

Anonymisierungsbeschreibung zur zweiten europäischen Erhebung zur beruflichen Weiterbildung (CVTS 2)

1. Vorbemerkungen

Im Jahr 1987 wurde mit § 16 Abs. 6 des Bundesstatistikgesetzes¹ 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 hohem Aufwand reidentifiziert werden können. „Unverhältnismäßig“ bedeutet hier, dass der Aufwand einer Reidentifikation deren Nutzen übersteigt (faktische Anonymität). Die Deanonymisierung von Einzelangaben in einem faktisch anonymen Datensatz kann nicht mit absoluter Sicherheit ausgeschlossen werden. Der vorliegende Leitfaden behandelt faktisch anonymisierte Datensätze für die Wissenschaft (ein so genannter Scientific-Use-File), generiert aus den Daten der CVTS 2-Erhebung aus dem Jahre 2000 mit Berichtsjahr 1999.

2. Basismaterial

In der Erhebung liegen Angaben von 3184 Unternehmen zur Teilnahme von Beschäftigten an Maßnahmen zur beruflichen Weiterbildung im Jahr 1999 vor. Besonders im Dienstleistungsbereich ist der Anteil der Unternehmen, die geantwortet haben, sehr niedrig; für manche Schichten liegt er im Verhältnis zur Grundgesamtheit bei unter 1 : 600. In anderen Wirtschaftsbereichen war zwar für manche Schichten eine Vollerfassung geplant; da die Teilnahme an der CVTS freiwillig ist, wurde diese aber nicht erreicht. Ein Auswahlsatz von über 50% wird nur in manchen Schichten der Bereiche Mit dem Kredit- und Versicherungsgewerbe verbundene Tätigkeiten und Bergbau und Gewinnung von Steinen und Erden erreicht. Somit besteht schon ein gewisser Schutz für den Datensatz darin, dass ein potentieller Angreifer keine Gewissheit hat, ob das von ihm gesuchte Unternehmen tatsächlich an der Erhebung teilgenommen hat.

Eine vollständige Liste der im Scientific-Use-File enthaltenen Merkmale findet sich im Anhang.

3. Anonymisierungsmaßnahmen

Regionale Gliederung

Es wird keine Regionalinformation weitergegeben. Regionale Angaben sind besonders geeignet für Reidentifikationen. Ein Erhalt solcher Merkmale in einem Scientific-Use-File stellt daher für die Anonymisierung ein schwieriges Unterfangen dar. Hinzu kommt, dass aufgrund der geringen Fallzahl in den neuen Bundesländern (585 Unternehmen) für die meisten Fragestellungen keine belastbaren Ergebnisse getrennt nach alten und neuen Bundesländern zu erzielen wären, so dass

¹ Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 16 des Gesetzes vom 21. August 2002 (BGBl. I S. 3322).

der Wegfall des Regionalmerkmals keine wesentliche Einschränkung des Analysepotentials bedeutet.

Wirtschaftszweigklassifikation

Ausgangspunkt sind die 30 Wirtschaftsbereiche, nach denen die Stichprobenauswahl erfolgte.

Auf Basis der Hochrechnungsfaktoren für ganz Deutschland haben sich einige Wirtschaftsbereiche als besonders gefährdet herausgestellt. Diese mussten daher mit anderen Wirtschaftsbereichen wie folgt zusammengelegt werden:

1. Mit dem Kredit- und Versicherungsgewerbe verbundene Tätigkeiten. Wurde mit Versicherungsgewerbe zusammengelegt.
2. Bergbau, Gewinnung von Steinen und Erden. Wurde mit Kokerei, Mineralölverarbeitung, Chemische Industrie zusammengelegt.
3. Nachrichtenübermittlung. Wurde mit Verkehr zusammengelegt.
4. Kraftfahrzeughandel, -instandhaltung, -reparatur, Tankstellen. Wurde mit Einzelhandel, Reparatur von Gebrauchsgütern zusammengelegt.
5. Holzgewerbe; Herstellung von Möbeln, Schmuck, Musikinstrumenten, Spielwaren, Sportgeräten und sonstigen Erzeugnissen; Recycling. Wurde mit Papiergewerbe zusammengelegt.

Für die zusammengelegten Wirtschaftsbereiche wurden die Hochrechnungsfaktoren neu berechnet als Quotient aus der Anzahl der Unternehmen des zusammengefassten Bereichs in der Grundgesamtheit und der Anzahl der Unternehmen des zusammengefassten Bereichs in der Erhebung.

Eine Übersicht der in der anonymisierten Datei vorhandenen Wirtschaftsbereiche (im Wesentlichen die NACE 30-Klassifikation) samt Fallzahlen ist in nachfolgender Tabelle gegeben:

Tabelle 1: Verteilung der CVTS 2-Unternehmen auf Wirtschaftsbereiche

Wirtschaftsgliederung	NACE 30-Angabe	Anzahl	Grundgesamtheit
Ernährungsgewerbe u. Tabakverarbeitung	02	184	16886
Textil- u. Bekleidungsgewerbe; Ledergew.	03	193	3549
Verlags- u. Druckgewerbe; Vervielfältigung von besp. Ton-, Bild- und Datenträgern	06	125	5737
H. v. Gummi- u. Kunststoffwaren; Glasgewerbe, Keramik, Ver. v. Steinen u. Erden	08	148	8207
Metallerzg. u. -bearbeitung; H. v. Metallerzeugnissen	09	188	16107
Maschinenbau	10	96	10449
H. v. Büromasch., Dv-Gerät. u. einr.; Elekt.-technik; Feinmech. und Optik	11	77	10658
Fahrzeugbau	12	163	1806
Energie- und Wasserversorgung	13	228	1363
Baugewerbe	14	202	61083
Handelsvermittlung und Großhandel	16	115	20727
Gastgewerbe	18	125	16543
Kreditgewerbe	21	198	2788

Grundstücks- und Wohnungswesen; Vermietung bewegl. Sachen; F&E	24	22	27940
Datenverarb. und Datenbanken	25	22	4169
Rechts-, Steuer- u. Unternehmensber.; Markt- u. Meinungsforschung; Beteiligungsgesellschaften	26	25	15167
Architektur- und Ingenieurbüros; Technische, physikalische und chemische Untersuchung; Werbung	27	26	9559
Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften; Detekteien und Schutzdienste; Reinigung von Gebäuden, Inventar und Verkehrsmitteln; Erbringung von sonstigen Dienstleistungen überwiegend für Unternehmen	28	17	7939
Kultur, Sport und Unterhaltung	29	16	8650
Abwasser- u. Abfallbes., sonst. Entsor- gung; Interessenvertretungen; kirchl. u. sonst. Relig. Einr.; Erbringung v. sonst. Dienstleistungen	30	30	15765
Bergbau, Gew. V. Steinen u. Erden; Kokerei, Mineralölver., H. u. Ver. V. Spalt- u. Brutstoffen, Chem. Industrie	01 und 07	254	2746
Holzgew., H. v. Möbeln, Schmuck, Musikinstr., Sportger., Spielw. u. sonst. Erzeugnissen; Recycling; Papiergew.	04 und 05	253	9710
KFZ-handel; Instandhaltung u. Reparatur v. KFZ; Tankstellen; Einzelhandel (ohne Handel mit Kraftfahrzeugen und ohne Tankstellen); Reparatur von Gebrauchsgütern	15 und 17	221	48343
Landverkehr; Transport in Rohrfernleitun- gen; Schiff- u. Luftfahrt; Tätigk. Für den Verkehr; Verkehrsvermittlung; Nachrichtenübermittlung	19 und 20	204	16138
Versicherungsgewerbe; mit dem Kredit- u. Versicherungsgew. verb. Tätigkeiten	22 und 23	52	905

Beschäftigte des Unternehmens

Für Unternehmen mit mehr als 2000 Beschäftigten Ende 1999 wurde die absolute Anzahl der Beschäftigten nicht ausgewiesen, sondern nur, dass das Unternehmen über 2000 Beschäftigte hat. Ferner wurden für diese Unternehmen für männliche und weibliche Beschäftigte nur die Anteile an den Gesamtbeschäftigten und anstelle der Beschäftigtenzahl Ende 1998 die Veränderung von 1998 zu 1999 in Prozent angegeben. Bei den Merkmalen, die abhängig von der Anzahl der Beschäftigten sind (geleistete Arbeitsstunden, Personalaufwendungen, Kosten, Teilnahmestunden, Teilnehmer), wurden die Originalwerte auf Pro-Kopf-Werte (Division durch Anzahl der Beschäftigten Ende 1999) bzw. auf Werte pro Teilnehmer umgerechnet. Von diesen Änderungen sind 85 Unternehmen betroffen, also knapp 3% aller Fälle.

Weitere Maßnahmen

Einige Merkmale, bei denen entweder die Qualität sehr schlecht war oder nur für sehr wenige Unternehmen Angaben vorlagen, wurden aus dem Datensatz entfernt oder modifiziert. Im einzelnen betraf dies die folgenden Merkmale:

Jahresdurchschnitt Beschäftigte 1999: Diese Angabe wurde nur für Unternehmen mit starken saisonbedingten Schwankungen gemacht. Stattdessen wurde ein Merkmal „saison“ ausgewiesen mit saison = 1, falls es saisonbedingte Schwankungen der Beschäftigung gibt.

Anteil der indirekten Kosten: Fällt weg wegen mangelhafter Qualität.

Zahl der ganz oder teilweise mit Lehrveranstaltungen beschäftigten Personen: Fällt weg wegen mangelhafter Qualität, die mitunter darauf zurückzuführen ist, dass die erfragte Information nicht in dieser Form in der Buchhaltung der Unternehmen vorlag.

Fonds, an die Beiträge gezahlt werden: Diese Merkmale fallen weg, da es bei regionalen Fonds nur 77 Ja-Antworten gibt, bei nationalen 7 und bei Sonstigen 47.

Einnahmen, Einnahmequellen, Saldo: Diese Merkmale fallen weg, da es im Datensatz nur 57 Unternehmen gibt, die Einnahmen aus Lehrveranstaltungen haben. Sollte ein potenzieller Datenangreifer Kenntnisse über diese Merkmale haben, so können in Kombination mit den anderen Überschneidungsmerkmalen (Wirtschaftsbereich und Beschäftigtenangabe) eindeutige Fälle entstehen. Es wird daher ein neues Merkmal berechnet, welches den Wert 1 annimmt, wenn ein Unternehmen Einnahmen aus Lehrveranstaltungen hat und 0 sonst.