

Über die Berechnung von Durchschnittswerten

Von Dr. H. Wyss, Mathematiker der eidg. Versicherungskasse

I.

Die beobachtende Statistik untersucht ein bestimmt umschriebenes Material nach zwei Merkmalen und ordnet das eine anhand des zweiten. Im allgemeinen braucht sich die Gliederung nicht auf zwei Merkmale zu beschränken, doch kann die Unterscheidung von mehreren jeweils als Kette von zweidimensionalen Darstellungen aufgefasst und auf diesen einfachen Fall zurückgeführt werden. Bei Unterscheidung von zwei Merkmalen kann prinzipiell sowohl das eine wie das andere als Ordnungsgrundlage auftreten. Praktisch wird sich aber meist das eine als ausgesprochenes Ordnungsmerkmal darbieten und zur Gliederung des Materials dienen. Zu diesem ersten Merkmal, das die Rolle der unabhängigen Basis spielt, steht das zweite, beobachtete Merkmal in einer gewissen Abhängigkeit. Seine fortlaufende Zuordnung zum Ordnungsmerkmal ergibt eine für das Material charakteristische Beobachtungsreihe.

Bildlich lässt sich eine solche Gliederung etwa darstellen, indem das Ordnungsmerkmal mit einer bestimmten Streckeneinheit als Basis aufgetragen und jeder Stufe das abhängige Beobachtungsmerkmal als Höhe zugeordnet wird.

Soll beispielsweise aus einem Personalbestande die Lohnsumme nach dem Alter des Arbeiters geordnet werden, so tritt das Alter als Ordnungsmerkmal, die Lohnhöhe als Beobachtungsmerkmal auf. Es entsteht, gestützt auf die Feststellung der schweizerischen Unfallversicherungsanstalt, dass

ein Arbeiter vom Alter	einen durchschnittlichen Taglohn bezieht von Fr.
25	10, 96
26	11, 16
27	11, 34
28	11, 51
29	11, 67
30	11, 82
31	11, 96
32	12, 08
33	12, 20
34	12, 30
35	12, 40

die nachstehende graphische Darstellung (Fig. 1):

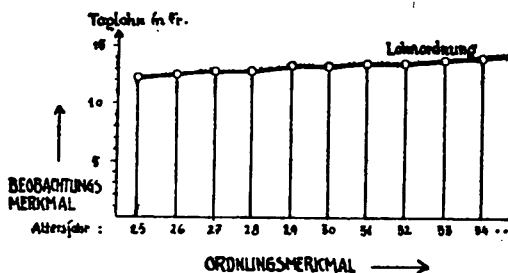


Fig. 1

Die Reihe des vom Ordnungsmerkmal abhängigen Beobachtungsmerkmals (in der Figur durch die ausgezogene Linie hervorgehoben) stellt die Lohnordnung des untersuchten Personalbestandes dar.

II.

Es können dabei zwei verschiedene Arten von statistischen Beobachtungen unterschieden werden, die zu zwei besonderen Darstellungstypen führen:

Typus A: Fortgesetzte Beobachtung an ein und demselben Material zeichnet die Veränderung des Beobachtungsmerkmals von der ersten zur letzten Messung auf, liefert also eine *Entwicklungsaufnahme*.

Typus B: Eine einzige Beobachtung gliedert den Materialbestand nach einem Ordnungsmerkmal und liefert eine *Gliederungsaufnahme*.

Jedes Beobachtungsergebnis lässt sich einem dieser zwei Typen unterordnen und entweder als Entwicklungsaufnahme oder als Gliederungsaufnahme darstellen. Anhand von zwei Beispielen sollen vorerst einige Besonderheiten dieser beiden Darstellungstypen festgestellt werden.

A. Beispiel einer Entwicklungsaufnahme. Das eidgenössische Getreideamt beherbergte während des Jahres 1926 in seinen Lagerhäusern folgende Getreidemengen:

Am 1. Januar	53.000 Tonnen	= l_0
Ende Januar	64.000 »	= l_1
» Februar	60.000 »	= l_2
» März	49.000 »	= l_3
» April	37.000 »	= l_4
» Mai	33.000 »	= l_5
» Juni	17.000 »	= l_6
» Juli	33.000 »	= l_7
» August	36.000 »	= l_8
» September	28.000 »	= l_9
» Oktober	41.000 »	= l_{10}
» November	43.000 »	= l_{11}
» Dezember	49.000 »	= l_{12}
Summe . . .	543.000 Tonnen	= $\sum_{t=0}^{12} l_t$

Während des ganzen Jahres wurde in Abständen von 1 Monat die Lagermenge festgestellt. Das Ergebnis gibt also eine Entwicklungsaufnahme. Ihre graphische Darstellung ist aus Figur 2 ersichtlich:

Die Lagermenge ist nur in Abständen von ganzen Monaten bekannt; für zwischenliegende Zeiten wurde sie nicht beobachtet. Ist es wünschbar, dennoch ihre kontinuierliche Entwicklung zu verfolgen, so kann die Veränderung von einem Stande zum nächstfolgenden als gleichmässig, als geradlinig gedacht werden. Dann wird näherungsweise die Entwicklung der Lagermenge durch die ausgezogene Linie der Figur 2 dargestellt. Aus dieser Aufzeichnung ist auch ersichtlich, welche Getreidemenge während des ganzen Jahres 1926 in den Lagerhäusern lag. Diese wird nämlich dargestellt durch die Fläche, die begrenzt ist von der Entwicklungslinie zwischen der ersten und letzten Beobachtung. Geometrisch setzt sie sich zusammen aus 12 Trapezen von der Breite 1, deren parallele Seiten jeweils durch Beobachtung bekannt sind.

Aus der Figur 2 kann also sofort die algebraische Darstellung der gelagerten Menge abgelesen werden. Bezeichnet nämlich, wie in der Aufstellung oben bereits angedeutet wurde, l_t den Lagerbestand im Zeitpunkt t , dann ist der Inhalt des t^{ten} Trapezes gegeben durch:

$$\frac{l_{t-1} + l_t}{2}$$

Die Gesamtfläche aller Trapeze, welche die im Jahre beherbergte Lagermenge darstellt, wird also berechnet durch:

$$L = \sum_{t=1}^{12} \frac{l_{t-1} + l_t}{2} = \sum_{t=0}^{12} l_t - \left(\frac{l_0}{2} + \frac{l_{12}}{2} \right)$$

Nach Einsetzung der Zahlwerte aus der oben angegebenen Feststellung lässt sich im vorliegenden Beispiel die Lagermenge für das Jahr 1926 berechnen:

	$\sum_{t=0}^{12} l_t$	= 543.000 Tonnen
abzüglich:		
halber Anfangsbestand	$= \frac{l_0}{2} = 26.500$	
» Endbestand	$= \frac{l_{12}}{2} = 24.500$	
	$= L$	$= 492.000 \text{ Tonnen}$

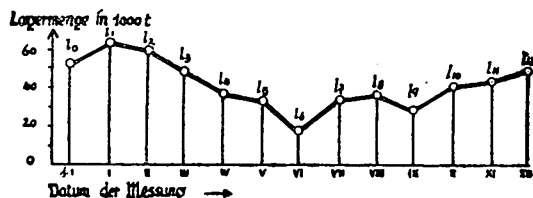


Fig. 2. Entwicklung der Lagermenge in eidgenössischen Getreidelagern, 1926

Durch diese Darlegung wird der häufige Irrtum, als Lagermenge einfach die Summe der je auf Monatsende festgestellten Lagerbestände (im vorliegenden Fall 490.000 Tonnen) zu betrachten, ausgeschlossen.

B. Beispiel einer Gliederungsaufnahme. Auf Ende 1926 wurde die Dienstaltersverteilung der schweizerischen Grenzwächter festgestellt. Als Ordnungsmerkmal diene dabei die Zahl der zurückgelegten Dienstjahre, als Beobachtungsmerkmal die zugehörige Personenzahl. Werden näherungsweise die Personen innerhalb einer Altersgruppe als gleichmässig verteilt vorausgesetzt, so ist das mittlere Dienstalter der Gruppe mit n zurückgelegten Dienstjahren (also für die zwischen dem n^{ten} und $(n + 1)^{\text{ten}}$ Dienstalter Stehenden) $n + \frac{1}{2}$ Jahre. Die Beobachtungen im Bestande der Grenzwächter ergaben auf Ende 1926 folgende Dienstaltersverteilung:

zurückgelegte Dienstjahre	Dienstalter	Anzahl Grenzwächter
t	$t + \frac{1}{2}$	$k_{t+\frac{1}{2}}$
0	$\frac{1}{2}$	4 = $k_{\frac{1}{2}}$
1	$1\frac{1}{2}$	49 = $k_{1+\frac{1}{2}}$
2	$2\frac{1}{2}$	42 = $k_{2+\frac{1}{2}}$
3	$3\frac{1}{2}$	31 = $k_{3+\frac{1}{2}}$
4	$4\frac{1}{2}$	35 = $k_{4+\frac{1}{2}}$
5	$5\frac{1}{2}$	68 = $k_{5+\frac{1}{2}}$
6	$6\frac{1}{2}$	106 = $k_{6+\frac{1}{2}}$
7	$7\frac{1}{2}$	184 = $k_{7+\frac{1}{2}}$
8	$8\frac{1}{2}$	310 = $k_{8+\frac{1}{2}}$
9	$9\frac{1}{2}$	86 = $k_{9+\frac{1}{2}}$
10	$10\frac{1}{2}$	29 = $k_{10+\frac{1}{2}}$
—	—	—
—	—	—
—	—	—
Total (0—55)		$1947 = \sum_{t=0}^{55} k_{t+\frac{1}{2}}$

Wird in jedem aufgeführten Dienstalter $\left(t + \frac{1}{2}\right)$ die entsprechende Personenzahl $k_{t+\frac{1}{2}}$ als Höhe aufgetragen, so ergibt sich nachstehende graphische Darstellung (Fig. 3, Seite 374).

Der Voraussetzung gleichmässiger Verteilung der Personen auf die Dienstaltersgruppe entspricht die Darstellung durch Rechtecke. Die aus den Rechtecken zusammengesetzte Beobachtungsfläche kommt der Gesamtzahl der Ende 1926 beobachteten Grenzwächter gleich.

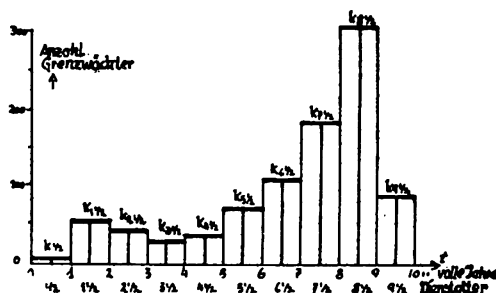


Fig. 3. Gliederung der schweizerischen Grenzwächter nach dem Dienstalter, Ende 1926

Wird zur algebraischen Darstellung, wie in der Tabelle oben bereits angedeutet, die beobachtete Zahl für das Ordnungsmerkmal t mit k_t bezeichnet, so wird der Inhalt der ganzen Beobachtungsfläche als Summe der Rechtecke berechnet nach:

$$K = \sum_{t=0}^{\omega} k_{t+\frac{1}{2}}$$

wo ω das höchste Ordnungsmerkmal bedeutet.

Im gewählten Beispiele ist

$$K = \sum_{t=0}^{55} k_{t+\frac{1}{2}} = 1947,$$

die Gesamtzahl der Grenzwächter.

III.

Zwischen den beiden Aufnahmetypen besteht ein Zusammenhang, so dass der eine in den andern übergeführt werden kann, vorausgesetzt, dass die nötigen Einzelheiten der Beobachtung bekannt sind. So ist es beispielsweise möglich, eine Gliederungsaufnahme durch Rekonstruktion in eine Entwicklungsaufnahme umzuformen. Eine solche Umwandlung wird andeutungsweise am zweiten Beispiele des vorigen Abschnittes gezeigt.

Dort wurde festgestellt, dass Ende 1926

$k_{t+\frac{1}{2}}$ Grenzwächter im Dienstalter $t + \frac{1}{2}$ stehen

$k_{t+1+\frac{1}{2}}$ » » » $t + 1 + \frac{1}{2}$ usw.

Wird aber als Ordnungsmerkmal nicht das Dienstalter, sondern jeweils das rekonstruierte Stichjahr gewählt, dann lässt sich das Beobachtungsmaterial anders ordnen.

Wenn sich nämlich die Eintritte gleichmässig über das Jahr verteilen, wie es der oben getroffenen Voraussetzung entspricht, dann gehörten von den Ende 1926 Beobachteten der Grenzpolizei an:

vor t Jahren	also auf Jahresende	? Grenzwächter	
0	1926	l_ω	(alle Beobachteten) $= \sum_{t=0}^{\omega} k_{t+\frac{1}{2}}$
1	1925	$l_{\omega-1}$	(alle ohne die 1926 Eingetretenen) $= \sum_{t=1}^{\omega} k_{t+\frac{1}{2}}$
2	1924	$l_{\omega-2}$	$= \sum_{t=2}^{\omega} k_{t+\frac{1}{2}}$
..	...		..
..	...		..
t	$1926 - t$	$l_{\omega-t}$	$= \sum_t^{\omega} k_{t+\frac{1}{2}}$
..	...		..
$\omega-1$	$1926 - (\omega-1)$	l_1	(bloss die Dienst- ältesten) $= \sum_{t=\omega-1}^{\omega} k_{t+\frac{1}{2}}$
ω	$1926 - \omega$	l_0	(keine) $= \sum_{\omega}^{\omega} k_{t+\frac{1}{2}} = 0$

Wird nach dieser Aufstellung mit dem Stichjahre als Ordnungsmerkmal die Figur konstruiert, wobei gemäss der Voraussetzung linearer Verteilung der Eintritte über das Jahr die aufgetragenen Höhen geradlinig verbunden werden können, so ergibt sich die *Entwicklungsdarstellung* in Figur 4:

Die ausgezogene Linie stellt die Entwicklung des Endbestandes dar, der durch die Endordinate l_ω wiedergegeben wird. Dessen Aufteilung in Gruppen von Personen vom Dienstalter $\frac{1}{2}$, $1\frac{1}{2}$, $2\frac{1}{2}$ usw. wird in der Figur durch die Projektionen der frühern Ordinaten $l_{\omega-1}$, $l_{\omega-2}$ usw. auf die Endordinate angedeutet.

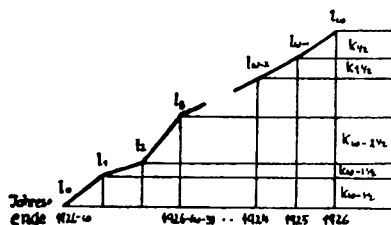


Fig. 4

Die Überführung der *Gliederungsaufnahme* in eine Entwicklungsaufnahme ist also in allgemeiner Form wirklich gelungen. Dabei ist aber zu bemerken, dass die rekonstruierte Entwicklung natürlich bloss jene Elemente berücksichtigt, die in deren Endpunkt dem Materiale noch angehören, nicht aber solche, die wohl vor einigen Jahren zum Bestande gehörten, seither aber ausgeschieden sind.

Die Umformung wurde im vorliegenden Beispiele für ein Zeitmass als Ordnungsmerkmal durchgeführt. Es ist aber leicht einzusehen, dass die Möglichkeit einer Umwandlung ebenso für ein ganz beliebiges Ordnungsmerkmal besteht, wenn auch in manchen Fällen die Rekonstruktion der Entwicklung keinen praktischen Sinn bekommt.

IV.

Im soeben behandelten Beispiel (Figur 4) stellt die letzte Beobachtung l_ω die Gesamtzahl der Grenzwächter auf Ende 1926 dar. Die ganze Fläche aber, die Summe aller Trapeze:

$$L = \sum_{t=0}^{\omega} l_t - \left(\frac{l_0}{2} + \frac{l_\omega}{2} \right)$$

oder wegen $l_0 = 0$

$$L = \sum_{t=1}^{\omega} l_t - \frac{l_\omega}{2}$$

gibt die vom Endbestande im ganzen zugebrachte Dienstzeit an.

Es soll nun auch aus der nämlichen, aber als Gliederungsaufnahme dargestellten Beobachtung die gesamte Dienstzeit berechnet werden.

In jener Darstellung (vgl. Fig. 3) wird durch die Fläche als Summe der Rechtecke die Zahl der Mitglieder wiedergegeben.

Zur Ermittlung der gesamten Dienstzeit des Bestandes kommt jedem Rechteck verschiedene Wichtigkeit zu. Es umfasst nämlich das erste Rechteck jene Wächter, die durchschnittlich $\frac{1}{2}$ Jahr im Dienste gestanden sind; die Gruppe des zweiten Rechteckes hat schon $1\frac{1}{2}$ Dienstjahre zurückgelegt, die Gruppe des dritten schon $2\frac{1}{2}$ usw. Zur Berechnung der ganzen zurückgelegten Dienstzeit muss also das zweite Rechteck stärker ins Gewicht fallen als das erste, das dritte aber stärker als das zweite usw. Den verschiedenen Rechtecken muss ein statistisches Gewicht beigeordnet werden, ihrer Wichtigkeit für die betreffende Rechnung entsprechend. Im vorliegenden Beispiel dient das Dienstalder der jeweiligen Gruppe als Gewicht.

Algebraisch lässt sich die Summe der auf solche Weise gewogenen Rechtecke darstellen durch:

$$L^1 = \sum_{t=0}^{\omega} \left(t + \frac{1}{2} \right) k_{t+\frac{1}{2}}$$

Dieser Wert soll der soeben aus der rekonstruierten Entwicklungsdarstellung berechneten gesamten Dienstzeit L gleichkommen.

Dieser Zusammenhang muss durch algebraische Umformung nachweisbar sein:

Der Ausdruck für L^1 kann umgewandelt werden:

$$\begin{aligned} L^1 &= \sum_{t=0}^{\omega} \left(t + \frac{1}{2} \right) k_{t+\frac{1}{2}} \\ &= \frac{1}{2} k_{\frac{1}{2}} + 1 \frac{1}{2} k_{1+\frac{1}{2}} + 2 \frac{1}{2} k_{2+\frac{1}{2}} + \dots + \left(\omega - \frac{1}{2} \right) k_{\omega-\frac{1}{2}} \end{aligned}$$

$$\begin{aligned}
&= k_{1+\frac{1}{2}} + 2 k_{2+\frac{1}{2}} + 3 k_{3+\frac{1}{2}} + \dots + (\omega - 1) k_{\omega-\frac{1}{2}} \\
&\quad + \frac{1}{2} \left[k_{\frac{1}{2}} + k_{1+\frac{1}{2}} + k_{2+\frac{1}{2}} + \dots + k_{\omega-\frac{1}{2}} \right] \\
&= \sum_{t=1}^{\omega} k_{t+\frac{1}{2}} + \sum_{t=2}^{\omega} k_{t+\frac{1}{2}} + \dots + \sum_{t=\omega-1}^{\omega} k_{t+\frac{1}{2}} \\
&\quad + \frac{1}{2} \sum_0^{\omega} k_{t+\frac{1}{2}}
\end{aligned}$$

Aus der Aufstellung des vorigen Abschnittes IV geht hervor, dass nach der gewählten Bezeichnungsweise gilt:

$$l_{\omega-t} = \sum_t^{\omega} k_{t+\frac{1}{2}}$$

Wird diese Einsetzung durchgeführt, so entsteht:

$$L^1 = l_{\omega-1} + l_{\omega-2} + \dots + l_1 + \frac{1}{2} l_{\omega}$$

oder:

$$L^1 = \sum_{t=1}^{\omega} l_t - \frac{1}{2} l_{\omega}$$

Die rechte Seite dieser Gleichung ist aber nach dem früheren Ergebnis $= L$, so dass wirklich besteht:

$$L^1 = L$$

V.

Aus den bisherigen Ergebnissen der vorliegenden Untersuchung kann über die beiden für die Darstellung statistischer Beobachtungen zu unterscheidenden Typen folgende Charakterisierung zusammengestellt werden.

A. *Entwicklungsaufnahme*: kommt zustande durch fortgesetzte Beobachtungen; führt zu einer Beobachtungsfläche, die sich aus Trapezen vom gleichen statistischen Gewicht zusammensetzt.

B. *Gliederungsaufnahme*: ordnet die Ergebnisse einer einzigen Beobachtung nach einem Ordnungsmerkmal: führt zu einer Beobachtungsfläche aus Rechtecken, die in bezug auf das Ordnungsmerkmal verschiedene statistische Gewichte besitzen.

Die Gliederungsaufnahme kann durch Rekonstruktion in eine Entwicklungsaufnahme umgeformt werden, so dass deren Trapezfläche der gewogenen Recht-

ecksfläche gleichkommt. Mit genügend detaillierten Beobachtungselementen ist auch umgekehrt die Umwandlung einer Entwicklungsaufnahme in eine Gliederungsaufnahme möglich.

VI.

Liegen nun die Ergebnisse verschiedener Beobachtungen vor, die miteinander verglichen werden sollen, so erweisen sich die unregelmässigen, aus Rechtecken oder Trapezen zusammengefügtten Beobachtungsflächen zu Gegenüberstellungen als ungeeignet. Da nur Verschiedenheiten in einer Richtung einfach abgeschätzt werden können, müssen die Beobachtungsflächen vorerst auf eine vergleichbare Form gebracht, am besten in ein inhaltsgleiches Rechteck umgewandelt werden. Wird dann die eine Ausdehnung der Vergleichsrechtecke übereinstimmend gewählt, so sind sie bloss noch nach der anderen Dimension zu vergleichen, was nun ohne weiteres geschehen kann. Die nicht gewählte Ausdehnung des Vergleichsrechtecks stellt aber jeweilen einen für die Beobachtung charakteristischen Durchschnittswert dar.

Wird vorerst die Vereinfachung einer Entwicklungsaufnahme in Betracht gezogen, so ist nach den Ausführungen weiter oben die Beobachtungsfläche bekannt als L . Der Inhalt des zu Vergleichen gesuchten Rechteckes muss eben so gross sein. Da aber weder über die Höhe noch über die Basis eine Vorschrift besteht, sind eine ganze Reihe von Rechtecken denkbar, die alle die Bedingung erfüllen, dass ihr Inhalt L ist. Von all diesen Kombinationen sind aber bloss deren zwei gebräuchlich: Entweder wird als Seitenbedingung für das Vergleichsrechteck die höchste in der Beobachtung verzeichnete Ausdehnung des Ordnungsmerkmals oder der Höchstwert des Beobachtungsmerkmals gewählt.

Im ersten Falle wird die Basis des Rechteckes der ganzen Ausdehnung der Beobachtungsordnung gleichgesetzt. Dann stellt die Höhe des Rechteckes vom vorgeschriebenen Inhalt L das mittlere Beobachtungsmerkmal dar.

Im zweiten Falle wird als Höhe des Rechteckes die grösste Beobachtung festgesetzt. Seine Basis gibt dann das mittlere Ordnungsmerkmal wieder.

Die Umwandlung der Darstellungsfläche einer Gliederungsaufnahme kann nicht in so einfacher Weise durchgeführt werden, da sie sich ja aus Rechtecken von verschiedenem Gewicht zusammensetzt. Doch ist ihre Umformung in eine Entwicklungsdarstellung und deren weitere Vereinfachung bekannt. Es besteht also auch die Möglichkeit, eine Gliederungsaufnahme zu reduzieren.

Die algebraische Formulierung dieser Ergebnisse verlangt die Einführung einiger Bezeichnungen:

l_H sei der höchste Beobachtungswert der Entwicklungsaufnahme,

T das grösste in Betracht gezogene Ordnungsmerkmal,

L wie früher die Beobachtungsfläche der Entwicklungsdarstellung. Der Inhalt des Vergleichsrechteckes muss dann ebenfalls L sein.

Im ersten Falle wird als dessen Basis T gewählt, so dass seine Höhe l_d (das mittlere Beobachtungsmerkmal) bestimmt wird durch:

$$l_d = \frac{L}{T}$$

Im zweiten Falle wird als Höhe des Rechteckes l_H gewählt, so dass dessen Basis t_d (das mittlere Ordnungsmerkmal) dargestellt wird als:

$$t_d = \frac{L}{l_H}$$

Liegen die Beobachtungsergebnisse in Form einer Gliederungsaufnahme vor, so wird die Berechnung der entsprechenden Durchschnittswerte folgendermassen erreicht:

Die Beobachtungsfläche misst ohne Berücksichtigung der statistischen Gewichte $K = \sum_{t=0}^{\omega} k_{t+\frac{1}{2}}$; die gewogene Fläche, welche nach den früheren Ausführungen der Beobachtungsfläche der rekonstruierten Entwicklungsdarstellung entspricht, misst $L = \sum_{t=0}^{\omega} \left(t + \frac{1}{2}\right) k_{t+\frac{1}{2}}$. Der höchste beobachtete Wert der rekonstruierten Aufnahme ist deren Endwert:

$$l_H = l_{\omega} = \sum_{t=0}^{\omega} k_{t+\frac{1}{2}} = K$$

Das grösste berücksichtigte Ordnungsmerkmal ist wieder T . Diese Einsetzungen, die aus der bereits eingeführten Bezeichnungsweise folgen, berücksichtigen nun schon die notwendige Umformung der Gliederungsaufnahme, so dass für sie die Mittelwerte wie folgt zu berechnen sind:

Mittleres Beobachtungsmerkmal =

$$l_d = \frac{L}{T} = \frac{\sum_{t=0}^{\omega} \left(t + \frac{1}{2}\right) k_{t+\frac{1}{2}}}{T}$$

Mittleres Ordnungsmerkmal =

$$t_d = \frac{L}{l_H} = \frac{L}{K} = \frac{\sum_{t=0}^{\omega} \left(t + \frac{1}{2}\right) k_{t+\frac{1}{2}}}{\sum_{t=0}^{\omega} k_{t+\frac{1}{2}}}$$

Die Beziehung zwischen mittlerem Beobachtungsmerkmal und mittlerem Ordnungsmerkmal ist allgemein hergestellt durch:

$$l_d \cdot T = t_d \cdot l_H$$

Zur Erläuterung wird das im Abschnitt II angeführte Beispiel einer Entwicklungsaufnahme herangezogen (Getreidelagerung 1926).

Wie dort bereits berechnet wurde, ist $L = 492.000$ Tonnen. Die Aufstellung zeigt, dass die höchste Lagermenge beträgt $l_H = l_1 = 64.000$. Das grösste Ordnungsmerkmal ist $T = 12$. Das mittlere Beobachtungsmerkmal (die mittlere Lagermenge) berechnet sich demnach zu:

$$l_A = \frac{L}{T} = \frac{492.000}{12} = 41.000 \text{ Tonnen.}$$

Das mittlere Ordnungsmerkmal (die mittlere Lagerdauer) beträgt:

$$t_A = \frac{L}{l_H} = \frac{492.000}{64.000} = 7,7 \text{ Monate.}$$

An diesem Beispiele wird auch besonders deutlich, weshalb als Seitenbedingung des Vergleichsrechteckes die Höchstwerte der Merkmale verwendet werden. Die mittlere Lagermenge beantwortet die Frage: Welche gleichbleibende Getreidemenge muss während der ganzen Beobachtungszeit gelagert werden, so dass im ganzen die beobachtete Menge L beherbergt wird? In dieser Frage ist eben schon die ganze Beobachtungszeit T als Bedingung enthalten.

Die mittlere Lagerdauer aber löst die Frage: Wie lange muss das Lager gefüllt sein, bis es die beobachtete Menge L beherbergt hat? Als Bedingung in dieser Frage tritt die Menge auf, welche das Lager füllt; dies ist aber offenbar die höchst beobachtete Menge l_H . Denn so viel muss das Lager fassen, mehr braucht es (wenigstens während der in Betracht gezogenen Beobachtungszeit) nicht zu fassen.

Die Bildung der besprochenen Durchschnittswerte kann schematisch folgendermassen dargestellt werden:

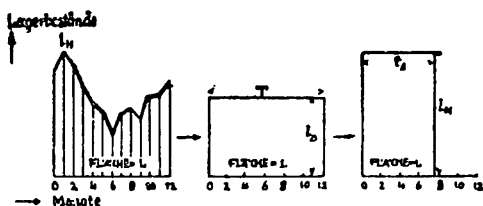


Fig. 5. Beobachtete Fläche $\longrightarrow$ umgewandelte Fläche $\longrightarrow$ umgewandelte Fläche
 (Höhe des Rechteckes $\longrightarrow$ Basis des Rechteckes
 $=$ mittlere Lagermenge) $\longrightarrow$ $=$ mittlere Lagerdauer)

VII.

Aus den bisherigen Ausführungen geht hervor, dass im allgemeinen der Wert des mittleren Merkmales nicht durch ein einfaches arithmetisches Mittel aus den Beobachtungen dargestellt wird. Es hat sich nämlich ergeben, dass die Fläche L einer Entwicklungsaufnahme aus einer Summe von Trapezen besteht und zu den Beobachtungen in der Beziehung steht:

$$L = \sum_{t=0}^{\omega} l_t - \left(\frac{l_0}{2} + \frac{l_{\omega}}{2} \right)$$

Die Durchschnittswerte einer Gliederungsaufnahme jedoch berechnen sich aus der gewogenen Fläche nach der Beziehung:

$$L = \sum_{t=0}^{\omega} \left(t + \frac{1}{2} \right) k_{t+\frac{1}{2}}$$

(wo l_t bzw. k_t gemäss der früheren Bezeichnungsweise die Beobachtungsergebnisse bedeuten).

Durch eine genaue Darstellung dieser an und für sich einfachen Verhältnisse können eine Reihe von häufig übersehenen Irrtümern vermieden werden. Wir möchten nur einen davon erwähnen, da er sich in dem verbreiteten Lehrbuche von Schär ¹⁾ findet und als Schema zu solchen Durchschnittsberechnungen benützt wird.

In jenem Werke (S. 298) soll nämlich die mittlere Lagerdauer berechnet werden aus Anfangsbestand, Endbestand und Eingang. Wir benützen zur Wiedergabe von Schärs Methode die oben benützten Bezeichnungen:

Eingangsinventur	= l_0
Schlussinventur	= l_{ω}
Durchschnittswert des Lagers.	= $\frac{1}{2} (l_0 + l_{\omega})$
Eingang im Jahr	= E
Einstandswert der verkauften Ware.	= $l_0 + E - l_{\omega}$

Dabei wurde als Beobachtungsmerkmal der Wert des Lagers, nicht die Lagermenge, wie in unserem Beispiele der Getreidelagerung, gewählt; dadurch wird aber die Berechnungsmethode nicht beeinflusst. Unter Benützung obiger Bezeichnungen lautet die von Schär angegebene Formel zur Berechnung der mittleren Lagerdauer:

$$t_d = \frac{365 \frac{l_0 + l_{\omega}}{2}}{l_0 + E - l_{\omega}} \text{ Tage}$$

Vorerst ist zu bemerken, dass das Rechnungsergebnis nicht genauer sein kann als die Messung der Beobachtungen. Aus bloss zwei Beobachtungen zum Beginn und Schluss des Jahres kann die mittlere Lagerdauer nur angenähert, jedenfalls nicht auf Tage genau berechnet werden.

Ferner braucht Schärs Formel bloss für einige charakteristische Fälle ausgewertet zu werden, um ihre Unvollkommenheit deutlich zu machen.

1. Fall: Das Lager öffne sich erst im Laufe des Jahres und werde wieder ausverkauft. Es ist also: $l_0 = l_{\omega} = 0$. Trotzdem offensichtlich das Lager eine Zeitlang belegt ist durch den Eingang E , und also eine mittlere Lagerdauer zu verzeichnen ist, ergibt Schärs Formel:

¹⁾ Schär, Buchhaltung und Bilanz (Berlin 1921).

$$t_d = \frac{365 \frac{0+0}{2}}{0+E-0} = \frac{0}{E} = 0$$

2. Fall: Weder Eingang noch Ausgang ist zu verzeichnen: $E = 0$. Anfangs- und Endbestand des Lagers sind also gleich: $l_0 = l_\omega$. Die mittlere Lagerdauer ist offensichtlich 1 Jahr. Schärs Formel aber ergibt:

$$t_d = \frac{365 \frac{l_0 + l_0}{2}}{l_0 + 0 - l_0} = \frac{365 l_0}{0} = \infty$$

Diese Hinweise mögen die Unbrauchbarkeit jener Formel illustrieren. Andererseits verstärkt ein solcher Irrtum gerade die Bedeutung einer einfachen Darstellung und Vergegenwärtigung der Verhältnisse, wie sie in den vorhergehenden Abschnitten versucht wurde.

VIII.

Für die Rechnung mit mittleren Merkmalen können aus den hergeleiteten Ergebnissen einige Grundsätze aufgestellt werden.

Setzen sich zwei oder mehrere Beobachtungsbestände zu einem neuen Bestande zusammen, so bildet dessen Beobachtungsfläche die Summe aus den entsprechenden Flächen der Einzelbeobachtungen. Es gilt also z. B. für zwei Bestände:

$$L = L_{(1)} + L_{(2)}$$

Die mittleren Werte werden dann dargestellt für das Beobachtungsmerkmal durch:

$$l_d = \frac{L}{T}; \quad l_{d(1)} = \frac{L_{(1)}}{T_{(1)}}; \quad l_{d(2)} = \frac{L_{(2)}}{T_{(2)}}$$

für das Ordnungsmerkmal durch:

$$t_d = \frac{L}{l_H}; \quad t_{d(1)} = \frac{L_{(1)}}{l_{H(1)}}; \quad t_{d(2)} = \frac{L_{(2)}}{l_{H(2)}}$$

Zwischen den mittleren Merkmalen der einzelnen und des gesamten Bestandes besteht also die Beziehung:

$$l_d T = l_{d(1)} T_{(1)} + l_{d(2)} T_{(2)}$$

bzw.

$$t_d l_H = t_{d(1)} l_{H(1)} + t_{d(2)} l_{H(2)}$$

Daraus folgt, dass das mittlere Merkmal des Gesamtbestandes nur dann gleich ist der Summe der entsprechenden Mittel für die Teilbestände, wenn sich die Durchschnittswerte auf gleiche Basis beziehen. Denn nur wenn

$$T = T_{(1)} = T_{(2)} \quad \text{bzw.} \quad l_H = l_{H(1)} = l_{H(2)}$$

gilt die einfache Beziehung

$$l_A = l_{A(1)} + l_{A(2)} \quad \text{bzw.} \quad t_A = t_{A(1)} + t_{A(2)}$$

In der Praxis wird es meist zutreffen, dass die Ausdehnung des Ordnungsmerkmals (z. B. die Dauer der Beobachtung) für Vergleichsmaterialien gleich gross ist, weil sie vom Beobachter einheitlich gewählt wird. Im allgemeinen wird tatsächlich

$$T = T_{(1)} = T_{(2)}$$

sein. Das mittlere Beobachtungsmerkmal besitzt dann die additive Eigenschaft, so dass gilt:

$$l_A = l_{A(1)} + l_{A(2)}$$

Im Gegensatz dazu kann das höchst beobachtete Merkmal nicht willkürlich gewählt werden, weshalb allgemein gilt:

$$l_H \neq l_{H(1)} \neq l_{H(2)}$$

Daraus folgt aber, dass das mittlere Ordnungsmerkmal die additive Eigenschaft im allgemeinen nicht besitzt, so dass gilt:

$$t_A \neq t_{A(1)} + t_{A(2)}$$

Vergleiche zwischen Durchschnittswerten verschiedener Beobachtungsmaterialien sind bloss stichhaltig, wenn sich die Mittel auf gleiche Basen beziehen. Aus dem soeben erreichten Ergebnis folgt, dass sich im allgemeinen Durchschnittswerte des Ordnungsmerkmals nicht zu Vergleichen eignen, wohl aber solche des Beobachtungsmerkmals.

Diese Folgerungen ergeben sich auch sofort aus der Summation bzw. Vergleichung der oben zur Darstellung benützten, zu Rechtecken reduzierten Beobachtungsflächen.

IX.

Bevor wir dazu übergehen, die Bedeutung der besprochenen Durchschnittswerte in einem allgemeinen Kollektiv darzustellen, möchten wir eine Bemerkung über die näherungsweise Integration beifügen, welche sich an die Ergebnisse des II. Abschnittes fügt.

Die kontinuierliche Betrachtungsweise, deren Anwendung in der mathematischen Statistik und der Versicherungsrechnung eine Reihe von besonders schönen Ergebnissen gebracht hat, ersetzt die einzelnen Beobachtungswerte durch eine stetige Funktion $f(t)$. Die Fläche, welche eine solche Funktion zwischen einer Anfangsgrenze a und einer obern Grenze b einschliesst, wird mathematisch dargestellt durch das Integral:

$$F = \int_a^b f(t) dt$$

In Fig. 6 werden diese Verhältnisse skizziert:

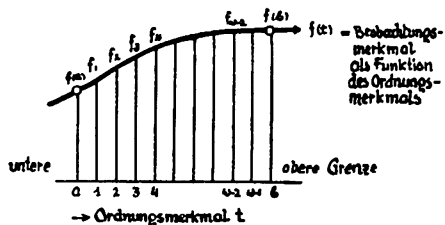


Fig. 6

Integrals näherungsweise als Fläche, aus deren Begrenzung nur einzelne Punkte bekannt sind, berechnet werden. In Figur 6 wurden solche beobachtete Punkte $f_a, f_{a+1}, f_{a+2} \dots f_b$ usw. schematisch angedeutet. Im Abschnitt II der vorliegenden Ausführungen hat sich ergeben, dass die bekannten Ordinaten f_t entweder als parallele Seiten von Trapezen oder als Höhen von Rechtecken auftreten, je nachdem es sich um eine Entwicklungs- oder eine Gliederungsaufnahme handelt. Aus diesen Flächenelementen kann die gesamte Fläche als Summe berechnet werden.

Aus den früheren Darstellungen folgen also zwei näherungsweise Lösungen für ein bestimmtes Integral:

1. Stellt $f(t)$ eine Entwicklung dar, dann gilt angenähert:

$$\int_a^b f(t) dt = \sum_{t=a}^b f_t \cdot \left(\frac{ta}{2} + \frac{fb}{2} \right)$$

2. Stellt $f(t)$ eine Gliederung dar, dann gilt angenähert:

$$\int_a^b f(t) dt = \sum_{t=a}^b f_t$$

X.

Aus unserer Charakterisierung der Durchschnittswerte geht hervor, dass solche nur für endlich begrenzte Kollektive bestehen; denn bloss für eine endliche Höchstbeobachtung oder bloss für eine endliche Ausdehnung der Beobachtungsreihe können die Mittel endliche Grössen sein.

Als Beispiele möchten wir nur zwei einfache Verteilungsreihen anführen.

1. Die *Gauss'sche Fehlerfunktion*, als Beispiel einer symmetrischen Verteilung (vgl. Fig. 7, Seite 385).

Die kontinuierliche Gleichung dieser Kurve lautet nach Gauss:

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$$

wo h eine Materialkonstante π und e die bekannten Naturkonstanten sind.

Die Kurve verläuft asymptotisch zur x -Achse, so dass das Ordnungsmerkmal x variiert von $-\infty$ bis $+\infty$; es gilt somit $T = \infty$. Das Maximum der Kurve, die höchste Beobachtung, beträgt

$$l_H = f(0) = \frac{h}{\sqrt{\pi}}$$

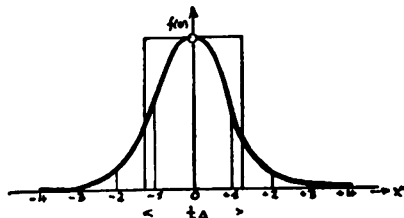


Fig. 7

Die Beobachtungsfläche misst, wie aus der Fehlerrechnung bekannt ist:

$$L = \int_{-\infty}^{\infty} f(x) dx = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-h^2 x^2} dx = 1$$

Aus diesen Elementen lassen sich die beiden Durchschnittswerte des Ordnungs- und Beobachtungsmerkmals berechnen. Für den ersteren gilt:

$$t_A = \frac{L}{l_H} = \frac{\sqrt{\pi}}{h}$$

Er ist als Basis des reduzierten Rechteckes in Fig. 7 angedeutet. Für das durchschnittliche Beobachtungsmerkmal ergibt sich:

$$l_A = \frac{L}{T} = \frac{1}{\infty} = 0$$

d. h. es gibt bei voller Ausdehnung des Kollektivs nach der Gaussischen Verteilungskurve keinen mittleren Beobachtungswert. Ein solcher ist erst möglich bei endlicher Begrenzung der Beobachtungen.

2. Die Überlebensordnung für einen Personenbestand sei bekannt als Funktion $l(t)$ (vgl. Fig. 8).

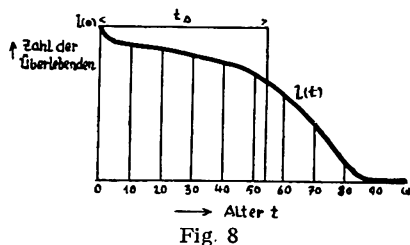


Fig. 8

Bilden wir die zur Berechnung der Durchschnitte nötigen Elemente, so ergibt sich:

$$\text{Beobachtungsfläche } L = \int_0^{\omega} l(t) dt$$

Höchster Beobachtungswert $l_H = l(0)$
Ausdehnung der Beobachtungen über $T = \omega$.

Dann folgt als Durchschnitt des Ordnungsmerkmals:

$$t_A = \frac{L}{l_H} = \frac{\int_0^{\omega} l(t) dt}{l(0)}$$

Fasst man die Überlebensordnung als Entwicklungsaufnahme auf, so wird:

$$t_d = \frac{\sum_{t=0}^{\omega} l_t - \left(\frac{l_0}{2} + \frac{l_{\omega}}{2} \right)}{l_0}$$

und wegen $l_{\omega} = 0$

$$t_d = \frac{1}{l_0} \sum_{t=0}^{\omega} l_t - \frac{1}{2}$$

t_d stellt die mittlere künftige Lebensdauer eines Neugeborenen dar und wird in Fig. 8 als Basis des flächengleichen Rechteckes angedeutet.

Stellt $l(t)$ die Verteilung eines Personenbestandes dar, so wird die Beobachtungsfläche als Gliederungsaufnahme näherungsweise berechnet zu:

$$L = \int_0^{\omega} l(t) dt = \sum_{t=0}^{\omega} l_t$$

Aus der gewogenen Fläche ergibt sich das mittlere Alter des Bestandes zu:

$$\frac{\sum_{t=0}^{\omega} t \cdot l_t}{\sum_{t=0}^{\omega} l_t}$$

Der Durchschnittswert des Beobachtungsmerkmals folgt aus:

$$l_d = \frac{L}{T} = \frac{\int_0^{\omega} l(t) dt}{\omega};$$

für eine Entwicklungsaufnahme wird

$$l_d = \frac{1}{\omega} \left(\sum_{t=0}^{\omega} l_t - \frac{l_0}{2} \right)$$

und stellt den durchschnittlich unter Risiko gestandenen Bestand dar. Für eine

Gliederungsaufnahme aber wird $l_d = \frac{1}{\omega} \sum_{t=0}^{\omega} l_t$ und stellt die durchschnittliche

Besetzung der Altersstufen im Moment der Beobachtung dar.

Diese beiden Beispiele sollen genügen als Hinweise auf die Verwendung von Durchschnittswerten bei der Betrachtung allgemeiner Kollektive.