3 Einige Ausgewählte Aspekte der Bevölkerungsstatistik

3.1 Gegenstand und Grundbegriffe

Literatur: Schaich & Schweitzer (1995)

3.1.1 Gegenstand

- Aufgabe der Bevölkerungsstatistik:
  Beschreibung und Analyse von
  - Umfang
  - Zusammensetzung und
  - räumliche Verteilung einer Bevölkerung sowie deren
  - Veränderungen im Zeitablauf

- Unterteilbar in statische und dynamische Komponente
  - Bevölkerungsstrukturstatistik:
    Betrachtung der Zusammensetzung einer Bevölkerung zu einem bestimmten Zeitpunkt
  - Bevölkerungsprozessstatistik:
    Betrachtung von Umfang, Zusammensetzung und räumlicher Verteilung im Zeitablauf
- Demographie (Bevölkerungswissenschaft)
  - modellbasiert
  - Suche nach Hintergründen/Erklärungen von Strukturen und Prozessen
  - interdisziplinär (v.a. Einbezug von Wirtschafts- und Sozialwissenschaften)

- Typische bevölkerungsstatistische/demographische Variablen:
  - Geschlecht
  - Alter
  - Familienstand
  - Religions- (bzw. Konfessions-) zugehörigkeit
  - Beteiligung am Erwerbsleben
  - Zugehörigkeit zu bestimmter geographischer Einheit

  in Demographie zusätzlich:
  - Bildungsstand
  - „Soziale Schicht“
  - Einkommen etc.

- Typische Themen in der Bevölkerungsprozessstatistik:
  - Mortalität (Sterblichkeit)
  - Fertilität (Fruchtbarkeit)
  - Migration (Wanderungstätigkeit)
  - Nuptialität (Heiratsverhalten)
- Bedeutung des Themas:

3.1.2 Die verlaufsstatistische Struktur/Fortschreibung

$B(t)$	(Bevölkerungs-)Bestand zum Zeitpunkt t
$Z(t_1, t_2)$	(Gesamt-)Zugänge im Intervall $(t_1, t_2]$
$A(t_1, t_2)$	(Gesamt-)Abgänge im Intervall $(t_1, t_2]$

Fortschreibungsformel

$$B(t_2) = B(t_1) + Z(t_1, t_2) - A(t_1, t_2)$$

Differenziert man Zugänge und Abgänge weiter, so ist

$$Z(t_1, t_2) = G(t_1, t_2) + I(t_1, t_2)$$

mit

$G(t_1, t_2)$	Anzahl der Lebendgeborenen in $(t_1, t_2]$
$I(t_i, t_2)$	Anzahl der Zuwanderungen in $(t_1, t_2]$

und

$$A(t_1, t_2) = S(t_1, t_2) + E(t_1, t_2)$$

mit

$S(t_i, t_2)$	Anzahl der Gestorbenen in $(t_1, t_2]$
$E(t_i, t_2)$	Anzahl der Ausgewanderten in $(t_1, t_2]$,

also

$$B(t_2) = B(t_1) + G(t_1, t_2) + I(t_1, t_2) - S(t_1, t_2) - E(t_1, t_2).$$

Die Differenz

$$G(t_1, t_2) - S(t_1, t_2)$$

heißt Geburtenüberschuss bzw. -defizit im Intervall $(t_1, t_2]$;

$$I(t_1, t_2) - E(t_1, t_2)$$

heißt Wanderungssaldo des Intervalls $(t_1, t_2]$.

3.1.3 Gliederungen, rohe und spezifische Verhältniszahlen

Oft werden die Größen und Bestandszahlen weiter „gegliedert“, oder „spezifiziert“, d.h. sie werden in spezielle Subgruppen aufgespalten, insbesondere nach dem

- Geschlecht (w,m), man schreibt dann $B_w(t)$, $B_m(t)$, $S_w(t)$, $S_m(t)$, usw.

- Alter (x), man schreibt dann $B_x(t)$, $S_x(t)$, für $x = 0, 1, \ldots, 101$.

  Vorsicht, der Laufindex x kann sich neben dem erreichten Alter (in Jahren, hier) auch auf das x-te Lebensjahr beziehen.

- auch tiefergehende Gliederung durch 2 Merkmale üblich, z.B. $B_{w,x}(t)$ Anzahl der Frauen im Alter x usw.

- weitere gängige Gliederungsmerkmale sind z.B. Familienstand, Haushaltsgröße, Erwerbsstatus, geographische Herkunft.

Zwei Arten von Verhältniszahlen

- rohe (globale, unspezifizierte, allgemeine, übliche) Verhältniszahlen:
  für die gesamte Bevölkerung ausgewiesen

- spezifizierte Verhältniszahlen:
  gruppenspezifisch, v.a. wichtig später bei Raten.

3.2 Alterspyramiden

3.2.1 Gliederung nach dem Geschlecht

- $\gamma^{(t)} = \dfrac{B_w(t)}{B_m(t)} \cdot 1000$ heißt Sexualproportion (Geschlechtsrelation)
- Alternativ durch Gliederungszahlen ausdrückbar

$$\alpha_m^{(t)} := \frac{B_m(t)}{B(t)} \cdot 100$$

$$\alpha_w^{(t)} := \frac{B_w(t)}{B(t)} \cdot 100$$

- Überlicherweise: $\gamma > 1000$: „Frauenüberschuss“, wegen größerer Sterblichkeit v.a. der älteren Männer

- Altersspezifische Sexualproportionen

$$\gamma_x(t) = \frac{B_{w,x}(t)}{B_{m,x}(t)} \cdot 1000,\ x = 0, 1, \ldots, 101$$

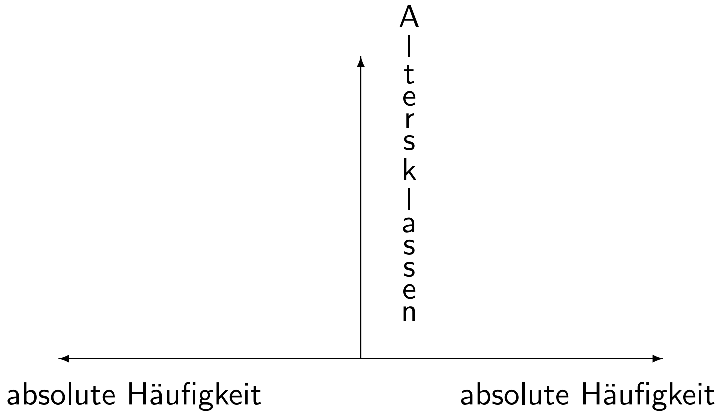
$\gamma_0^{(t)}$ Sexualproportion der Lebendgeborenen

- γ lässt sich als gewogene Summe der g_i schreiben

$$\begin{aligned}\gamma^{(t)} &= \frac{B_w(t)}{B_m(t)} \cdot 1000 = \frac{\sum_{x=1}^{101} B_{w,x}(t)}{\sum_{x=1}^{101} B_{m,x}(t)} \cdot 1000 = \sum_{x=1}^{101} \frac{B_{w,x}(t)}{B_m(t)} \cdot 1000 = \\ &= \sum_{x=1}^{101} \frac{B_{w,x}(t)}{B_m(t)} \cdot \frac{B_{m,x}(t)}{B_{m,x}(t)} = \sum_{x=1}^{101} \gamma_x \frac{B_{m,x}(t)}{B_m(t)}.\end{aligned}$$

3.2.2 Alterspyramiden I - Aufbau

- Veranschaulichung der Geschlechts- und Altersstruktur durch „geeignet zusammengelegte, gedrehte Histogramme“ der Altersverteilung der männlichen und weiblichen Bevölkerung.



- Man wählt üblicherweise wegen der größeren Anschaulichkeit hier
  - absolute Häufigkeiten
  - Altersklassen der Größe 1 (Höhe der Histogrammbalken proportional zu Häufigkeiten)

- Altersspezifische Sexualproportionen über das Flächenverhältnis der zu dem entsprechenden Alter gehörenden Säulen; die Sexualproportion ist ablesbar aus dem Verhältnis der Gesamtflächen

3.2.3 Die Alterspyramiden der BRD

3.2.4 Grundtypen von Alterspyramiden

- Pyramidenförmiger Altersaufbau
- Glockenförmiger Altersaufbau
- Urnenförmiger Altersaufbau

Wichtige Kenngröße *Abhängigkeitsverhältnis*:

$$A := \frac{\text{Umfang der Bevölkerung im typischerweise nicht-erwerbsfähigen Alter}}{\text{Umfang der Bevölkerung im typischerweise erwerbsfähigen Alter}}$$

$$\text{z.B. } A(t) = \frac{\sum_{x=0}^{19} B_x(t) + \sum_{x=65}^{101} B_x(t)}{\sum_{x=20}^{64} B_x(t)}$$

3.3 Bevölkerungsprozessstatistik: Raten und Tafeln

3.3.1 Vorbemerkungen

- Dynamik durch
  - Zugänge (Geburt, Zuwanderung)
  - Abgänge (Tod, Abwanderung)
  - Bewegung zwischen Sektoren (ledig $\longrightarrow$ verheiratet, erwerbstätig $\longrightarrow$ nichterwerbstätig)
- Beschreibung Dynamik durch
  - Anzahlen
  - Raten (Anzahl bezogen auf Umfang)

3.3.2 Tafeln

- Sterbetafeln, Geburtentafeln, Heiratstafeln, Erwerbspersonentafeln
- „Wahrscheinlichkeiten“ geschlechtsspezifisch, entlang einjähriger Altersklassen
- sind eigentlich keine Wahrscheinlichkeiten, sondern bedingte Häufigkeiten
- prognostisch verwendbar (z.B. Prämienkalkulation)
- Kohorten- versus Periodensicht, entspricht „Längsschnitt-“ versus „Querschnittbetrachtung“.
- Kohorte: reale Gesamtheit von Personen, z.B. 100000 Lebendgeborene eines Jahrgangs $\longrightarrow$ Kohortentafel
- Periodentafel: beobachte Prozess in kurzer Spanne (z.B. 3 Jahre) bei allen Personen aller Altersklassen

Def. 3.1. [Abgangsrate]

Sei $\tilde{A}(t)$ die Anzahl der Abgänge eines bestimmten Typs (nachfolgend Todesfälle) in $(0, t]$. Setzt man, wie im Folgenden, $\tilde{A}(t)$ als differenzierbar voraus, so heißt

$$\tilde{a}(t) = \frac{d\tilde{A}(t)}{dt}$$

$Abgangsfunktion\ (Abgangsintensität)$.
Mit $B(t)$ als Bestand zum Zeitpunkt t, heißt dann

$$\tilde{\alpha}(t) = \frac{\tilde{a}(t)}{B(t)}$$

$Abgangsrate$.

Typischerweise sind weder Zähler noch Nenner in dieser infinitesimalen Form ermittelbar.

Für eine operationale (berechenbare) Version ersetzt man den Zähler durch den Differenzenquotienten

$$\frac{\tilde{A}(t+1)-\tilde{A}(t)}{(t+1)-t}=\tilde{A}(t+1)-\tilde{A}(t).$$

Im Nenner ersetzt man in einem ersten Schritt den Durchschnittsbestand $B(t)$ durch

$$\overline{B}(t,t+1):=\int_t^{t+1} B(u)du,$$

der aber noch weiter zu approximieren ist, um ihn berechnen zu können.

Man setzt üblicherweise

$$\overline{B}(t,t+1)\approx\frac{B(t)+B(t+1)}{2}$$

oder

$$\overline{B}(t_1, t_2) \approx B\left(\frac{t_1 + t_2}{2}\right)$$

(Vorsicht bei saisonalen Schwankungen!)

Bem. 3.2.

Damit heißt unter Verwendung dieser Approximationen

$$m(t) = \frac{S(t, t+1)}{\overline{B}(t_1, t_2)} \quad \text{bzw.} \quad m(t) = \frac{S(t, t+1)}{\overline{B}(t_1, t_2)} \cdot 1000 \tag{3.14}$$

(operationale Form der) *rohe(n) Sterberate* (Sterbeziffer)

Beispiel: rohe Sterbeziffern, 2003

Deutschland	10.3
Vietnam	6.4

3.3.3 Sterbetafeln

- Kohorten oder Periodentafel
- Argument t hier weggelassen
- Tabellarische Darstellung der „Abgangsordnung“ (Survivorfunktion) eines sich durch Todesfälle ständig reduzierenden Bestandes
- wie viele (von z.B. 100000 Lebendgeborenen) erreichen das Alter x?
- *Sterbewahrscheinlichkeit*

$$q_x = \frac{\ell_x - \ell_{x+1}}{\ell_x}$$

mit ℓ_x als Anzahl der betrachteten Lebenden im Alter x

- *„Tafelfunktionen"*
  - Anzahl der Überlebenden
  - Anzahl der Gestorbenen

$$d_x = \ell_x - \ell_{x+1}$$

  - Sterbe- bzw. Überlebenswahrscheinlichkeit

$$q_x = \frac{\ell_x - \ell_{x+1}}{\ell_x} = 1 - \frac{\ell_{x+1}}{\ell_x} = \frac{d_x}{\ell_x} \quad \text{bzw.} \quad p_x = 1 - q_x$$

vom Alter x bis zum Alter $x + 1$

– die Anzahl

$$L_x = \ell_{x+1} + 0.5 d_x = \frac{1}{2}\left(\ell_x + \ell_{x+1}\right)$$

der von den Überlebenden im Alter x bis zum Alter $x+1$ durchlebten Personenjahre (unter der vereinfachenden Annahme der Gleichverteilung der Sterbefälle)

– die Anzahl

$$T_x = \sum_{y=x}^{100} L_y$$

der von den Überlebenden im Alter x insgesamt durchlebten Jahre.

– Restlebenserwartung

$$e_x^0 = \frac{T_x}{e_x}$$

der Überlebenden im Alter x

- verschiedene Verfahren zur empirischen Schätzung von Sterbewahrscheinlichkeiten
  - Geburtsjahrmethode nach Becker-Zeuner (alle Sterbefälle eines Geburtsjahrganges, jedoch mit unterschiedlichen Gewichten)
  - Sterbejahrmethode von Raths (Sterbefälle eines Jahres, also auf zwei Geburtsjahrgänge bezogen)
  - Farr (geschätzte Sterbewahrscheinlichkeiten durch Umrechnung altersspezifischer Sterbeziffern unter Zusatzannahmen, gilt als am unempfindlichsten gegenüber Wanderungen, aktuell verwendet)

$$\begin{aligned} q_x &= \frac{\ell_x - \ell_{x+1}}{\ell_x} = \frac{d_x}{\ell_x} \\ m_x &= \frac{d_x}{\overline{B}_x} = \frac{\ell_x - \ell_{x+1}}{\overline{B}_x} \end{aligned}$$

Wegen $\overline{B}_x < \ell_x$ gilt: (ceteris paribus)

$$m_x > q_x.$$

Annahme

$$\overline{B}_x \approx \frac{\ell_x + \ell_{x+1}}{2} \qquad \left(= \ell_x - \frac{1}{2} d_x \right)$$

dann

$$m_x = \frac{\ell_x - \ell_{x+1}}{\overline{B}_x} \approx \frac{\ell_x - \ell_{x+1}}{\frac{\ell_x + \ell_{x+1}}{2}}$$

$$\begin{aligned} m_x \quad &\approx \quad \frac{\ell_x - \ell_{x+1}}{\frac{\ell_x + \ell_{x+1}}{2}} = \frac{2\ell_x - 2\ell_{x+1}}{\ell_x + \ell_{x+1}} = \frac{2(\ell_x - \ell_{x+1})}{\ell_x + \ell_{x+1}} \cdot \frac{\ell_x}{\ell_x} = \frac{2 \cdot q_x \cdot \ell_x}{\ell_x + \ell_{x+1}} = \\ &= \quad \frac{2q_x}{\ell_x + \underbrace{\frac{\ell_{x+1}}{\ell_x}}_{q_x = 1 - \frac{\ell_{x+1}}{\ell_x}}} = \frac{2q_x}{1 + 1 - q_x} = \frac{2q_x}{2 - q_x} \end{aligned}$$

also

$$q_x \quad = \quad \frac{2m_x}{2 + m_x} = \frac{m_x}{1 - \frac{1}{2} m_x}$$

- die Schätzungen der Sterbewahrscheinlichkeiten werden in der Praxis meist noch (z.B. durch Splines) geglättet.