

Die Bildung repräsentativer Stichproben bei Massenakten

Zur Methode und Praxis der Bewertung
mittels Zufallsauswahl

Irene Amstutz, Staatsarchiv Basel-Stadt

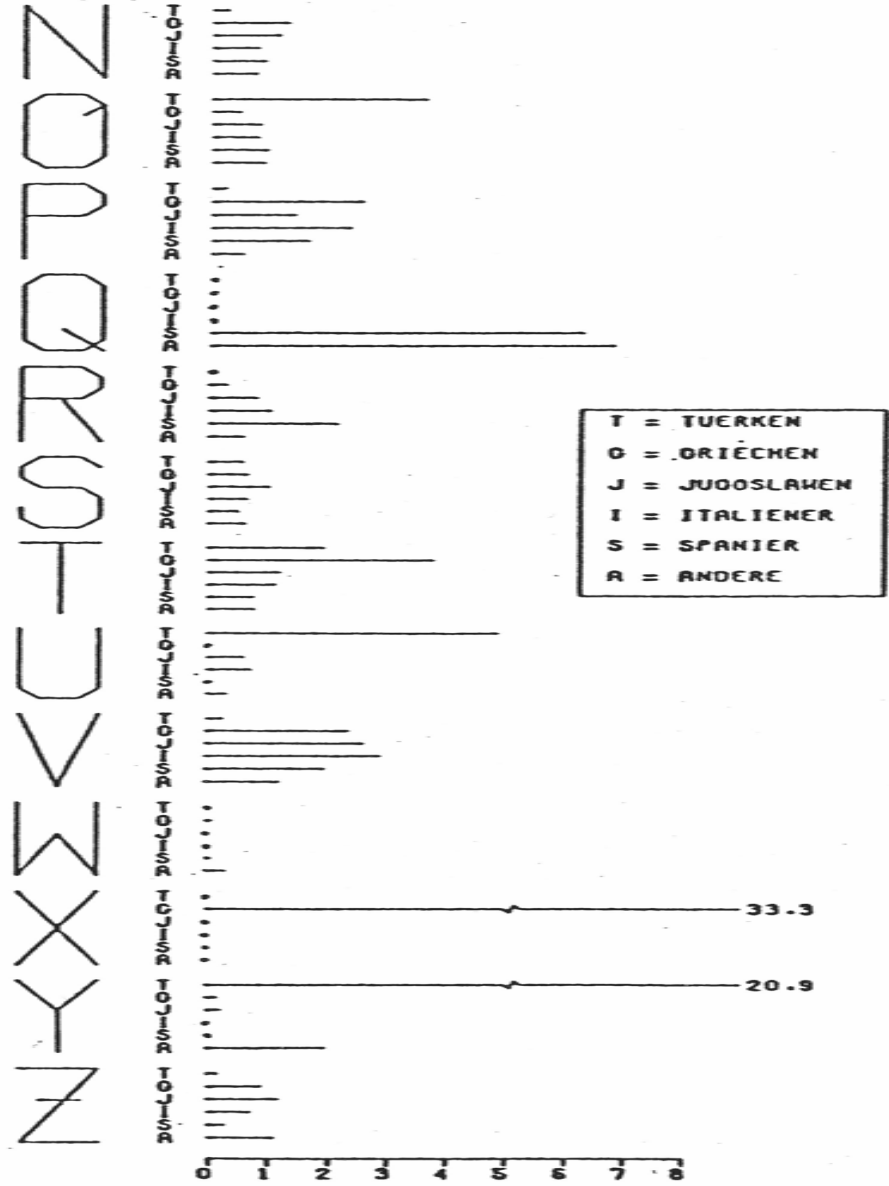


Meilenstein

- Buchholz Matthias
- Überlieferungsbildung bei massenhaft gleichförmigen Einzelfallakten im Spannungsverhältnis von Bewertungsdiskussion, Repräsentativität und Nutzungsperspektive
- Köln 2001



Verteilung von Nationalitäten



T = TUERKEN
 G = ORIÉCHEN
 J = JUOSLAHEN
 I = ITALIENER
 S = SPANIER
 A = ANDERE



Auswertungsoffenheit

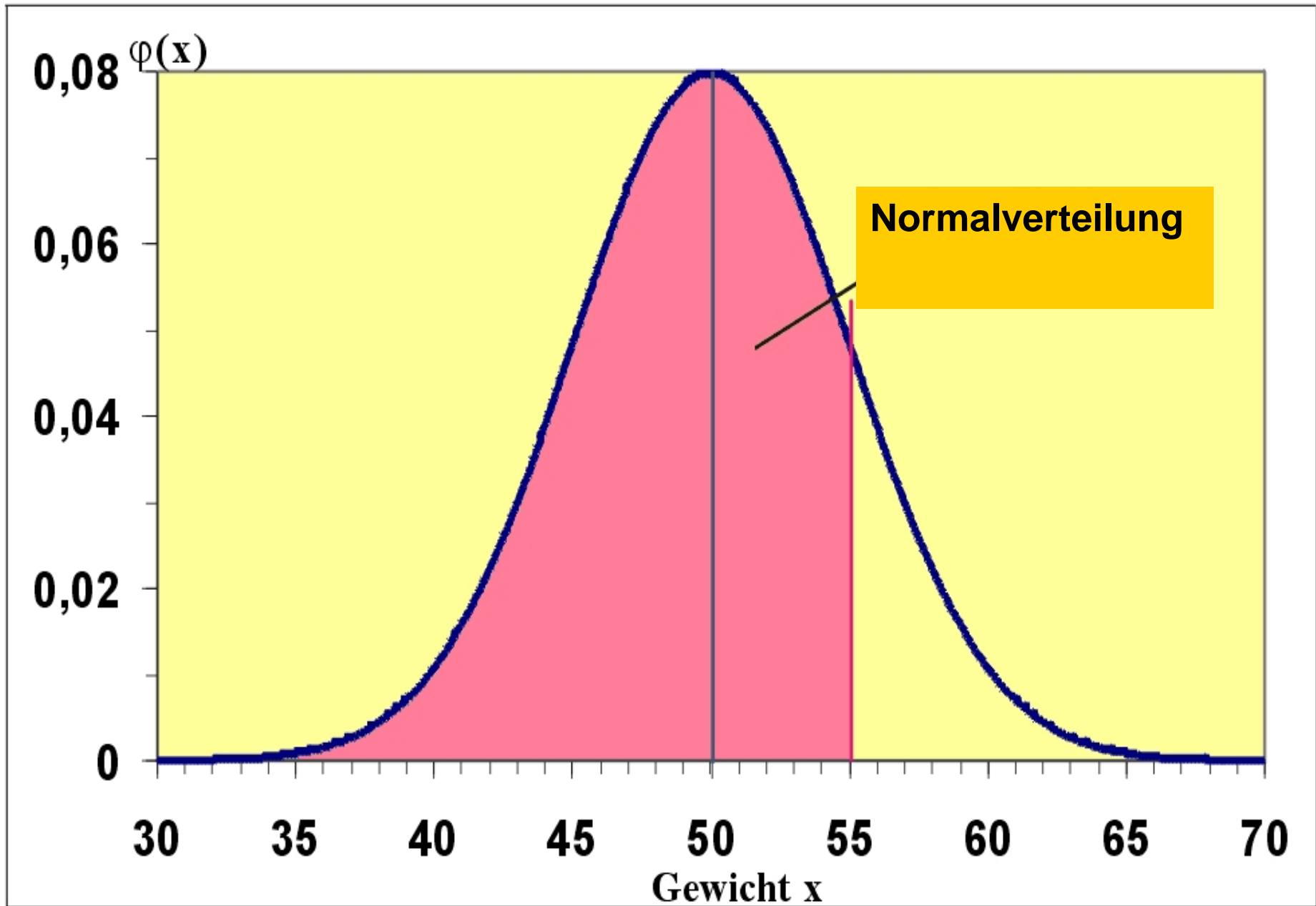
- Aufgrund des Zufallsprinzips
- Jede Einheit verfügt über dieselbe Wahrscheinlichkeit, in die Stichprobe zu gelangen
- Alle Merkmale der Grundgesamtheit sind in der Stichprobe im gleichen Masse enthalten



Wahrscheinlichkeitstheorie

- Normalverteilung
- Erfahrungswissen
- Gaussche Glockenkurve, 1794
- Ereignisse gruppieren sich um den Mittelwert und werden je seltener, je stärker sie vom Mittelwert abweichen





Grad der Repräsentativität

- **Konfidenzintervall**
 - Ausdruck davon, wieviele Stichproben innerhalb eines bestimmten Bereichs um den Mittelwert liegen
 - Bsp.: bei einem Konfidenzintervall von 95% liegen 95 von 100 Stichproben innerhalb eines bestimmten Bereichs.
- **Fehlertoleranz**
 - Aussage über den oben genannten Bereich
 - Bsp.: Eine Fehlertoleranz von 2% bedeutet, dass die obigen 95 von 100 Stichproben $\pm 2\%$ um den Mittelwert liegen.



Formel für die Berechnung des Stichprobenumfangs, gekürzt

- Für ein Konfidenzintervall von 95% und eine Fehlertoleranz von 5% lässt sich die Formel gekürzt so darstellen:

$$\textit{Stichprobe numfang} = \sqrt{\frac{N - 384}{N - 1}} * 384$$

Konfidenzintervall 95%, Fehlertoleranz 5%

■ Grundgesamtheit

- 800
- 1000
- 2000
- 4000
- 6000
- 8000
- 10'000

Stichprobengrösse

277	(34,6%)
302	(30,2%)
345	(17,2%)
365	(9,1%)
372	(6,2%)
375	(4,7%)
377	(3,8%)

Buchholz, S. 299ff.



Übernahmequoten

- Übernahmequoten sind nicht mehr starr
- Sie hängen ab von
 - Grundgesamtheit
 - Konfidenzintervall
 - Fehlertoleranz
- Je grösser die Grundgesamtheit, je kleiner das Konfidenzintervall, je grösser die Fehlertoleranz → je kleiner die Übernahmequote



Beispiel Stiftungsaufsicht

- Grundgesamtheit 3000 Dossiers
- Konfidenzintervall 95%, Fehlertoleranz 5%



Beispiel Stiftungsaufsicht

$$\textit{Stichprobenumfang} = \sqrt{\frac{3000 - 384}{3000 - 1}} * 384$$

- Resultat 359
- Das heisst, man übernimmt aus 3000 Dossiers 359 Dossiers, ausgewählt nach dem Zufallsprinzip
- Übernahmequote: 12%



Vorgehensschritte

- Grundgesamtheit ermitteln (Anzahl Dossiers)
- Konfidenzintervall und Fehlertoleranz festlegen
- Anzahl Dossiers der Stichprobe berechnen
- Zufallsgenerator anwerfen (www.random.org)
- Dossiers mit entsprechenden Nummern aussondern



Beispiele

Bestand	Konfidenzintervall (%)	Fehler-toleranz (%)	Grund-gesamtheit	Übernahme- quote (%)
Ehe- und Erbverträge	90	5	48 012 in Schichten von ca. 10'000	2,8
Lehrfirmen-dossiers	95	5	4826	7,4
Admini-strativakten Polizei	95	4	6332	9
Stiftungs-aufsicht	95	5	3000	12
Gerichtsfall-dossiers	95	5	Verschieden	kombiniert mit B & besond. Fällen



Buchstabenauswahl

- Buchstabe B
- Ermöglicht Rekonstruktion familialer Zusammenhänge über Serien hinweg
- Quote: ca. 10 %
- Reduktion: die Hälfte = 5%
- Klumpen bilden → die erste Hälfte B
- NICHT: jedes 2. B

