

Rundung als Geheimhaltungsverfahren bei Veröffentlichungen der Daten der Berufsbildungsstatistik der statistischen Ämter des Bundes und der Länder

– Ein kurzer Erfahrungsbericht des BIBB –

Dr. Alexandra Uhly

Bonn, Mai 2010

Das BIBB entschied sich im Rahmen eines Datenschutzkonzeptes dafür, die Daten der Berufsbildungsstatistik der statistischen Ämter des Bundes und der Länder (kurz: Berufsbildungsstatistik) nur noch gerundet zu veröffentlichen. Die Rundungsmethode wurde auf Basis einer Literaturrecherche sowie von Beratungen mit dem Statistischen Bundesamt gewählt. Im Folgenden wird eine knappe Übersicht über das Vorgehen und die Erfahrungen, die wir mit diesem Verfahren gemacht haben, dokumentiert. Vor- und Nachteile, insbesondere im Vergleich zu Verfahren von Zellsperrungen (Sperrung kleiner Zellwerte oder von Verteilungen mit 100% bzw. 100%-1 in einer Ausprägung), werden dargestellt.

Das Rundungsverfahren des BIBB

Gewählt wurde ein *konventionelles Rundungsverfahren*, bei dem alle Absolutzahlen der Berufsbildungsstatistik auf ein Vielfaches von 3 (auf- bzw. ab-)gerundet werden; Echtzahlen werden nicht mehr veröffentlicht. Folgende Übersicht stellt Beispiele für Echtwerte und veröffentlichte (gerundete) Werte gegenüber:

Echtwerte	0	1	2	3	4	5	6	7	8	9	10	...	1.000	1.001	...
Gerundete Werte	0	0	3	3	3	6	6	6	9	9	9	...	999	1.002	...

Diese Rundungsregel wird für alle Fälle und alle Variablen angewandt; weder wird fallweise entschieden, ob auf- oder abgerundet wird, noch wird entschieden, für welche Merkmale keine Echtzahlen veröffentlicht werden. Die Rundung erfolgt als Tabellengeheimhaltung und nicht im verwendeten Datensatz, sondern erst mit bzw. nach der Auswertung/Abfrage von einzelnen Datentabellen.

Dieses Rundungsverfahren erfüllt Geheimhaltungskriterien, da sich sowohl hinter allen gerundeten Absolutwerten ≥ 3 jeweils drei verschiedene Echtwerte verbergen können und zugleich eine ausgewiesene 0 immer eine echte 0 oder eine echte 1 sein kann. Die Möglichkeit der Rekonstruktion von Echtwerten kann nicht vollständig vermieden werden; i. d. R. ist jedoch keine Rekonstruktion von Echtwerten zu Einzelpersonen möglich (siehe hierzu Anlage). Auch wenn man dies nicht vollständig ausschließen kann (insbesondere, wenn viele unterschiedlich gegliederte Tabellen veröffentlicht werden), so kann man aber annehmen, dass i. d. R. der Aufwand und Sachverstand zur Rekonstruktion der Echtwerte von Einzelpersonen hoch ist.

Nachteil des Rundungsverfahrens

Das Rundungsverfahren erfordert die Verfälschung aller Echtzahlen. Die Abweichungen von den Echtzahlen betragen allerdings maximal den Wert 1. Da die Rundung erst nach der Abfrage von fallweise erzeugten Datentabellen erfolgt, ergeben sich auch keine weiteren Kummulationen von Abweichungsbeträgen. Die maximale Abweichung vom jeweiligen Echtwert der Erhebung ist somit vernachlässigbar gering, es muss nicht von nennenswerten Verzerrungen ausgegangen werden. Bei tiefer gegliederten Betrachtungen von z. B. Einzelberufen können Abweichungen um $+/-1$ deutliche Verzerrungen hervorrufen, in dem Bereich solcher Größenordnungen sind jedoch Absolutwerte, Veränderungsraten oder Verteilungen auch im Falle der Echtzahlen i. d. R. nicht interpretierbar.

Vorteile der Rundungsmethode

- Es ist keine Entscheidung darüber erforderlich, welche Merkmale überhaupt zusätzliche Informationen über Personen darstellen und welche Merkmale datenschutzrechtlich sensibel sind. Dies erleichtert das Vorgehen erheblich, insbesondere, da Einschätzungen über die Sensibilität personenbezogener Merkmale subjektiv sind.
- Die Verfahrensregeln lassen sich in der Praxis unkompliziert umsetzen (z. B. Excel-Funktion =VRUNDEN(Zahl; Vielfaches)), der Umsetzungsaufwand ist gering. Außerdem ist das Rundungsverfahren auch einfach kommunizierbar, sodass es für Datennutzer oder Leser von Veröffentlichungen (auf Basis der gerundeten Daten) leicht nachvollziehbar ist.
- Es sind keine weiteren Zellprüfungen (Gegensperren) und Abstimmungsprozesse zwischen verschiedenen Veröffentlichungsprogrammen erforderlich. Das Verfahren kann somit auch flexibel bei beliebigen Sonderaufbereitungen (nicht nur bei Standardtabellen) angewandt werden. Beim Verfahren der Zellsperrung muss dagegen sichergestellt werden, dass bei Veröffentlichung von Zwischensummen die Werte gesperrter Zellen nicht zu errechnen sind. Dies erfordert eine aufwendige Prüfung aller Spalten und Zeilen aller Tabellen sowie ggf. die Sperrung weiterer Zellen. Ebenso ist beim Zellsperrungsverfahren eine Abstimmung zwischen verschiedenen Veröffentlichungen erforderlich (Zwischensummen). All dies entfällt bei der Rundungsmethode.

- Alle Aggregatdaten können in verschiedenen Differenzierungsgraden – soweit Differenzierungen fachlich sinnvoll sind – veröffentlicht bzw. weitergegeben werden (z. B. nach Berufen und/oder Zuständigkeitsbereichen, nach Staatsangehörigkeit, nach Vorbildung etc.). Dagegen muss beim Zellsarringsverfahren häufig auf tiefere Gliederungen verzichtet werden, um das Ausmaß und den Aufwand der Zellsarrings und Gegensperrung zu begrenzen. Somit ist der Informationsverlust trotz Verfälschung aller Einzelzahlen beim Verfahren der Rundung geringer einzuschätzen als beim Verfahren der Zellsarrings.

Fazit

Trotz der prinzipiellen Verfälschung aller Echtzahlen ist der Verzerrungseffekt bei der Rundungsmethode vernachlässigbar gering (vermutlich geringer als übliche Ungenauigkeiten in Datensätzen), der Informationsverlust ist im Vergleich zum Zellsarringsverfahren sogar geringer. Auch der Aufwand ist erheblich geringer (für jeden, der Daten veröffentlicht, und vor allem im Vergleich zu den erforderlichen Abstimmungen, die bei anderen Verfahren erforderlich sind). Zudem ist die Funktionsweise des Verfahrens leicht nachvollziehbar. Aus diesen Gründen sehen wir auf Basis unserer bisherigen Erfahrungen mit dem Rundungsverfahren ausschließlich Vorteile. Allerdings sind Echtzahlen von Einzelfällen nur dann nicht rekonstruierbar, wenn in keiner Veröffentlichung Echtzahlen verwendet werden.

Anlage Fallbeispiele

Bei folgenden ausgewählten Fallbeispielen können relevante Geheimhaltungsfälle vorliegen. Das einfache Rundungsverfahren berücksichtigt die Geheimhaltungserfordernisse, wenn an keiner Stelle Echtzahlen veröffentlicht werden. Bei den Fällen 1 - 3 sind aus den gerundeten Werten die Echtzahlen nicht eindeutig rekonstruierbar (verschiedene potenzielle Echtwertverteilungen führen zu den gleichen gerundeten Werten). Insbesondere bei kleinen Fallzahlen können unter Umständen die Echtzahlen rekonstruiert werden. Z. B. trifft dies bei den Fällen 4 und 5 zu, allerdings ist auch hierbei die Merkmalsausprägung nicht für bestimmte Einzelfälle ersichtlich. Liegt in einem Beruf nur ein Neuabschluss vor, werden alle Ausprägungen der betrachteten Merkmale sowie die Neuabschlusszahl in diesem Beruf mit Null ausgewiesen.

Fall 1: Alle haben bei einem Merkmal die gleiche Ausprägung: z. B. bei allen neu abgeschlossenen Ausbildungsverträgen (kurz: Neuabschlüsse) in einem Beruf (z. B. 50 Verträge Echtzahl) liegt bei den Jugendlichen als höchster allgemeinbildender Schulabschluss ein Hauptschulabschluss vor.

Fall 2: Mit Ausnahme von einem Vertrag haben alle bei einem Merkmal die gleiche Ausprägung, ein Mal liegt eine andere Ausprägung vor: z. B. bei den neu abgeschlossenen Ausbildungsverträgen (kurz: Neuabschlüsse) in einem Beruf (z. B. 30 Verträge Echtzahl) liegt bei 29 Jugendlichen als höchster allgemeinbildender Schulabschluss ein Hauptschulabschluss vor, bei 1 ein Realschulabschluss. Einschränkend ist zu erwähnen, dass im Falle von einer Variablen mit nur zwei Ausprägungen, bei einer solchen Verteilung die eine Person mit der spezifischen Ausprägung erkennen kann, dass alle anderen die andere Ausprägung aufweisen; die Einzelperson ist jedoch auch in diesem Fall nicht zu identifizieren.

Fall 3: Insgesamt liegt eine kleine Fallzahl vor. Z. B. 4 Neuabschlüsse in einem Beruf, davon 1 ohne, 3 mit Hauptschulabschluss und 1 mit Realschulabschluss.

Fall 4: Insgesamt liegen 4 Fälle vor und jeweils 2 in einer bestimmten Ausprägung, z. B. 4 Neuabschlüsse in einem Beruf, davon 2 ohne und 2 mit Hauptschulabschluss und Fall 5: Insgesamt liegen 4 Fälle vor, jeweils 1 in jeder Ausprägung. Bei Fall 4 und 5 können die Echtwerte aus den gerundeten Werten rekonstruiert werden, nicht aber ein Einzelfall.

Fall 1:

Neuabschlüsse insgesamt	Ohne Hauptschulabschluss	Hauptschulabschluss	Realschulabschluss	Studienberechtigung	Im Ausland erworber Abschluss, der nicht zugeordnet werden kann
Echtwerte					
50	0	50	0	0	0
Gerundete Werte					
51	0	51	0	0	0
Potenzielle Echtwerte					
50	0	50	0	0	0
51	0	51	0	0	0
52	0	52	0	0	0
51	0	50	1	0	0
51	0	50	0	1	0
51	0	50	0	0	1
51	1	50	0	0	0
52	1	50	1	0	0
etc. ...					

Fall 2:

Neuabschlüsse insgesamt	Ohne Hauptschulabschluss	Hauptschulabschluss	Realschulabschluss	Studienberechtigung	Im Ausland erworber Abschluss, der nicht zugeordnet werden kann
Echtwerte					
30	0	29	1	0	0
Gerundete Werte					
30	0	30	0	0	0
Potenzielle Echtwerte					
29	0	29	0	0	0
30	0	30	0	0	0
31	0	31	0	0	0
30	0	29	1	0	0
30	0	29	0	1	0
30	0	29	0	0	1
30	1	29	0	0	0
31	1	29	1	0	0
etc. ...					

Fall 3:

Neuabschlüsse insgesamt	Ohne Hauptschulabschluss	Hauptschulabschluss	Realschulabschluss	Studienberechtigung	Im Ausland erworber Abschluss, der nicht zugeordnet werden kann
Echtwerte					
4	1	2	1	0	0
Gerundete Werte					
3	0	3	0	0	0
Potenzielle Echtwerte					
2	0	2	0	0	0
3	0	3	0	0	0
3	0	2	1	0	0
3	0	2	0	1	0
3	0	2	0	0	1
3	1	2	0	0	0
4	0	4	0	0	0
4	1	2	1	0	0
etc. ...					

Fall 4:

Neuabschlüsse insgesamt	Ohne Hauptschulabschluss	Hauptschulabschluss	Realschulabschluss	Studienberechtigung	Im Ausland erworber Abschluss, der nicht zugeordnet werden kann
Echtwerte					
4	2	2	0	0	0
Gerundete Werte					
3	3	3	0	0	0
Potenzielle Echtwerte					
4	2	2	0	0	0

Fall 5:

Neuabschlüsse insgesamt	Ohne Hauptschulabschluss	Hauptschulabschluss	Realschulabschluss	Studienberechtigung	Im Ausland erworber Abschluss, der nicht zugeordnet werden kann
Echtwerte					
5	1	1	1	1	1
Gerundete Werte					
6	0	0	0	0	0
Potenzielle Echtwerte					
5	1	1	1	1	1

Herausgeber:

Bundesinstitut für Berufsbildung, Bonn

Internet: www.bibb.de

Kontakt:

Dr. Alexandra Uhly, BIBB, Arbeitsbereich 2.1, E-Mail: uhly (at) bibb. de

© Copyright: Die veröffentlichten Inhalte sind urheberrechtlich geschützt.

Namentlich gekennzeichnete Beiträge stellen nicht unbedingt die Meinung des Herausgebers dar.

Veröffentlichung im Internet: 24.01.2012