

Tags: archives évaluation,

## **Die Bildung repräsentativer Stichproben bei Massenakten. Zur Methode und Praxis der Bewertung mittels Zufallsauswahl**

Die Bewertung von Massenakten beschäftigt Archivarinnen und Archivare seit den 1950er Jahren intensiv. Allzu lange wurden Stichproben anhand von Kriterien gebildet, bei denen von falschen Voraussetzungen in Bezug auf deren Qualität ausgegangen wurde. Sie stellen somit eher «Samples der Hilflosigkeit» dar, wie Robert Kretzschmar sie einmal genannt hat (Kretzschmar Robert: Aussonderung und Bewertung von sogenannten Massenakten. Erfahrungen der staatlichen Archivverwaltung Baden-Württemberg. In: Historische Überlieferung aus Verwaltungsunterlagen. Zur Praxis der archivischen Bewertung in Baden-Württemberg. Hg. Robert Kretzschmar, Stuttgart 1997, S. 103–118, S. 111). Einen Meilenstein in der theoretischen Auseinandersetzung und der praktischen Anwendung stellt die 2001 erschienene Dissertation von Matthias Buchholz dar (Buchholz Matthias: Überlieferungsbildung bei massenhaft gleichförmigen Einzelfallakten im Spannungsverhältnis von Bewertungsdiskussion, Repräsentativität und Nutzungsperspektive. Köln 2001).

Buchholz hat sich mit der Überlieferungsbildung bei massenhaft gleichförmigen Einzelfallakten befasst. Er legt theoretisch und empirisch dar, dass im Bereich der Auswahlverfahren einzig die Zufallsauswahl eine repräsentative Stichprobenauswahl erzeugt. Wenden wir eine Klumpenstichprobe, wie etwa eine Buchstabenauswahl des Buchstabens B, an oder bilden wir eine systematische Auswahl, wie sie jedes 10. Dossier darstellt, haben wir zwar relativ hohe Übernahmequoten von ca. 10%, sind jedoch hinsichtlich der Auswertungsoffenheit eingeschränkt. Eingeschränkt in einem kaum einschätzbaren Mass.

Zur Illustration: Wird der Buchstabe B einer Dossierserie übernommen, welche nach Nachnamen geordnet ist, so sind in dieser Auswahl die Nationalitäten nicht in gleichem Masse vertreten, wie sie es in der Grundgesamtheit sind. Denn die Verteilung der Nachnamen über das Alphabet präsentiert sich nicht in allen Kulturräumen auf dieselbe Weise. Griechinnen und Griechen etwa haben signifikant häufiger Nachnamen, welche mit dem Buchstaben X beginnen, Türkinnen und Türken etwa haben im Vergleich zum deutschsprachigen Raum weitaus häufiger Nachnamen mit Ypsilon. Folglich sind sie in einer Stichprobenauswahl, die auf dem Buchstaben B basiert, untervertreten. Dies hat weitreichende Konsequenzen für die Auswertungsmöglichkeiten der Stichprobenauswahl. Interessiert zum Beispiel die Frage, ob Türkinnen und Türken in der Schweiz andere Automarken als Schweizerinnen und Schweizer bevorzugen, kann diese Frage mit der Stichprobenauswahl B nicht adäquat beantwortet werden. Man wird eine hohe Fehlerquote haben, wobei sich nicht genau bestimmen lässt, wie hoch die Fehlerquote ist.

Ähnlich ungenaue Ergebnisse wird die quantifizierende Forschung erhalten, wenn beispielsweise bestimmte Geburtsmonate übernommen werden. Der wohl ungünstigste Monat dürfte der Januar sein, werden doch von vielen Behörden Personen, welche ihr Geburtsdatum nicht angeben können oder wollen, mit dem Geburtsdatum 1. Januar versehen. In der wohl nachvollziehbaren Annahme, dass es sich dabei meist um ausländische Personen wie etwa Flüchtlinge handelt, werden diese mit hoher Wahrscheinlichkeit in meiner Stichprobe Januar übervertreten sein. Damit werden nicht nur alle Fragestellungen betreffend Geburtsdatum keine quantifizierbaren Resultate ergeben, sondern auch alle damit in irgendeiner Form zusammenhängenden Fragestellungen, von denen sich oft deren Kausalitäten nicht erschliessen lassen.

Diese Beispiele sollen ansatzweise verdeutlichen, welche Qualität man sich einhandelt, bewertet man Massenakten mittels Zufallsauswahl: Die Auswahl bleibt auswertungsoffen, weil alle Merkmale der Grundgesamtheit in der Stichprobe in gleichem Mass enthalten sind. Dies ist der Fall, weil jede Einheit über dieselbe Wahrscheinlichkeit verfügt, in die Stichprobe zu gelangen. Es wird mit der Zufallsauswahl – weil sie, wie das Wort sagt, auf dem Zufall basiert – verunmöglicht, der Grundgesamtheit innewohnende Strukturen oder Eigenschaften verzerrt zu überliefern, da kein Element bewusst oder unbewusst bevorzugt oder benachteiligt wird.

## **Die Berechnung des Stichprobenumfangs**

Die Formel, welche für die Auswahl mittels Zufall angewendet wird, basiert auf der Normalverteilung, welche wiederum einen zentralen Bestandteil der Wahrscheinlichkeitstheorie darstellt. Die Normalverteilung wird nach ihrem Entdecker und gemäss ihrer Form auch Gauss'sche Glockenkurve genannt (C. F. Gauss, 1794).

Sie basiert auf dem Erfahrungswissen, dass bei vielen Ereignissen, welcher Art auch immer, eine bestimmte Anzahl von Ereignissen sich symmetrisch um den Mittelwert gruppieren und umso seltener werden, je stärker sie vom Mittelwert abweichen. An einem Beispiel lässt sich das veranschaulichen: Auf einer Hühnerfarm mit sehr vielen Hühnern werden eine Woche lang alle gelegten Eier gewogen. Ein Ei wiegt im Durchschnitt 50 Gramm, da die meisten Eier 50 Gramm wiegen, sich also in der Grafik um den Mittelwert 50 gruppieren. Einige Eier wiegen 45 Gramm, andere 55 Gramm. Jedoch ist die Wahrscheinlichkeit klein, dass ein Ei 35 Gramm oder 65 Gramm wiegt. Je grösser die Abweichung in Gramm, je kleiner ist die Wahrscheinlichkeit, vom Durchschnittsgewicht abzuweichen. Dies ist die Beschreibung der Normalverteilung.

### *Das Konfidenzintervall*

Der Ausdruck dafür, wie viele Eier innerhalb eines bestimmten Bereichs um den Mittelwert gruppiert sind, wird Konfidenzintervall genannt. Bei einem Konfidenzintervall von 95 haben 95 von 100 Fällen die Eigenschaft des Mittelwerts mit einer bestimmten Abweichung.

### *Die Fehlertoleranz*

Die bestimmte Abweichung vom Mittelwert wird Fehlertoleranz genannt. Eine Fehlertoleranz von 2 Prozent und ein Konfidenzintervall von 95 Prozent sagen aus, dass 95 Prozent der Eier plus/minus 2 Prozent um den Mittelwert liegen, also 95 von 100 Eiern ein Gewicht zwischen 49 und 51 Gramm aufweisen.

Konfidenzintervall und Fehlertoleranz bestimmen den Grad der Repräsentativität. Je näher das Konfidenzintervall bei 100 liegt und je kleiner die Fehlertoleranz angelegt wird, je mehr Elemente wird meine Stichprobe umfassen. Werte ich die so erhaltene Stichprobe aus, wird die Abweichung des Fehlers umso kleiner sein, je grösser die Stichprobe ist. Die Grösse der Abweichung wird klar definiert über das Konfidenzintervall und die Fehlertoleranz.

Die Normalverteilung lässt sich in einer Formel ausdrücken<sup>1</sup>. Die Formel setzt sich zusammen aus der Grundgesamtheit, dem Konfidenzintervall, der Fehlertoleranz, der Merkmalsverteilung innerhalb der Grundgesamtheit und einem Korrekturfaktor.

Die *Grundgesamtheit* wird  $N$  genannt. Wenn für das *Konfidenzintervall*  $t$  eingesetzt wird, ergibt das bei einem 95%igen Konfidenzintervall den Tabellenwert 1,96, der aussagt, dass in der Gaus'schen Glockenkurve 95% der schraffierten Fläche innerhalb der Intervallgrenzen der Abweichung liegen. Für ein Konfidenzintervall von 90 Prozent lautet der Tabellenwert 1,645. Die *Fehlertoleranz* wird mit  $d$  angezeigt und wird beispielsweise mit 5% festgesetzt. Das ergibt einen Wert von 0,05, wenn 100% 1 ist. Die *Merkmalsverteilung* wird mit  $p$  und  $q$  ausgedrückt.  $P$  und  $q$  ergeben zusammen immer 100% bzw. 1. Wenn für das Merkmal Geschlecht der männliche Anteil  $p$  und der weibliche Anteil  $q$  sind, könnte das beispielsweise bedeuten, dass  $p$  den Wert 0,6 hat und  $q$  den Wert 0,4, also 60% Männer und 40% Frauen in der Grundgesamtheit vertreten sind. Ist über die Verteilung nichts bekannt oder handelt es sich wie in den Fällen aus dem archivischen Alltag um eine Vielzahl unterschiedlicher Merkmale, wird in der Statistik von einer Gleichverteilung ausgegangen.  $P$  und  $q$  werden mit dem Wert 0,5 – also 50% – versehen. Damit wird die Fehlerwahrscheinlichkeit auf dem kleinstmöglichen Niveau gehalten, da innerhalb einer Spannbreite von 0 bis 100 Prozent ein Wert von 50% den Mittelwert darstellt. Man erhält damit den grösstmöglichen Stichprobenumfang. Die Stichprobenuntersuchung erhält in keinem Fall eine geringere Wahrscheinlichkeit als gefordert. Den *Korrekturfaktor* benötigt man, um den sich aus dem Bruch ergebenden Wert  $n$  für unendlich grosse Grundgesamtheiten zu berichtigen, damit er auf die in Archiven anzutreffenden endlich grossen Grundgesamtheiten anwendbar ist. Der Korrekturfaktor lässt sich mit dieser Formel berechnen:

$$\text{Korrekturfaktor} = \sqrt{\frac{N-n}{N-1}}$$

Die Formel für die Berechnung des Stichprobenumfangs lautet:

$$\text{Stichprobenumfang} = \text{Korrekturfaktor} * \frac{\text{Konfidenzintervall}^2 * \text{Varianz}}{\text{Fehlertoleranz}^2}$$

Mit den oben eingeführten Variablen sieht sie so aus:

$$\text{Stichprobenumfang} = \sqrt{\frac{N-n}{N-1}} * \frac{t^2 * (p * q)}{d^2}$$

Dabei gilt:

$$n = \frac{t^2 * (p * q)}{d^2}$$

Füllt man die Formel für n mit den Zahlen, welche sich auf ein 95%iges Konfidenzintervall und eine 5%ige Fehlertoleranz beziehen, sieht sie wie folgt aus:

$$n = \frac{(1,96)^2 * (0,5 * 0,5)}{0,05^2}$$

Capture D'Écran 2019 06 09 À 20 35 35

berechnet man diese Formel, bekommt man das Resultat n = 384. Geht man von der Grundvoraussetzungen aus, dass das Konfidenzintervall 95% beträgt und die Fehlertoleranz 5% lässt sich die ausführliche Formel wie folgt vereinfachen:

$$\text{Stichprobenumfang} = \sqrt{\frac{N-384}{N-1}} * 384$$

Capture D'Écran 2019 06 09 À 20 39 12

Mit dieser Formel lässt sich der Stichprobenumfang von allen beliebigen Grundgesamtheit berechnen, wenn mit einem Konfidenzintervall von 95% und einer Fehlertoleranz von 5% gearbeitet wird. In der Statistik sind diese beiden Werte sehr gebräuchlich§§

Grundgesamtheit	Stichprobengrösse	Grundgesamtheit	Stichprobengrösse
800	277 (34,6%)	3000	359 (12%)
1000	302 (30,2%)	4000	365 (9,1%)
1200	317 (26,4%)	5000	369 (7,4%)
1400	327 (23,4%)	6000	372 (6,2%)
1600	335 (20,9%)	7000	373 (5,3%)
1800	341 (18,9%)	8000	375 (4,7%)
2000	345 (17,2%)	9000	376 (4,2%)
2200	349 (15,9%)	10 000	377 (3,8%)

Konfidenzintervall 95%, Fehlertoleranz 5% <sup>5</sup>

Capture D'Écran 2019 06 10 À 08 39 40

Es wird deutlich, dass die Überlieferungsquote mit steigender Grösse der Grundgesamtheit abnimmt.

Die Übernahmequote ist somit nicht starr. Sie hängt ab von Grundgesamtheit, Konfidenzintervall und Fehlertoleranz. Aus diesen Abhängigkeiten ergeben sich gewisse Steuerungsmöglichkeiten. Je grösser die Grundgesamtheit, je kleiner das Konfidenzintervall und je grösser die Fehlertoleranz, desto kleiner wird die Übernahmequote. Damit wird der Grad der Repräsentativität festgelegt. Es wird also festgelegt, wie hoch die Genauigkeit allfälliger Auswertungen aus der Stichprobe ausfallen wird.

Eine Randbemerkung: Bei der Erschliessung von Dossiererien, die nach diesem Modell gesammelt wurden, sind neben der Methode unbedingt auch die beiden Werte Konfidenzintervall und Fehlertoleranz der Beschreibung des Bestandes anzugeben.

## Das Praxisbeispiel Stiftungsaufsicht

Im Jahr 2005 standen Dossiers der Stiftungsaufsicht zur Übernahme ins Staatsarchiv Basel an. Die Stiftungsaufsicht beaufsichtigt alle Stiftungen, welche im Kanton erreicht werden. Es handelt sich dabei um klassische Stiftungen, wie sie etwa im kulturellen Bereich gängig sind, und um Vorsorgestiftungen von Firmen. Als Bewertungsziel wurde formuliert, die Tätigkeit der Stiftungsaufsicht und die Stiftungen zu dokumentieren. Im Folgenden werde ich mich auf die Auswahl der klassischen Stiftungen und der ausserobligatorischen Vorsorgestiftungen beschränken. Weggelassen werden die obligatorischen Vorsorgestiftungen, welche in einer separaten Serie abgelegt werden.

In einem ersten Schritt wurde überprüft, in welchem Kontext sie entstehen. Dann wurde ermittelt, ob innerhalb der Verwaltung horizontale oder vertikale Bezüge zur Tätigkeit der Stiftungsaufsicht bzw. zu anfallenden Unterlagen existieren. So werden Stiftungen auch im Handelsregister via Errichtungsurkunde, Reglement und formale Änderungen dokumentiert. Bei der Steuerverwaltung fallen Steuerakten zu den Stiftungen an. Diese Akten wurden einige Jahre zuvor vom Staatsarchiv bereits bewertet. Es wurde damals eine Auswahl festgelegt und festgehalten, dass die substanzielle Überlieferung via Stiftungsaufsicht erfolge.

Bei der Analyse der Dossiers wurde festgestellt, dass keine Detailinformationen über die Tätigkeit der Stiftung abgelegt wurden. So lässt sich anhand der Vorsorgestiftung kaum ermitteln, wie die Firma selbst sich entwickelt hat. Bei den Kulturstiftungen wird die inhaltliche Tätigkeit nicht fassbar. Die Dossiers enthalten die Stiftungsurkunde, Statuten, Jahresrechnungen und Korrespondenzen mit verschiedenen Verwaltungsstellen. Anzuführen ist, dass das Stiftungswesen in Basel-Stadt aufgrund seiner gesellschaftlichen und wirtschaftlichen Entwicklung eine besondere Rolle spielt. Es wurde entschieden, eine repräsentative Stichprobe zu ziehen. Zudem werden besondere Fälle langfristig archiviert. Für besondere Fälle wurden folgende Kriterien definiert: pressenotorische Vorfälle im Umfeld von Stiftungen und Fälle, in denen die Aufsicht mit strafrechtlichen Verfahren verbunden ist. Der Bestand umfasste über 3000 Dossiers. Die Dossiers wurden seit den 1930er Jahren nach Nummern abgelegt. Die Grundgesamtheit wurde deshalb auf der Basis der vorhandenen Dossiers, unabhängig von der Aufbewahrungsfrist und damit vom Ablieferungszeitpunkt, gebildet.

Der Schnitt wurde bei der runden Zahl 3000 gezogen. Als Konfidenzintervall wurde 95% und als Fehlertoleranz 5% festgelegt. Daraus ergab sich, dass 359 Dossiers übernommen wurden. Das entspricht einer Quote von 12%.

Da die Dossiers nummeriert sind, gestaltet sich die Auswahl relativ einfach. Mittels Zufallsgenerator wurden aus dem Nummernbereich 1 bis 3000 359 Nummern gezogen. Der Stiftungsaufsicht werden die Nummern der zu überliefernden Dossiers mitgeteilt. Sie werden nach Löschung der Stiftung und Ablauf der Aufbewahrungsfrist von der Stiftungsaufsicht ausgesondert und dem Archiv übergeben.

Besonders hingewiesen sei noch auf die Ermittlung der Zufallszahlen. Die Zufallszahlen müssen die Qualität von echten Zufallszahlen aufweisen. Sie dürfen also nicht mit Zufallsgeneratoren aus Computerprogrammen generiert werden. Computerprogramme arbeiten mit Algorithmen, die zwar weitläufig sein können und Zahlen errechnen, die zufällig scheinen. Da sie jedoch berechnet werden, handelt es sich um Pseudozufallszahlen, welche vorhersehbar sind, da sie auf einer Formel beruhen. Reine Zufallszahlen sind nur auf der Basis von natürlichen Ereignissen zu generieren, wie etwa atmosphärisches Rauschen oder radioaktiver Zerfall. Im Internet sind Zufallszahlengeneratoren für echte Zufallszahlen zugänglich. Eine verlässliche Webseite ist: [www.random.org](http://www.random.org).

## Die einzelnen Vorgehensschritte

Folgende Vorgehensschritte sind nötig, um die Zufallsauswahl zu erhalten: Zuerst wird die Grundgesamtheit ermittelt. Sie soll möglichst gross sein, was zum Beispiel mit der Ablieferungskadenz und der Anwendung prospektiver Bewertung gesteuert werden kann. Anlässlich der Bewertung werden Konfidenzintervall und Fehlertoleranz festgelegt. Üblich sind ein 95%iges Konfidenzintervall und eine 5%ige Fehlertoleranz. Daraus wird mit der oben vorgestellten Formel die Anzahl Dossiers ermittelt und mit dem Zufallsgenerator gezogen.

Zufallsauswahlen können, wie im vorangehenden Beispiel gezeigt, auch geschichtet und mit qualitativen Verfahren kombiniert werden. Allerdings muss zuerst die Zufallsauswahl getroffen und danach die qualitative Auswahl z.B. besonderer Fälle vollzogen werden.

## Die Übernahmequote in der Praxis

Im Staatsarchiv Basel-Stadt arbeiten wir seit 2004 mit der Zufallsauswahl. Angewendet wurde sie bei der Bewertung von Ehe- und Erbverträgen der Vormundschaftsbehörden, Lehrfirmendossiers des Amtes für Berufsbildung und Gerichtsfalldossiers, um einige Beispiele zu nennen. In Zahlen vorgestellt, präsentieren sich die genannten Serien wie folgt:

Bestand	Konfidenzintervall (%)	Fehlertoleranz (%)	Grundgesamtheit	Übernahmequote (%)
Lehrfirmendossiers	95	5	4826	7,4
Stiftungsaufsicht (kombiniert mit besonderen Fällen)	95	5	3000	12
Administrativakten Polizei- und Militärdepartement	95	4	6332	9
Ehe- und Erbverträge	90	5	48 012 in Schichten von ca. 10 000	2,8
Beitragsdossiers Kulturpauschale Ressort Kultur	90	5	2400	10,6
Gerichtsfalldossiers (kombiniert mit B und besonderen Fällen)	95	5	Verschieden	



Bei anderen Serien, wo es uns wichtig scheint, dass familiäre Zusammenhänge rekonstruiert werden können, wenden wir nach wie vor die Methode der sektoriellen Klumpenstichproben an und übernehmen die Auswahl des Buchstabens B. Mit dieser Auswahl können also einzelne Personen oder Familien in mehreren Zusammenhängen untersucht werden. Der Vorteil der Methode besteht zudem darin, sehr praktikabel zu sein, wenn Dossiers alphabetisch abgelegt sind. Neben der fehlenden Repräsentativität liegt der Nachteil im Verlust der Rekonstruktionsfähigkeit in all jenen Fällen, wo ein Namenswechsel vollzogen wurde, etwa durch Heirat oder Scheidung. Zudem ergibt sich eine hohe Überlieferungsquote von ca. 10%. Dieser hohen Quote begegnen wir heute damit, dass wir jeweils nur die erste Hälfte des Buchstabens B übernehmen. Wir reduzieren also die Klumpenstichprobe wiederum mit einer Klumpenstichprobe. Methodisch falsch wäre es, aus der Buchstabenauswahl jedes zweite Dossier zu übernehmen, um sie auf die geforderten 5% zu reduzieren.

Was wir ausdrücklich vermeiden, ist die systematische Stichprobe, welche allen genannten Auswahlverfahren deutlich unterlegen ist.

<sup>1</sup> Siehe dazu Buchholz, S. 156 ff.



### **Irene Amstutz**

Seit 2009 Leiterin des Schweizerischen Wirtschaftsarchivs und der Universitätsbibliothek Wirtschaft in Basel. Mitglied der AG Archive der privaten Wirtschaft und des Bildungsausschusses des VSA. Zwischen 1999 und 2009 im Staatsarchiv Basel-Stadt vorarchivisch und in der Erschliessung tätig, ab 2008 als Leiterin Vorarchiv.

## Résumé

### Français

L'archivage des dossiers s'effectue toujours selon différentes procédures: une conservation exhaustive, la constitution d'échantillons, des tris qualitatifs ou l'élimination. La méthode retenue varie selon les besoins et les conditions d'évaluation des séries versées. Toutefois, seule la constitution d'échantillons aléatoires autorise des analyses quantitatives. C'est la méthode qui est la plus adaptée à toute exploitation. En plus, elle fournit au chercheur des intervalles de confiance et des marges d'erreur qui affinent la précision de l'analyse statistique.

Un avantage de la méthode tient au fait que plus la population des dossiers est grande, plus le nombre de dossiers à archiver peut être restreint. Et ceci sans perte de qualité de l'échantillon. En revanche, en cas d'effectif réduit du total des dossiers, il faudra archiver une proportion élevée de ceux-ci.

Les limites sont toutefois connues: en procédant à des échantillons, il n'est par exemple plus possible de reconstituer des relations familiales ou le suivi temporel d'une personne inscrite dans plusieurs séries. Ces méthodes stochastiques permettent toutefois aux archivistes de ne plus se cantonner aux seuls échantillons élaborés avec naïveté et sans fondement théorique.