DIE REKURSIONSFORMELN FÜR ARITHMETISCHE MITTEL UND VARIANZEN

von Shayle R. Searle, Cornell University
Originaltitel in "Teaching Statistics" Vol. 1 (1983):
The Recurrence Formulae for Means and Variances
Übersetzung: H. Lerche

Zusammenfassung: Die einfachen und bekannten Formeln, die das arithmetische Mittel (und die Varianz) von n+1 Werten auf das arithmetische Mittel (und die Varianz) von n Werten zurückführen, werden abgeleitet und veranschaulicht.

Walter (1981) zeigt, daß die Mittelung von bereits berechneten Mitteln mit einem neuen Wert nicht zu arithmetischen Mitteln, sondern zu Mitteln mit anderer Wichtung führt. Eine nützliche Möglichkeit zur rekursiven Berechnung von arithmetischen Mitteln bietet dagegen die bekannte und einfache Rekursionsformel, mit der man das Mittel von n+1 Werten direkt aus dem Mittel von n Werten und dem neu hinzugekommenen Wert berechnen kann. Eine entsprechende Rekursionsformel gibt es für die Summe der Quadrate der Abweichungen vom Mittel und somit auch für die daraus gebildete erwartungstreue Schätzgröße der Varianz. Diese Rekursionsformeln haben folgende Vorteile:

a) Falls ein neuer Wert auftritt, muß man bei der Berechnung von Mittel und Varianz nicht immer wieder ganz von vorne anfangen.
b) Sie können leicht bei programmierbaren Taschenrechnern angewandt werden.
c) Die Berechnung der Varianz führt nicht zu den Rundungsfehlern, die beim Subtrahieren zweier großer fast gleicher Zahlen, wie z. B. bei $n\sum x_i^2-(\sum x_i)^2$ auftreten können.

Formel für das arithmetische Mittel

Sei

$$M_n = (x_1 + x_2 + \cdots + x_n)/n$$

das arithmetische Mittel der n Werte $x_1, x_2, \ldots, x_n$. Dann ist das arithmetische Mittel dieser Werte zusätzlich eines weiteren Wertes x_{n+1}

$$\begin{aligned} M_{n+1} &= (x_1 + x_2 + \cdots + x_n + x_{n+1})/(n+1) \\ &= (nM_n + x_{n+1})/(n+1) \qquad (0) \\ &= [(n+1)M_n + x_{n+1} - M_n]/(n+1) \end{aligned}$$

Also gilt

$$M_{n+1} = M_n + (x_{n+1} - M_n)/(n+1) \qquad (1)$$

mit $M_0 = 0$ und $n = 0, 1, 2, 3, \ldots$

Mit dieser Formel erhält man das Mittel von n+1 Werten aus dem Mittel von n Werten durch einfaches Addieren der durch n+1 dividierten Differenz zwischen dem (n+1)ten Wert und dem Mittel der n Werte - kurz der Differenz zwischen dem neuen Wert und dem alten Mittel.

Beispiel:

Angenommen, ein Student erhält bei fünf aufeinanderfolgenden Tests die Punktzahlen 66, 72, 63, 83 und 66. Tabelle 1 zeigt diese Punktzahlen und ihre aufeinanderfolgenden Summen sowie ihre Mittel, sowohl direkt als auch mit Formel (1) nach jedem Test berechnet.

Tabelle 1: BERECHNUNG AUFEINANDERFOLGENDER MITTEL

$n+1$	x_{n+1}	Summen $x_1 + \ldots + x_{n+1}$	Mittel M_{n+1}	Berechnung mit (1) $M_{n+1} = M_n + (x_{n+1} - M_n)/(n+1)$
1	66	66	66	$66 = 0 + (66 - 0)/1$
2	72	138	69	$69 = 66 + (72 - 66)/2$
3	63	201	67	$67 = 69 + (63 - 69)/3$
4	83	284	71	$71 = 67 + (83 - 67)/4$
5	66	350	70	$70 = 71 + (66 - 71)/5$

Formel für die Summe der quadratischen Abweichungen

Die Summe der quadratischen Abweichungen der n Werte vom arithmetischen Mittel ist

$$S_n^2 = (x_1 - M_n)^2 + (x_2 - M_n)^2 + \ldots + (x_n - M_n)^2$$

bzw. bekanntlich auch

$$S_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2 - nM_n^2 .$$

Somit ergibt sich für n+1 Werte

$$\begin{aligned} S_{n+1}^2 &= x_1^2 + x_2^2 + \cdots + x_n^2 + x_{n+1}^2 - (n+1)M_{n+1}^2 \qquad (2) \\ &= x_1^2 + x_2^2 + \cdots + x_n^2 - nM_n^2 + nM_n^2 + x_{n+1}^2 - (n+1)M_{n+1}^2 \\ &= S_n^2 + nM_n^2 + x_{n+1}^2 - (n+1)M_{n+1}^2 . \qquad (3) \end{aligned}$$

Aus (0) folgt

$$x_{n+1} = (n+1)M_{n+1} - nM_n .$$

In (3) eingesetzt ergibt sich

$$S_{n+1}^2 = S_n^2 + nM_n^2 + [(n+1)M_{n+1} - nM_n]^2 - (n+1)M_{n+1}^2 .$$

Nach einer kleinen Vereinfachung führt dies zu

$$S_{n+1}^2 = S_n^2 + n(n+1)\,(M_{n+1} - M_n)^2 \qquad (4)$$

mit $S_0^2 = 0$ und $n = 0, 1, 2, 3, \ldots$

Die neue Summe S_{n+1}^2 der quadratischen Abweichungen ist somit einfach die alte Summe S_n^2 der quadratischen Abweichungen vermehrt um das n(n+1)fache des Quadrats der Differenz zwischen dem neuen und dem alten Mittel. Dies bedeutet bei Verwendung von (1) eine sehr einfache Berechnung verglichen mit der Benutzung von Formel (2), bei der man wieder von Anfang an rechnen muß. Zu beachten ist auch, daß (4) zeigt, daß S_{n+1}^2 immer S_n^2 übertrifft, abgesehen von dem Fall, daß S_{n+1}^2 und S_n^2 gleich groß sind, d. h. daß $M_{n+1} = M_n$ oder wegen (1) gleichbedeutend $x_{n+1} = M_n$ gilt.

Fortsetzung des Beispiels:

Für das Beispiel von Tabelle 1 zeigt Tabelle 2 die Berechnungen von S_{n+1}^2 mit Hilfe von (2) und auch mit Hilfe der Rekursionsformel (4).

Tabelle 2: BERECHNUNG DER AUFEINANDERFOLGENDEN SUMMEN DER QUADRATISCHEN ABWEICHUNGEN

$n+1$	x_{n+1}^2	Summe $x_1^2 + \ldots + x_{n+1}^2$ (a)	$(n+1)M_{n+1}^2$ (b)	Gleichung (2) für S_{n+1}^2 (a)–(b)	Berechnung mit (4) $S_{n+1}^2 = S_n^2 + n(n+1)(M_{n+1} - M_n)^2$
1	4356	4356	4356	0	$0 = 0 + 0(1)(66-0)^2$
2	5184	9540	9522	18	$18 = 0 + 1(2)(69-66)^2$
3	3969	13509	13467	42	$42 = 18 + 2(3)(67-69)^2$
4	6889	20398	20164	234	$234 = 42 + 3(4)(71-67)^2$
5	4356	24754	24500	254	$254 = 234 + 4(5)(70-71)^2$

Formel für die Schätzgröße der Varianz

Repräsentieren $x_1, x_2, \ldots, x_n$ eine Stichprobe, so ist

$$\hat{\sigma}_n^2 = S_n^2/(n-1)$$

eine erwartungstreue Schätzung für die Varianz ($n \geqslant 2$).
Wendet man (4) auf

$$\hat{\sigma}_{n+1}^2 = S_{n+1}^2/n \quad \text{an,}$$

so ergibt sich

$$\hat{\sigma}_{n+1}^2 = S_n^2/n + (n+1)(M_{n+1} - M_n)^2 \qquad (5).$$

Setzt man $\hat{\sigma}_1^2 = 0$, so folgt mit $\hat{\sigma}_n^2 = \frac{S_n^2}{n-1}$ schon für $n \geqslant 1$

$$S_n^2 = (n-1)\hat{\sigma}_n^2 .$$

In (5) eingesetzt, ergibt sich also für $n \geqslant 1$:

$$\hat{\sigma}_{n+1}^2 = (1 - 1/n)\hat{\sigma}_n^2 + (n+1)(M_{n+1} - M_n)^2 \qquad (6)$$

Dies ist eine Rekursionsformel zur Berechnung aufeinanderfolgender $\hat{\sigma}^2$, wenn Schritt für Schritt zusätzliche Werte verfügbar werden.
Man beachte: Während aus (4) $S_{n+1}^2 \geq S_n^2$ gefolgert werden konnte, zeigt Formel (6), daß

$$\hat{\sigma}_{n+1}^2 > \hat{\sigma}_n^2$$

nur gilt, wenn $n(n+1)(M_{n+1} - M_n)^2 > \hat{\sigma}_n^2$ ist.

Fortsetzung des Beispiels:

Tabelle 3 zeigt die Berechnung von $\hat{\sigma}_{n+1}^2$ (für $n+1 \geq 2$) sowohl mit Hilfe der Definition, als auch mit Hilfe der Rekursionsformel (6).

Tabelle 3: BERECHNUNG AUFEINANDERFOLGENDER $\hat{\sigma}^2$

$n+1$	$\hat{\sigma}_{n+1}^2 = S_{n+1}^2/n$	Berechnung mit (6) $\hat{\sigma}_{n+1}^2 = (1-1/n)\hat{\sigma}_n^2 + (n+1)(M_{n+1} - M_n)^2$
2	$18/1 = 18$	$18 = (1-1/1)0 + 2(69-66)^2$
3	$42/2 = 21$	$21 = (1-1/2)18 + 3(67-69)^2$
4	$234/3 = 78$	$78 = (1-1/3)21 + 4(71-67)^2$
5	$254/4 = 63\frac{1}{2}$	$63\frac{1}{2} = (1-1/4)78 + 5(70-71)^2$

Ergebnisse

Ist ein neuer Wert mit n bereits vorhandenen Werten zu verarbeiten, so können für die n+1 Werte die Berechnungen des arithmetischen Mittels, der Summe der quadratischen Abweichungen und die Berechnungen der erwartungstreuen Schätzung der Varianz mit Hilfe der entsprechenden Größen für die n Werte durchgeführt werden. Man muß nur die Rekursionsformeln (1), (4) und (6) benutzen und braucht nicht immer wieder von Anfang an Summen berechnen.

LITERATUR

Walter, M. (1981): How to get a higher grade: beware of averages. Teaching Statistics, 3, 77 - 79.