

Statistische Geheimhaltung in den neuen webbasierten Anwendungen der Statistik der BA

Auswahl und Anwendung eines Verfahrens der deterministischen Rundung



Impressum

Produktlinie/Reihe:	Grundlagen: Hintergrundinfo
Titel:	Statistische Geheimhaltung in den neuen webbasierten Anwendungen der Statistik der BA
Veröffentlichung:	August 2020
Herausgeberin:	Bundesagentur für Arbeit Statistik/Arbeitsmarktberichterstattung
Rückfragen an:	Statistik/Arbeitsmarktberichterstattung, CF 3 Regensburger Straße 104 90478 Nürnberg
E-Mail:	Zentrale.CF3@arbeitsagentur.de
Telefon:	0911 179-6507 / -8311
Fax:	0911 179-1383

Weiterführende statistische Informationen:

Internet:	https://statistik.arbeitsagentur.de
Zitierhinweis:	Statistik der Bundesagentur für Arbeit, Grundlagen: Hintergrundinfo – Statistische Geheimhaltung in den neuen webbasierten Anwendungen der Statistik der BA, Nürnberg, August 2020
Nutzungsbedingungen:	© Statistik der Bundesagentur für Arbeit

Sie können Informationen speichern, (auch auszugsweise) mit Quellenangabe weitergeben, vervielfältigen und verbreiten. Die Inhalte dürfen nicht verändert oder verfälscht werden. Eigene Berechnungen sind erlaubt, jedoch als solche kenntlich zu machen.

Im Falle einer Zugänglichmachung im Internet soll dies in Form einer Verlinkung auf die Homepage der Statistik der Bundesagentur für Arbeit erfolgen.

Die Nutzung der Inhalte für gewerbliche Zwecke, ausgenommen Presse, Rundfunk und Fernsehen und wissenschaftliche Publikationen, bedarf der Genehmigung durch die Statistik der Bundesagentur für Arbeit.

Inhaltsverzeichnis

Zusammenfassung	4
Zellspernung für neue webbasierte Anwendungen überwiegend nicht geeignet.....	4
Statistische Geheimhaltung durch Rundung auf die Basis zehn.....	4
Vorteile und Nachteile der Verfahren	5
Summen.....	6
Arithmetische Mittel	6
Kennzahlen und Quoten	7
Absolute Veränderungen	7
Relative Veränderungen und Anteilswerte	7
Fazit	9
Literaturverzeichnis.....	9

Zusammenfassung

Die statistische Geheimhaltung in den neuen webbasierten Anwendungen der Statistik der BA wird mittels Rundung auf Vielfache der Basis zehn sichergestellt. Anders als das Zellsperverfahren lässt sich die Rundung vollautomatisiert umsetzen und bietet Schutz vor der Offenlegung gesperrter Werte durch Rückrechnungen. Vorteile für Nutzer sind eine hohe Flexibilität beim Tabellenaufbau und eine große regionale und fachliche Auswertungstiefe. Wie beim Zellsperverfahren treten Informationsverluste bei sehr kleinen Fallzahlen auf. Anders als bei der Zellsperung bleiben Größenverhältnisse innerhalb der Tabelle immer erkennbar. Die Ergebnisse sind auswertungsübergreifend konsistent. Für die Darstellung von Anteilen und relativen Veränderungen werden Mindestfallzahlen festgelegt. In den übrigen Veröffentlichungen der Statistik der BA wird weiterhin das Zellsperverfahren für die Geheimhaltung verwendet.

Zellsperung für neue webbasierte Anwendungen überwiegend nicht geeignet

Die meisten Statistiken der BA stellen personenbezogene Daten dar. Die Verpflichtung zur statistischen Geheimhaltung gemäß § 16 des Bundesstatistikgesetzes und weiterer Gesetze gilt für die Statistik und Arbeitsmarktberichterstattung der BA entsprechend. Auch die in den neuen webbasierten Anwendungen der Statistik der BA veröffentlichten Daten beziehen sich überwiegend auf Personen und liegen zudem in großer fachlicher und regionaler Tiefe vor. Im Vergleich zu technisch vergleichbaren Angeboten anderer Anbieter stellt sich hier die Frage eines geeigneten Geheimhaltungsverfahrens in besonderem Maße.

Die Statistik der BA verwendet in ihren Veröffentlichungen bisher übergreifend das Zellsperverfahren zur Sicherstellung der statistischen Geheimhaltung. Für die meisten Ergebnisdarstellungen im Rahmen der neuen webbasierten Anwendungen ist dieses Verfahren nicht geeignet. Aufgrund der Komplexität insbesondere des Ausschlusses von Möglichkeiten zur Rückrechnung stehen keine Algorithmen für eine automatisierte Umsetzung des Zellsperverfahrens zur Verfügung. Selbst wenn diese verfügbar wären, würden sie die Performanz des Systems enorm einschränken, weil für jede vom Nutzer individuell erstellte Tabelle umfangreiche Berechnungen erforderlich wären. Das würde zu nicht akzeptablen Reaktionszeiten für die Nutzer führen. Zudem könnten Nutzer des geplanten Self-Service jene Werte, die in einzelnen Ergebnistabellen gesperrt sind, einfach mittels weiterer Auswertungen offenlegen (Beispiel: erste Auswertung – Frauen 98, Männer *; zweite Auswertung – Insgesamt 100).

Statistische Geheimhaltung durch Rundung auf die Basis zehn

Deshalb wird für die statistische Geheimhaltung in den neuen webbasierten Anwendungen die (automatisierte) deterministische Rundung als Geheimhaltungsverfahren verwendet.¹ Bei der deterministischen Rundung wird jeder Wert einer Tabelle auf das jeweils nächstliegende Vielfache einer festgelegten Basis

¹ Die Rundung bietet keine Lösung für die Geheimhaltung von Dominanzfällen. Diese wird in den neuen webbasierten Anwendungen durch ein eigenes automatisiertes Verfahren umgesetzt.

gerundet. Das Verfahren ist anerkannt und in der Literatur ausführlich beschrieben.² Beispiele für Anwendungsfälle sind die Dreierroundung in der Berichterstattung des BIBB, die Fünferroundung in der Ausländerstatistik nach dem Ausländerzentralregister und die Zehnerroundung beim Zensus 2011.³

Rundungsbasis für die Ergebnisdarstellungen ist zehn. Die Abweichung von den Originalwerten ist bei dieser Basis höher als bei kleineren Basen. Dieses Argument tritt jedoch hinter den folgenden beiden zurück.

Sicherheit vor Aufdeckung: Für kleinere Rundungsbasen – hier am Beispiel von drei und fünf – lassen sich Konstellationen finden, die eine Aufdeckung der wahren Werte mithilfe der Randsummen zulassen.

Runden auf drei: $0 + 0 = 3$ kann nur aus den Originaldaten $1 + 1 = 2$ hervorgegangen sein.

Runden auf fünf: Hinter $0 + 0 + 0 + 0 = 10$ kann nur die Konstellation $2 + 2 + 2 + 2 = 8$ stehen.

Runden auf zehn: Dieses Problem kann nicht auftreten, falls zwischen einem auf null abgerundeten Wert und einem nicht vorhandenen Wert nicht unterschieden wird.

Transparenz: Anhand der dargestellten Werte ist vermutlich für viele Nutzer offensichtlich, dass es sich um gerundete Werte handelt. Selbst wenn Nutzer die entsprechenden methodischen Hinweise nicht beachten oder in der Weiterverwendung der Daten nicht an Auswertungen anbringen, bleibt das Rundungsverfahren leicht erkennbar.

Jeder Wert wird also auf das jeweils nächstliegende Vielfache von zehn gerundet. Bis zur Endziffer vier wird abgerundet, ab fünf aufgerundet (kaufmännisches Runden). Die größte mögliche Abweichung eines dargestellten Werts vom Originalwert beträgt somit fünf (vgl. Tabelle 1).

Originalwert	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Gerundeter Wert	0	0	0	0	0	10	10	10	10	10	10	10	10	10	10	20	20
Abweichung (absolut)	0	-1	-2	-3	-4	5	4	3	2	1	0	-1	-2	-3	-4	5	4

Tabelle 1: Beispiel – Rundung auf Vielfache von zehn (**fett**: angezeigte Werte)

Vorteile und Nachteile der Verfahren

Das Verfahren der Rundung kann ohne hohen algorithmischen Aufwand und ohne größere Performanzverluste in automatisierten Verfahren wie den neuen webbasierten Anwendungen implementiert werden. Dabei ist eine hohe Flexibilität der Nutzer beim individuellen Tabellenaufbau sichergestellt. Die Auswertungstiefe (z. B. regional – Kreise, Gemeinden oder fachlich – Berufe, Branchen) ist praktisch beliebig.

Ähnlich wie beim Zellsperverfahren ist von höheren Informationsverlusten bei sehr kleinen Fallzahlen auszugehen. Allerdings bleiben auch bei überwiegend kleinen Fallzahlen die Größenverhältnisse innerhalb der Tabelle erkennbar, weil kein Wert mit eventuell höherer Fallzahl zum Ausschluss von Rückrechnungen gesperrt werden muss (Beispiel: Insgesamt 100, Männer 98, Frauen 2 wird bei Zellsperung zu

² vgl. z. B. Rohde et al. (2018), S. 90 ff.

³ vgl. Uhly (2010), S. 1 ff.; Brückner et al. (2018), S. 105 ff.; Gießing et al. (2014), S. 643.

Insgesamt 100, Männer *, Frauen * oder zu Insgesamt *, Männer 98, Frauen *; bei Rundung auf Vielfache der Basis zehn bleibt Insgesamt 100, Männer 100, Frauen 0). Die Ergebnisse sind auswertungsübergreifend konsistent, weil jeder Originalwert immer auf den gleichen Wert gerundet wird. Gleichzeitig ist die auswertungsübergreifende Geheimhaltung sichergestellt, weil auch eine Auswertung in anderer Struktur – wie bei der Zellspernung häufig möglich – keine Deanonymisierung erlaubt.

Summen

Summen werden auf Basis der Originalwerte berechnet und anschließend gerundet. Das führt möglicherweise zu einer Abweichung zwischen den dargestellten Summanden und der zugehörigen Summe. Dafür liegt die angezeigte Summe grundsätzlich näher am Originalwert (vgl. Tabelle 2a). Zusätzlich wird jeder Wert immer gleich angezeigt, egal ob er in der aktuellen Darstellung als Einzelwert oder als Summe erscheint (vgl. graue Felder in den Tabellen 2a und 2b).

	Anzahl Original	Summe gerundet	Anzahl gerundet
A	74		70
B	13		10
C	11		10
D	3		0
Summe (A-D)	101	100	90

Tabelle 2a: Beispiel – Summenbildung (**fett**: angezeigte Werte)

	Anzahl Original	Summe gerundet	Anzahl gerundet
A-D	101		100
E-H	134		130
I-M	113		110
N-Z	232		230
Summe (A-Z)	580	580	570

Tabelle 2b: Beispiel – Summenbildung (**fett**: angezeigte Werte)

Arithmetisches Mittel

Bei der Durchschnittsbildung wird die gerundete Summe der Originalwerte verwendet, um größere Abweichungen des angezeigten Durchschnitts von dem auf Basis der Summe der Originalwerte errechneten Durchschnitt zu vermeiden (vgl. Tabelle 3).

	Anzahl Original	Summe gerundet	Anzahl gerundet
A	74		70
B	13		10
C	11		10
D	3		0
Summe	101	100	90
Durchschnitt	25,3	25,0	22,5

Tabelle 3: Beispiel – Durchschnittsbildung (**fett**: angezeigte Werte)

Kennzahlen und Quoten

Kennzahlen und Quoten (z. B. Arbeitslosen- /Unterbeschäftigungsquoten) werden grundsätzlich auf Basis der Originalwerte berechnet.

Absolute Veränderungen

Absolute Veränderungen werden auf Basis der gerundeten Werte errechnet (vgl. Tabelle 4).

	Aktueller Monat	Vormonat	Veränderung absolut
Original	254	250	4
Gerundet	250	250	0

Tabelle 4: Beispiel – Berechnung von absoluten Veränderungen (**fett**: angezeigte Werte)

Relative Veränderungen und Anteilswerte

Relative Veränderungen und Anteilswerte werden auf Basis der Originalwerte berechnet und anschließend auf ganze Zahlen gerundet (vgl. Tabellen 5 und 6). Dieses Verfahren wird der Berechnung von Anteilen und Veränderungen auf Basis gerundeter Werte vorgezogen.

Die Vorteile einer Berechnung auf Basis gerundeter Werte bestünden vor allem darin, dass sie vom Nutzer anhand der dargestellten Absolutwerte nachvollzogen werden könnten. Zudem lägen die dargestellten Werte bei großen Fallzahlen oft näher am Originalwert als bei einer Berechnung auf Basis der Originalwerte und anschließender Rundung. Das Verfahren hat jedoch auch Nachteile; so wichen die dargestellten Werte bei kleineren Fallzahlen (unterhalb 1.000) oft stärker ab als bei einer Berechnung mit Originalwerten. Die ausgewiesene Kommastelle suggerierte eine Genauigkeit, die de facto (zumindest bei kleineren Fallzahlen) nicht gegeben ist. Schließlich ließen sich die dargestellten Anteile meist nicht zu 100 % aufsummieren. Anteile von über 105 % oder unter 95 % in der Summe könnten Nutzer irritieren.

Die Vorteile einer Berechnung auf Basis der Originalwerte liegen vor allem darin, dass die dargestellten Werte bei kleineren Fallzahlen oft deutlich näher am Originalwert liegen als bei der Berechnung mit gerundeten Werten. Anteilswerte lassen sich zudem meist zu 100 % (Abweichungen von höchstens einem Prozentpunkt) aufsummieren. Ein Nachteil liegt darin, dass die Berechnung vom Nutzer nicht nachvollzogen werden kann, weil er die Originalwerte nicht kennt und das Ergebnis als „falsch“ fehlinterpretiert. Diese Gefahr ist jedoch auch bei der Berechnung von Summen auf Basis der Originalwerte gegeben, und das einheitliche Vorgehen bei der Summenbildung und bei der Berechnung von Anteilen und Veränderungen ist aus Nutzersicht ein Vorteil. Nachteilig ist zudem, dass absolute und relative Veränderungen auf Basis unterschiedlicher Ausgangswerte berechnet werden sowie dass die dargestellten Werte bei großen Fallzahlen (über 1.000) mitunter stärker von den Originalwerten abweichen als bei einer Berechnung mit gerundeten Werten (ohne anschließende Rundung des Ergebnisses). Wir nehmen jedoch an, dass für Nutzer durchgängig geringe Abweichungen in ähnlicher Größenordnung akzeptabler sind als hohe Abweichungen bei kleinen Fallzahlen und zunehmende Präzision bei hohen Fallzahlen. Dies kommt Nutzern von Daten auf regionaler Ebene, die tendenziell kleinere Werte betrachten, entgegen und mindert den Nutzen für Betrachter eher großer Zahlen nicht deutlich.

	Aktueller Monat	Vormonat	Veränderung	
			absolut	relativ in %
Original	254	250	4	1,6
Gerundet	250	250	0	2,0

Tabelle 5: Beispiel – Berechnung von relativen Veränderungen (**fett**: angezeigte Werte)

	Gesamt	Unter 25	25 bis unter 50	50 und älter	Anteile in %			Summe Anteile
					Unter 25	25 bis unter 50	50 und älter	
Original	255	8	246	1	3,1	96,5	0,4	100
Gerundet	260	10	250	0	3,0	96,0	0,0	99

Tabelle 6: Beispiel – Anteilsberechnung (**fett**: angezeigte Werte)

Um die Aufdeckung geschützter Werte zu verhindern, ist bei der Berechnung von relativen Veränderungen und Anteilen auf Basis der Originalwerte und anschließender Rundung des Ergebnisses auf ganze Zahlen eine Mindestfallzahl für den Nenner festzulegen. Für Veröffentlichungen im Rahmen der neuen webbasierten Anwendungen wird diese auf 250 festgelegt und bietet damit Sicherheit vor Aufdeckung auch in sehr seltenen Wertekonstellationen (vgl. Tabellen 7 und 8).

	Aktueller Monat	Vormonat	Veränderung	
			absolut	relativ in %
Original	3	2	1	50,0
Gerundet	0	0	0	(50,0) x

 Tabelle 7: Beispiel – Darstellung von Veränderungen bei Unterschreitung der Mindestfallzahl 250 (**fett**: angezeigte Werte)

	Gesamt	U25	25 bis U50	50 und älter	Anteile in %		
					U25	25 bis U50	50 und älter
Original	94	7	86	1	7,4	91,5	1,1
Gerundet	90	10	90	0	(7,0) x	(91,0) x	(1,0) x

 Tabelle 8: Beispiel – Darstellung von Anteilen bei Unterschreitung der Mindestfallzahl 250 (**fett**: angezeigte Werte)

Fazit

Aus den hier angestellten Überlegungen geht das Runden auf Vielfache der Basis zehn als realisierbares und mit Blick auf Sicherheit und Nutzerbedarfe sinnvolles Verfahren hervor. Dies gilt für die Ergebnisdarstellungen in den neuen webbasierten Anwendungen der Statistik der BA, bei denen es als Lösungsweg für die Anforderung, die statistische Geheimhaltung sicherzustellen, implementiert wird. In den übrigen Veröffentlichungen der Statistik der BA wird weiterhin das Zellsperverfahren für die Geheimhaltung verwendet.

Literaturverzeichnis

- Brückner, G. et al. (2018): Statistische Geheimhaltung in Bevölkerungsstatistiken. Einsatz der Fünfermethode in der Ausländerstatistik nach dem Ausländerzentralregister. In: Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik, Ausg. 3, S. 105-117.
- Gießing, S. et al. (2014): Geheimhaltung beim Zensus 2011. Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik, Ausg. 11, S. 641-647.
- Rohde, J. et al. (2018): Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens. In: Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik, Ausg. 3, S. 90-104.
- Uhly, A. (2010): Rundung als Geheimhaltungsverfahren bei Veröffentlichungen der Daten der Berufsbildungsstatistik der statistischen Ämter des Bundes und der Länder – Ein kurzer Erfahrungsbericht des BIBB. Bundesinstitut für Berufsbildung (Hrsg., 2012), Bonn, S. 1-6. https://www.bibb.de/dokumente/pdf/a21_dazubi_rundung_erfahrungsbericht_bibb_04-05-2010.pdf, abgerufen am 04.08.2020.

Statistik-Infoseite

Im Internet stehen statistische Informationen unterteilt nach folgenden Themenbereichen zur Verfügung:

Fachstatistiken:

[Arbeitsuche, Arbeitslosigkeit und Unterbeschäftigung](#)
[Ausbildungsmarkt](#)
[Beschäftigung](#)
[Einnahmen/Ausgaben](#)
[Förderung und berufliche Rehabilitation](#)
[Gemeldete Arbeitsstellen](#)
[Grundsicherung für Arbeitsuchende \(SGB II\)](#)
[Leistungen SGB III](#)

Themen im Fokus:

[Berufe](#)
[Bildung](#)
[Corona](#)
[Demografie](#)
[Eingliederungsbilanzen](#)
[Entgelt](#)
[Fachkräftebedarf](#)
[Familien und Kinder](#)
[Frauen und Männer](#)
[Langzeitarbeitslosigkeit](#)
[Menschen mit Behinderungen](#)
[Migration](#)
[Regionale Mobilität](#)
[Wirtschaftszweige](#)
[Zeitarbeit](#)

Die [Methodischen Hinweise der Statistik](#) bieten ergänzende Informationen.

Die [Qualitätsberichte](#) der Statistik erläutern die Entstehung und Aussagekraft der jeweiligen Fachstatistik.

Das [Glossar](#) enthält Erläuterungen zu allen statistisch relevanten Begriffen, die in den verschiedenen Produkten der Statistik der BA Verwendung finden.

Abkürzungen und Zeichen, die in den Produkten der Statistik der Bundesagentur für Arbeit vorkommen, werden im [Abkürzungsverzeichnis](#) bzw. der [Zeichenerklärung](#) der Statistik der BA erläutert.