



Anonymisierung quantitativer Daten in der Kriminologie

1. Was sind quantitative Daten (in der Kriminologie)?

In den Sozialwissenschaften werden die in empirischen Forschungsvorhaben produzierten Daten in aller Regel in zwei verschiedene Bereiche abgegrenzt. So wird unterschieden zwischen qualitativen und quantitativen Daten. Der Sammelbegriff *quantitative Daten* umfasst dabei alle Daten, die in irgendeiner Weise auf die numerische Erfassung von Sachverhalten und/oder Zusammenhängen abzielen. Unter Zuhilfenahme von statistischen Methoden und Instrumenten werden so in der Regel vorab festgelegte Hypothesen überprüft.

Zu den in der empirischen Forschung weitläufig angewandten quantitativen Methoden zählen dabei vor allem standardisierte Befragungen, die schriftlich, telefonisch oder auch online durchgeführt werden können und mit Hilfe von vorrangig geschlossenen und Multiple-Choice-Fragen Daten zu einem bestimmten Thema bzw. einer bestimmten Forschungsfrage erheben. Daneben gehören aber auch Daten aus standardisierten Beobachtungen oder Experimenten zu den Methoden der quantitativen Datenerhebung.

In der deutschsprachigen Kriminologie stehen bei quantitativen empirischen Studien neben Aktenanalysen vor allem Datenerhebungen im Rahmen von Befragungen im

Fokus. Da gerade im Kontext der Erforschung von abweichendem und strafbarem Verhalten eine Vielzahl an personenbezogenen und höchstpersönlichen Daten erhoben und verarbeitet werden, kommt der gewissenhaften Anonymisierung dieser Daten folglich eine hohe Bedeutung zu – insbesondere (aber nicht nur) im Kontext einer etwaigen Nachnutzung der Daten.

2. Datenschutz als rechtliche Grundlage für die Anonymisierung

Ethisch empirische Forschungsarbeiten setzen voraus, dass mit den so produzierten Daten verantwortungsvoll umgegangen wird. Während das für alle empirisch erzeugten Daten gleichermaßen gilt, erhält der Datenschutz bei Forschungsarbeiten zu abweichendem oder strafbarem Verhalten eine besondere Brisanz. So enthalten viele der generierten Daten Angaben, die nach § 3 Abs.1 BDSG zu den personenbezogenen Daten, also „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person“ zählen. Daneben sind die erhobenen Daten häufig auch „personenbeziehbar“, sodass diese „eine bestimmte Person zwar nicht eindeutig oder unmittelbar identifizieren, [...] aber erlauben, die Identität der Person mit Hilfe anderer Informationen festzustellen“.¹ Aus diesem Grund zielt eine Anonymisierung der Forschungsdaten nach § 3 Abs.6 BDSG darauf ab, personenbezogene Daten so zu verändern, dass die einzelnen Angaben „nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“. Um diesen gesetzlichen wie forschungsethischen Vorgaben nachzukommen, soll die Anonymisierung der Daten im Forschungsprozess daher frühestmöglich erfolgen.

Aus forschungsethischen und datenschutzrechtlichen Gründen ist es erforderlich, dass die Forschungsteilnehmenden über den Sinn und Zweck der Datenerhebung sowie über den Umgang mit und die Verwendung der Daten umfassend aufgeklärt werden. Das Einverständnis zur Verarbeitung und Erhebung der Daten ist daher noch vor Beginn der betreffenden Datenerhebung von den Studienteilnehmenden schriftlich einzuholen (Stichwort: „informed consent“²). Sofern eine Nachnutzung der Daten

¹ Metschke/Wellbrock (2002), S. 19.

² Detailliertere Informationen zur Einwilligungserklärung („informed consent“) finden sich in den „FDM-Infos to go“ unter dem Stichwort „Informierte Einwilligung“.

vorgesehen ist, muss dies entsprechend bereits im Zuge der Datenerhebung bedacht und den Studienteilnehmenden kommuniziert werden. Voraussetzung für die Nachnutzung ist, dass das Einverständnis der Teilnehmenden – bevorzugt schriftlich – eingeholt wurde.

3. Vorgehen bei der Anonymisierung quantitativer Daten

Um die Identität der befragten Personen zu schützen und um einer etwaigen Re-Identifizierung der Befragten vorzubeugen, werden auch quantitative Forschungsdaten umfassend anonymisiert. Da insbesondere die in standardisierten Befragungen erhobenen Daten in aller Regel jedoch kaum personenbezogene und auch nur wenige personenbeziehbare Angaben abfragen, wodurch die Daten in vielen Fällen bereits „quasi-anonymisiert“ erhoben werden, ist die Anonymisierung quantitativer Daten deutlich einfacher zu vollziehen als beispielsweise bei sehr kontextintensiven qualitativen Interviewdaten. Wenn im Rahmen einer standardisierten quantitativen Befragung keinerlei personenbezogene und/oder personenbeziehbare Daten erhoben werden, sodass eine Identifizierung der Befragten in jedem Fall ausgeschlossen werden kann, handelt es sich um eine absolute Anonymisierung der Daten.

Insbesondere vor der Archivierung der Daten und auch vor dem Hintergrund einer etwaigen Nachnutzung muss dennoch auch bei quantitativen Daten sichergestellt werden, dass die Daten so anonymisiert wurden, dass auch durch eine mögliche Kombination verschiedener Angaben keine Identifikation der Befragten möglich ist. Da jede Form der Anonymisierung auch mit einem Informationsverlust verbunden ist, ist dabei stets genau zu prüfen, inwiefern die unternommenen Anonymisierungsmaßnahmen einer etwaigen Nachnutzung entgegenstehen könnten. Bei der Anonymisierung quantitativer Daten sind die folgenden Aspekte zu beachten.

3.1 Formale Anonymisierung

Bei der formalen Anonymisierung von quantitativen Daten werden personenbezogene Daten wie Name, Anschrift, Telefonnummer und E-Mailadresse, Berufsbezeichnungen oder sonstige persönliche Daten, die eine Identifizierung der Person ermöglichen könnten, unverzüglich entfernt. Hierzu zählt beispielsweise auch, dass die IP-Adressen

bei Onlinebefragungen im besten Fall gar nicht erst gespeichert und/oder im Nachgang zumindest sofort gelöscht werden. Außerdem erhalten alle Studienteilnehmenden von Beginn an eine numerische ID, die anstelle des Namens zur Identifikation der Person im Datensatz verwendet wird. Daneben gibt es auch personenbeziehbare Angaben, die auf eine etwaige erforderliche Anonymisierung hin geprüft werden müssen. So sollten – je nach Fragestellung und der zugrundeliegenden Population beispielsweise auch Angaben zum Beschäftigungsverhältnis und zur Berufsbezeichnung oder bei der Nennung von Institutionen – darauf geprüft werden, ob diese nicht ggf. eine Identifikation der befragten Person ermöglichen könnten und daher zwingend anonymisiert werden müssen. Folglich stellen vor allem Variablen mit offenen Angaben eine Herausforderung dar, weshalb es sich vielfach anbietet, diese aus dem Datensatz zu entfernen.³

3.2 Weitere Anonymisierungsstrategien

3.2.1 Aggregieren einzelner Werte und Kategorien

Durch die Aggregation einzelner, sehr seltener Werte einer Variablen kann einer möglichen Re-Identifizierung der befragten Personen entgegengewirkt werden. Dies ist insbesondere deshalb angeraten, weil entsprechende sensible Informationen (beispielsweise mit Blick auf das Alter, das Einkommen, die Herkunft oder die Muttersprache) unter Berücksichtigung weiterer im Rahmen der Studie erhobener Daten gegebenenfalls Rückschlüsse auf die betreffende Person ermöglichen könnten. Um solchen Fällen vorzubeugen, empfiehlt es sich die potenziell „problematischen“ Werte bzw. Kategorien der Variable zu aggregieren und diese in nach oben oder unten offene Kategorien zusammenzufassen bzw. umzucodieren.

Beispiel:

Eine Person aus einer Kleinstadt in Baden-Württemberg ist 106 Jahre alt und nimmt an einer Befragung teil. Da die befragte Person alleine aufgrund des außergewöhnlich hohen Alters und der höchstwahrscheinlich sehr geringen Population der über 100-jährigen Personen in der besagten Kleinstadt beinahe zweifelsfrei zu identifizieren wäre, wird die Angabe aus Datenschutzgründen in eine nach oben offene Kategorie

³ Vgl. Ebel/Meyermann (2015), S. 5.

(etwa „über 90 Jahre alt“) überführt, in der mehrere Fälle zusammengefasst werden können.⁴

3.2.2 Aggregieren aller Werte

Bei Variablen, deren Informationen unter datenschutzrechtlichen Gesichtspunkten insgesamt als problematisch eingestuft werden, können die gesamten Variablen recodiert werden. Dies kann mit einem großen Arbeitsaufwand einhergehen, sodass vorab geprüft werden sollte, ob dieser Schritt erforderlich ist.

Beispiel:

Die Teilnehmenden sollten im Rahmen der Umfrage die ersten drei Stellen der Postleitzahl angeben. Da auch dies unter Umständen (und in Kombination mit anderen Informationen) Rückschlüsse auf die befragten Personen zulassen kann, bietet es sich an, die erhobenen Werte auf das betreffende Bundesland oder zumindest den Landkreis zu aggregieren.⁵

3.2.3 Löschung von Variablen

Der mit Blick auf die Anonymisierung der Daten mit Abstand größte Eingriff, um eine Re-Identifizierung der befragten Personen zu verhindern, betrifft die Löschung ganzer Variablen. Da ein solcher Schritt im Vergleich zu den anderen, bereits vorgestellten Anonymisierungsoptionen den umfassendsten Informationsverlust mit sich bringt, sollte die Löschung ganzer Variablen nur erfolgen, wenn dies aus Datenschutzgründen unvermeidbar erscheint.⁶

4. Dokumentation der Anonymisierungsbemühungen

Im Zuge der Anonymisierung von empirischen Daten ist stets darauf zu achten, dass konsistent und sorgfältig gearbeitet wird. Hierzu gehört im Sinne der guten wissenschaftlichen Praxis auch, dass alle Entscheidungen und Codierungen in geeigneter Form dokumentiert werden müssen. Der Dokumentation sämtlicher methodischen

⁴ Vgl. Ebel (o.J.), S. 3.; Ebel/Meyermann (2015), S. 7.

⁵ Vgl. Ebel (o.J.), S. 4.; Ebel/Meyermann (2015), S. 8.

⁶ Vgl. Ebel (o.J.), S. 4.; Ebel/Meyermann (2015), S. 8 f.

Entscheidungen im Rahmen der zugrundeliegenden Studie kommt dabei insbesondere im Kontext einer Datenarchivierung und einer etwaigen Nachnutzung der Daten durch Dritte eine besondere Bedeutung zu.⁷

Literaturverzeichnis

Ebel, T./Meyermann, A. (2015): Hinweise zur Anonymisierung von quantitativen Daten. Frankfurt am Main: DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. 10.25656/01:21970

Metschke, R. & Wellbrock, R. (2002). Datenschutz in Wissenschaft und Forschung Verfügbar unter: https://userpage.fu-berlin.de/~makracht/mes/ws_2003/Unterlagen/Kursunterlagen/metadaten/mat_28_Datenschutz_Wissenschaft_Forschung.pdf [Letzter Aufruf: 16.4.2025]

Trixa, J./Ebel, T./Harzenetter, K. (2019). Hinweise zur Aufbereitung quantitativer Daten. Frankfurt am Main: DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. 10.25656/01:21967.

Weiterführende Literatur

Schwartmann, R./Jaspers, A./Lepperhoff, N./Weiß, S./Meier, M. (2022). Praxisleitfaden zum Anonymisieren personenbezogener Daten. Anforderungen, Einsatzklassen und Vorgehensmodell. Stiftung Datenschutz. Verfügbar unter: https://stiftung-datenschutz.org/fileadmin/Redaktion/Dokumente/Anonymisierung_personenbezogener_Daten/SDS_Studie_Praxisleitfaden-Anonymisieren-Web_01.pdf [Letzter Aufruf: 5.6.2025].

Stand: 28.7.2025

⁷ Eine Checkliste, in der alle nötigen Arbeitsschritte zur Vorbereitung quantitativer Daten für eine Archivierung und ggf. Nachnutzung aufgeführt sind, findet sich bei *Trixa/Ebel/Harzenetter* (2019), S. 5.