

Daten- und
Methodenbericht
April 2023

Elke Middendorff

Das soziale Profil in der Begabtenförderung

Daten- und Methodenbericht zur Befragung der
Geförderten 2008

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

Projektmitarbeiter*in

Cagla Belgin Varol

Impressum

Herausgeber

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

Geschäftsführung:

Prof. Dr. Monika Jungbauer-Gans

Vorsitzender des Aufsichtsrats:

Ministerialdirigent Peter Greisler

Registergericht:

Amtsgericht Hannover | B 210251
Umsatzsteuer-Identifikationsnummer:
DE291239300

April 2023

Inhaltsverzeichnis

Tabellenverzeichnis	I
1 Einleitung	2
2 Datenaufbereitung	4
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
3 Gewichtung	7
4 Anonymisierung	9
Literaturverzeichnis	12

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: Überblick zur Anonymisierung der Befragung der Geförderten der Begabtenförderungswerke	11

1 Einleitung

Hintergrund des Projektes war die vom Bildungsministerium gemeinsam mit den – damals elf – Begabtenförderungswerken¹ 2006 gestartete Informationskampagne „stipendium plus“². Mit ihr soll der Anteil der Studierenden, die als Begabte gefördert werden, bis zum Ende der sechszehnten Legislaturperiode (2005-2009) auf 1 Prozent erhöht werden. Im Jahr 2005 betrug der Anteil an Geförderten eines Begabtenförderungswerkes unter allen Studierenden etwa 0,7 Prozent. Um das Ziel von 1 Prozent erreichen zu können, waren die Mittel für die Begabtenförderung um fast 33 Millionen auf 113,2 Millionen Euro angehoben worden. So konnten 2008 schon knapp 0,9 Prozent der Studenten als Begabte gefördert werden. Eine weitere Steigerung der Fördermittel war für 2009 geplant. In diesem Zusammenhang sind von einigen Begabtenförderungswerken bereits verschiedene Maßnahmen ergriffen worden, die das Ziel haben, Begabungsreserven auch aus hochschulfernen Schichten zu entdecken und in die Förderung einzubeziehen. Das Bundesministerium für Bildung und Forschung (BMBF) und die elf Begabtenförderungswerke in Deutschland benötigten deshalb Informationen zur sozialen Zusammensetzung der mit Stipendien und Büchergeld geförderten Studierenden und Promovierenden.

Methodisch hatte die Studie einmalig gute Voraussetzungen: Sie konnte als Vollerhebung aller Geförderten mit aktiver Unterstützung der elf Begabtenförderungswerke durchgeführt werden. Die Werke leiteten die Einladungs-E-Mail der mit der Durchführung beauftragten Hochschul-Information System GmbH an ihre Geförderten weiter und motivierten mit einer eigenen Mail zur Teilnahme an der Online-Befragung.

Als Referenzrahmen für den Fragekatalog und zur Einordnung der Befunde diente die 18. Sozialerhebung aus dem Jahr 2006 (s. Isserstedt et al. 2007). Mit ihr standen belastbare Daten zum sozialen Profil des Querschnitts aller Studierenden in Deutschland zur Verfügung. Der Fragekatalog der Befragung der Geförderten orientierte sich deshalb bei den nicht unmittelbar auf die Förderung bezogenen Fragen an dem Instrumentarium der 18. Sozialerhebung. Dass die förderungsspezifischen Fragen zutreffend, relevant und geeignet sind, stellten Expertentests mit Angehörigen der Begabtenförderungswerke sicher.

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zum Scientific Use File (SUF) zur Online-Vollerhebung der Geförderten in der Studien- bzw. der Promotionsförderung der Begabtenförderungswerke 2008 (doi: 10.21249/DZHW:gif2008:1.0.0). Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen im 2. Kapitel. Kapitel 3 beschreibt Aspekte der Datenaufbereitung,

¹ Studienstiftung des deutschen Volkes, Friedrich-Ebert-Stiftung, Friedrich-Naumann-Stiftung für die Freiheit, Hanns-Seidel-Stiftung, Heinrich-Böll-Stiftung, Konrad-Adenauer-Stiftung, Rosa-Luxemburg-Stiftung, Hans-Böckler-Stiftung des Deutschen Gewerkschaftsbundes sowie die Stiftung der Deutschen Wirtschaft, Evangelisches Studienwerk e. V. Villigst, Bischöfliche Studienförderung Cusanuswerk. Nicht einbezogen waren Geförderte des 2009 gegründeten Ernst Ludwig Ehrlich Studienwerkes sowie des Avicenna-Studienwerkes, das 2012 gegründet wurde.

² <https://www.stipendiumplus.de/startseite.html>

Kapitel 4 und 5 enthalten die Beschreibung der vorgenommenen Gewichtung bzw. der Anonymisierung.

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.

Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebogen, Variablen-Fragbogen etc.) können frei im Metadaten-System des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) 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 das SUF der Online-Vollerhebung unter den Geförderten der elf Begabtenförderungswerke 2008 angewandt wurde. Es entspricht dem harmonisierten Variablennamenschema des kumulierten Datensatzes der 10. - 21. Sozialerhebung.³ Auch die Namen der einzelnen Variablen orientieren sich an diesem Scientific Use File (ssy-poolv3) bzw. entsprechen ihm, wenn die Items übereinstimmen.

³ Für den 2019 erstellten Einzel-SUF der 18. Sozialerhebung (<https://doi.org/10.21249/DZHW:ssy18:1.0.0>) wurden die mnemotechnischen Variablenamen des Primärforschungsprojektes übernommen. Es empfiehlt sich deshalb, für den direkten Vergleich den kumulierten Datensatz der 10. - 21. Sozialerhebung zu verwenden (<https://doi.org/10.21249/DZHW:ssypool:3.0.0>).

Tabelle 2: Themengebiete in den Variablennamen

Nr.	Themengebiete-Kürzel (= Stamm)	Themengebiet (englisch)	Themengebiet (deutsch)
1	dem	socio-demographic characteristics	sozio-demographische Merkmale
2	par	characteristics of parents	Merkmale der Eltern
3	stu	characteristics of study	Merkmale des Studiums
4	ped	prior education and entry into HE	Vorbildung und Hochschulzugang
5	fin	financing (of living during studies)	Finanzierung (des Lebensunterhalts während des Studiums)
6	job	job during studies	Erwerbstätigkeit während des Studiums
7	abr	studying abroad	studienbezogener Auslandsaufenthalt
8	liv	living (accommodation)	Wohnsituation

2.1.2 Präfix und Suffix

Das Präfix kennzeichnet die Welle, mit der eine Wiederholungsmessung erfolgt ist. Da die Befragung Geförderten der Begabtenförderungswerke eine Studie 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: -2 = „keine Angabe“ und -1 = „trifft nicht zu“. Es wurde darauf verzichtet, Fragen, die ausschließlich an Geförderte der Studienförderung (auch als Grundförderung bezeichnet) gerichtet waren, mittels Missing von Fragen zu unterscheiden, die nur Geförderten in der Promotionsförderung vorgelegt wurden. Nicht vorhandene Angaben wurden unterschiedslos als -2 = „k. A.“ ausgewiesen. Für den SUF wurden diese Werte als -987 = „designbedingt fehlend“ gekennzeichnet, weil ein Teil der Fragen via Split des Online-Erhebungsinstrumentes der jeweils nicht betroffenen Gruppe nicht gestellt wurden (s. Tabelle 3).

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

Primärforschungsprojekt		SUF gif2008	
Code	Wertelabel	Code	Wertelabel
-2	k. A.	-998	keine Angabe
-1	tnz	-989	filterbedingt fehlend
-2	k. A.	-987	designbedingt fehlend (Split, Filter, Einblendbedingung)

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.

[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.⁴ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Missing Not at Random), der einer Modellierung bedarf.⁵

[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 Designs der Befragung der Geförderten konnte auf ein Designgewicht verzichtet werden, denn aufgrund der Vollerhebung war ein Ausgleich ungleicher Inklusionswahrscheinlichkeiten nicht erforderlich. Da auf individueller Ebene keine Informationen zu Nichtteilnehmer(inne)n vorliegen, kann keine Nonresponse-Adjustierung auf Individualebene erfolgen. Das Gewicht wird in einem anhand einer Merkmalsverteilung der Grundgesamtheit ermittelt. Da hier Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf aggregierter Ebene vorliegen, erfolgt hier zugleich eine Form der Nonresponse-Adjustierung.

⁴ Das trifft dann zu, wenn die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

⁵ Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

Nach Förderstatus (Studien- bzw. Promotionsförderung) getrennt, wurde folgende Informationen in die Gewichtung einbezogen:

1. Begabtenförderungswerk (s. o.)
2. Fächergruppe (Sprach-, Kulturwissenschaften; Rechts-, Wirtschafts-, Sozialwissenschaften; Mathematik/Naturwissenschaften; Ingenieurwissenschaften; Medizin; Agrarwissenschaften; Kunst, Kunstwissenschaften)
3. Geschlecht (männlich, weiblich)

Der Scientific Use File enthält dem entsprechend zwei Gewichte:

- *ggrundrl* – für Auswertungen der Geförderten in der Studien- bzw. Grundförderung
- *gpromrl* – für Auswertungen der Geförderten in der Promotionsförderung.

[Hinweise zur Anwendung der Gewichte] Bei dem erstellten Gewicht handelt es sich um probability weights, die in Stata mit Hilfe ado-spezifischer Optionen berücksichtigt werden können.⁶

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.

⁶ Siehe hierzu die Stata-Hilfe (Befehl: help weights).

4 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten⁷, 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 2017.⁸ 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 Daten der Befragung der Geförderten der elf Begabtenförderwerke 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 Befragung der Geförderten 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 identifizieren.⁹ Für die Befragung der Geförderten wurden beispielsweise folgende Merkmale als Quasi-Identifikatoren eingestuft: Studienfach bzw. Promotionsfach, Abschlussart, Alter. Um eine eindeutige

⁷ „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).

⁸ 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.

⁹ 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.

Zuordnung der Daten zu betroffenen Personen zu unterbinden, wurden diese Schlüsselmerkmale aggregiert oder gelöscht (s. Tabelle 5).

Die Anonymisierung orientiert sich an den zu diesem Zweck gebildeten Aggregaten des kumulierten Datensatzes der 10. – 21. Sozialerhebung (ssypoolv3), um auch hier eine Vergleichbarkeit der Befunde zwischen beiden Studien(reihen) zu erleichtern.

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 10. 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.

Tabelle 4: Überblick zur Anonymisierung der Befragung der Geförderten der Begabtenförderungswerke

Merkmal	Download-SUF
Original-ID	Löschung und Vergabe einer zufälligen ID
Studienfächer	Aggregation zu amtlichen Fächergruppen ^a
Bundesland des aktuellen Studiums/ Bundesland der Promotion	Aggregation zu Regionen ^b
Typ der Hochschule	Zusammenfassung kirchlicher und privater Hochschulen zu „nicht-staatlichen“
Alter (in Jahren)	bis 39 einzeln ausgewiesen, ansonsten Aggregation: „40 bis 49“, „50 Jahre und älter“
Anzahl der Kinder	Aggregation: „3 Kinder und mehr“
(vorherige) Staatsangehörigkeit	Aggregation der Staaten zu Wirtschafts- bzw. Weltregionen: EU25, sonstiges Europa, Afrika, Asien, Ozeanien, Latein-/Mittelamerika, Nordamerika
Anzahl der Geschwister in Ausbildung	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 2006/2006, vgl. Codierliste cl-dzhw-70, die hier hinterlegt ist: [https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-gif2008-ins1\\$?version=1.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-gif2008-ins1$?version=1.0.0)
- b Aggregation zu den Regionen Süd-, West- und Norddeutschland, vgl. Codierliste cl-dzhw-69, die hier hinterlegt ist: [https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-gif2008-ins1\\$?version=1.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-gif2008-ins1$?version=1.0.0)

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.
- Issersedt, W., Middendorff, E., Fabian, G., Wolter, Ä. (2007). Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2006. 18. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem. Bonn, Berlin.
- 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., Isserstedt, W., Kandulla, M. (2009). Das soziale Profil in der Begabtenförderung. Ergebnisse einer Online-Befragung unter allen Geförderten der elf Begabtenförderungswege im Oktober 2008. HIS-Projektbericht. HIS Hochschul-Informationssystem GmbH. Hannover (<https://www.dzhw.eu/pdf/21/Begabte-Bericht.pdf>)
- Middendorff, E. (2022). Die Sozialerhebungen des Deutschen Studentenwerks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen, Methoden, Themen und projektbezogene Publikationen (Working Paper). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). (<https://t.co/kLF4R0Wkcp>)
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(2), 581–592.