

# Hinweise für die Informationserfassung und -verwaltung und die statistische Auswertung von menschlichem Skelettmaterial

FRANK SIEGMUND

*Heinrich Heine Universität Düsseldorf, Institut für Geschichtswissenschaften III*

## Zusammenfassung

Der Beitrag entwickelt Empfehlungen für die Datenerfassung an menschlichem Skelettmaterial und die anschließende Auswertung der Daten. Es werden einige einfache Regeln in Erinnerung gerufen und einfache Wege und Werkzeuge für die Auswertung skizziert. Der Beitrag möchte eine Grundlage für die eingehende Diskussion dieses Themas in der Arbeitsgruppe bieten, damit im Ergebnis ein gemeinsamer Leitfaden entstehen kann. Dieser künftige Leitfaden kann allen zur Orientierung dienen, einen gemeinsam vereinbarten Mindeststandard formulieren und helfen, vergleichendes Arbeiten und den Austausch von Daten zu erleichtern. Der Beitrag versucht, wenige und einfache Regeln und Instrumente als Minimalstandard vorzuschlagen, ohne weiter gehende und aufwendigere Projekte auszuschliessen.

*Schlüsselwörter: Empfehlungen, gute Praxis, Datenerfassung, Datenverwaltung, statistische Auswertung*

## Abstract

The article develops recommendations for the data collection of human skeletal material and the subsequent analysis of it. Simple tips and tools will be outlined for the analysis and the evaluation of the data. The paper aims to provide to the working group a basis for the development of specific guidelines. These future guidelines could assist the researchers to formulate standards in order to compare and share their data. The paper attempts to propose few and simple rules and tools as a minimum standard, without excluding further and more complex projects.

*Key words: guidelines, best practice, data acquisition, data management, statistical analysis*

## Einleitung

Der folgende Text möchte das mit dem Workshop *“Preservation and representation of human bones and its importance for anthropological analysis”* am 31. Mai 2013 im IRM Bern begonnene, umfassendere Projekt in Richtung auf eine optimale und standardisierte Auswertbarkeit der Informationen begleiten. Der Text ist nicht als Feststellung gültigen Wissens oder als Lehrbuch gedacht, sondern enthält Hinweise und formuliert Thesen, damit die aktive Gruppe über die hier angeführten Aspekte frühzeitig diskutieren und die nötigen Entscheidungen treffen kann. Denn ein Erfassungskonzept, das die spätere Nutzung der Informationen bis ans Ende der üblichen Auswertungen mitdenkt, hat eine höhere Chance, sich langfristig zu bewähren. Nach den nötigen Diskussionen, Verbesserungen und Ergänzungen könnte der folgende Text die Grundlage für einen Teil des

resultierenden gemeinsamen Erhebungs- und Auswertungsleitfadens bilden.

## Datenerfassung eher konventionell denn technikgestützt

Die im Mai 2013 im IRM Bern zusammengekommene Arbeitsgruppe diskutiert die nähere Definition eines Aufnahmesystems für menschliches Skelettmaterial, damit Einzelprojekte effizienter arbeiten können und eine hohe Vergleichbarkeit von Informationen und ein unkomplizierter Austausch von Daten möglich sind. In diesem Zusammenhang liegt es nahe, zugleich auch an eine computergestützte Datenerfassung zu denken, d.h. an eine komplexe Datenbank samt Eingabemaske. Ohne ein solches Vorhaben grundsätzlich anzulehnen, sei davor gewarnt, es in den

Mittelpunkt des Vorhabens dieser Gruppe zu stellen. Denn die anhaltend schnelle Entwicklung auf dem Feld der IT zwingt dazu, ein solches bereits in der Erstellung aufwändige System anschliessend dauerhaft zu pflegen. Die dafür notwendige langfristig verfügbare Kapazität von IT-Spezialisten kann in Projektgruppen wie der jetzt Zusammengekommenen nur selten garantiert werden. Daher sollte man erwägen, ob nicht die Erfassung der Informationen bei der Arbeit am Knochen weiterhin mit "Papier und Bleistift" erfolgt, d.h. auf wohlüberlegten und standardisierten Formblättern, die anschliessend in eine einfache Tabellenkalkulation wie z.B. LibreOffice Calc oder Microsoft Excel übertragen werden. Es ist ein sicheres und m.E. letztlich effizienteres Vorgehen. Denn eine vergleichsweise einfache Datentechnik erfordert einen geringen Aufwand bei ihrer ersten Erstellung und ist bei den von aussen hereinkommenden Weiterentwicklungen der Betriebssysteme und Programme erfahrungsgemäss weniger anpassungsbedürftig. Gelänge es, sich gemeinsam auf geeignete leere Tabellen in einfachen und weit verbreiteten Standardformaten wie z.B. \*.ods, \*.xls oder \*.csv zu einigen, wäre ein Austausch der normierten Informationen und deren Einlesen in Datenbanken und auswertende Programme leicht und auch in absehbarer Zukunft ohne Aufwand möglich.

Für das Anlegen solcher Tabellen sei kurz an einige bewährte Regeln erinnert:

- Jeder Fall bildet eine Zeile.
- Jeder Fall benötigt eine eindeutige Identifizierung ähnlich einer Inventarnummer. Anhand dieser Identifizierungsnummer können später mehrere unterschiedliche Tabellen, die Informationen zu einem Individuum oder einem archäologischen Befund enthalten (z.B. Langknochenmasse, Schädelmasse, Zähne), zusammengefasst werden.
- Jede Spalte enthält Information entweder als Zahl oder als Buchstabe. Dabei sind Zahlen technisch stets vorzuziehen. Erscheint eine Kombination von Zahlen und Buchstaben notwendig, ist es in der Regel besser, für die Zahl eine Spalte und den Buchstaben eine separierte weitere Spalte vorzusehen. Beispiel: Wird eine Strecke gemessen als 43 mm, wird die Zahl "43" in der entsprechenden Spalte notiert. Soll vermerkt werden, dass der Wert nur ungenau beobachtet werden konnte, ist die Notation als "43?" oder "c. 43" technisch ungeschickt. Besser ist es, zwei Spalten vorzusehen, die erste mit der Zahl "43" und als zweite eine Textspalte z.B. mit der Bedeutung "circa".
- "Beobachtet Null" und "keine Beobachtung" sind unterschiedliche Informationen. Wenn das

Leerlassen einer Zelle bedeutet, dass eine Beobachtung nicht gemacht werden konnte, muss im anderen Fall die Beobachtung "null" ausdrücklich als Zahl Null eingegeben werden. Alternativ ist die Definition eines realiter nie vorkommenden Wertes wie z.B. "999" als Kodierung für "nicht beobachtet / nicht vorhanden" sinnvoll.

- Kommas oder Punkte für die Erfassung von Zehnteln erhöhen den Eingabeaufwand. Ein 43,5 cm langer Femur kostet bei der Erfassung vier Anschläge, der 435 mm lange Femur nur drei Anschläge.
- Das Ablegen von Informationen in kodierter Form ist weitaus effizienter als ausgeschriebene Texte. Kaum Jemand wird in eine Spalte "nicht vorhanden" schreiben, eine gleich bedeutende Kodenummer "9" ist schneller getippt als der Text "n.v."

Früh im Verlauf von Studien ist festzulegen, was im jeweiligen Datensatz eigentlich ein Fall ist. In vielen Fällen bildet das Skelettindividuum einen Fall. Bei der Erfassung anderer Informationen wie z.B. der Zähne hat es sich bewährt, jeden einzelnen Zahn oder Knochen als Fall aufzunehmen und in eine Tabelle zu übertragen. Beim Entwurf solcher Tabellen ist es wesentlich, von Anbeginn an alle Wünsche zur späteren Zusammenfassung und Zusammenführung der Daten mitzudenken und den einzelnen Fällen entsprechende Fallkennzeichen mit zu geben. Nur so können z.B. die zahnweise erhobenen Informationen später zu einem ganzen Gebiss zusammengefasst werden und die Kennzahlen das ganzen Gebisses wiederum anderen Informationen vom gleichen Skelett gegenübergestellt werden.

Für die spätere Auswertung ist es wesentlich, zu allen Merkmalen, die erhoben werden, festzuhalten, ob sie an dem betreffenden Individuum überhaupt beobachtbar waren. Da dies im Workshop eingehend insbesondere von Chr. Papageorgopoulou ausgeführt wurde, kann hier auf ihren Beitrag verwiesen werden. Aus Sicht des Statistikers sind die beliebten farbig ausgemalten Skelettschemata, in denen die noch vorhandenen Reste eines Individuums markiert werden, zwar anschaulich, aber für die weitere Auswertung bedarf es zusätzlich einer elektronisch lesbaren Erfassung dieser Informationen.

### Pilotstudie

Jedes grössere Erfassungsvorhaben sollte mit einer Pilotstudie beginnen, d.h. mit einer Erfassung einer

kleinen Untermenge an Fällen, die anschliessend wie in der geplanten Studie auch in der EDV erfasst und statistisch ausgewertet wird. Inhaltliche Ergebnisse werden dabei nicht abgestrebt, sondern Erfahrungen mit den Formblättern, Tabellen und Auswertungsprogrammen sowie zum benötigten Zeitaufwand. Am Ende einer Pilotstudie können Schwächen des Versuchsdesigns bereinigt werden, bevor Korrekturen angesichts der bereits erfassten Daten sehr aufwendig werden.

### Dokumentation

Die Datenerhebung mit allen Merkmalsdefinitionen, Konventionen und Kodierungen und die Übertragung in ein EDV-System sollten frühzeitig und fortlaufend protokolliert werden. Gerade ad-hoc-Ergänzungen an einem Erfassungssystem während der Datenerfassung, die nicht notiert werden, da sie in diesem Stadium eines Projekts auswendig gewusst sind, neigen dazu, nach dem vorläufigen Abschluss eines Projekts überraschend schnell vergessen zu werden. Nur eine frühe und fortwährende Dokumentation bewahrt davor.

Wer ein bewährtes Erfassungssystem von Dritten übernehmen kann, spart viele schon gemachte und nunmehr ausgemerzte Fehler, kann auf eine Pilotstudie verzichten und die nötige Dokumentation übernehmen. Unabhängig von der Frage eines elaborierten gemeinsamen EDV-Systems liegt bereits hierin ein spürbarer Gewinn der angestrebten gemeinsamen Konvention zur Datenerfassung.

### Beobachterfehler

In vernünftigem Rahmen, d.h. begrenzt auf den Fokus der späteren Studie, sollte eine Datenerhebung auch die Ermittlung des Beobachterfehlers ermöglichen. Je nach Studie und Fragestellung müssen dazu Stichproben wiederholt aufgenommen werden. Den eigenen Erhebungsfehler (*“intra-observer error”*) sollte jeder Wissenschaftler von Zeit zu Zeit selbst erfassen und dokumentieren. Bei Routineuntersuchungen müssen Beobachterfehler nicht an jeder Serie erhoben werden, sondern es genügt, sie im jeweiligen Labor einmal und dann in grösseren Abständen gelegentlich zu erheben. Nehmen mehrere Wissenschaftler an der Primärerhebung von Daten teil oder werden gepoolte Daten gemeinsam ausgewertet, sollte auch eine gezielte Studie zum möglichen Fehler zwischen unterschiedlichen Erfassern erfolgen (*“inter-observer*

*error”*). Bei der Ermittlung des Intra- oder Interobserver Errors ist die Güte der erreichten Aussage nicht abhängig vom relativen Anteil an Wiederholungsmessungen an der gesamten Serie, sondern von der absoluten Anzahl an Wiederholungsmessungen.

Ergänzend sei darauf hingewiesen, dass sich eine wirksame Qualitätskontrolle bei den 14C-Labors etabliert hat, deren Vorgehen als Modell genutzt werden kann. Dort haben sich mit dem Aufkommen einiger ausnehmend preiswert datierender Labors – der Begriff Dumpingpreise lief um – viele qualitätsbewusste Labors weltweit zusammengeschlossen zu einer Gruppe, die sich gegenseitig kontrolliert. Blindproben identischen Ausgangsmaterials werden verschickt, gemessen und die Ergebnisse zusammengeführt, ausgewertet und veröffentlicht. Ähnliches wäre auch mit menschlichen Skeletten denkbar und anzustreben. Im Nebeneffekt würde man wertvolle Informationen über den üblichen *Inter-observer error* gewinnen. Es wird angeregt, in der zusammengekommenen Gruppe parallel zum gemeinsam akkordierten Erfassungsschema auch eine gemeinsame Qualitätskontrolle zu planen.

### Fehlerprüfung und Datensicherheit

Die Prüfung auf Erfassungs- und Eingabefehler ist notwendiger Bestandteil jeder Datenerfassung. Sie muss zeitnah unmittelbar nach der Erfassung erfolgen. Eine einfache und zugleich effiziente Form der Datenprüfung ist es, die Daten einer einfachen statistischen Auswertung zu unterziehen. Tippfehler, die bei Textvariablen vorkommen, fallen bei einer Auszählung nach Häufigkeit als Einzelfälle auf. Typische Fehler bei der Eingabe von Zahlen fallen später oft als deutlich zu kleine oder zu große Werte auf, mit Histogrammen lässt sich das Vorhandensein solcher Fehler schnell aufdecken. An diese erste Prüfung sollte sich eine weitere, etwas tiefere Überprüfung anschließen, die auf dem Zusammenhang zwischen den verschiedenen Informationen eines Skelettes beruht. Eine relativ lange Tibia sollte mit einer relativ langen Ulna kombiniert sein, ein relativ kurzer Femur lässt eine relativ kurze Tibia erwarten usw. Wenn sich die Gruppe auf gemeinsame elektronische Erfassungsblätter einigt, lassen sich mit vertretbarem Aufwand Skripte anlegen, die solche Fehlerprüfungen weitgehend automatisch durchführen und in den Datensätzen die betreffenden Zahlen als erneut zu prüfen markieren.

Die erfassten Daten sind kostbar, sie sollten regelmäßig auf externen Medien gespeichert werden. Wird statt einer Sicherungskopie, die stets die ältere Sicherung überschreibt, ein System mit zwei externen

Medien gewählt und zwar so, dass nie die letzte, sondern stets nur die vorletzte Sicherung von der aktuell durchgeführten Sicherung überschrieben wird, entsteht eine hohe Sicherheit gegen Ausfälle der Technik. Ein besonders sensibler Moment in der Datensicherheit stellt das Ende von Projekten dar, da schon bald danach erfahrungsgemäss die Sorge um die Datensicherheit deutlich nachlässt. Da die Daten aber für Folgeprojekte, Überprüfungen oder Kooperationen wichtig bleiben, sollte rechtzeitig über deren weitere Aufbewahrung und Pflege über das Ende des aktuellen Vorhabens hinaus nachgedacht werden. Die sicherste Form der Nachhaltigkeit ist dann die Publikation der Daten, die zugleich im Gegensatz zur beliebten Weitergabe auf Nachfrage ihre Zitierfähigkeit herstellt.

### **Programme für die statistische Auswertung**

Die weit verbreiteten Tabellenkalkulationsprogramme eignen sich für die Datenerfassung und auch für deren Verwaltung. Die meisten Tabellenkalkulationen beinhalten die Möglichkeit, zu sortieren und Filter zu setzen, d.h. fragestellungsorientiert Untermengen eines Datensatzes zu bilden, und einfache Kennzahlen zu errechnen wie z.B. Mittelwert und Standardabweichung. Bei umfassenderen statistischen Auswertungen geraten Tabellenkalkulationsprogramme jedoch schnell an die Grenzen ihrer Möglichkeiten beziehungsweise das Arbeiten mit Ihnen wird zunehmend ineffizient. Die Anwendung von Statistikpaketen wie z.B. SPSS oder dem Open-Source-Programm R ist empfehlenswert, der Anfangsaufwand des Erlernens zahlt sich schnell aus. Wenn das mächtige und relativ leicht erlernbare, aber teure Bezahlprogramm SPSS aus finanziellen Gründen unerreichbar ist und der Lernaufwand für das kostenlos verfügbare, sehr leistungsfähige Programm R – kombiniert z.B. mit dem R-Commander – zu hoch erscheint, sollte man kostenfreie Programme wie PAST, Gnumeric oder PSPP evaluieren; sie leisten auf jeden Fall mehr als das weit verbreitete Excel.

### **Statistische Auswertung: gemessene Werte**

Die grundlegenden statistischen Auswertungen sind nicht schwierig. Da immer die gleichen oder ähnliche Informationen an einander ähnlichen Objekten und Serien erhoben werden, lässt sich ein klarer Kanon von statistischen Parametern und Methoden benennen, der zu Anwendung kommen sollte. Bei gemessenen Werten

wie z.B. den Längen oder Durchmessern von Knochen sollten folgende zusammenfassende Kennzahlen ermittelt werden: Anzahl der Beobachtungen, Mittelwert und Standardabweichung. Oft werden noch das beobachtete Minimum und Maximum einer Serie hinzugefügt, doch ist dies letztlich eine weitgehend wertlose Information, da sie vor allem von den Ausreissern bestimmt wird. Sollte sich die Datenreihe bei einem entsprechenden Test, z.B. dem Kolmogorov-Smirnov-Test auf Normalverteilung, als nicht normalverteilt erweisen, sind statt Mittelwert und Standardabweichung oder ergänzend der Median als Zentralmass und der Interquartilabstand als Streuungsmass angemessen. Ob eine untersuchte normalverteilte Datenreihe in ihren Kennzahlen signifikant von anderen Datenreihen abweicht, oder ob sie aus der gleichen Grundgesamtheit stammt, testet man mit dem t-Test. Liegt keine Normalverteilung vor, was auch bei biologischen Daten häufig der Fall ist, dienen parameterfreie Tests zur Prüfung auf signifikante Unterschiede; hier bietet sich der Kruskal-Wallis-Test an.

### **Statistische Auswertung: multivariate Verfahren**

Erst nach der Basisauswertung einer Serie sollten komplexere Verfahren angewendet werden. Deren Wahl hängt von den vorhandenen Informationen und vor allem von der Fragestellung ab. Als Einstieg in die Überlegungen zum Vorgehen können zwei Fragen dienen: (1) Erwartet man das Vorliegen von Gruppen oder von Populationen? (2) Folgen die Beobachtungen einem linearen Modell oder einem unimodalen Modell?

Ein gutes Beispiel für die Erwartung von Gruppen wäre das biologische Geschlecht, bei dem man von zwei einander unterscheidenden Gruppen ausgeht, ohne dass es regelhaft eine stark besetzte fließende Übergangszone zwischen diesen Gruppen gibt. Erwartet man hingegen zwischen Serien von Individuen eher fließende Übergänge bei deren Unterschieden, die sich z.B. durch unterschiedliche Umwelteinflüsse, Arbeitsbelastungen o.ä. ergeben haben, sollte man dem Populationskonzept folgen und entsprechende Verfahren wählen.

Ein gutes Beispiel für lineare Zusammenhänge ist die Körperhöhe. Für alle Langknochen erwarten wir einen weitgehend linearen Zusammenhang zwischen der Länge des Knochens und der Körperhöhe. Je kürzer der Knochen, desto geringer die Körperhöhe und umgekehrt. Ein einprägsames Beispiel für ein uni-

modales Modell ist das Körpergewicht in Bezug zum Lebensalter von Individuen. Wie die Alltagserfahrung zeigt, wächst das Körpergewicht eines Individuums nach der Geburt mit dem Lebensalter an; die meisten Menschen erreichen irgendwann in ihrem Erwachsenenleben ein maximales Körpergewicht und werden danach mit weiter zunehmendem Alter wieder etwas leichter. In Bezug zum Lebensalter folgt das Körpergewicht dem unimodalen Modell.

Möchte man Populationen untersuchen und erwartet vorwiegend lineare Zusammenhänge, sind Faktorenanalysen geeignete Verfahren. Sie ordnen Fälle und Merkmale und zeigen multifaktorielle Zusammenhänge auf. Möchte man Populationen untersuchen und erwartet vorwiegend unimodale Zusammenhänge, ist die Korrespondenzanalyse das geeignete multivariate Verfahren.

Bei der Untersuchung von biologischen Gruppen werden normalerweise zwei unterschiedlich gerichtete Fragestellungen verfolgt: Im einen Fall sollen bereits bekannte Gruppen reproduziert werden, in anderen Fall sollen noch nicht bekannte Gruppen gefunden werden. Ein in der Anthropologie geläufiger Fall für das Reproduzieren bereits bekannter Gruppen ist die Geschlechtsbestimmung. Das geeignete statistische Verfahren ist dann eine Diskriminanzanalyse: An einer Referenzserie geschlechtsbekannter Individuen wird untersucht, ob ein oder meist mehrere gemessene Werte tauglich sind, diese vorab bekannte Gruppierung nachzuvollziehen. Ist dies aufgrund der eingegebenen Informationen mit brauchbarer Trennschärfe der Fall, kann die gewonnene Formel auch auf andere Serien übertragen und zur Geschlechtsbestimmung genutzt werden. Sollen in einem Material noch unbekannte Gruppen nachgezeichnet werden, sind Clusteranalysen die Verfahren der Wahl. Sie fügen einzelne Fälle zu Gruppen einander ähnlicher Individuen zusammen. Es steht eine grosse Anzahl unterschiedlicher Clusterverfahren zur Auswahl. Normalerweise sind hierarchische Clusterverfahren für anthropologische Daten und Fragestellungen eine gute Wahl. Unter ihnen ergibt ein Single-Linkage-Clustering meist weniger überzeugende Ergebnisse. Andere hierarchische Clusterverfahren wie Complete-Linkage, Average-Linkage oder Ward's Method sollten jeweils Ergebnisse erbringen, die zu denen der Alternativen sehr ähnlich sind, weshalb die Wahl des speziellen Verfahrens innerhalb dieser Familie nicht sehr wesentlich ist. Insgesamt ist bei allen multivariaten Verfahren die Fragestellung und Modellbildung ein wesentlicher Schritt, und meist ist die gut begründete Auswahl der in die Berechnungen eingehenden Fälle und Merkmale weitaus entscheidender als manches statistische Detail.

### Statistische Auswertung: gezählte Häufigkeiten

Viele Beobachtungen in der Anthropologie sind jedoch keine Messwerte, sondern Ja-Nein-Beobachtungen (Knochenbruch vorhanden / nicht vorhanden) oder Beobachtungen, bei denen nach wenigen Kategorien klassifiziert wird, z.B. die Nahtverschlüsse am Schädel in die Zustände Null bis Vier oder die Altersbestimmung Erwachsener in die üblichen Klassen adult, matur und senil. Solche Informationen resultieren bei der Auswertung auf Populationsniveau in gezählten Häufigkeiten. Um einen schnelleren Überblick zu gewinnen und Zahlen besser vergleichen zu können, werden zu diesen beobachteten Häufigkeiten normalerweise die Prozentwerte berechnet, um die untersuchte Population mit den Werten anderer Populationen vergleichen zu können.

Aus statistischer Sicht ist hierbei wesentlich, dass stets auch die tatsächlich beobachteten Häufigkeiten nachgewiesen und publiziert werden, nie nur die Prozentwerte allein. Da aus taphonomischen und anthropologischen Gründen die Bezugssumme von Prozentangaben häufig wechselt, muss diese jeweils auch klar ausgewiesen werden – was wiederum bereits bei der Befunderhebung am Skelett zu berücksichtigen ist. Das mag mühsam sein, aber nur so ergeben sich solide Grundlage für eine saubere Statistik und tragfähige wissenschaftliche Aussagen. Ist das Problem der Grundgesamtheit bedacht und gelöst, sind Prozentwerte einfach zu berechnen und sehr anschaulich. Jeder Bearbeiter wird ein gewisses Gefühl für die Belastbarkeit von Aussagen in Prozentwerten haben. Beruhen sie auf vielen Beobachtungen, erscheint der Prozentwert als verlässlich und aussagekräftig, beruht der Prozentwert auf wenigen Beobachtungen, ist er zwar rechnerisch richtig, aber erscheint als nur unsichere Basis von weitergehenden Aussagen. Für dieses Problem der gefühlten Aussagekraft von Prozentzahlen bietet die Statistik eine objektive Hilfe, nämlich Konfidenzintervalle. Das Konfidenzintervall besagt, innerhalb welcher Spanne mit hoher Wahrscheinlichkeit der wahre Prozentwert in einer Population liegt, wenn die beobachtete (meist kleine) Serie als Stichprobe aus dieser Population gesehen wird. Drei kariöse Zähne bei zehn beobachtbaren Zähnen ergeben einen Anteil von 30 %, ebenso ergibt sich ein Anteil von 30 % bei sechzig kariösen Zähnen aus 200 beobachtbaren Zähnen. Im Fall 3 aus 10 beträgt das Konfidenzintervall 6,67 bis 65,25, d.h. die Hintergrundpopulation weist eine Kariesfrequenz zwischen etwa 7 % und 65 % auf. Im Fall 60 aus 200 beträgt das Konfidenzintervall 23,73 bis 36,86, d.h. die Hintergrundpopulation weist eine Kariesfrequenz

zwischen etwa 24 % und 37 % auf. Solche Konfidenzintervalle lassen sich in gedruckten Tabellenwerken schnell nachschlagen oder mit Hilfe einer Tabellenkalkulation leicht errechnen. Sie veranschaulichen eindrucksvoll, wie unsicher die aus kleinen Stichproben resultierenden Werte sind. Da wirklich grosse Serien, die zu spürbar engeren Konfidenzintervallen führen, in der Anthropologie selten sind, machen die Intervalle zugleich deutlich, dass das Argumentieren mit Prozentwerten zwar anschaulich ist, aber meist keine verlässliche Grundlage für wissenschaftlich tragfähige Aussagen bietet.

Ein angemesseneres Vorgehen ist in solchen Fällen die Analyse der tatsächlich beobachteten Häufigkeiten, nicht der Prozentwerte, und ihr Vergleich mit einer Referenz- oder Vergleichspopulation mit Hilfe des Chi-Quadrat-Verfahrens und des Chi-Quadrat-Tests. Dieses Vorgehen erbringt verlässliche Aussagen darüber, ob sich die untersuchte Population statistisch signifikant von einer anderen Serie unterscheidet oder ob die beobachteten Unterschiede noch im Rahmen der üblichen Zufallsstreuung liegen. Das Chi-Quadrat-Verfahren ist in seiner Umsetzung nicht kompliziert und kann mit Hilfe eines Tabellenkalkulationsprogramms schnell und zuverlässig durchgeführt werden. Daher kann der möglicherweise schwierige Zugriff auf Statistikprogramme auch nicht als Begründung dienen, auf Tests und Signifikanzangaben bei Populationsvergleichen zu verzichten. Obwohl solche Tests letztlich leicht umzusetzen sind und Statistik heutzutage im Rahmen eines wissenschaftlichen Studiums normalerweise Bestandteil der Ausbildung ist, steht es jedem sich unsicher Fühlenden frei, zusätzlich externen Rat und Expertise einzuholen.

### Ausblick

Die gemeinsame Entwicklung einer standardisierten Datenerfassung kann in der Folgezeit viel Arbeit sparen, da dann getestete und in der Praxis bewährte Formulare zur Verfügung stehen und die Menge der sonst anzuschliessenden Überprüfungen, Korrekturen und Dokumentationen deutlich reduziert wird. Dabei ist die Erhaltung und Beobachtbarkeit von Merkmalen stets mit zu erfassen, da deren Kenntnis für jede Auswertung wesentlich ist. Die Einigung auf einen gemeinsamen Datenstandard würde es erlauben, Fehlerprüfungen zumindest teilweise zu automatisieren und Daten zu Vergleichszwecken leichter auszutauschen oder im Sinne von *Open data* zitierfähig zu publizieren.

Da Studien ohne eine saubere statistische Auswertung kaum noch eine Chance auf eine gut

platzierte Publikation haben, bietet es sich an, das angestrebte Standardprotokoll um eine Standardauswertung zu ergänzen. Eine mit dem neuen Protokoll erfasste, reale und typische Serie könnte modellhaft mit geeigneten statistischen Verfahren ausgewertet werden, um ergänzend zum Protokoll als Vorlage und Referenz auch für einen statistischen Mindeststandard zu dienen.

### Dank

Ich danke den Organisatorinnen des Workshops für die Einladung und für die perfekte Organisation des Treffens, und allen Teilnehmerinnen und Teilnehmern für die anregenden Diskussionen in kollegialer Atmosphäre. Christina Papageorgopoulou danke ich für wertvolle fachliche Hinweise zur Endfassung dieses Textes.

#### Adresse:

PD Dr. Frank Siegmund  
Heinrich Heine Universität Düsseldorf  
Institut für Geschichtswissenschaften III  
Universitätsstraße 1  
40225 Düsseldorf  
Deutschland  
E-mail: mail@frank-siegmund.de