

**Konzept zur absoluten Anonymisierung statistischer Einzeldaten aus dem
„AFiD-Panel Industrieunternehmen“
und dem
„Monatsbericht für Betriebe im Verarbeitenden Gewerbe, Bergbau sowie der
Gewinnung von Steinen und Erden“ der Erhebungsjahre 2003 - 2007
sowie aus dem
„Jahresbericht für Betriebe für Betriebe im Gewerbe, Bergbau sowie der Gewin-
nung von Steinen und Erden“ des Erhebungsjahres 2007
zur Nutzung als CAMPUS-File (CAF).**

CAF AFiD-Panel Industrieunternehmen

Version 1.1

Version control

Version	Datum	Beschreibung	Autor/-in
1.1	08/2017	Korrektur Merkmalsliste	A. Malchin, FDZ

Vorbemerkung

Im Jahr 1987 wurde mit § 16 Abs. 6 des Bundesstatistikgesetzes 1 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 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 des AFiD-Panel Industrieunternehmen und Angaben der Jahresergebnisse des Monatsberichts für Betriebe im Verarbeitenden Gewerbe (VG) der Erhebungsjahre 2003 bis 2007 sowie des Jahresberichts für Betriebe im VG des Erhebungsjahres 2007.

Basismaterial

AFiD-Panel Industrieunternehmen der Erhebungsjahre 2003 - 2007, mit Angaben aus

- der Kostenstrukturerhebung im VG,
- dem Jahresbericht für Unternehmen im VG und
- der Investitionserhebung im VG sowie

Zusammengefasste Jahresergebnisse des Monatsberichtes für Betriebe im VG der Erhebungsjahre 2003 – 2007 und des **Jahresberichts für Betriebe im VG des Erhebungsjahres 2007¹**.

¹ Zum Erhalt des Berichtskreises ergänzt der Jahresbericht für Betriebe im VG 2007 die Ergebnisse der Monatsberichte für Betriebe im VG 2007.

Erstellung der Mikrodatendateien

Die Datensätze der genannten Unternehmenserhebungen werden als integrierter Datensatz bereitgestellt. Dabei werden die vollen Querschnitte der einzelnen Statistiken verwendet. Die Verknüpfung erfolgt über die in allen Erhebungen enthaltenen Unternehmensnummern. Für die im Datensatz enthaltenen Unternehmen werden anschließend die jahres- und unternehmensweise summierten Umsatzangaben der zugehörigen Betriebe aus den Monats- und Jahresberichten für Betriebe ebenfalls über die Unternehmensnummern zugespielt.

Anonymisierung der Datensätze

Folgende Anonymisierungsmaßnahmen werden, entsprechend der genannten Reihenfolge, auf den Datenbestand angewandt:

- Lediglich die in **Anlage 1** genannten **Merkmale** verbleiben im Datensatz, alle anderen werden ersatzlos gelöscht.
- Die **Regionalangaben** für Unternehmen werden auf die Regionen „West“ und „Ost“ vergrößert. Die Regionen umfassen dabei:

West	Ost
Schleswig-Holstein	Berlin
Hamburg	Brandenburg
Niedersachsen	Mecklenburg-Vorpommern
Bremen	Sachsen
Nordrhein-Westfalen	Sachsen-Anhalt
Hessen	Thüringen
Rheinland-Pfalz	
Baden-Württemberg	
Bayern	
Saarland	

- Die Angaben zu den **Wirtschaftszweigen** der Unternehmen werden auf den 2-Steller vergrößert und durch **zufällige 2-Steller-Nummern** im Wertebereich [10;37] ersetzt. Die systemfreien Nummern sind über alle Unternehmen und Erhebungsjahre für gleiche Wirtschaftszweige identisch.
- Für die Anonymisierung werden die Unternehmen in drei Beschäftigtengrößenklassen eingeteilt, wobei die größte in einem beliebigen Erhebungsjahr auftretende Anzahl an Beschäftigten ausschlaggebend ist². Daraus ergeben sich folgende Größenklassen:

² Weist bspw. ein Unternehmen in vier Erhebungsjahren 400 Beschäftigte und im fünften Erhebungsjahr 600 Beschäftigte auf, wird es für alle Jahre entsprechend der Größenklasse „500 bis unter 1000 Beschäftigte“ behandelt.

- mehr als 1000 Beschäftigte
 - 500 bis unter 1000 Beschäftigte
 - weniger als 500 Beschäftigte
- Alle Unternehmen, die in mindestens einem Jahr **mehr als 1.000 Beschäftigte** haben, werden aus dem Datensatz **entfernt**.
- **Stichprobenziehung:**
- Aus der Gruppe der Unternehmen, die in mindestens einem Jahr **500 bis unter 1.000 Beschäftigte** aufweisen, werden diejenigen **entfernt**, die nicht für alle Jahre im Panel vertreten sind. Aus den verbleibenden Unternehmen wird eine **50% Stichprobe** gezogen.
 - Aus der Gruppe der Unternehmen, die in allen Jahren **weniger als 500 Beschäftigte** haben, wird eine **75%-Stichprobe** gezogen.
- **Mikroaggregation:** Alle noch verbliebenen **Unternehmen**, die in mindestens einem Jahr **mehr als 500 Beschäftigte** (aber immer unter 1.000) aufweisen, werden nach (anonymisiertem) Wirtschaftszweig, Region und Jahr (=Schicht) **mikroaggregiert**, wobei die Gruppenmitglieder in jedem Jahr gleich bleiben. Dies bedeutet, dass die Angaben zu allen Wertmerkmalen durch den jeweiligen Durchschnitt des Merkmals der jeweiligen Gruppe ersetzt werden. Maßgeblich für die Bildung der Aggregationsgruppen ist die durchschnittliche jährliche Beschäftigtenzahl. Dabei wird wie folgt vorgegangen:
- Für jedes Unternehmen der genannten Größenklasse wird die durchschnittliche jährliche Beschäftigtenzahl über alle Jahre gebildet.
 - Für die Mikroaggregation werden die Unternehmen nach Wirtschaftszweig und Region des Berichtsjahres 2007 sowie durchschnittlicher Beschäftigtenzahl sortiert. Sofern in einer Schicht weniger als 3 Unternehmen enthalten sind, werden alle Unternehmen dieser Schicht gelöscht.
 - Anschließend erfolgt die Mikroaggregation in 3er-Gruppen nach absteigender Beschäftigtenzahl. Sollte die Anzahl der Unternehmen in einer Schicht keinem Vielfachen von 3 entsprechen, wird die Gruppe der beschäftigungsärmsten Unternehmen entsprechend auf 4 bzw. 5 Einheiten erweitert.
- **Zufallsüberlagerung:** Für alle noch verbliebenen Unternehmen werden die metrischen Merkmale mit einem Zufallfaktor multipliziert. Der einem Unternehmen zugewiesene Faktor ist für alle metrischen Merkmale über alle Jahre gleich. Die Hälf-

te der Unternehmen bekommt einen Faktor zwischen 0,6 und 0,8 und die andere Hälfte einen Faktor zwischen 1,2 und 1,4 zugewiesen.

- Die **Beschäftigtengrößenklasse** (jbu_ef7) wird an den durch die Mikroaggregation und die Zufallsüberlagerung veränderten Wert der Beschäftigtenzahl **angeplichen**.
- Alle **Unternehmensnummern** werden durch **systemfreie Nummern** ersetzt. Dies geschieht durch die Sortierung nach einer Zufallszahl, so dass kein Rückschluss auf die Originalnummern mehr erfolgen kann. Die systemfreien Nummern sind über alle Erhebungsjahre für die gleiche Einheit identisch.

Fazit

Der nachfolgenden Tabelle ist noch einmal zusammengefasst zu entnehmen, welche Anonymisierungsmaßnahmen für welche Beschäftigtengrößenklasse angewendet werden.

Beschäftigtengrößenklasse	aus Datensatz gelöscht		Mikroaggregation		Zufallsüberlagerung	
	ja	nein	ja	nein	ja	nein
mehr als 1000 Beschäftigte	x					
500 bis unter 1000 Beschäftigte (nicht alle Jahre im Panel)	x					
500 bis unter 1000 Beschäftigte (alle Jahre im Panel)		x	x		x	
weniger als 500 Beschäftigte		x		x	x	

Alle großen Unternehmen sind aus dem Datensatz entfernt. Für alle noch verbliebenen Unternehmen sind nach der Anonymisierung für alle metrischen Merkmale keine Originalwerte mehr vorhanden. Neben einer Zufallsüberlagerung fand für einige Unternehmen zusätzlich eine Mikroaggregation statt.

Zudem wurden die Regional- und Wirtschaftszweigangaben vergrößert und den Unternehmensnummern systemfreie Nummern zugewiesen.

Das erzeugte CAF für das AFiD-Panel Industrieunternehmen ist als absolut anonym einzustufen.

Anlage 1: Liste der enthaltenen Merkmale

Merkmalsliste AFiD-Panel Industrieunternehmen Campus-File 2003-2007		
Variable	Typ	Beschreibung
unr	Num	Pseudo-Unternehmensnummer
jahr	Num	[jahr] Erhebungsjahr
region	Num	Region: Ost/West (1 = West, 2 = Ost)
wz2	Num	Pseudo-WZ-2-Steller nach WZ2003
ade	Num	Art der Einheit (1 = Einbetriebsunternehmen ; 2 = Mehrbetriebsunternehmen, 3 = Mehrländerunternehmen)
Jahresbericht für Unternehmen im VG		
jbu_ef7	Num	[jbu_ef7] Beschäftigtengrößenklasse 1 = 1 bis 49 Beschäftigte 2 = 50 bis 99 Beschäftigte 3 = 100 bis 249 Beschäftigte 4 = 250 bis 499 Beschäftigte 5 = 500 bis 999 Beschäftigte
jbu_ef14	Num	[jbu_ef14] tätige Personen insgesamt
jbu_ef18u3	Num	[jbu_ef18u3] darunter: weibliche tätige Personen (nur 2003 - 2005)
jbu_ef15	Num	[jbu_ef15] Lohn- und Gehaltssumme
jbu_ef16	Num	[jbu_ef16] Umsatz insgesamt
jbu_ef17	Num	[jbu_ef17] darunter: Umsatz aus sonstiger Tätigkeit
jbu	Num	[jbu] Teilnahme Jahresbericht (1 = ja, 0 = nein)
Monatsbericht für Betriebe im VG (Jahresergebnisse) bzw. Jahresbericht für Betriebe im VG		
mb_27sum	Num	[mb_27sum] Umsatz insgesamt (aggregiert auf Unternehmensebene)
mb_26sum	Num	[mb_26sum] Auslandsumsatz (aggregiert auf Unternehmensebene)
mb_23sum	Num	[mb_23sum] Inlandsumsatz (aggregiert auf Unternehmensebene)
mb	Num	[mbb] Unternehmenssatz enthält Betriebsangaben aus Monatsbericht oder Jahresbericht
Investitionserhebung im VG (Unternehmensdatensätze)		
ieu_ef46	Num	[ieu_ef46] Bruttozugänge insgesamt: davon Investitionen in Grundstücke mit Bauten
ieu_ef47	Num	[ieu_ef47] Bruttozugänge insgesamt: davon Investitionen in Grundstücke ohne Bauten
ieu_ef48	Num	[ieu_ef48] Bruttozugänge insgesamt: davon Investitionen in Maschinen, maschinelle Anlagen, Ausstattung

Merkmalsliste AFiD-Panel Industrieunternehmen Campus-File 2003-2007		
Variable	Typ	Beschreibung
ieu_ef49	Num	[ieu_ef49] Bruttozugänge insgesamt (Bruttoanlageinvestitionen)
ieu_ef50	Num	[ieu_ef50] Bruttozugänge insgesamt: darunter selbsterstellte Anlagen
ieu_ef51	Num	[ieu_ef51] Bruttozugänge insgesamt: darunter gebraucht erworbene Bauten
ieu_ef52	Num	[ieu_ef52] Bruttozugänge insgesamt: darunter aktivierte Sachanlagen an Dritte vermietet
ieu_ef53	Num	[ieu_ef53] Verkaufserlöse aus Sachanlagen insgesamt
ieu_ef54	Num	[ieu_ef54] Verkaufserlöse aus Sachanlagen insgesamt: darunter Erlöse aus dem Abgang von Grundstücken ohne Bauten
ieu_ef121	Num	[ieu_ef121] Zugang gemietete/gepachtete Sachanlagen: davon Grundstücke mit neuen Bauten
ieu_ef122	Num	[ieu_ef122] Zugang gemietete/gepachtete Sachanlagen: davon maschinen, maschinelle Anlagen, Ausstattung
ieu_ef123	Num	[ieu_ef123] Zugang gemietete/gepachtete Sachanlagen insgesamt
ieu_ef55	Num	[ieu_ef55] Zugang gemietete/gepachtete Sachanlagen: darunter Wert der mittels Finanzierungsleasing neu beschafften Sachanlagen
ieu	Num	[ieu] Teilnahme Investitionserhebung (1 = ja, 0 = nein)
Kostenstrukturerhebung im VG		
kse_ef4	Num	[kse_ef4] Rechtsform (1 = Einzelfirma; 7 = Kapitalgesellschaft, 8 = sonstige)
kse_ef21	Num	[kse_ef21] Anzahl der tätigen Inhaber
kse_ef28	Num	[kse_ef28] Anzahl der tätigen Inhaber: darunter weiblich
kse_ef22	Num	[kse_ef22] Anzahl der Arbeitnehmer insgesamt
kse_ef29	Num	[kse_ef29] Anzahl der Arbeitnehmer: darunter weiblich
kse_ef24	Num	[kse_ef24] Anzahl der Arbeitnehmer: darunter Teilzeitbeschäftigte.
kse_ef25	Num	[kse_ef25] Anzahl der Arbeitnehmer: darunter Teilzeitbeschäftigte, umgerechnet in Vollzeiteinheiten
kse_ef27	Num	[kse_ef27] Anzahl der tätigen Personen insgesamt (kse_ef21+kse_ef22)
kse_ef87	Num	[kse_ef87] Anzahl der tätigen Personen insgesamt: darunter mit Aufgaben im Bereich FuE
kse_ef35	Num	[kse_ef35] Umsatz: aus eigenen Erzeugnissen
kse_ef37	Num	[kse_ef37] Umsatz: aus Handelsware
kse_ef38	Num	[kse_ef38] Umsatz: aus Provisionen aus Handelsvermittlung
kse_ef39	Num	[kse_ef39] Umsatz: aus sonstige Tätigkeiten
kse_ef40	Num	[kse_ef40] Umsatz insgesamt (summe aus kse_ef35 bis kse_ef39)
kse_ef41	Num	[kse_ef41] Bestände an unfertigen/ fertigen Erzeugnissen aus eigener Produktion am Anfang des Geschäftsjahres
kse_ef42	Num	[kse_ef42] Bestände an unfertigen/ fertigen Erzeugnissen aus eigener Produktion am Ende des Geschäftsjahres

Merkmalsliste AFiD-Panel Industrieunternehmen Campus-File 2003-2007		
Variable	Typ	Beschreibung
kse_ef43	Num	[kse_ef43] Veränderung der Bestände an unfertigen/ fertigen Erzeugnissen aus eigener Produktion
kse_ef44	Num	[kse_ef44] Wert der im Geschäftsjahr aktivierten selbsterstellten Anlagen
kse_ef46	Num	[kse_ef46] Gesamtleistung des Unternehmen im Geschäftsjahr
kse_ef50	Num	[kse_ef50] Bestände an Roh-, Hilfs- und Betriebsstoffen zu Beginn des Geschäftsjahr
kse_ef51	Num	[kse_ef51] Bestände an Roh-, Hilfs- und Betriebsstoffen am Ende des Geschäftsjahres
kse_ef52	Num	[kse_ef52] Eingänge (Einkäufe) von Roh-, Hilfs- und Betriebsstoffen während des Geschäftsjahres
kse_ef53	Num	[kse_ef53] Verbrauch von Roh-, Hilfs- und Betriebsstoffen während des Geschäftsjahr
kse_ef55	Num	[kse_ef55] darunter Verbrauch von Brenn- und Treibstoffen, Elektrizität, Gas, Wärme u. dgl. (ohne Wasser)
kse_ef56	Num	[kse_ef56] Bestand an Handelswaren zu Beginn des Geschäftsjahres
kse_ef57	Num	[kse_ef57] Bestand an Handelswaren am Ende des Geschäftsjahres
kse_ef58	Num	[kse_ef58] Eingänge (Einkäufe) an Handelswaren während des Geschäftsjahres
kse_ef59	Num	[kse_ef59] Gesamteinsatz an Handelswaren (kse_57 - kse_56 + kse_58)
kse_ef60	Num	[kse_ef60] geleistete Bruttoentgelte (ohne Arbeitgeberanteile)
kse_ef61	Num	[kse_ef61] geleistete Bruttoentgelte (Arbeitgeberanteile); gesetzliche Sozialkosten
kse_ef62	Num	[kse_ef62] sonstige Sozialkosten
kse_ef63	Num	[kse_ef63] Kosten für Leiharbeiter
kse_ef64	Num	[kse_ef64] Kosten für durch andere Unternehmen durchgeführte Lohnarbeiten
kse_ef65	Num	[kse_ef65] Kosten für durch Dritte erbrachte/durchgeführte Reparaturen, Instandhaltungsarbeiten, Installationen, Montagen u.ä.
kse_ef66	Num	[kse_ef66] Kosten für Mieten, Pachten und Leasing
kse_ef67	Num	[kse_ef67] darunter Kosten für langfristig gemietete oder mittels Operating-Leasing beschaffte Sachanlagen
kse_ef68	Num	[kse_ef68] sonstige Kosten
kse_ef69	Num	[kse_ef69] darunter gezahlte Versicherungsbeiträge
kse_ef71	Num	[kse_ef71] Steuern sowie öffentliche Gebühren und Beiträge
kse_ef72	Num	[kse_ef72] darunter Verbrauchsteuern auf selbst hergestellte Erzeugnisse
kse_ef74	Num	[kse_ef74] steuerliche Abschreibungen auf Sachanlagen
kse_ef75	Num	[kse_ef75] Fremdkapitalzinsen
kse_ef78	Num	[kse_ef78] Kosten insgesamt

Merkmalsliste AFiD-Panel Industrieunternehmen Campus-File 2003-2007		
Variable	Typ	Beschreibung
kse_ef86	Num	[kse_ef86] darunter Aufwendungen für innerbetriebliche Forschung und Entwicklung insgesamt
kse_ef80	Num	[kse_ef80] Subventionen für die laufende Produktion im Geschäftsjahr
kse_ef82	Num	[kse_ef82] Umsatzsteuer, die Kunden in Rechnung gestellt wurde
kse_ef83	Num	[kse_ef83] abzugsfähige Umsatzsteuer sowie Vorsteuer
kse_ef84	Num	[kse_ef84] darunter abzugsfähige Umsatzsteuer für den Kauf von Sachanlagen
kse	Num	[kse] Teilnahme Kostenstrukturerhebung (1 = ja, 0 = nein)
		sonstiges
AggKZ	Num	[AggKZ] Unternehmen wurde mikroaggregiert (1 = ja, 0 = nein)