

## **Leitfaden zur absoluten Anonymisierung für die Erstellung eines Campus-File aus den Einzeldaten der Verdienststrukturerhebung 2006**

### 1. Vorbemerkungen

Im Jahr 1987 wurde mit § 16 Abs. 6 des Bundesstatistikgesetzes<sup>1</sup> der Wissenschaft ein privilegierter Zugang zu Mikrodaten der amtlichen Statistik eingeräumt. Hiernach ist die Übermittlung von Einzeldaten an die Wissenschaft erlaubt, sofern diese nur mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft reidentifiziert werden können (faktische Anonymität). „Unverhältnismäßig“ bedeutet hier, dass der Aufwand einer Reidentifikation deren Nutzen übersteigt. Die Deanonymisierung von Einzelangaben in einem faktisch anonymen Datensatz kann nicht mit absoluter Sicherheit ausgeschlossen werden. Die Schutzmaßnahmen für einen, im Rahmen der Wissenschaft und Lehre frei zugänglichen Datensatz, dem so genannten Campus-File müssen weitaus höher angesetzt werden. In diesem Fall ist es nur zulässig, absolut anonymisierte Daten weiterzugeben, d.h. die eindeutige Identifikation von Fällen ist ausgeschlossen. Der vorliegende Leitfaden behandelt absolut anonymisierte Datensätze für die Wissenschaft und Lehre, generiert aus den Daten der Verdienststrukturerhebung (VSE) 2006.

### 2. Basismaterial

Die Verdienststrukturerhebung 2006 wurde als Stichprobe bei knapp 28.700 Betrieben mit 10 und mehr Beschäftigten durchgeführt. Dabei wurden nach Ländern differenzierte Auswahlsätze vorgegeben. In kleineren Ländern wurden höhere und in großen Ländern niedrigere Auswahlsätze angelegt. Die ausgewählten Betriebe bezogen bundesweit rund 1,8 Millionen Beschäftigte ein. Es handelt sich um ein zweistufiges Auswahlverfahren mit Betrieben in der 1. Stufe und Beschäftigten in der 2. Stufe. In den kleineren Betrieben wurden alle Beschäftigten erfasst, während in den größeren Betrieben nach einem Zufallsverfahren nur ein Teil der Beschäftigten ausgewählt wurde. Für den Wirtschaftsabschnitt M "Erziehung und Unterricht" werden die Daten nicht erhoben, sondern auf Grundlage der Personalstandstatistik und von Tarifinformationen geschätzt. Alle ca. 1,4 Millionen Arbeitnehmer im Datensatz dieser Statistik, die dem NACE-Abschnitt M zuzuordnen sind, wurden in die Verdienststrukturerhebung übernommen. Somit sind in der VSE insgesamt ca. 3,2 Millionen Beschäftigte enthalten.

Bei der VSE werden Angaben über Beschäftigte im Individualverfahren erhoben. Damit können individuelle Merkmale der Arbeitnehmer – mit dem Arbeitsplatz verbundene Merkmale einerseits und persönliche Merkmale andererseits – zum Verdienst in Bezie-

---

<sup>1</sup> Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 3 des Gesetzes vom 7. September 2007 (BGBl. I S. 2246).

hung gesetzt werden. Neben den Bruttoverdiensten werden auch Nettoverdienste und Sozialversicherungsbeiträge der Arbeitnehmer nachgewiesen. Brutto- und Nettoverdienste sowie Sonderzahlungen beziehen sich auf das gesamte Jahr 2006, Brutto- und Nettomonatsverdienste und Arbeitszeiten auf den Berichtsmonat Oktober 2006.

Eine vollständige Liste der im vorgestellten Campus-File enthaltenen Merkmale findet sich in der Datensatzbeschreibung im Anhang.

### 3. Anonymisierungsmaßnahmen

Nachfolgende Maßnahmen wurden durchgeführt:

#### **Stichprobenziehung**

Die Stichprobenziehung erfolgt für alle Wirtschaftskategorien (außer Erziehung und Unterricht) in 2 Stufen:

- (1) Zunächst werden 5% der Betriebe des Originalmaterials gezogen, geschichtet nach den unten angegebenen Wirtschaftskategorien (ohne Erziehung und Unterricht), alten und neuen Bundesländern und 6 Beschäftigtengrößenklassen.
- (2) Aus den im Basismaterial enthaltenen Beschäftigten dieser ausgewählten Betriebe wird dann im zweiten Schritt eine Zufallsstichprobe gezogen, wobei der Auswahlsatz in Abhängigkeit von der Größenklasse des Betriebes zufällig ermittelt wird. Für Größenklassen 1 und 2 liegt der Faktor zwischen 0.6 und 0.9; für Größenklassen 3 und 4 zwischen 0.3 und 0.7 und für Größenklassen 5 und 6 zwischen 0.2 und 0.8. Der genaue Auswahlsatz pro Betrieb wird nicht weitergegeben. Damit soll verhindert werden, dass von der Anzahl der Beschäftigten des Betriebes in der Stichprobe Rückschlüsse auf die Anzahl der Beschäftigten des Betriebes möglich sind, was insbesondere für die größten Betriebe eine Reidentifikation ermöglichen könnte.

Für den Bereich „Erziehung und Unterricht“ erfolgt eine einfache Stichprobenziehung von 0,7% der Beschäftigten.

#### **Vergrößerung der Regionalangabe**

Es werden die beiden Regionen alte Bundesländer (mit Berlin) und neue Bundesländer (ohne Berlin) ausgewiesen.

#### **Vergrößerung der Wirtschaftsgruppen**

Es werden 10 Wirtschaftskategorien ausgewiesen. Diese entsprechen (bis auf die neu hinzu gekommenen Wirtschaftsbereiche) denjenigen in den Campus-Files der Gehalts- und Lohnstrukturerhebung 2001 und der Europäischen Erhebung zur beruflichen Weiterbildung (CVTS2 + CVTS33) ausgewiesenen. Damit wird die Verwendung mehrerer

wirtschaftsstatistischer Campus-Files und die Kombination der Ergebnisse von Analysen erleichtert.

### **Anzahl der Beschäftigten eines Betriebes**

Es werden nur Beschäftigtengrößenklassen ausgewiesen, wobei die letzte Gruppe Betriebe mit 500 und mehr Beschäftigten umfasst. Da es in den neuen Bundesländern weniger große Betriebe gibt, werden hier zusätzlich die beiden höchsten Klassen zusammengefasst, so dass hier die letzte Gruppe aus den Betrieben mit 250 und mehr Beschäftigten besteht.

### **Anzahl der Beschäftigten des Unternehmens**

Da Analysen ergeben haben, dass die Qualität dieses Merkmals nicht besonders gut ist, wird es nicht ausgewiesen.

### **Beschäftigte des Betriebes nach Geschlecht**

Die originalen Anteile in Prozent werden ausgewiesen.

### **Einfluss der öffentlichen Hand auf das Unternehmen**

Die Tatsache, dass Betriebe, die unter einem Einfluss der öffentlichen Hand stehen, sich auf einige wenige Wirtschaftszweige konzentrieren, erleichtert eine mögliche Reidentifikation. Unter der Auflage, dass die Angaben in einem Campus-File absolut anonym sein müssen, lässt sich dieses Merkmal nicht beibehalten. Es wird daher aus dem Datensatz entfernt.

### **Tarifvertragsschlüssel**

Vom Tarifvertragsschlüssel wird nur die zweite Stelle ausgewiesen, die angibt, ob es sich um einen Kollektiv- oder Firmentarifvertrag oder eine Betriebsvereinbarung handelt. Dabei werden zusätzlich die Ausprägungen *Firmentarifvertrag* und *Betriebsvereinbarung* zusammengefasst. Für die Beschäftigten einiger Betriebe erfolgte aus Datenschutzgründen im Scientific Use – File der Verdienststrukturerhebung 2006 nur die Angabe, ob ein Tarifvertrag angewandt wird oder nicht. Für diese Beschäftigte wird im Campus-File bei der Variablen *tarifart* „keine Angabe“ ausgewiesen.

### **Ausgeübte Tätigkeit**

Die ausgeübten Tätigkeiten werden zu 20 Berufsgruppen zusammengefasst (s. Datensatzbeschreibung im Anhang).

### **Brutto-Monats- und Jahresverdienst**

Für hohe Verdienste ab 7000 Euro brutto im Monat und ab 84000 Euro brutto im Jahr wird ein Topcoding durchgeführt; d. h. für diese Verdienste werden nur 7000 Euro und mehr bzw. 84000 Euro und mehr ausgewiesen.

### **Sonstige Verdienstmerkmale / Sozialabgaben**

Für die sonstigen Verdienstangaben wird der Anteil am Bruttomonatsverdienst (bzw. bei Angaben, die sich auf das Jahr beziehen, am Bruttojahresverdienst) in Prozent ausgewiesen. Analog wird bei den Sozialabgaben verfahren.

### **Geburtsjahr**

Für Beschäftigte, die 17 Jahre und jünger sind, wird *1989 oder später* ausgewiesen; für Beschäftigte, die 60 Jahre und älter sind, wird *1946 oder früher* ausgewiesen.

### **Hochrechnungsfaktoren**

Jeder Betrieb und jeder Beschäftigte einer Größenklasse einer Wirtschaftskategorie erhält pro Region einen einheitlichen Hochrechnungsfaktor. Der Hochrechnungsfaktor für die Betriebe ergibt sich als Quotient aus der Anzahl der Betriebe der Größenklasse der Wirtschaftskategorie in der Region in der Grundgesamtheit und der entsprechenden Anzahl in der Stichprobe. Die Berechnung des Hochrechnungsfaktors für die Beschäftigten erfolgt analog unter Verwendung der entsprechenden Beschäftigtenanzahlen.

## Anhang : Merkmalsliste des Campus-Files

<i>Merkmal</i>	<i>Inhalt</i>	<i>Bemerkung</i>
REGION	Region	1 = alte Bundesländer (mit Berlin) 2 = neue Bundesländer (ohne Berlin)
BETR_ID	Nummer für Betrieb	systemfrei
BES_ID	Nummer für Beschäftigten	Fortlaufend je Betrieb
WZGRUPPE	Wirtschaftsgruppe, in der Haupttätigkeitsbereich des Betriebes liegt	1 = Ernährungsgewerbe und Tabakverarbeitung; Textil- und Bekleidungsgewerbe; Ledergewerbe 2 = Verarbeitendes Gewerbe ohne Ernährungsgewerbe, Tabakverarbeitung, Textil- und Bekleidungsgewerbe, Ledergewerbe 3 = Energie- und Wasserversorgung 4 = Baugewerbe 5 = Handelsvermittlung und Großhandel; Einzelhandel; Reparatur von Gebrauchsgütern; Kraftfahrzeughandel; Instandhaltung und Reparatur von Kraftfahrzeugen; Tankstellen 6 = Gastgewerbe 7 = Verkehr; Nachrichtenübermittlung 8 = Kreditgewerbe; Versicherungsgewerbe; Mit dem Kredit- und Versicherungsgewerbe verbundene Tätigkeiten 9 = Grundstücks- und Wohnungswesen; Vermietung beweglicher Sachen ohne Bedienungspersonal; Erbringung von sonstigen Dienstleistungen überwiegend für Unternehmen; Gesundheits- und Veterinärwesen; Sozialwesen; Abwasser- und Abfallbeseitigung und sonstige Entsorgung; Interessenvertretungen sowie kirchliche und sonstige Vereinigungen; Kultur, Sport und Unterhaltung; Erbringung von sonstigen Dienstleistungen 10 = Erziehung und Unterricht
EF9	Leistungsgruppe	0 = keine Angabe 1 = Leitende Arbeitnehmer 2 = angelernter Arbeiter 3 = Arbeitnehmer mit besonderen Erfahrungen 4 = Arbeitnehmer ohne eigene Entscheidungsbefugnis

		5 = Arbeitnehmer in einfacher Tätigkeit
EF10	Geschlecht	1 = männlich 2 = weiblich
EF11	Geburtsjahr	1989 = 1989 und später 1946 = 1946 und früher
EF12U2	Jahr des Eintritts in das Unternehmen	1960 = 1960 und früher
BERUF	Berufsgruppe	1 = Pflanzenbauer, Tierzüchter, Fischereiberufe (KdB 01 – 06) 2 = Bergleute, Mineralgewinner; Herstellung und Verarbeitung von Stein, Baustoff, Keramik, Glas, Kunststoff, Papier, Holz und Metall (KdB 07 – 24) 3 = Schlosser, Mechaniker, Montierer, Elektriker, Maschinisten (KdB 25 – 32 + 54) 4 = Textil- und Ernährungsberufe (KdB 33 – 43) 5 = Bau- und Raumausstattungsberufe (KdB 44 – 50) 6 = Warenprüfer, Versandfertigmacher; Lager- und Transportarbeiter (KdB 52 + 74) 7 = Ingenieure / Naturwissenschaftler (KdB 60 + 61) 8 = Techniker / Technische Sonderfachkräfte (KdB 62 + 63) 9 = Warenkaufleute (KdB 68) 10 = Dienstleistungskaufleute (KdB 69 + 70) 11 = Berufe des Verkehrs und Nachrichtenwesens (KdB 71 + 73) 12 = Unternehmer, Abgeordnete u. ä. (KdB 75 + 76) 13 = Rechnungskaufleute; Datenverarbeitungsfachleute (KdB 77) 14 = Bürokräfte (KdB 78) 15 = Ordnungs-/Sicherheitsberufe (KdB 79-81) 16 = Publizisten, Künstler u. ä. (KdB 82 + 83) 17 = Gesundheitsdienstberufe; Sozial- und Erziehungsberufe (KdB 84 – 87) 18 = Sonstige geistes- und naturwissenschaftliche Berufe (KdB 88) 19 = Allgemeine Dienstleistungsberufe

		(KdB 90 – 93) 20 = Sonstige Berufe (KdB 53, 89, 97 – 99)
EF16U1	Stellung im Beruf	0 = Auszubildende 1 = Arbeiter, nicht als Facharbeiter tätig 2 = Facharbeiter 3 = Meister, Polier 4 = Angestellter 5 = Beamter in Vollzeit 6 = Beamter in Teilzeit 7 = Heimarbeiter 8 = Teilzeit mit weniger als 18 Wochenstunden 9 = Teilzeit mit mindesten 18 Wochenstunden
EF16U2	Ausbildung	1 = Volks-/Hauptschule, mittlere Reife ohne abgeschlossene Berufsausbildung 2 = Volks-/Hauptschule, mittlere Reife mit abgeschlossener Berufsausbildung 3 = Abitur ohne abgeschlossene Berufsausbildung 4 = Abitur mit abgeschlossener Berufsausbildung 5 = Abschluss einer Fachhochschule 6 = Hochschulabschluss 7 = unbekannt, Angabe nicht möglich
EF17	Art des Arbeitsvertrages	1 = unbefristet 2 = befristet ohne Auszubildende und Praktikanten 3 = Ausbildungsvertrag 4 = Altersteilzeit 5 = geringfügig Beschäftigte 6 = Beamte
EF18	Arbeitsvertraglich vereinbarte wöchentliche Arbeitszeit	
EF19	Bezahlte Stunden Insgesamt	
EF20	Bezahlte Mehrarbeitsstunden	
EF21	Bruttoverdienst Insgesamt im Berichtsmonat in Euro	7000 = 7000 Euro und mehr
EF22	Verdienst aus Mehrarbeitszeit in Prozent von EF21	
EF23	In Monatsverdienst enthaltene Zulagen für Schicht- / Nacht- und Sonntagsarbeit in Prozent von EF21.	
EF24	Lohnsteuer incl. Solidaritäts-	

	zuschlag ohne Kirchensteuer in Prozent von EF21	
EF25	Beiträge der Arbeitnehmer zur Sozialversicherung in Prozent von EF21	
EF26	Sozialversicherungspflichtige Arbeitstage im Berichtsjahr	
EF27	Bruttogehaltsverdienst insgesamt in Euro	84000 = 84000 Euro und mehr
EF28	Sonderzahlungen für das ganze Jahr in Prozent von EF27	
EF29	Urlaubsanspruch für das Berichtsjahr	
EF36	Grundlage der Urlaubstageberechnung	4 = 4-Tage-Woche 5 = 5-Tage-Woche 6 = 6-Tage-Woche 7 = 7-Tage-Woche
EF44	Nettomonatsverdienst in Prozent von EF21	
TARIFART	Art des Tarifvertrages	0 = kein Tarifvertrag 1 = Kollektivtarifvertrag 2 = Firmentarifvertrag oder Betriebsvereinbarung 9 = unbekannt oder keine Angabe
B_EF13	Beschäftigtengrößenklassen des Betriebes	1 = 0 bis 19 2 = 20 bis 49 3 = 50 bis 99 4 = 100 bis 249 5 = 250 bis 499 (West) 6 = 500 und mehr (West) 56 = 250 und mehr (Ost)
B_EF11	Anteil der männlichen Beschäftigten des Betriebes in Prozent	
B_EF12	Anteil der weiblichen Beschäftigten des Betriebes in Prozent	
B_EF30	Hochrechnungsfaktor der Betriebe	Neuberechnung
B_EF31	Hochrechnungsfaktor der Arbeitnehmer	Neuberechnung