Tim Hochgürtel

studierte Soziologie in Mainz und ist seit 2008 im Statistischen Bundesamt tätig, seit 2016 als Referent im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“. Schwerpunkte seiner Arbeit sind Analysen sowie Publikationen zu privaten Haushalten und Lebensformen.

Clarissa Wilke

hat in Mainz Politik- und Wirtschaftswissenschaften studiert. Seit 2015 ist sie im Statistischen Bundesamt tätig und seit April 2018 im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“. Schwerpunkte ihrer Arbeit sind Auswertungen und Publikationen zu privaten Haushalten und Lebensformen.

MESSFEHLER-BEDINGTER BIAS BEI KLEINEN TEILMENGEN

Tim Hochgürtel, Clarissa Wilke

Schlüsselwörter: Messfehler – Bias – binäres Merkmal – Aggregat – Teilpopulation

ZUSAMMENFASSUNG

Die Messungen von dichotomen Individualeigenschaften können fehlerbehaftet sein. Diese Messfehler sind in der Regel nicht erkennbar und können somit nicht korrigiert werden. Damit besteht die Möglichkeit, dass Messfehler der Individualebene Aggregate wie Häufigkeiten und Anteilswerte verzerren. Der vorliegende Beitrag zeigt, dass besonders kleine Häufigkeiten und Anteilswerte von einer solchen Verzerrung betroffen sein können. Die Quantifizierung von kleinen Anteilswerten und Häufigkeiten stellt damit besondere Anforderungen an die Güte der Messinstrumente.

Keywords: *Measurement error – bias – binary variable – aggregate – subpopulation*

ABSTRACT

The measurement of dichotomous individual traits can be defective. These measurement errors are generally undetectable and therefore cannot be corrected. There is the possibility that measurement errors on an individual level can distort aggregates such as frequencies and shares. This article shows that small frequencies and shares are especially affected by distortion. The quantification of small shares and frequencies places special demands on the quality of the measuring instruments.

1

Messfehler in Befragungen

Die Beobachtung sozialer Phänomene basiert auf der Verwendung von Unterscheidungen (Luhmann, 1990). Messungen von Individualeigenschaften dienen der Erfassung von Unterscheidungen. Bei der Realisierung von Messungen wird einem empirischen Relativ ein numerisches Relativ zugeordnet (Fischer, 1974). Dies bedeutet, dass den Merkmalsausprägungen der Merkmale von Objekten (empirisches Relativ) Werte (numerisches Relativ) über eine Messanweisung zugewiesen werden (Schumann, 2006). Über das Skalenniveau der Messung wird das Verhältnis der Merkmalsausprägungen abgebildet.

So sind etwa für Personen verschiedene Merkmalsausprägungen eines Merkmals unterscheidbar. Welche Merkmalsausprägungen es grundsätzlich geben kann, ist abhängig von Annahmen über das zu messende Phänomen. Beispielsweise erfolgt in Abhängigkeit von den Annahmen über die Eigenschaften des Merkmals „Geschlecht“ eine Erfassung der Ausprägungen anhand von zwei (männlich, weiblich) oder mehr Kategorien (zum Beispiel mit den Kategorien „männlich“, „weiblich“, „divers“ und „Kein Eintrag im Personenstandsregister“). Die Menge der möglichen Merkmalsausprägungen, welche eine Messung zulässt, entspricht dem numerischen Relativ. Das empirische Relativ besteht aus der Menge der Personen, bei denen die Messung durchgeführt wird, und der Relation der Personen hinsichtlich des zu messenden Merkmals. Durch die Messung wird den Personen ein Messwert aus der Menge der zulässigen Merkmalsausprägungen des numerischen Relativs zugeordnet.

Unabhängig von ihrem Skalenniveau sind Messungen fehleranfällig. Bei einem Messvorgang kann die Zuordnungsregel, die der Übertragung des empirischen Relativs in das numerische Relativ dient, verletzt werden, sodass sich der gemessene Wert vom wahren Wert unterscheidet (Dieckmann, 2014). Die Eigenschaften der Messfehler sind abhängig von der Art des betrachteten Merkmals. Für stetige Merkmale, wie beispielsweise eine Einkommensmessung in Euro, eignen sich Modelle, bei welchen der Messfehler als additive oder multiplikative Überlagerung des wahren Werts verstanden wird (siehe zum Beispiel Grabe, 2011). Hierbei wird die Merkmalsmessung von stetigen Merkmalen immer als fehlerbehaftet betrachtet. Für die im Folgenden diskutierte Messung von dichotomen, also zweigeteilten Merkmalen erweisen sich probabilistische Ansätze als zielführend (Biemer/Stokes, 1991). Bei einer Messung, die einem Merkmalsträger eine von zwei möglichen Merkmalsausprägungen zuweist, kann diese Zuweisung richtig oder falsch sein. Die Messung kann damit fehlerfrei sein, sofern die Messung dem Merkmalsträger die richtige Merkmalsausprägung zuweist. Wird die falsche Merkmalsausprägung zugewiesen, ist die Messung fehlerhaft.

Zu Messfehlern in sozialwissenschaftlichen Erhebungen liegt eine umfangreiche Methodenforschung vor. Hierbei wurde eine Vielzahl von Ursachen für Messfehler identifiziert. Eine Fehlerquelle liegt in der möglichen Fehlinterpretation der Fragen durch die Befragten. Des Weiteren können sich Fragen sowie die dazugehörigen Antworten auf die Beantwortung von Folgefragen auswirken (sogenannter Halo-Effekt). Bei einer persönlichen Befragung können Eigenschaften der Interviewerin oder des Interviewers sowie die Anwesenheit weiterer Personen das Antwortverhalten der Befragten beeinflussen. Daneben findet sich bei Befragten die Tendenz, Fragen zustimmend zu beantworten sowie das Antwortverhalten in Richtung einer sozialen Erwünschtheit zu verzerren (Dieckmann, 2014; Schnell und andere, 2013).

Um Messfehler möglichst zu reduzieren, gibt es eine Vielzahl von Empfehlungen zur Konstruktion von Frageinstrumenten (Porst, 2011; Raab-Steiner/Benesch, 2015). Hierzu gehört beispielsweise, in den Fragebogen zu schriftlichen Befragungen einfache und präzise Sprache zu verwenden (Dieckmann, 2014; Schumann, 2006).

Aber auch wenn diese Empfehlungen zur Erstellung von Messinstrumenten eingehalten werden, ist davon auszugehen, dass Messungen fehlerbehaftet sind. Nach dem Messvorgang liegt zwar das Messergebnis vor, es ist aber nicht möglich, eine fehlerhafte Messung als solche zu erkennen. Aus den Messergebnissen einer Population lassen sich keine Rückschlüsse auf die Güte der Messungen ziehen. Bei der Messung von dichotomen Variablen bleibt der Anteil der fehlerhaften Messungen unbekannt. Daneben ist es auch nicht möglich zu identifizieren, bei welchen Merkmalsträgern eine fehlerhafte Messung vorliegt.

Fehlerhafte Messungen können auffallen, wenn mindestens zwei Angaben einer Person aus einer Befragung miteinander im Widerspruch stehen (Schumann, 2006). Personen, die als Familienstand „ledig" angegeben haben, können nicht zugleich mit ihrem Ehepartner in einem Haushalt leben. In einem solchen Fall ist aber nicht bekannt, welche Messung der beiden widersprüchlichen Angaben fehlerbehaftet ist.

Liegen aus verschiedenen Messvorgängen eines Merkmals abweichende Ergebnisse vor, ist ebenfalls evident, dass mindestens eine der Messungen fehlerhaft sein muss. Es lassen sich aber nicht ohne Weiteres Rückschlüsse darauf ziehen, welche der beiden Messungen vom wahren Wert abweicht.

Da fehlerhafte Messungen als solche nicht erkennbar sind, können Messfehler nicht korrigiert werden und gehen somit in die weiteren datenverarbeitenden Prozesse mit ein. Messfehler auf der Individualebene haben das Potenzial, systematische Verzerrungen von Aggregaten zu verursachen. Diese Verzerrungen im Aggregat können ebenfalls nicht als solche erkannt werden. Während etwa zufällige Fehler einer Stichprobenerhebung quantifiziert und damit kontrolliert werden können (siehe zum Beispiel Krug und andere, 2001), bleiben systematische Fehler unbekannt.

Der folgende Beitrag zeigt am Beispiel der Messung eines binären Merkmals, wie Messfehler auf der Individualebene zu einem Bias, also zu einer Verzerrung auf der Aggregatebene führen können. Es wird gezeigt, dass Messfehler auf der Individualebene insbesondere bei kleinen Häufigkeiten und Anteilswerten einen stark verzerrenden Effekt haben. Für die Quantifizierung kleiner Teilmengen fallen Messfehler besonders ins Gewicht. Die Wahrscheinlichkeit einer deutlichen Verzerrung von Ergebnissen ist daher hier besonders hoch.

Im zweiten Kapitel des Beitrags wird zunächst die Messung eines binären Merkmals auf der Individualebene beschrieben. Anschließend werden basierend auf der in Kapitel 2 entwickelten Wahrscheinlichkeitsmatrix der Erwartungswert und die Varianz von Häufigkeiten und Anteilswerten abgeleitet (Kapitel 3). Im darauffolgenden vierten Kapitel wird analysiert, welche Verzerrungen von Aggregaten aus fehlerhaften Messungen auf der Individualebene folgen. Dabei wird ersichtlich, dass insbesondere kleine Anteilswerte und Häufigkeiten anfällig für einen ausgeprägten Bias sind. Der Beitrag schließt mit einem Fazit.

2

Modellierung von Messfehlern von dichotomen Variablen

Im Folgenden wird untersucht, welche systematischen Effekte fehlerhafte Messungen der Individualebene auf Aggregate aufweisen. Die systematischen Abweichungen der Aggregate sind hierbei von der Messgüte der Individualdaten abhängig. Unter Messgüte wird der Anteil der korrekten Messungen der Merkmalsausprägungen verstanden.

In diesem Beitrag werden ausschließlich dichotome Merkmale betrachtet. Der verwendete Ansatz ist aber grundsätzlich für alle diskreten Merkmale anwendbar, da sich diskrete Merkmale mit k Ausprägungen auch als k dichotome Merkmale transformieren lassen.

↘ **Tabelle 1** beschreibt die Messung eines dichotomen Merkmals X durch ein Messinstrument Y. Die Wahrscheinlichkeitsmatrix zeigt, mit welcher Wahrscheinlichkeit das Messinstrument Y bei der Messung eines binären Merkmals X mit den Ausprägungen 1 und 2 auf der Individualebene einen Messwert $\tilde{X}$ misst. Die Messungen der Messwerte $\tilde{X}$ sind dabei von den Ausprägungen des Merkmals X abhängig.

Tabelle 1
Wahrscheinlichkeitsmatrix: $P(\tilde{X} = i|X = j)$

		$\tilde{X}$	
		1	2
X	1	α_1	$1-\alpha_1$
	2	$1-\alpha_2$	α_2

Messergebnisse werden als bedingte Wahrscheinlichkeiten $P(\tilde{X} = i|\, X = j)$ verstanden. Die Wahrscheinlichkeit, ein bestimmtes Messergebnis $\tilde{X}$ zu erzielen, ist dabei von den Ausprägungen des Merkmals X abhängig. Mit einer Wahrscheinlichkeit von $P(\tilde{X} = 1|\, X = 1) = \alpha_1$

misst das Messinstrument Y für eine Person den Wert $\tilde{X} = 1$, wenn der wahre Wert $X = 1$ ist. Die Wahrscheinlichkeit, dass bei einem wahren Wert von $X = 1$ das fehlerhafte Messresultat $\tilde{X} = 2$ ermittelt wird, ist somit $P(\tilde{X} = 2 | X = 1) = 1 - \alpha_1$. Die Wahrscheinlichkeit, bei einem vorliegenden Wert von $X = 2$ für eine Person auch den Messwert $\tilde{X} = 2$ zu erhalten, ist $P(\tilde{X} = 2 | X = 2) = \alpha_2$. Entsprechend ist die komplementäre Wahrscheinlichkeit der fehlerhaften Messung $P(\tilde{X} = 1 | X = 2) = 1 - \alpha_2$.

Zu beachten ist, dass die Wahrscheinlichkeiten der Messung der wahren Werte von den Merkmalsausprägungen abhängen. Ein Messinstrument Y kann eine Ausprägung 1 mit einer anderen Wahrscheinlichkeit richtig messen als eine Ausprägung 2.

Die Merkmalsausprägungen von dichotomen Variablen können ungleich gut zu messen sein. Für eine Methodenstudie befragten Preisendörfer und Wolter Personen, die bereits in einem Gerichtsverfahren verurteilt worden waren (Preisendörfer/Wolter, 2014). Eine Verurteilung gaben nur rund 63 % der befragten Personen in der Erhebung auch an. Dies entspricht einem α-Wert von 0,63. Die Frage nach einer Verurteilung generiert bei verurteilten Personen einen großen Messfehler, da die Verurteilungen in mehr als einem Drittel der Fälle verschwiegen werden. Die Wahrscheinlichkeit eines Messfehlers ist damit $1 - \alpha = 0{,}37$. Dies lässt sich als Effekt der sozialen Erwünschtheit verstehen. Ein Teil der Befragten verschweigt eine Verurteilung, um sich selbst positiver darzustellen.

Personen, die nicht verurteilt wurden, haben entsprechend keine Motivation, bei einer Frage nach einer Verurteilung die Unwahrheit zu sagen. Daher ist für diese Messung ein deutlich geringerer Fehler zu erwarten.

Im Folgenden wird der systematische Bias von Häufigkeiten und Anteilswerten in Abhängigkeit von $P(\tilde{X} = i | X = j)$ diskutiert. Dabei lässt sich zeigen, dass die Quantifizierung kleiner Häufigkeiten und Anteilswerte einen systematischen Bias begünstigt. In einem empirischen Anwendungsfall wird der Bias jedoch nicht messbar sein, solange α_1 und α_2 unbekannt bleiben. Dies ist fast immer der Fall.

3

Häufigkeit und Anteilswert

Bei Befragungen ist im Allgemeinen der wahre Wert X von Personen unbekannt. Lediglich der Messwert $\tilde{X}$ einer Person liegt vor. Daher müssen die Messwerte $\tilde{X}$ verwendet werden, um weitere Aggregatsgrößen zu bilden. Gemäß dem Modell aus Kapitel 2 sind Messungen diskreter Merkmale Zufallsprozesse, die richtige oder falsche Ergebnisse erzeugen können. Damit sind auch abgeleitete Aggregatsgrößen wie Häufigkeiten und Anteile Zufallsvariablen.

Im Folgenden wird die Bildung von Häufigkeiten und Anteilswerten eines Merkmals X betrachtet, welches bei einer Population des Umfangs n gemessen wurde.

Eine gemessene Häufigkeit entspricht der Realisation einer Zufallsvariable (siehe hierzu auch Anhang I) mit folgendem Erwartungswert und Varianz:

$$E\left(n \middle| \tilde{X} = 1\right) = (n|X = 1) \cdot \alpha_1 + (n|X = 2) \cdot (1 - \alpha_2)$$

$$VAR\left(n \middle| \tilde{X} = 1\right) = (n|X = 1) \cdot \alpha_1 \cdot (1 - \alpha_1) + (n|X = 2) \cdot (1 - \alpha_2) \cdot \alpha_2$$

Hierbei bezeichnet $n|\tilde{X} = 1$ die Anzahl der Merkmalsträger aus der betrachteten Population, bei denen der Wert $\tilde{X} = 1$ gemessen wurde. Die Teilpopulation $n|\tilde{X} = 1$ setzt sich aus zwei Mengen zusammen. Zum einen besteht sie aus Personen, bei denen der Wert $X = 1$ richtig gemessen wurde, für die also gilt $X = 1$ und $\tilde{X} = 1$. Zum anderen gehen Personen in die Teilpopulation mit ein, bei denen zwar $X = 2$ gilt, aber $\tilde{X} = 1$ gemessen wurde.

Hingegen sind Personen mit $X = 1$ und $\tilde{X} = 2$ nicht in der Teilpopulation enthalten. Personen, bei denen trotz eines wahren Werts von $X = 1$ ein Wert von $\tilde{X} = 2$ gemessen wurde, tragen nicht zur Häufigkeit $n|\tilde{X} = 1$ bei.

Ähnliches gilt für den Anteilswert $\frac{n|\tilde{X}=1}{n}$. Auch gemessene Anteilswerte können als Zufallsvariable verstanden werden, für die gilt (siehe Anhang II):

$$E\left(\frac{n|\tilde{X}=1}{n}\right)=\left(\frac{n|X=1}{n}\right)\cdot\alpha_1+\left(\frac{n|X=2}{n}\right)\cdot(1-\alpha_2)$$

$$Var\left(\frac{n|\tilde{X}=1}{n}\right)=\left(\frac{n|X=1}{n^2}\right)\cdot\alpha_1\cdot(1-\alpha_1)+\left(\frac{n|X=2}{n^2}\right)\cdot(1-\alpha_2)\cdot\alpha_2$$

Damit beeinflussen auch Fehlmessungen die gebildeten Häufigkeiten und Anteilswerte. So gehen in ein Aggregat jene Personen nicht mit ein, welche zwar zur Teilpopulation gehören, die durch das Aggregat beschrieben werden soll, denen jedoch fälschlicherweise ein anderer Messwert zugewiesen wurde. Hingegen werden bei der Aggregation Personen berücksichtigt, welche nicht zur zu beschreibenden Teilpopulation gehören, denen aber fälschlicherweise ein Messwert zugewiesen wurde, welcher sie als der Teilpopulation zugehörig kennzeichnet.

4

Systematische Fehler von Aggregaten

Im Allgemeinen bleiben Messfehler von Individualangaben unbekannt. Im Falle der Messung dichotomer Merkmale lässt sich nicht ermitteln, welche Messungen im Einzelnen fehlerhaft sind und welche Häufigkeiten und Anteile somit fehlerbehaftet sind. Es lässt sich jedoch zeigen, dass kleine Häufigkeiten und Anteile auch unabhängig von der Messgüte eines Messinstruments im besonderen Maße von potenziellen Verzerrungen betroffen sind.

Als Bias wird die systematische Verzerrung des gemessenen Aggregatswerts gegenüber dem wahren Wert bezeichnet (Faulbaum, 2014). Aus Messfehlern auf der Individualebene können systematische Verzerrungen (Bias) auf der Aggregatsebene resultieren. Der Erwartungswert der gemessenen Häufigkeit $n|\tilde{X}=1$ und des Anteilswerts $\frac{n|\tilde{X}=1}{n}$ sind durch Messfehler um den Faktor ω verzerrt.

Es gilt:

$$E\left(n|\tilde{X}=1\right)=(n|X=1)\cdot\omega$$

$$E\left(\frac{n|\tilde{X}=1}{n}\right)=\left(\frac{n|X=1}{n}\right)\cdot\omega$$

wobei (siehe Anhang III):

$$\omega=\alpha_1+\frac{n|X=2}{n|X=1}\cdot(1-\alpha_2)$$

Der verzerrende Faktor ω ist für Häufigkeiten und Anteilswerte identisch. Häufigkeiten und Anteilswerte sind unverzerrt, wenn $\omega=1$. In drei Szenarien ist dies der Fall:

I. Werden die Merkmalsausprägungen eines dichotomen Merkmals X ohne Messfehler erhoben, sind die Aggregate ebenfalls frei von einer systematischen Verzerrung. Damit sind Häufigkeiten und Anteilswerte unverzerrt, wenn gilt:

   $1-\alpha_1=1-\alpha_2=0$

II. Messfehler auf der Individualebene gleichen sich im Aggregat aus, wenn die Merkmalsausprägungen eines dichotomen Merkmals mit gleicher Güte gemessen werden und die Anzahl der Merkmalsträger je Merkmalsausprägung identisch ist. Die Messung ist nicht durch einen systematischen Fehler beeinflusst, wenn:

   $(n|X=1)=(n|X=2)$ und $1-\alpha_1=1-\alpha_2$

III. Der Fall II. lässt sich weiter verallgemeinern. Verzerrungen gleichen sich aus, wenn der Messfehler $1-\alpha_1$ die gleiche Größe aufweist wie das Produkt aus dem Verhältnis der beiden Teilmengen und dem Messfehler $1-\alpha_2$. Dies ist beispielsweise dann der Fall, wenn der Messfehler $1-\alpha_1$ doppelt so groß ist wie der Messfehler $1-\alpha_2$, zur Teilmenge mit $X=2$ aber doppelt so viele Personen beitragen wie zur Teilmenge $X=1$. Häufigkeiten und Anteilswerte sind frei von einem Bias, wenn gilt:

   $$1-\alpha_1=\frac{n|X=2}{n|X=1}\cdot(1-\alpha_2)$$

In der Regel sind bei der Bestimmung von Häufigkeiten oder Anteilswerten zwar $\left(n|\tilde{X}=1\right)$ und $\left(n|\tilde{X}=2\right)$ bekannt, aber nicht $1-\alpha_1$ und $1-\alpha_2$. Daher kann der durch Messfehler auf der Individualebene verursachte

Bias nicht bestimmt werden. Damit ist auch eine Korrektur der messfehlerbedingten Verzerrung nicht möglich.

Es ist aber zu beachten, dass der Bias einer Häufigkeit oder eines Anteilswerts nicht nur von den Messfehlern $1-\alpha_1$ und $1-\alpha_2$ abhängt. Im Fall der Analyse von dichotomen Merkmalen ist der Bias ω auch unmittelbar vom Verhältnis $\frac{(n|X=2)}{(n|X=1)}$ der beiden Teilmengen $n|X=1$ und $n|X=2$ abhängig. Besonders bei den Analysen kleiner Teilmengen kann dies zu einer beträchtlichen Verzerrung der Ergebnisse führen.

↘ Tabelle 2 weist den verzerrenden Faktor ω bei der Messung eines Anteilswerts auf Basis eines dichotomen Merkmals für verschiedene Kombinationen von Messfehlern und Anteilswerten aus. Um das Beispiel übersichtlich zu halten, gilt $\alpha_1=\alpha_2$. Vereinfachend wird davon ausgegangen, dass die Ausprägungen 1 und 2 mit gleicher Güte gemessen werden. In den Spalten von Tabelle 2 wird der Bias für verschiedene Messgüten angenommen. Der Anteil der zutreffenden Messungen variiert zwischen 99,5 und 90%.

Tabelle 2
Bias ω für ausgewählte Anteilswerte und α_k (k=1,2)

$\frac{n\|X=1}{n}$	α_k			
	0,995	0,99	0,95	0,9
99%	1,00	0,99	0,95	0,90
90%	1,00	0,99	0,96	0,91
70%	1,00	0,99	0,97	0,94
50%	1,00	1,00	1,00	1,00
30%	1,01	1,01	1,07	1,13
10%	1,04	1,08	1,40	1,80
5%	1,09	1,18	1,90	2,80
1%	1,49	1,98	5,90	10,80
0,5%	1,99	2,98	10,90	20,80
0,2%	3,49	5,98	25,90	50,80

In den Zeilen von Tabelle 2 werden verschiedene Anteilswerte von 99 bis 0,2% dargestellt. In den Zellen wird der jeweilige Bias für die Kombination aus Anteilswert und Messgüte ausgewiesen.

Für einen Anteilswert von 50% beträgt der Bias 1, bei identischer Messgüte von α_1 und α_2. Die Anteilswerte sind damit nicht verzerrt. Dies entspricht einem Anwendungsfall des in der oberen Aufzählung unter II. diskutierten Szenarios. Da die Menge der Merkmalsträger in beiden Teilpopulationen identisch ist und die Messgüte der Erfassung der Merkmalsausprägungen nicht variiert, gleichen sich die Messfehler aus. Die zu erwartende Menge der Merkmalsträger mit $X=1$, die aufgrund einer fehlerhaften Messung fälschlicherweise der Population $X=2$ zugewiesen wurde, entspricht genau der Menge, bei der $X=2$ vorliegt, aber $X=1$ gemessen wurde.

Für die folgenden kleineren Anteilswerte gilt dies nicht. Je geringer die Messgüte, desto größer wird der Bias. Wenn ein Anteilswert von 30% (wahrer Wert) gemessen werden soll, beträgt der Bias 1,01 bei $\alpha_1=\alpha_2=0{,}995$. Damit ist ein Messwert von 30,3% zu erwarten. Bei einer geringeren Messgüte von $\alpha_1=\alpha_2=0{,}9$ steigt der Bias auf 1,13. Damit ist ein Messwert von 33,9% erwartbar.

Tabelle 2 zeigt deutlich, dass der Bias neben der Messgüte vom zu messenden Anteilswert abhängig ist. So sind kleine Anteilswerte im besonderen Maße von der verzerrenden Wirkung des messfehlerbedingten Bias betroffen. Ein Anteilswert von 0,2% wird selbst bei einer Messgüte von $\alpha_1=\alpha_2=0{,}995$ um den Faktor 3,49 verändert. Damit ist ein gemessener Anteilswert von 0,7% zu erwarten. Nimmt die Messgüte weiter ab, so ist dies mit einem Anstieg des Bias verbunden. Wenn eine von zehn Messungen fehlerhaft ist, wird der gemessene Anteilswert um mehr als den Faktor 50 überhöht. Damit kann die Messung kleiner Populationen um ein Vielfaches größer ausfallen als es dem wahren Wert entspricht.

Bei großen Anteilswerten zeigt sich der verzerrende Effekt in deutlich geringerem Maße. Ein Anteil von 99% wird bei einer Messgüte von 0,95 um den Faktor 0,95 verzerrt. Damit ist ein gemessener Anteil von rund 94% zu erwarten. Bei gleicher Messgüte wird ein Anteil von 1% um den Faktor 5,9 verzerrt, sodass ein gemessener Anteil von rund 6% erwartet werden kann.

Damit zeigt sich, dass insbesondere kleine Populationen als Häufigkeit oder Anteilswert nur verzerrt abgebildet werden können, wenn die zugrundeliegenden Messungen der Individualangaben keine hinreichende Messgüte aufweisen. Ursache hierfür ist, dass ein geringer Anteil von fehlerhaften Messungen in der komplementären Population den Umfang der zu messenden Population deutlich zu groß erscheinen lässt.

Dies bringt besonders Konsequenzen für die Interpretation von Ergebnissen der Quantifizierung kleiner Teilpopulationen mit sich. Werden kleine Teilpopulationen gemessen, muss der mögliche verzerrende Effekt berücksichtigt werden, der mit Messungen eingeschränkter Messgüte verbunden ist. Insbesondere gilt dies bei der Messung von Phänomenen, bei denen keine Vergleichswerte aus anderen Erhebungen vorliegen oder die Messgüte der Erfassungsinstrumente vor dem Hintergrund der sozialwissenschaftlichen Methodenforschung als eingeschränkt zu betrachten ist.

Dies zeigt sich etwa bei der Erfassung von Menschen mit diversem Geschlecht auf Basis von Befragungen der amtlichen Statistik wie dem Mikrozensus. Es kann erwartet werden, dass nur eine sehr kleine Teilpopulation dem diversen Geschlecht zugehörig ist. Wenn nun ein kleiner Anteil der Menschen, welche nicht dem diversen Geschlecht angehören, aber „divers" als eigenes Geschlecht benennt, ist der gemessene Anteil deutlich verzerrt. Damit würde die Gruppe der Personen mit diversem Geschlecht als zu groß dargestellt werden. Erschwerend kommt hierbei hinzu, dass zur Anzahl der Menschen diversen Geschlechts keine Vergleichszahlen aus anderen Erhebungen vorliegen.

5

Fazit

Bei der Messung von Individualwerten kann es zu Messfehlern kommen. In der Regel bleiben diese Messfehler unerkannt. Damit bestehen auch keine Möglichkeiten, Messfehler auf der Individualebene zu korrigieren.

Die sozialwissenschaftliche Methodenforschung hat eine Vielzahl potenzieller Ursachen für Messfehler identifiziert. Daneben werden Empfehlungen zur Ausgestaltung von Befragungsinstrumenten formuliert. Damit lassen sich Messfehler zwar reduzieren, aber nicht ausschließen.

Die unerkannten Messfehler werden in der weiteren Ergebniserstellung mitgeführt. Bei der Erstellung von Aggregaten, wie Häufigkeiten oder Anteilen, gehen die fehlerhaften Messwerte in die Berechnungen mit ein.

Wenn die Messung von dichotomen Merkmalen bei einem Teil der Merkmalsträger fehlerhaft ist, kann dies auch zu einer Verzerrung von Aggregaten führen. Der Grad der Verzerrung hängt dabei einerseits von der Messgüte ab, mit welcher das dichotome Merkmal gemessen wurde. Andererseits ist das Verhältnis der Teilmengen $\frac{(n|X=2)}{(n|X=1)}$ unmittelbar ausschlaggebend für den Grad der Verzerrung. Damit sind besonders kleine Häufigkeiten und Anteile von Verzerrungen betroffen.

Um kleine Häufigkeiten und Anteile mit hinreichender Genauigkeit quantifizieren zu können, sind Messinstrumente notwendig, die nur wenige Messfehler erzeugen. Eine Bewertung, ob ein Messinstrument das genannte Kriterium erfüllt, ist jedoch kaum möglich. Dies würde wiederum voraussetzen, dass die Messfehler als solche erkennbar sind.

Ist der Anteil der fehlerhaften Messungen eines Messinstruments unbekannt, sollte besonders bei kleinen Häufigkeiten oder Anteilen nur eine vorsichtige Interpretation der Ergebnisse erfolgen. Besonders bei kleinen Häufigkeiten und Anteilswerten kann ein kleiner Messfehler von Individualangaben das Aggregat um ein Vielfaches zu groß erscheinen lassen.

ANHANG I

Gemessen wird an n Personen, für die gilt X=k (mit k=1,2) das Merkmal $\tilde{X}$. Sei P($\tilde{X} = k|X=k$)= α_k und P($\tilde{X} \neq$ k|X=k)= $1 - \alpha_k$. Die Häufigkeiten $n|\tilde{X} = k \cap X = k$ und $n|\tilde{X} = k \cap X \neq k$ folgen daher jeweils einer Binomialverteilung mit dem Erwartungswert E($n|\tilde{X} = k \cap X = k$)= $n \cdot \alpha_k$ beziehungsweise E($n|\tilde{X} = k \cap X \neq k$)=$n \cdot (1 - \alpha_{k^c})$ und der Varianz $VAR\left(n\middle|\tilde{X} = k \cap X = k\right) = (n|X = k) \cdot \alpha_k \cdot (1 - \alpha_k)$ beziehungsweise $VAR\left(n\middle|\tilde{X} = k \cap X \neq k\right) = (n|X \neq 1) \cdot \alpha_{k^c} \cdot (1 - \alpha_{k^c})$.

Daher ist

$$\mathrm{E}(n|\tilde{X} = 1 \cap X = 1)+\mathrm{E}(n|\tilde{X} = 1 \cap X = 2)=E\left(n\middle|\tilde{X} = k\right) = (n|X = k) \cdot \alpha_1 + (n|X \neq k) \cdot (1 - \alpha_2).$$

Da die Varianzen unabhängig sind, gilt

$$VAR\left(n\middle|\tilde{X} = k \cap X = k\right) + VAR\left(n\middle|\tilde{X} = k \cap X \neq k\right) =$$

$$VAR\left(n\middle|\tilde{X} = k\right) = (n|X = k) \cdot \alpha_k \cdot (1 - \alpha_k) + (n|X \neq k) \cdot (1 - \alpha_{k^c}) \cdot \alpha_{k^c}.$$

ANHANG II

Da der Anteil $\frac{n|\tilde{X} = k}{n}$ einer mit konstantem Faktor $\frac{1}{n}$ normierten Häufigkeit $n|\tilde{X} = k$ entspricht, gilt

$$\frac{1}{n} \cdot E\left(n\middle|\tilde{X} = k\right) = E\left(\frac{n|\tilde{X} = k}{n}\right) = \left(\frac{n|X = k}{n}\right) \cdot \alpha_k + \left(\frac{n|X \neq k}{n}\right) \cdot (1 - \alpha_{k^c}) \text{ und}$$

$$\frac{1}{n} \cdot VAR\left(n|\tilde{X} = k\right) = \left(\frac{n|X = 1}{n^2}\right) \cdot \alpha_1 \cdot (1 - \alpha_1) + \left(\frac{n|X = 2}{n^2}\right) \cdot (1 - \alpha_2) \cdot \alpha_2.$$

ANHANG III

Es gilt $E\left(n\middle|\tilde{X}=1\right) = (n|X=1)\cdot\alpha_1 + (n|X=2)\cdot(1-\alpha_2) = (n|X=1)\cdot\omega$.

Dies lässt sich umformen zu

$$(n|X=1)\cdot\alpha_1 + (n|X=2)\cdot(1-\alpha_2) = (n|X=1)\cdot\omega \qquad \Big|\cdot\frac{1}{n|X=1}$$

$$\alpha_1 + \frac{(n|X=2)}{(n|X=1)}\cdot(1-\alpha_2) = \omega.$$

Für den Anteilswert gilt

$$E\left(\frac{n|\tilde{X}=1}{n}\right) = \left(\frac{n|X=1}{n}\right)\cdot\alpha_1 + \left(\frac{n|X=2}{n}\right)\cdot(1-\alpha_2) = \left(\frac{n|X=1}{n}\right)\cdot\omega.$$

Dies lässt sich ebenfalls umformen zu

$$\left(\frac{n|X=1}{n}\right)\cdot\alpha_1 + \left(\frac{n|X=2}{n}\right)\cdot(1-\alpha_2) = \left(\frac{n|X=1}{n}\right)\cdot\omega \qquad \Big|\cdot\frac{n}{n|X=1}$$

$$\alpha_1 + \frac{(n|X=2)}{(n|X=1)}\cdot(1-\alpha_2) = \omega.$$

LITERATURVERZEICHNIS

Biemer, Paul P./Stokes, Lynne S. *Approaches to the modeling of measurement error.* In: Biemer, Paul P./Groves, Robert M./Lyberg, Lars E./Mathiowetz, Nancy A./Sudman, Seymour. Measurement Errors in Surveys. New York 1991.

Dieckmann, Andreas. *Empirische Sozialforschung. Grundlagen Methoden Anwendungen.* 9. Auflage. Hamburg 2014.

Grabe, Michael. *Grundriss der Generalisierten Gauß'schen Fehlerrechnung.* Berlin 2011.

Faulbaum, Frank. *Total Survey Error.* In: Baur, Nina/Blasius, Jörg (Herausgeber). Handbuch Methoden der empirischen Sozialforschung. Wiesbaden 2014, Seite 439 ff.

Fischer, Gerhard H. *Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen.* Wien 1974.

Krug, Walter/Nourney, Martin/Schmidt, Jürgen. *Wirtschafts- und Sozialstatistik. Gewinnung von Daten.* 6. Auflage. München/Wien 2001.

Luhmann, Niklas. *Die Wissenschaft der Gesellschaft.* Frankfurt 1990.

Porst, Rolf. *Fragebogen.* 3. Auflage. Wiesbaden 2011.

Preisendörfer, Peter/Wolter, Felix. *Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys.* In: Public Opinion Quarterly. Band 78, Seite 126 ff.

Raab-Steiner, Elisabeth/Benesch, Michael. *Der Fragebogen.* 4. Auflage. Wien 2015.

Schnell, Rainer/Hill, Paul B./Esser, Elke. *Methoden der empirischen Sozialforschung.* 10. Auflage. München 2013.

Schumann, Siegfried. *Repräsentative Umfrage. Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren.* 4. Auflage. München 2006.

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktionsleitung: Juliane Gude
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Oktober 2020
Das Archiv älterer Ausgaben finden Sie unter www.destatis.de

Artikelnummer: 1010200-20005-4, ISSN 1619-2907