (s. Tabelle 3). Darüber hinaus weisen die Daten für viele Variablen eine Reihe unbelabelter (System-)Missings auf, deren Be-

deutung ex post nicht vollständig geklärt werden konnte. Einige dieser (System-)Missings erwiesen sich als filterbedingt fehlend und wurden im Rahmen der Datenaufbereitung entsprechend vercodet. Konnte die Bedeutung nicht ermittelt werden, wurden diese Missings als „unbekannter fehlender Wert“ gekennzeichnet.

Tabelle 3: Systematik fehlender Werte im Datensatz des Primärforschungsprojektes und im SUF

Primärforschungsprojekt		SUF ssy11	
Code	Wertelabel	Code	Wertelabel
0	[ohne Label]		
0	keine Angabe		
-1	keine Angabe	-998	keine Angabe
-2	keine Angabe		
9	[ohne Label]		
0	[ohne Label]		
9	[ohne Label]		
-1	[ohne Label]	-989	filterbedingt fehlend
-2	[ohne Label]		
.	[ohne Label]		
0	[ohne Label]		
1	[ohne Label]		
6	[ohne Label]		
998	[ohne Label]	-969	unbekannter fehlender Wert
999	[ohne Label]		
.	[ohne Label]		



## 3 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe aufgrund des Stichprobendesigns sowie unterschiedlicher Mitwirkungsbereitschaft verschiedener Gruppen in der Grundgesamtheit. Sie erfolgt im Vergleich zur definierten Grundgesamtheit. Nach einer allgemeinen Einführung in die Vorgehensweise und einer Darstellung der erstellten Gewichte wird die Gewichtungsprozedur im Detail beschrieben.

### 3.1 Vorgehen und Anwendungshinweise

**[Ursachen für die Verzerrungen der Stichproben]** Maßgeblich für die Verzerrungen von Stichproben sind zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um für bestimmte relevante Subgruppen die Fallzahlen zu erhöhen.
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe. Wenn diese Ausfallsprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.<sup>8</sup> Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Missing Not at Random), der einer Modellierung bedarf.<sup>9</sup>

**[Konzeptuelles Vorgehen]** Im Zuge einer Gewichtungsprozedur sollten idealerweise zunächst designbedingte Disproportionalitäten ausgeglichen werden. Die hierfür benötigten *Designgewichte* ergeben sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Im Anschluss sollte eine Adjustierung der Designgewichte mit Hilfe von *Nonresponsegewichten* im Quer- und Längsschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf Individualebene erzeugt werden.

In einem letzten Schritt können die nonresponseadjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden (Kalibrierung).

Aufgrund des Stichprobendesigns der 11. Sozialerhebung wird in einem ersten Schritt ein Designgewicht gebildet, um die ungleichen Inklusionswahrscheinlichkeiten auszugleichen. Da auf individueller Ebene keine Informationen zu Nichtteilnehmer(inne)n vorliegen, kann keine Nonresponse-Adjustierung des Designgewichts auf Individualebene erfolgen. Das Designgewicht wird in einem letzten Schritt anhand einer Merkmalsverteilung der Grundgesamtheit kalibriert. Da hier Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf aggregierter Ebene vorliegen, erfolgt

---

<sup>8</sup> Das trifft dann zu, wenn die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

<sup>9</sup> Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

hier zugleich eine Form der Nonresponse-Adjustierung. In Tabelle 4 sind die erstellten Gewichte dargestellt.

Tabelle 4: Bereitgestellte Gewichte zur 11. Sozialerhebung (1985)

Variablenname	Beschreibung
<i>gewireg</i>	Gewicht für Analysen auf Regionalebene - Süd, West Nord (inkl. Berlin-West)
<i>gewibund</i>	Gewicht für Analysen auf der Ebene Westdeutschlands (inkl. Berlin-West)

**[Hinweise zur Anwendung der Gewichte]** Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe ado-spezifischer Optionen berücksichtigt werden können.<sup>10</sup> Es stehen zwei Gewichte zur Verfügung für Analysen auf verschiedenen Ebenen innerhalb der elf alten Länder. Das Gewicht *gewireg* ist für Analysen auf Regionalebene (Süd-, West-, Norddeutschland - inklusive Berlin-West). und das Gewicht *gewibund* für bundesweite Auswertungen (bezogen auf Westdeutschland inkl. Westberlin) gedacht. Beide Gewichte berücksichtigt die Merkmale Fächergruppe, Hochschulart, Geschlecht. Bei 30 Fällen fehlte die Angabe zum Bundesland; sie erhalten bei beiden Gewichten den Gewichtungsfaktor 1,0.

Grundlegend ist zu beachten, dass Gewichte nur dann sinnvolle Korrekturgrößen darstellen, wenn das verwendete Analysemodell die zur Gewichtung herangezogenen Variablen enthält oder mit diesen in einem Zusammenhang steht. Aus diesem Grund müssen Gewichte immer mit Fokus auf die analysierte Fragestellung verwendet werden. Im Folgenden wird die Vorgehensweise bei der Erstellung des Gewichtes näher dargestellt.

## 3.2 Gewichtung des Datensatzes

**[Designgewichtung]** Aufgrund des Stichprobendesigns sind Studierende einiger Hochschulen überrepräsentiert. Die deshalb bestehende höhere Wahrscheinlichkeit für Studierende dieser Hochschulen, in die Stichprobe zu gelangen, wurde durch eine Designgewichtung ausgeglichen. Elemente, die mit höherer Wahrscheinlichkeit als andere in die Stichprobe eingehen, erhalten somit ein niedrigeres Gewicht und umgekehrt.

**[Kalibrierung der Designgewichte]** Eine Nonresponse-Adjustierung der Designgewichte war auf Individualebene nicht möglich. Es lagen jedoch Informationen zu folgenden Merkmalen der Grundgesamtheit<sup>11</sup> vor, die zur Kalibrierung der Gewichte verwendet werden konnten: Region der Hochschule, Geschlecht, Fächergruppe, Hochschultyp, Personen mit deutscher Staatsangehörigkeit versus Bildungsinländer\*innen.<sup>12</sup> Bei der Redressment-Gewichtung wurden einzelne, sehr kleine Zellbesetzungen, die zu sehr hohen Gewichten führen würden, zusammengefasst.

<sup>10</sup> Siehe hierzu die Stata-Hilfe (Befehl: help weights).

<sup>11</sup> Alle Informationen, die zur Kalibrierung der Designgewichte verwendet wurden, leiten sich aus Daten des Statistischen Bundesamtes zum Semester vor der Erhebung (WiSe 1984/1985) ab, da die aktuelle Statistik zum Zeitpunkt der Gewichtung noch nicht vorlag.

<sup>12</sup> Die Gewichtung wurde entlang folgender Ausprägungen durchgeführt: Geschlecht: weiblich versus männlich; Region: Nord (Bremen, Hamburg, Niedersachsen, Schleswig-Holstein, Berlin-West), Süd (Baden-Württemberg, Bayern), West (Hessen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland); Hochschultyp: Universität (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen) versus Fachhochschule; Fächergruppe.

Die Kalibrierung erfolgte sowohl auf Bundesebene als auch gesondert für jedes Bundesland. Im Prozess der Anonymisierung wurden die Bundesländer zu drei Regionen aggregiert (Süd-, West-, Norddeutschland).

Da die Merkmalsträger in der Grundgesamtheit ebenfalls Informationen über Nichtteilnehmer(innen) enthielten, erfolgte durch die Verwendung der Redressmentgewichte zusätzlich eine Art Nonresponse-Adjustierung im Hinblick auf die verwendeten Merkmale.

## 4 Anonymisierung

**[Datenschutzrechtlicher Rahmen]** Für personenbezogene Daten<sup>13</sup>, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017.<sup>14</sup> Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

**[Datenzugang, Anonymisierungsgrad und Analysepotential]** Das FDZ des DZHW stellt für die 11. Sozialerhebung ein SUF für die wissenschaftliche Sekundärnutzung zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt.

Das SUF wird via Download angeboten. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen für den Zugangsweg Download-SUF erläutert.

**[Statistische Anonymisierungsmaßnahmen]** Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Von diesen sogenannten *direkten Identifikatoren*, wie Namen, Adressen oder E-Mail-Adressen, wurde im Rahmen der 11. Sozialerhebung keine erfasst. Um einen Rückbezug auf die Originaldaten zu verhindern, wurde die Original-Identifikationsnummer aus dem Datensatz entfernt und durch eine neue, nach dem Zufallsprinzip vergebene Identifikationsnummer ersetzt.

Anschließend wurden die *Quasi-Identifikatoren* bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizie-

---

<sup>13</sup> „Personenbezogene Daten (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 DSGVO, S. 1).

<sup>14</sup> Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn.

ren.<sup>15</sup> Für die 11. Sozialerhebung wurden beispielsweise folgende Merkmale als Quasi-Identifikatoren eingestuft: Name sowie Art und Ort der Hochschule, Studienfach, Abschlussart, Alter. Um eine eindeutige Zuordnung der Daten der 11. Sozialerhebung zu betroffenen Personen zu unterbinden, wurden diese Schlüsselmerkmale aggregiert oder gelöscht (s. Tabelle 5).

Ebel und Meyermann (2015) empfehlen, offene Angaben in jedem Fall zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“ (Ebel & Meermann, 2015, S. 5). Die offenen Angaben waren größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet worden und werden in dieser Form zur Verfügung gestellt. Teilweise wurden jedoch – in Abhängigkeit von der Sensibilität der enthaltenen Informationen – die vom Primärforschungsprojekt vorgenommenen Codierungen zusätzlich aggregiert. Nicht codierte offene Angaben wurden im SUF gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung oder zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht unmittelbar zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (Koberg, 2016, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). In der 11. Sozialerhebung wurden gesundheitsbezogene Informationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten gelöscht. Die nachfolgende Tabelle 5 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen dar. Variablen, die im SUF aus Datenschutzgründen *nicht* verfügbar sind, sind im Variablenfragebogen mit einem entsprechenden Hinweis gekennzeichnet.

---

<sup>15</sup> Dabei ist darauf hinzuweisen, dass die Identifikation einer Person bereits durch die Stichprobenauswahl erschwert wird, da eine Ungewissheit darüber besteht, ob eine befragte Person eine einzigartige Merkmalskombination in der Population aufweist.

Tabelle 5: Überblick zur Anonymisierung der 11. Sozialerhebung (1985)<sup>16</sup>

Merkmal	Download-SUF
Original-ID	Löschung und Vergabe einer zufälligen ID
Studienfächer	Aggregation zu Fächergruppen <sup>a</sup>
Abschlussart	Zusammenfassung: „kirchliche Prüfung“ und „sonstiger Abschluss“ zu „sonst. Abschluss, inkl. kirchl. Abschl.“
Hochschule	Aggregation zu Hochschulart <sup>b</sup>
Bundesland des aktuellen Studiums	Aggregation zu Regionen <sup>b</sup>
Art der vorherigen Hochschule	Aggregation zu Hochschulart <sup>b</sup>
Bundesland der vorherigen Hochschule	Aggregation zu Regionen <sup>b</sup>
Gründe für Studienunterbrechung	Zusammenfassung: „gesundheitliche Probleme“ zu „sonstige Gründe“
Wartezeit bis Studienbeginn (in Monaten)	Aggregation: „bis 1 Jahr“, „mehr als 1 Jahr bis 2 Jahre“, „mehr als 2 Jahre bis 3 Jahre“, „mehr als 3 bis 5 Jahre“, „mehr als 5 bis 10 Jahre“ und „mehr als 10 Jahre“
Alter (in Jahren)	bis 39 einzeln ausgewiesen, ansonsten Aggregation: „40 bis 49“, „50 Jahre und älter“
Anzahl der Kinder	Aggregation: „3 Kinder und mehr“
Anzahl der Geschwister	Aggregation: „3 Geschwister und mehr“
(Sonstige) gesundheitliche bzw. private Informationen	Löschung
(Sonstige) offene Angaben	Vercodung/Löschung

a Aggregation orientiert an Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 1984/1985 und SoSe 1985, vgl. Codierliste cl-dzhw-63, die hier hinterlegt ist: [https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy11-ins1\\$?version=1.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy11-ins1$?version=1.0.0)

b Aggregation der Stichprobenhochschulen (s. S. 2) zu den beiden Hochschularten Fachhochschule und Universität (inklusive Pädagogische Hochschulen, Kunst- und Musikhochschulen) sowie Aggregation zu den Regionen Süd-, West- und Norddeutschland, vgl. Codierliste cl-dzhw-64, die hier hinterlegt ist: [https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy11-ins1\\$?version=1.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy11-ins1$?version=1.0.0).

<sup>16</sup> Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadatensuchsystem ([https://metadata.fdz.dzhw.eu/#!/de/studies/stu-ssy11\\$?version=1.0.0](https://metadata.fdz.dzhw.eu/#!/de/studies/stu-ssy11$?version=1.0.0)) zu entnehmen.

# Literaturverzeichnis

- Daniel, A. & Weber, A. (2017). Einheitliches Variablennamenschema für das FDZ des DZHW. Gold- und Silberstandard. Version 3.0. Projektbericht. Hannover: FDZ-DZHW.
- Ebel, T. & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten (Forschungsdaten Bildung informiert Nr. 3). Verbund Forschungsdaten Bildung.
- Griesbach, H., Schnitzer, K. (1987). 11. Sozialerhebung – Methodenfragen. Kurzinformation A2/87. Hochschul-Informationssystem GmbH (HIS).
- Isserstedt, W. (1988). Die soziale Herkunft der Studierenden. Unveröffentlichter Projektbericht. Hochschul-Informationssystem GmbH (HIS).
- HIS-Projekt Sozialerhebung (1988). Sozialstrukturelle Indikatoren der Bildungsbeteiligung – Überlegungen zu Ausprägung, Datenbasis, Bedeutung und Aussagekraft. Unveröffentlichter Projektbericht. Hochschul-Informationssystem GmbH (HIS).
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2
- Middendorff, E. & Wallis, M. (2022). 13.-21. Sozialerhebung 1991-2016. Daten- und Methodenbericht zum gepoolten Datensatz der neun Studierendenbefragungen. Hannover: FDZ-DZHW.
- Middendorff, E. (2022). Die Sozialerhebungen des Deutschen Studentenwerks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen, Methoden, Themen und projektbezogene Publikationen (Working Paper). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). (<https://t.co/kLF4ROWkcp>)
- Schnitzer, K., Isserstedt, W., Leszczensky, M. (1986a). Das soziale Bild der Studentenschaft in der Bundesrepublik Deutschland. Studien zu Bildung und Wissenschaft 42. Bundesminister für Bildung und Wissenschaft (Hrsg.). (11. Sozialerhebung des Deutschen Studentenwerkes). Bonn.
- Schnitzer, K., Isserstedt, W., Leszczensky, M. (1986b). Students in the Federal Republic of Germany – A Social Report. The 11th Social Survey of the Deutsches Studentenwerk. Hochschul-Informationssystem GmbH (HIS).
- Schnitzer, K., Leszczensky, M., Isserstedt, W. (1986). 11. Sozialerhebung. Zusammenfassung der Ergebnisse der 11. Sozialerhebung des Deutschen Studentenwerks im Sommersemester 1985 durchgeführt von der Hochschul-Informationssystem GmbH (HIS). Kurzinformation A 10/86. Hochschul-Informationssystem GmbH (HIS).
- Weidner-Russell, B., Haase, K. (1986). Wie fahren die Studenten zur Hochschule? Kurzinformation Bau und Technik B4/86. Hochschul-Informationssystem GmbH (HIS).

Weidner-Russell, B., Haase, K. (1987). Hochschulische und außerhochschule Aktivitäten im Studienalltag. Kurzinformation Bau und Technik B2/87. Hochschul-Informationssystem GmbH (HIS).

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(2), 581–592.