

Konzept eines CAMPUS-Files der Studentenstatistik (EVAS 21311) des Wintersemesters 2000/2001

1. Vorbemerkungen

Bei so genannten CAMPUS-Files handelt es sich um vollständig anonymisierte Einzelmateriale, welche speziell für Lehrende und Studierende konzipiert werden und in den Hochschulen zu Lehrzwecken eingesetzt werden können. Diese Files stehen auf der Homepage des Forschungsdatenzentrums zum kostenlosen Download zur Verfügung. Durch ihren Einsatz in der praktischen Statistikausbildung können sich die Studierenden fundierte Methodenkenntnisse aneignen und sozial- und wirtschaftswissenschaftliche Fragestellungen auf Basis der amtlichen Einzeldaten analysieren.

2. Basismaterial

Bei der Studentenstatistik handelt es sich um eine auf den Verwaltungsdaten der Hochschulen basierende Sekundärstatistik. Ihre Daten werden halbjährlich jeweils für das Sommer- und Wintersemester innerhalb der Immatrikulations- und Rückmeldefristen erhoben. Während es sich im Wintersemester um eine Vollerhebung aller Studierenden an deutschen Hochschulen handelt, werden im Sommersemester dagegen lediglich Daten zu den Studienanfängern (Erst- und Neueinschreibungen; Exmatrikulationen) erfasst. Um ein umfassendes Bild über alle Studierenden zu gewinnen, bildet das Basismaterial zur Erstellung des CAMPUS-Files daher die Studentenstatistik des Wintersemesters 2000/2001. Dieses umfasst insgesamt 87 Merkmale (bis EF113) von 2.067.660 Studierenden an deutschen Hochschulen.

3. Anonymisierungsmaßnahmen

Nachfolgend werden die Maßnahmen zur Sicherstellung der absoluten Anonymität des Einzelmateriale erläutert.

3.1. Stichprobenziehung

Aus der Totalerhebung aller Studierenden wird eine 10%-Zufallsstichprobe gezogen, sodass sich der Gesamtdatensatz auf 206867 Fälle reduziert. Durch diese Stichprobenziehung wird das Re-Identifikationsrisiko deutlich eingeschränkt, da nicht bekannt ist, ob die gesuchte Person tatsächlich in der Stichprobe enthalten ist. Zugleich gewährleistet die trotz erfolgter Stichprobenziehung hohe Fallzahl jedoch ein – insbesondere für Lehrzwecke – ausreichendes Analysepotential. Tabelle 1 zeigt die Verteilung der in der Stichprobe enthaltenen Studierenden auf die einzelnen Bundesländer auf.

Tabelle 1: Verteilung der in der Stichprobe enthaltenen Studierenden auf das Bundesgebiet

Bundesland der meldenden Hochschule im Berichtsemester		
	Häufigkeit	Prozent
Schleswig-Holstein	4674	2,3
Hamburg	7433	3,6
Niedersachsen	16463	8,0
Bremen	2961	1,4
Nordrhein-Westfalen	55966	27,1
Hessen	17222	8,3
Rheinland-Pfalz	9624	4,7
Baden-Württemberg	23085	11,2
Bayern	24525	11,9
Saarland	2310	1,1
Berlin	15522	7,5
Brandenburg	3770	1,8
Mecklenburg-Vorpommern	4643	2,2
Sachsen	9768	4,7
Sachsen-Anhalt	4299	2,1
Thüringen	4602	2,2
Gesamt	206867	100,0

3.2. Alter des Datenmaterials

Eine weitere Schutzwirkung bietet das Alter des Datenmaterials. Statt der aktuellsten Ränder wird für das CAMPUS File eine zeitlich weiter zurückliegende Welle gewählt. Dieses Material sollte in der Regel keine Angaben zu Studierenden bereitstellen, welche ihr Studium noch nicht abgeschlossen haben, sodass die Angaben in gewisser Weise bereits überholt sind.

3.3. Reduktion und Klassifikation identifizierender und sensibler Merkmale

Weitere Anonymisierungsmaßnahmen setzen unmittelbar an den im Material enthaltenen Merkmalen an. Alle unmittelbaren Identifikatoren (Paginier- und Matrikelnummer EF5 und EF6) werden aus dem Bestand gelöscht. Gleiches gilt für Leerfelder sowie Felder mit der Kennung „frei für StLÄ“.

Um einen einzelnen Studenten re-identifizieren zu können, muss ein potenzieller Datenanreifer über ein bestimmtes Vorwissen zu Merkmalen, welche mit den Daten der Studententstatistik bereitgestellt werden, verfügen. Zu diesen meist in Kombination re-identifizierenden Merkmalen zählen soziodemographische Angaben wie das Geschlecht (EF7), der Geburtsmonat (EF8U1), das Geburtsjahr (EF8U2), Ort und Art der Hochschulzugangsberechtigung (EF111; EF110), Heimat- und Semesterwohnsitz (EF10, EF11) und die Staatsangehörigkeit (EF9), sowie Informationen zur Hochschule (EF4, EF16, EF56, EF66, EF74) und fachlich-organisatorische Merkmale zum Studienfach (viele EF-Nummern betroffen) sowie zum Prüfungsabschluss (EF96; EF97; EF105; EF106). Daher werden auch diese Merkmale entsprechend im Hinblick auf ihr Re-Identifikationspotenzial überprüft und entsprechend verändert. Während Geschlecht (EF7), Geburtsjahr (EF8U2) und Jahr des Prüfungsabschluss (EF97; EF106) für Analysezwecke interessante Informationen bereitstellen und somit unverändert in das CAMPUS-File aufgenommen werden, ist durch die Kenntnis des Geburtsmonats (EF8U2) sowie des Monats des Prüfungsabschlusses (EF96; EF105) kein Informationszugewinn zu erwarten, sodass diese Merkmale aus dem Datenmaterial entfernt werden. Darüber hinaus werden einige identifizierende Merkmale vergrößert. Die Merkmale Heimat- und Semesterwohnsitz (EF10, EF11) sowie der Ort des Erwerbs der Hochschulzugangsberechtigung (EF111) werden nur auf Bundeslandebene angegeben und die Art der Hochschulzugangsberechtigung (EF110) wird auf die vier Ausprägungen „allgemeine Hochschulreife (aHR)“, „fachgebundene Hochschulreife (fgHR)“, „Fachhochschulreife (FHR)“ und „Erwerb der HZB im Ausland“ reduziert. Ferner wird die

Staatsbürgerschaft (EF9) lediglich mit den Kategorien „deutsch“ vs. „ausländisch“ angeboten. Die zahlreichen einzelnen Studienfächer werden gemäß der Fächergruppensystematik des Statistischen Bundesamtes in die Kategorien „Sprach- und Kulturwissenschaften“, „Sport“, „Rechts-, Wirtschafts- und Sozialwissenschaften“, „Mathematik und Naturwissenschaften“, „Humanmedizin und Gesundheitswissenschaften“, „Veterinärmedizin“, „Agrar-, Forst- und Ernährungswissenschaften“, „Ingenieurwissenschaften“, „Kunst und Kunstwissenschaften“ und „Außerhalb der Studienbereichsgliederung“ zusammengefasst. Schließlich enthält das Material nicht mehr die einzelnen Hochschulen/-nummern, sondern lediglich Angaben zum Hochschultyp. Um dennoch regionalisierte Auswertungen mit dem CAMPUS-File zu ermöglichen, wird eine neue Variable zum Bundesland der Hochschule berechnet.

In Tabelle 2 werden die vorgenommenen Anonymisierungsmaßnahmen zusammenfassend dargestellt.

Tabelle 2: Übersicht über sensible und identifizierende Merkmale und entsprechende Anonymisierungsmaßnahmen beim CAMPUS-File

ursprüngliches Merkmal	Maßnahme	neues Merkmal
EF5 (Paginiernummer)	Löschung	entfällt
EF6 (Matrikelnummer)	Löschung	entfällt
EF7 (Geschlecht)	unverändert	unverändert
EF8U1 (Geburtsmonat)	Löschung	entfällt
EF8U2 (Geburtsjahr)	unverändert	unverändert
EF9 (Staatsangehörigkeit)	Vergrößerung	Staat 2 Kategorien: <ul style="list-style-type: none"> ○ deutsch ○ ausländisch
EF10 (Semesterwohnsitz)	Vergrößerung	SWohn 17 Kategorien: <ul style="list-style-type: none"> ○ Bundeslandebene Ausland
EF11 (Heimatwohnsitz)	Vergrößerung	HWohn 17 Kategorien: <ul style="list-style-type: none"> ○ Bundeslandebene Ausland
EF110 (Art der HZB)	Vergrößerung	ArtHZB 4 Kategorien: <ul style="list-style-type: none"> ○ allgemeine Hochschulreife ○ fachgebundene Hochschulreife ○ Fachhochschulreife ○ Erwerb der HZB im Ausland
EF111 (Ort der HZB)	Vergrößerung	OrtHZB 17 Kategorien: <ul style="list-style-type: none"> ○ Bundeslandebene ○ Ausland
EF32, EF34, EF36, EF47, EF49, EF51, EF60, EF61, EF62, EF70, EF71, EF72, EF78, EF79, EF80, EF92, EF93, EF94, EF101, EF102, EF103 (Studienfach)	Vergrößerung	Fach 10 Kategorien gemäß StBA Fächergruppensystematik <ul style="list-style-type: none"> ○ Sprach- und Kulturwissenschaften ○ Sport ○ Rechts-, Wirtschafts- und Sozialwissenschaften ○ Mathematik, Naturwissenschaften

		<ul style="list-style-type: none"> ○ Humanmedizin/Gesundheitswissenschaften ○ Veterinärmedizin ○ Agrar-, Forst- und Ernährungswissenschaften ○ Ingenieurwissenschaften ○ Kunst, Kunstwissenschaften ○ Außerhalb der Studienbereichsgliederung
EF4, EF16, EF56, EF66, EF74 (Hochschule)	Vergrößerung	HS 7 Kategorien: <ul style="list-style-type: none"> ○ Universitäten ○ Pädagogische Hochschulen ○ Theologische Hochschulen ○ Kunsthochschulen ○ Fachhochschulen ○ Verwaltungsfachhochschulen ○ Gesamthochschulen
EF96; EF105 (Monat Prüfungsabschluss)	Löschung	
EF97; EF106 (Jahr Prüfungsabschluss)	unverändert	
Bundesland der Hochschule entfällt im Originalmaterial	Neuberechnung aus EF4, EF16, EF56, EF66, EF74	HSLand 16 Kategorien: <ul style="list-style-type: none"> ○ Bundeslandebene

Durch die Kombination des Hochschultyps mit dem Bundesland der Hochschule können bundesweit zwar insgesamt 12 Hochschulen eindeutig identifiziert werden (Anhang 2 und Anhang 3),¹ doch gewährleistet die Stichprobenziehung sowie die vorgenommene Vergrößerung des Studienfachs in die Fächergruppen,² dass einzelne Studierende im Material nicht re-identifiziert werden können. Zudem handelt es sich bei vier der identifizierbaren Einrichtungen, um Hochschulen, welche zwischenzeitlich in andere Hochschulen eingegliedert wurden. Zwei weitere Hochschulen gehören den Kategorien „sonstige deutsche Hochschule“ bzw. „ohne Angabe“ an, welche eine eindeutige Zuordnung ausschließen.

Anreize zur Re-Identifikation gehen vor allem von sensiblen Merkmalen aus, welche dem Angreifer einen Nutzen versprechen. Zu diesen sensiblen Merkmalen, welche mit Hilfe geeigneter Schlüsselmerkmale nutzbringend aufgedeckt werden können, zählen in der Studentenstatistik vor allem die Prüfungsergebnisse, welche mit den Merkmalen EF98, EF99 und EF107, EF108 (Prüfungsergebnis und Gesamtnote für vor dem Berichtsemester abgelegte Prüfungen) bereitgestellt werden. Weiterhin besteht mit EF27U2 und EF42U2 (Grund der Exmatrikulation) ein Merkmal zum erfolgreichen bzw. nicht erfolgreichen Abschluss des Studiums in der Studentenstatistik. Eine sinnvoll anzunehmende Re-Identifikation (mit nutzbaren, verwertbaren Ergebnissen) sollte sich jedoch ausschließlich auf das Subset schlecht oder nicht bestandener Prüfungen richten, sodass nur Prüfungsergebnisse, welche unter den mittleren Ergebnissen liegen (z.B. ab „Befriedigend (Rite)“), als sensibel einzuschätzen sind. Ferner ist eine Sensibilität nur für exponierte, d.h. im gesellschaftlichen oder politischen Leben herausragende Personen gegeben. Eine derartige Konstellation ist bei der Studentenstatistik bereits aufgrund des gängigen Alters der Studierenden jedoch nahezu ausgeschlossen. Das verbleibende Restrisiko wird vor allem durch die Schutzwirkung der Stichproben-

¹ In Anhang 2 wird die Verteilung der Hochschultypen auf die einzelnen Bundesländer aufgezeigt. Anhang 3 listet alle eindeutig identifizierbaren Hochschulen auf.

² Auf die Ausweisung der Zwischenstufe „Studienbereich“ wurde zur Sicherstellung der absoluten Anonymität verzichtet.

ziehung – es ist nicht bekannt, ob die gesuchte Person tatsächlich in die Stichprobe gelangt ist – sowie die Veränderungen der identifizierenden Variablen ausgeschlossen.

3.4. Systemfreie Anordnung

Abschließend werden die im CAMPUS-File enthaltenen Fälle systemfrei nach dem Zufallsprinzip sortiert.

Anhang 1: Liste der bereitgestellten Merkmale und Anonymisierungsmaßnahmen für das CAMPUS File der Studentenstatistik des WS 2000/2001.

Merkmale		Anonymisierungsmaßnahme		
EF Nr.	Variablenname	Löschung	Rekodierung	neue Variable
EF1	Berichtsland			
EF2	Berichtssemester			
EF3	Berichtsjahr			
EF4	Hochschule		x	EF4: Vergrößerung auf Hochschultyp
EF5	Paginierungsnummer	x		
EF6	Matrikelnummer	x		
EF7	Geschlecht			
EF8U1	Geburtsmonat	x		
EF8U2	Geburtsjahr			
EF9	Staatsangehörigkeit		x	Staat: Vergrößerung „deutsch“ vs. „ausländisch“
EF10	Semesterwohnsitz		x	SWohn: Vergrößerung auf Bundeslandebene und Ausland
EF10U1	Bundesland bzw. Ausland	x		
EF10U2	Kreis bzw. Staat bei Ausland	x		
EF11	Heimatwohnsitz		x	HWohn: Vergrößerung auf Bundeslandebene und Ausland
EF11U1	Bundesland bzw. Ausland	x		
EF11U2	Kreis bzw. Staat bei Ausland	x		
EF12	Hörerstatus			
Ersteinschreibung in Deutschland/Hochschulsemester				
EF16	Hochschule		x	EF16: Vergrößerung auf Hochschultyp
EF17	Semester			
EF18	Jahr			
EF19	Anzahl Hochschulsemester an deutschen Hochschulen insgesamt			
EF20	davon Urlaubssemester			
EF21	davon Praxissemester			
EF22	davon Semester in DDR			
EF23	Nur bei Studium in der DDR - Art des Studiums			
EF24	Semester am Studienkolleg			
EF25	Anzahl Unterbrechungssemester insgesamt			
Studienunterbrechung im gleichen Studiengang				
EF26	Art der Einschreibung			
Studium im Berichtssemester				
EF27	Grund der Beurlaubung/ Exmatrikulation		x	EF27U1: Grund der Beurlaubung; EF27U2: Grund der Exmatrikulation
EF28	Art des Studiums			
EF29	Voll-/Teilzeitstudium/ Duales Studium			
EF30	Anzahl der Fachsemester im Hinblick auf angestrebte Abschlussprüfung			
EF31	Angestrebte Abschlussprüfung			
EF32	1. Studienfach		x	EF32: Vergrößerung auf Fächergruppen
EF34	2. Studienfach		x	EF34: Vergrößerung auf Fächergruppen
EF36	3. Studienfach		x	EF36: Vergrößerung auf Fächergruppen

2. Studiengang an der meldenden Hochschule			
EF41	Art der Einschreibung		
EF42	Grund der Beurlaubung/ Exmatrikulation	x	EF42U1: Grund der Beurlaubung; EF42U2: Grund der Exmatrikulation
EF43	Art des Studiums		
EF44	Voll-/Teilzeitstudium/ Duales Studium		
EF45	Anzahl der Fachsemester im Hinblick auf angestrebte Abschlussprüfung		
EF46	Angestrebte Abschlussprüfung		
EF47	1. Studienfach	x	EF47: Vergrößerung auf Fächergruppen
EF49	2. Studienfach	x	EF49: Vergrößerung auf Fächergruppen
EF51	3. Studienfach	x	EF51: Vergrößerung auf Fächergruppen
Einschreibung an einer anderen Hochschule			
EF56	Hochschule	x	EF56: Vergrößerung auf Hochschultyp
EF59	Angestrebte Abschlussprüfung		
EF60	1. Studienfach	x	EF60: Vergrößerung auf Fächergruppen
EF61	2. Studienfach	x	EF61: Vergrößerung auf Fächergruppen
EF62	3. Studienfach	x	EF62: Vergrößerung auf Fächergruppen
Studium im vorhergehenden Semester			
EF64	Kennziffer für jetzige oder andere Hochschule		
EF65U1	Studiengang gleich dem Berichtsemester		
1. Studiengang			
EF66	Hochschule	x	EF66: Vergrößerung auf Hochschultyp
EF69	Angestrebte Abschlussprüfung		
EF70	1. Studienfach	x	EF70: Vergrößerung auf Fächergruppen
EF71	2. Studienfach	x	EF71: Vergrößerung auf Fächergruppen
EF72	3. Studienfach	x	EF72: Vergrößerung auf Fächergruppen
2. Studiengang			
EF74	Hochschule	x	EF74: Vergrößerung auf Hochschultyp
EF77	Angestrebte Abschlussprüfung		
EF78	1. Studienfach	x	EF78: Vergrößerung auf Fächergruppen
EF79	2. Studienfach	x	EF79: Vergrößerung auf Fächergruppen
EF80	3. Studienfach	x	EF80: Vergrößerung auf Fächergruppen
Früheres auf das jetzige Studium in Deutschland bezogenes Studium im Ausland			
EF82	1. Land (Staat)		
EF83	Anzahl Monate		
EF84	2. Land (Staat)		
EF85	Anzahl (Monate)		
Bereits vor dem Berichtsemester abgelegte Abschlussprüfungen an Hochschulen			
Letzte Prüfung			
EF91	Art der Prüfung		
EF92	1. Studienfach	x	EF92: Vergrößerung auf Fächergruppe
EF93	2. Studienfach	x	EF93: Vergrößerung auf Fächergruppe
EF94	3. Studienfach	x	EF94: Vergrößerung auf Fächergruppe
EF96	Monat	x	
	Prüfungsabschluss		
EF97	Jahr		
	Prüfungsabschluss		
EF98	Prüfungsergebnis		
EF99	Gesamtnote		
Ggf. vorletzte Prüfung			
EF100	Art der Prüfung		
EF101	1. Studienfach	x	EF101: Vergrößerung auf Fächergruppe
EF102	2. Studienfach	x	EF102: Vergrößerung auf Fächergruppe
EF103	3. Studienfach	x	EF103: Vergrößerung auf Fächergruppe
EF105	Monat	x	
	Prüfungsabschluss		
EF106	Jahr		
	Prüfungsabschluss		
EF107	Prüfungsergebnis		
EF108	Gesamtnote		
Hochschulzugangsberechtigung (HZB)			
EF109	Jahr des Erwerbs der		

EF110	HZB Art der HZB		x	ArtHZB: Vergrößerung in vier Kategorien OrtHZB: Vergrößerung auf Bundeslandebene und Ausland
EF111	Erwerb der HZB		x	
EF111U1	Bundesland bzw Ausland	x		
EF111U2	Kreis bzw. Staat bei Ausland	x		
Berufspraktische Tätigkeit vor dem Studium				
EF112	Berufsausbildung mit Abschluss			
EF113	Praktikum oder Volontariat im Hinblick auf das derzeitige Studium			

Anhang 2: Verteilung der Hochschultypen auf die Bundesländer

	Hochschultyp								Gesamt
	Universität	Pädagogische Hochschule	Theologische Hochschule	Kunsthochschule	Fachhochschule	Verwaltungshochschule	sonstige Hochschule in Deutschland	Hochschule im Ausland	
Schleswig-Holstein	4	1	0	2	18	3	0	0	28
Hamburg	6	0	0	2	7	2	0	0	17
Niedersachsen	16	8	0	2	44	6	0	0	76
Bremen	2	1	0	1	8	1	0	0	13
Nordrhein-Westfalen	21	15	10	15	73	15	0	0	149
Hessen	11	0	3	4	23	11	0	0	52
Rheinland-Pfalz	8	4	2	0	20	5	0	0	39
Baden-Württemberg	16	14	0	12	57	7	0	0	106
Bundesland Bayern	21	8	5	9	47	6	0	0	96
Saarland	1	1	0	2	3	1	0	0	8
Berlin	13	1	0	6	17	3	0	0	40
Brandenburg	4	0	0	2	12	3	0	0	21
Mecklenburg-Vorpommern	5	2	0	1	5	1	0	0	14
Sachsen	16	3	0	8	17	5	0	0	49
Sachsen-Anhalt	9	3	2	2	9	5	0	0	30
Thüringen	6	0	1	1	5	5	0	0	18
Ausland	0	0	0	0	0	0	1	1	2
Gesamt	159	61	23	69	365	79	1	1	758

Anhang 3: Liste identifizierbarer Hochschulen

Bundesland	Hochschulnummer	Hochschule
Schleswig-Holstein	3110	Ehem. Pädagogische Hochschule Kiel eingegliedert in die Universität Kiel 1000
Bremen	3280	Ehem. Pädagogische Hochschule Bremen eingegliedert in die Universität Bremen 1050
	2610	Hochschule für Künste Bremen
	5670	Hochschule für öffentl. Verwaltung Bremen
Saarland	1360	Universität des Saarlandes Saarbrücken
	3920	Ehem. Pädagogische Hochschule des Saarlandes, Saarbrücken eingegliedert in die Universität Saarbrücken 1360
	7630	Fachhochschule für Verwaltung Saarbrücken
Berlin	3960	Ehem. Pädagogische Hochschule Berlin eingegliedert in die FU Berlin 1380
Mecklenburg-Vorpommern	3030	Hochschule für Musik und Theater Rostock
	7900	Verwaltungsfachhochschule Güstrow
Sonstiges/Ausland	9000	Sonstige deutsche Hochschule
	9900	Ehem. ohne Angabe jetzt Hochschule im Ausland 9990