Prof. Dr. Peter von der Lippe

Der folgende Text ist eine überarbeitete Fassung von Kap. 12 der Formel- und Aufgabensammlung zur Deskriptiven Statistik, erscheinen bei Oldenbourg (ähnlich, wenngleich ausführlicher auch Kap. 12 meines Buches "Deskriptiven Statistik" in der (roten) UTB-Reihe. Er gliedert sich in die folgenden drei Abschnitte

**1. Bestands- und Bewegungsmassen**
**2. Kennzahlen der Dynamik eines Bestands**
**3. Stationäre Bevölkerung und Tafelrechnung**

# Kapitel 12: Bestandsanalyse und Tafelrechnung

## 1. Bestands- und Bewegungsmassen

**Def. 12.1: (Bestandsmasse, Bewegungsmasse, Verweildauer)**

a) Eine statistische Masse, deren Einheiten (i=1,2,...,n) jeweils gemeinsam zu einem bestimmten Zeitpunkt $t_j$ in einem Bestand (über eine nicht näher bestimmte Dauer) verweilen, heißt **Bestandsmasse** (engl. stock).
Der Umfang der Bestandsmasse zum Zeitpunkt $t_j$ heißt **Bestand** $B(t_j) = B_j$. Er ist zu jedem Zeitpunkt $t = t_j$ durch die **Bestandsfunktion** B(t) gegeben. Die Zeit kann als diskrete ($t = t_0, t_1,...,t_j,...,t_m$) oder stetige Variable betrachtet werden.

b) Eine statistische Masse, deren Einheiten dadurch charakterisiert sind, daß sie zu einem bestimmten Zeitpunkt ihren Zustand ändern (was ein "Ereignis" darstellt), heißt **Bewegungsmasse** (Ereignismasse, Stromgröße, engl. flow). Der Umfang einer Bewegungsmasse ist die Anzahl derartiger Ereignisse in einem gegebenen Zeitraum (Zeitintervall). Zustandsänderung kann insbesondere bedeuten: Zugang zu oder Abgang von einer Bestandsmasse.

c) Jede Einheit einer Bewegungsmasse (i=1,2,...,n) ist durch Zugangszeit ($t_{Zi}$) und Abgangszeit ($t_{Ai}$) gekennzeichnet. Der Zeitraum zwischen Zu- und Abgangszeit $d_i = t_{Ai} - t_{Zi}$ heißt **Verweildauer**.

**Methoden der Erhebung von Bestands- und Bewegungsmassen:**

**1. Feststellung der Bewegungen (Bewegungsmassen)**

a) durch individualisierte Erhebung aller Verläufe, d.h. für jede Einheit werden Zugangs- und Abgangszeit festgestellt (= **Längsschnittsanalyse** oder Verlaufsanalyse);

b) laufende Registrierung aller Bestandsveränderungen und Auswertung der über ein Beobachtungsintervall (von $t_o$ bis $t_j$ ) kumulierten Zugänge ($Z_{01}$) und Abgänge ($A_{01}$), d. h. der **Bruttoströme**.

c) Feststellung der **Bestandsveränderungen** (d.h. der Salden- oder Nettoströme $Z_{01}$-$A_{01}$).

**2. Feststellung der Bestände (Bestandsmassen)**

a) durch periodische **Inventuren** (Zählen oder Messen)

b) durch **Fortschreibung** für das Intervall $[t_o,t_j]$:

(12.1) $B_j = B_0 + Z_{0j} - A_{0j} \quad (j = 0, 1, ..., m)$

In Gl. 12.1 ist $B_0$ der Anfangsbestand, $B_{0j}$ der Bestand zum Zeitpunkt $t_{0j}$, $Z_{0j}$ die Anzahl der Zugänge und $A_{0j}$ die Anzahl der Abgänge im Beobachtungsintervall $[t_0, t_1]$.

c) Bei Kenntnis sämtlicher individueller Verläufe (wie in 1a), also bei Längsschnittdaten, ist der Bestand zu jedem beliebigen Zeitpunkt bekannt.

> Unter Querschnittsanalysen versteht man die Kombination 1b + 2a und unter Längsschnittsanalysen die Kombination 1a + 2c.

**<u>Beckersches Diagramm, Bestandsfunktion und Zeitmengenfläche</u>**

<u>1. Beckersches Diagramm:</u>

Eine graphische Darstellung der individuellen Verläufe ist das Beckersche Diagramm (Abb. 12.1 für das Beispiel 12.1)[1].

<u>2. Bestandsfunktion:</u>

Es ist leicht zu sehen, wie aus dem Beckerschen Diagramm (oberer Teil von Abb. 12.1) die Bestandsfunktion B(t) (t stetig) herzuleiten ist. Mit jedem Zugang (Abgang) einer Einheit erhöht (verringert) sich die Bestandsfunktion um 1.

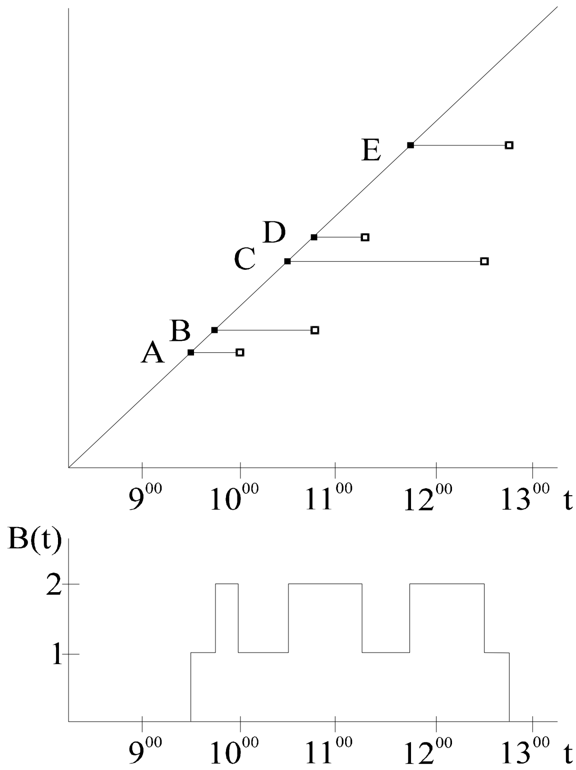
*Abb. 2.1 Beckersches Diagramm und Bestandsfunktion (Bsp.)*

<u>3. Zeitmengenfläche:</u>

Die schraffierte Fläche unter der Bestandsfunktion heißt Zeitmengenfläche F, oder genauer $F_{om}$ wenn die Fläche "über" dem Intervall $[t_o,t_m]$ betrachtet wird.

Zahlenbeispiel

| | Zeitpunkt | |
|---|---|---|
| | Zugang | Abgang |
| A | $09^{30}$ | $10^{00}$ |
| B | $09^{45}$ | $10^{45}$ |
| C | $10^{30}$ | $12^{30}$ |
| D | $10^{45}$ | $11^{15}$ |
| E | $11^{45}$ | $12^{45}$ |




**Def. 12.3: (offene-, geschlossene Masse)**

Eine Bestandsmasse heißt **geschlossen** bezüglich des Zeitintervalls $[t_0,t_m]$, wenn keine ihrer Einheiten vor $t_0$ zugegangen ist und nach $t_m$ abgeht (endgültig aus dem Bestand ausscheidet).
Es gilt folglich (speziell bei einer [beidseitig] geschlossenen Masse)

1. $B(0) = B(m) = 0$
2. $Z_{0m} = A_{0m}$ (alle Zugänge gehen auch wieder ab)
3. $F_{0m} = \Sigma d_i$ (Zeitmengenfläche = Verweilsumme)

Eine Masse, die nicht beidseitig geschlossen ist, heißt **offene** Masse. Man kann auch halbseitig und

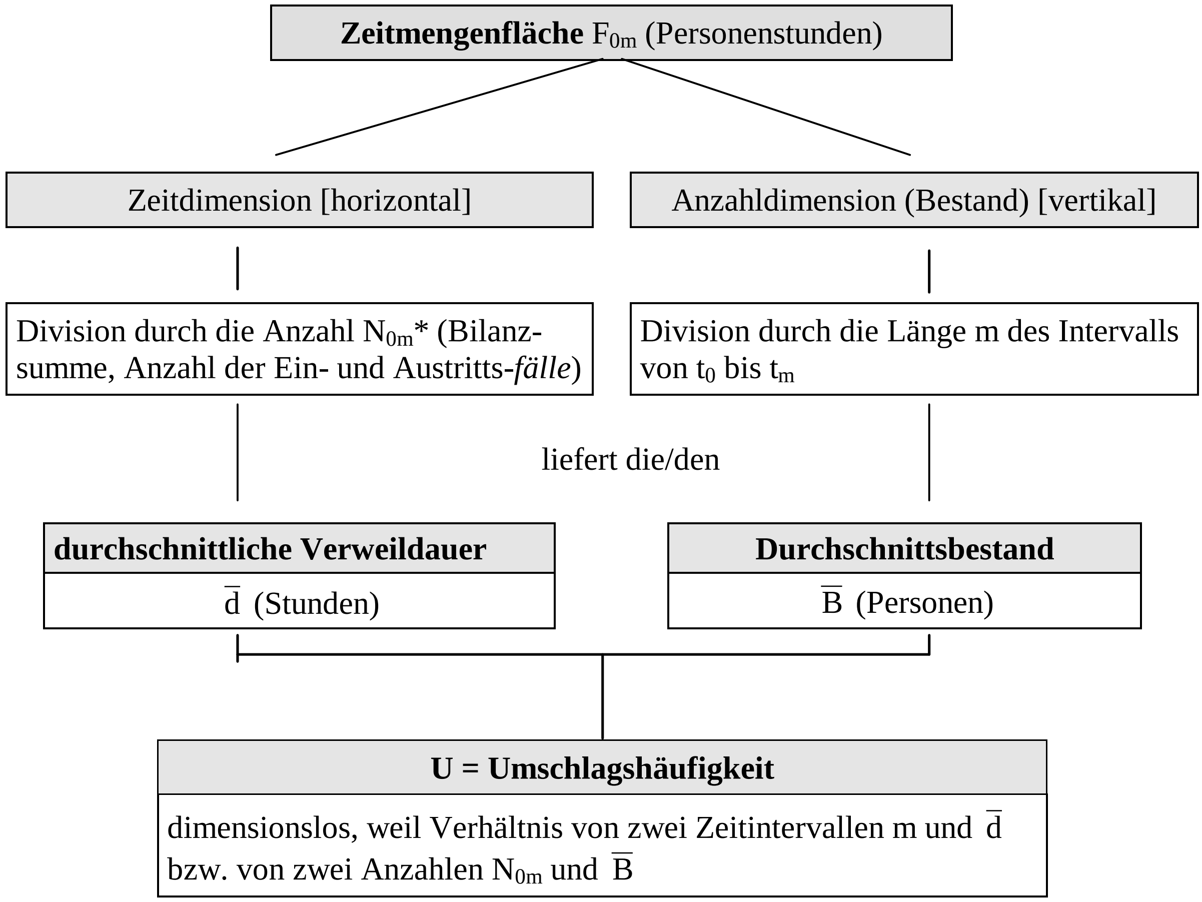
[1] vgl. Original des Textes (Lehrbuch im Oldenbourg Verlag), bzw. Demonstrationsbeispiel der Vorlesung.

beidseitig offene Massen unterscheiden.

## 2. Kennzahlen der Dynamik eines Bestands

### 1. Übersicht

*Übers. 12.1: Kennzahlen zur Beschreibung der Bestandsentwicklung*




*) Die Anzahl $N = N_{om}$ ist eine Anzahl von Fällen, nicht notwendig gleich der Anzahl n von Personen (Einheiten) die ein- und ausgetreten (zu- und abgegangen) sind.

### 2.1 Kennzahlen bei Kenntnis der individuellen Verläufe (Längsschnittsdaten)

| **Def. 12.4: (Durchschnittsbestand)** | **Def. 12.5: (durchschnittliche Verweildauer** | **Def. 12.5: (Umschlagshäufigkeit)** |
|---|---|---|
| (12.3) $\bar{B} = \frac{F_{om}}{m}$ | (12.4) $\bar{d} = \frac{\sum d_i}{N}$ ; und<br>(12.4a) $\bar{d} = \frac{F_{om}}{N}$ bei geschlossenen Massen | (12.5) $U = \frac{m}{\bar{d}}$<br>(12.5a) $U = \frac{N}{\bar{B}}$ * |

* folgt unmittelbar aus Gl. 12.3 und 12.4

### 2.2 Kennzahlen bei einer offenen Masse

Für die Definition des Durchschnittsbestands gilt weiterhin Gl. 12.3, die auch zur Berechnung von $\bar{B}$ herangezogen werden kann. Es ist aber nicht mehr von $\Sigma d_i = F_{0m}$ auszugehen. Vielmehr ist $F_{0m}$

zu korrigieren um die Zeiten, welche die

- $B_o$ Einheiten des Anfangsbestands vor to bereits dem Bestand angehört hatten und die Zeiten welche die
- $B_m$ Einheiten des Endbestands nach dem Ende des Beobachtungsintervalls, also nach $t_m$ dem Bestand noch angehören werden.

In diesem Sinne spricht man von Aufbauzeiten und Abbauzeiten und es ist davon auszugehen, daß für die Zeiten vor $t_o$ und nach $t_m$ keine Daten vorliegen, so daß die Auf- und Abbauzeiten nur geschätzt werden können. Die Verweilsumme ist unter diesen Voraussetzungen zu schätzen mit Gl.12.6. Hierin ist $B_0$, bzw. $B_m$ der Anfangs-, bzw. der Endbestand; $\bar{d}_o$ ist die mittlere **Auf**-bauzeit [d.h. die mittlere Verweildauer der $B_o$ Einheiten **vor** $t_o$] und $\bar{d}_m$ die mittlere **Ab**-bauzeit [mittlere restliche Verbleibdauer **nach** $t_m$ im Bestand]. Für die durchschnittliche Verweildauer gilt dann Gl. 12.8, woraus dann unter bestimmten weiteren Voraussetzungen Gl. 12.11 und 12.12 folgen:

| Geschätzte Verweilsumme $G_{0m}$ (durch Ergänzungen zur Zeitmengenfläche $F_{(0m)}$) | durchschnittliche Verweildauer |
|---|---|
| (12.6) $G_{om} = B_o\bar{d}_o + F_{om} + B_m\bar{d}_m$ | (12.8) $\bar{d} = G_{om}/N_{om}$ |

## 2.3 Kennzahlen bei Querschnittsdaten

### *2.3.1. Übersicht*

*Übersicht 12.2: Schätzprobleme bei den Kennzahlen*

| Daten / Masse | Zeitmengenfläche $F_{om}$ und damit Schätzung des Durchschnittsbestands | Verweilsumme $\Sigma d_i$ und damit Schätzung der durchschnittlichen Verweildauer |
|---|---|---|
| 1. Längsschnitt<br>a) geschlossen<br>b) offen | $F_{om}$ kein Problem, da Bestandsfunktion B(t) für jedes t bekannt | a) $\Sigma d_i$ identisch mit $F_{om}$<br>b) $\Sigma d_i$ aus $F_{om}$ zu schätzen mit Gl. 12.6 |
| 2. Querschnittsdaten | zu schätzen mit Gl. 12.10, da B(t) nur zu bestimmten Zeitpunkten bekannt ist | Übergang von $F_{om}$ zu $G_{om}$ wie im Fall 1b, Annahmen über Auf- und Abbauzeiten nötig |

### *2.3.2. Zeitmengenfläche und Durchschnittsbestand*

Wenn die Bestandsänderungen ausschließlich genau zu den Beobachtungszeitpunkten $t_j$ (j=1,2,...,m) stattfinden, dann ist die Zeitmengenfläche gegeben durch Gl. 12.9. :

Sind die Beobachtungszeitpunkte $t_j$ (mit j = 1,2,...,m) äquidistant, so daß $t_j$-$t_{j-1}$ = 1 (für alle j) und $t_m$-$t_o$ = m, so gilt Gl. 12.10, woraus dann $\bar{B}$ mit Gl. 12.3 zu errechnen ist:

| | |
|---|---|
| (12.9) $F_{om} = \sum_j B_{j-1}(t_j - t_{j-1})$ | (12.10) $F_{om} = \frac{1}{2}B_o + B_1 + \ldots + B_{m-1} + \frac{1}{2}B_m$ |

### *2.3.3. Durchschnittliche Verweildauer und Umschlagshäufigkeit*

Es ist analog zu Gl. 12.6 vorzugehen, d.h. $G_{om}$ ist die geeignete Schätzung von $\Sigma d_i$. Es sind jetzt Annahmen über die mittlere Aufbau- ($d_o$) und Abbauzeit ($d_m$) nötig. üblich ist die Annahme

$$\bar{d}_o = \delta\bar{d} \text{ und } d_m = (1-\delta)\bar{d} \text{ mit } 0 < \delta < 1$$

Eingesetzt in (12.6) liefert das Gl. 12.11 und für die sehr verbreitete Annahme $\delta$ = ½ die sehr bekannte Formel von Gl. 12.12:

| Geschätzte durchschnittliche Verweildauer bei offener Masse | Spezielle Annahme $\delta$ = ½ für die durchschnittliche Auf- und Abbauzeit |
|---|---|
| (12.11) $\bar{d} = \dfrac{F_{om}}{\delta Z_{om} + (1-\delta)A_{om}}$ | (12.12) $\bar{d} = \dfrac{2m\bar{B}}{Z_{om} + A_{om}}$ |

## 3. Stationäre Bevölkerung und Tafelrechnung

**Def. 12.7: (Kohorte, Abgangsordnung, stationäre Bevölkerung)**

a) Eine Zugangskohorte oder einfach **Kohorte** ist die Gesamtheit der gleichzeitig (zum gleichen Zeitpunkt $t_j$, bzw. im gleichen Intervall geringer Länge $[t_{j-1},t_j]$) zugehenden Einheiten. Der Umfang dieser Masse, d.h. die Anzahl der zugehenden Einheiten ist $l_o$.

b) Die **Abgangsordnung** $l_x$ (wobei x = 0, 1,..., $\omega$ das Alter, d.h. die Anzahl der vollendeten Jahre ist) ist die Anzahl der Überlebenden des Alters x. Es ist der Restbestand einer Geburtskohorte des Umfangs $l_o$ nach Vollendung von x Jahren.

c) Bei einer **stationären Bevölkerung** (Sterbetafelbevölkerung) wird jede Kohorte (jeder Geburtsjahrgang) in jedem aufeinanderfolgenden Intervall (in allen folgenden Jahren) durch eine

- **gleich große** Kohorte (so daß $Z_{j-1,j} = l_o$ für alle j)
- mit gleicher Abgangsordnung (d.h. **gleicher "Struktur"** ersetzt

so daß $l_x$ nicht von j sondern nur von x abhängig ist. Es wird im Modell von Wachstum und Strukturveränderung abstrahiert.

d) Bei einer **stabilen Bevölkerung** wird jede Kohorte durch eine $\lambda$–fache Kohorte ersetzt, so daß gilt: $Z_{01} = l_o$ , $Z_{12} = \lambda l_o$ $Z_{23} = \lambda^2 l_o$ ... (also weiter keine Strukturveränderung aber es ist Wachstum in der einfachsten Form [mit konstanter Wachstumsrate $\lambda$–1] zugelassen).

**Def. 12.8 (Tafelfunktionen $q_x$, $p_x$, $l_x$, $L_x$):**

a) Die einjährige **Sterbewahrscheinlichkeit** $q_x$ der x-jährigen ist die (bedingte) Wahrscheinlichkeit dafür, daß eine Person, die das Alter von x erreicht hat, das Alter von x+1 nicht mehr erreichen wird (mit x = 0,1,...,w für das Alter in vollendeten Jahren).

b) Die einjährige **Überlebenswahrscheinlichkeit** $p_x$ ist demzufolge $p_x = 1 - q_x$; $p_x$ ist eine bedingte Wahrscheinlichkeit.

c) **Sämtliche Sterbetafelfunktionen** sind allein Funktionen des Alters x und sie sind mit der Folge der Sterbewahrscheinlichkeiten $q_x$ und dem willkürlich gewählten Anfangsbestand (Geburten) $l_o$ eindeutig gegeben:

- die **Absterbeordnung** $l_x$ ist ausgehend von einem fiktiven Anfangsbestand von $l_o$ = 100.000 Personen **rekursiv** zu berechnen mit

  (12.18) $l_{x+1} = l_x p_x = l_x(1-q_x)$.

- Entsprechend ist die Anzahl $d_x$ der im Altersintervall (x,x+1) gestorbenen Personen

  (12.19) $d_x = l_x q_x = l_x - l_{x+1} \geq 0$.

d) Mit $L_x$ wird die Anzahl der von allen Überlebenden x-jährigen Personen bis zum Alter x+1 durchlebten Jahre (die Anzahl der im Intervall (x,x+1) verlebten Personenjahre) bezeichnet.

(12.20) $$L_x = \frac{1}{2}\left(l_x + l_{x+1}\right).$$

Man kann die Größe $L_x$ als Zeitmengenfläche interpretieren und auch als Mittelwert von $l_{x+1}$ Personen, die noch 1 Jahr leben werden und $d_x = l_x + l_{x+1}$ Personen, die noch 1/2 Jahr leben werden, denn $\frac{1}{2}\left(l_x - l_{x+1}\right) + 1 \cdot l_{x+1} = \frac{1}{2}\left(l_x + l_{x+1}\right)$.

**Def. 12.9: (Tafelfunktionen $T_x$, $e_x$)**

a) Die Tafelfunktion $T_x$, die Zahl der von den Überlebenden des Alters x noch zu durchlebenden Jahre ist die Summe der Größen $L_x, L_{x+1}, L_{x+2},...,L_w$.

(12.24) $$T_x = \sum_{y=x}^{w} L_y \qquad x \leq y \leq w.$$

(12.25) $$T_x^* = \sum_{y=x}^{w} l_y = T_x + \tfrac{1}{2} l_x$$

Die Größen $T_x$ und $T_x^*$ sind Verweilsummen; Maßeinheit: "Personenjahre".

b) Dividiert man $T_x$ bzw. $T_x^*$ durch die Anzahl der Überlebenden des Alters x, also durch $l_x$, so erhält man mit

(12.26) $$e_x = \frac{T_x}{l_x} = \frac{T_x^*}{l_x} - \frac{1}{2} = e_x^* - \frac{1}{2}$$

die (mittlere, durchschnittliche weitere) Lebenserwartung einer x-jährigen Person (spricht man von "der" Lebenserwartung, so ist $e_o$ gemeint).

**Beispiel einer Klausuraufgabe [Rechenaufgabe nicht Multiple ChoiceAufgabe]**

| 15 | 07 | 04 | |
|---|---|---|---|

a) Bei einer Sterbetafelbevölkerung gelte fürdie einjährige Sterbewahrscheinlichkeit von Frauen

$$q_x^w = \frac{1}{100 - x}$$

(x = Alter = Anzahl der vollendeten Lebensjahre, x = 0,1,...99)
Aus der Gestalt der Abgangsordnung ergibt sich, daß hier der folgende Spezialfall einer Abgangsordnung vorliegt:

Dabei ist die Anzahl der jedes Jahr (im Alter x = 0, 1, ...) gestorbenen Frauen $d_x$ =

b) bei welcher der beiden Absterbeordnungen ist die Lebenserwartung größer?

| Absterbeordnung 1 | |
|---|---|
| x = 0 | $l_x = l_0 = 50$ |
| x = 1 | $l_1 = 40$ |
| x = 2 | $l_2 = 30$ |
| x = 3 | $l_3 = 0$ |

| Absterbeordnung 2 | |
|---|---|
| x = 0 | $l_0 = 100$ |
| x = 1 | $l_1 = 60$ |
| x = 2 | $l_2 = 40$ |
| x = 3 | $l_3 = 0$ |

und wie groß ist $e_0$ in den beiden Fällen?