

Demographie: Vergangenheit, Gegenwart und Zukunft

Michel Truong Nhu
Seminarvortrag am Institut für Statistik, LMU München
Betreuer Prof. Dr. Thomas Augustin

13.Dezember 2011

Gliederung

0. Einführung
1. Populationsdenken
2. Bevölkerungsstatistik
 - 2.1 Alterspyramide
 - 2.2 Sterbetafel
3. Ereigniszeitanalyse
 - 3.1 Sterbetafelmethode
 - 3.2 Cox Regression
4. Diskussion

Begriff Demographie

„the study of the size, territorial distribution and composition of population, changes therein and the components of such changes“ (Hauser and Duncan 1959)

Einführung

- Rasanter Wachstum im 20. Jahrhundert z.B. PAA (Population Association of America) wuchs 1956 von 500 Mitglieder zu 3000 Mitgliedern im Jahr 1999
- Demographie liefert statistische Fakten & Prognosen über kommende Veränderungen in der Bevölkerungsstruktur
- Einfluss der Evolutionstheorie Darwins auf die Demographie
- Formale Demographie: Geburtenrate, Sterblichkeit, Altersstruktur und räumliche Verteilung der menschlichen Bevölkerung
- Bevölkerungsforschung: Bevölkerungsentwicklung hinsichtlich anderen Wissenschaften (Soziologie, Biologie usw.)

Populationsdenken

- Allgemein werden Variationen/Kombinationen von Objekten betrachtet
- Kombination selbst ist Gegenstand der Betrachtung und Objekte sind nebensächlich
- In der Biologie: Jedes Lebewesen ist einzigartig und weichen voneinander graduell ab →
- In der Demographie: eingeführt von Darwins Cousin in die Statistik
Korrelation und Regressionsanalyse

Essentialistisches Denken

- Auch: typologisches Denken, Denken in Typen
- Die Überzeugung, dass man alle natürlichen Variationen auf eine begrenzte Zahl grundlegender Kategorien zurückführen kann
- In der Biologie: Existenz eines Standard, von denen Mutationen davon abweichen
- Zielt auf die „Essenz“ oder „wahre Natur“ eines Menschen ab
 - Quetelets „mittlerer Mensch“

Bevölkerungsstatistik

Alterspyramide

- Auch Bevölkerungspyramide
- Verteilung der Altersklassen einer Bevölkerung
- Wird nach Männern und Frauen unterschieden
- Verschiedene Formen: Pyramide, Glocke, Urne usw.
- Abschätzung der Altersverteilung durch Exponential- oder Weibull-Verteilung

Alterspyramide

Bemerkung

- Prognosen über Bevölkerungsstrukturen basieren auf der 12. koordinierten Bevölkerungsvorausrechnung:
 - Außenwanderung: langsamer Anstieg des Wanderungssaldos bis 2014 auf + 100.000 Menschen, dann konstant
 - Binnenwanderung: _____
 - Geburtenhäufigkeit: ca. 1.4 Kindern pro Frau
 - Lebenserwartung: Zielwerte für Neugeborene im Jahr 2060

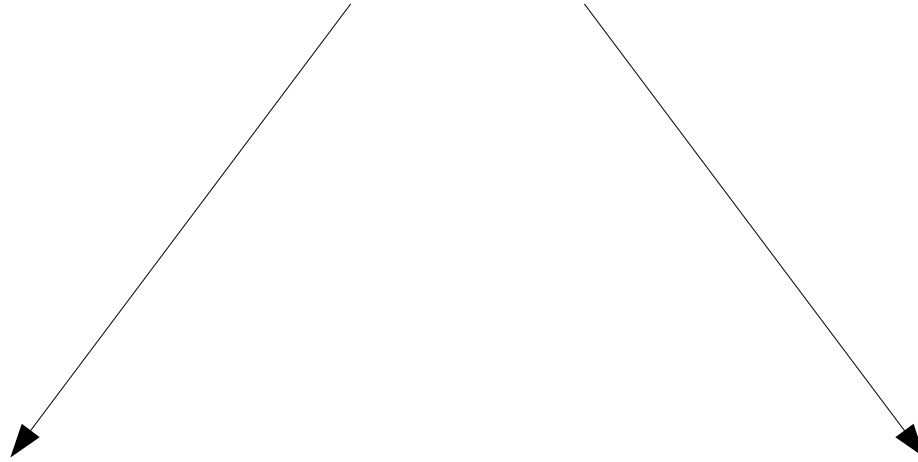


Sterbetafel

Allgemein

- Zeigt Sterblichkeitsverhältnisse einer Bevölkerung
- Getrennt nach Geschlechtern
- Gibt Auskunft über die durchschnittliche Lebenserwartung in den einzelnen Jahren
- Anwendung z.B. für Bevölkerungsvorausberechnungen, in der Medizin, Versicherungsmathematik, bei Rentenprognosen

Sterbetafel



Periodensterbetafel

Kohortensterbetafel

Periodensterbetafel

- Auch Querschnittsterbetafel
- Untersuchungseinheit: alle Geburtenjahrgänge in dieser Periode
- Beobachtungszeitraum: 3 Jahre
- Alle herrschenden Bedingungen in dem Zeitraum werden einbezogen

Kohortensterbetafel

- Auch Längsschnitt-, Generationensterbetafel
- Untersuchungseinheit: alle Personen eines Geburtsjahrgangs
- Beobachtungszeitraum: Von Geburt bis zum Tod der letzten Person

Bemerkung

- Statistische Bundesamt unterscheidet bei Perioden- und Kohortensterbetafeln zusätzlich zwischen

allgemeinen Sterbetafeln

- Jeweils im Anschluss an einer Volkszählung

abgekürzte Sterbetafeln

- Jeweils zwischen einer Volkszählung

- Endet bei Alterstufe 90

Sterbetafeln

- Im Gegensatz zur abgekürzten Sterbetafel geht die Altersstufe bis 100

Ereigniszeitanalyse

- Auch Verlaufsdatenanalyse, Ereignis(daten)analyse, Survival-Analyse, engl. Survival Analysis, Analysis of Failure Times, Event History Analysis
- Wie lange bleiben Untersuchungseinheiten in einem (Ausgangs-)Zustand?
- Häufig wichtiger Punkt: Einfluss erklärender Variablen auf die Verweildauer
- Zensierte Daten (links- und rechtszensiert): fehlende Informationen
- Auswertung von rechtszensierten Daten unter Voraussetzung: Zensierung hängt nicht kausal mit dem Prozess ab

Ereigniszeitanalyse

Grundbegriffe

- Dichtefunktion der Verweildauern

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t}$$

- Daraus ergibt sich die Verteilungsfunktion der Verweildauern

$$F(t) = P(T \leq t) = \int f(u) du$$

- Meistens wird stattdessen die Survivorfunktion (Gegenfunktion) benutzt

$$S(t) = P(T > t) = 1 - F(t)$$

- Eine entscheidende theoretische Größe ist die Hazardrate und in einigen Fällen die kumulierte Hazardrate

Sterbetafelmethode

- Nicht-parametrisches Verfahren
- Beobachtungszeitraum wird in Zeitintervalle unterteilt
- Auf Info über genauen Zeitpunkt des Ereignisses/ der Zensierung wird verzichtet
- Verwendete Bezeichnungen:

n_i = Anzahl an Individuen zu Beginn des Zeitintervalls $[a_i, b_i]$

berechnet sich: $n_i = n_{i-1} - d_{i-1} - c_{i-1}$

c_i = Anzahl der Zensierungen im Zeitintervall $[a_i, b_i]$

d_i = Anzahl der Gestorbenen im Zeitintervall $[a_i, b_i]$

Bedingte Sterbe- bzw. Überlebenswahrscheinlichkeit: $\hat{q}_i = \frac{d_i}{n_i}$ $\hat{p}_i = 1 - \hat{q}_i$

Anzahl der Personen unter Risiko: $\hat{r}_i = n_i - \frac{c_i}{2}$

Sterbetafelmethode

- Geschätzte Survivorfunktion: $\hat{S}_i = \prod_{j=1}^i \hat{p}_j$
(geschätzte Wahrscheinlichkeit, den Beginn des Intervalls i zu erleben)
- Geschätzte Dichtefunktion: $\hat{f}(t_i) = \frac{\hat{S}_i - \hat{S}_{i+1}}{h_i}$
(absolute Wahrscheinlichkeit, dass Ereignis im Intervall i eintritt)
- Aus den vorherigen Funktionen wird die Hazardrate berechnet: $\hat{r}_i = \frac{\hat{f}_i}{\hat{S}_i}$
(bedingte Wahrscheinlichkeit, dass Ereignis im Zeitintervall i eintritt)

Cox Regression

- Auch: proportionales Hazard Modell (engl. Proportional hazards model)
- Semi-parametrisches Verfahren
- Einsatz der Methode bei Untersuchung von Effekten mehrerer Einflussgrößen auf eine Zielvariable
- Voraussetzung: Effekte verschiedener Variablen auf das „Überleben“ sind über die Zeit konstant
- Grundlage: Hazardrate (bedingte Wahrscheinlichkeit, dass ein Ereignis eintritt, wenn bis dahin kein Ereignis eingetreten ist)

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(t < T < t + \Delta | t \leq T)}{\Delta}$$

- Cox-Modell ist gegeben durch

$$h(t) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m)$$

$h_0(t)$ ist eine unbekannte Hazardfunktion, X_i die Einflussvariablen, β_i die Regressionskoeffizienten (Stärke der Bedeutung der Effekte)

Diskussion