

Computer-Einsatz bei Rekursionsformeln für Mittelwert und Standardabweichung

von John S. Croucher, MacQuarie University

Originaltitel in "Teaching Statistics" Vol.6 (1984):  
A Computer Application of Recurrence Formulae for Means and Standard Deviations

Übersetzung: Andreas Horn

In einem Beitrag der Zeitschrift "Teaching Statistics" befaßt sich S.R. SEARLE [2] (siehe auch die Übersetzung in Stochastik in der Schule [3]) mit den bekannten Rekursionsformeln für die Berechnung von Mittelwert und Standardabweichung für n+1 Stichprobenwerte aus Mittelwert und Standardabweichung der ersten n Stichprobenwerte und einem neu hinzukommenden (n+1)-ten Wert.

Diese Formeln lauten: 
$$M_{n+1} = M_n + (x_{n+1} - M_n)/(n+1)$$
$$s_{n+1}^2 = (1-1/n)s_n^2 + (n+1)(M_{n+1} - M_n)^2,$$

wobei folgende Bezeichnungen gelten:

$x_{n+1}$  = (n+1)-ter Stichprobenwert

$M_{n+1}$  = Mittelwert von n+1 Stichprobenwerten

$s_{n+1}$  = Standardabweichung für n+1 Werte.

Einige Vorteile dieses Verfahrens wurden bereits genannt [2]:

- a) Bei neu hinzukommenden Stichprobenwerten ist die Berechnung von Mittelwert und Standardabweichung ohne vollständige Neuberechnung unter Einbeziehung der früheren Stichprobenwerte möglich.
- b) Mögliche Rundungsfehler werden vermieden.
- c) Der Einsatz programmierbarer Rechner bietet sich an.

Darüber hinaus hat dieses Verfahren noch andere Vorzüge. So weisen z.B. ANDERSON und BARNET (1983) [1] auf die Folgen des Computer-Einzuges in die Schule hin, der sich insbesondere auf die Didaktik des Statistik-Unterrichts auswirkt. Für die genannten Rekursionsformeln kann leicht

ein einfaches Programm erstellt werden, das tiefere Einsichten in die Problematik ermöglicht.

Ein solches Programm (in BASIC) ist weiter unten abgedruckt.

Einige Bemerkungen zum Programm:

Vorausgeschickt sei, daß die Eingabe des Stichprobenwertes 9999 das Programm beendet.

Jeweils nach Eingabe eines zusätzlichen Stichprobenwertes werden der neue Mittelwert und die Standardabweichung berechnet. Dadurch wird die Auswirkung einzelner Stichprobenwerte auf die gesamte Datenmenge verdeutlicht. Außerdem liefert das Programm jeweils die prozentuale Abweichung der betreffenden Parameter aufgrund des neu hinzutretenden Stichprobenwertes. Daher können Schüler nach Eingabe der Stichprobenwerte gefragt werden, wann ihrer Ansicht nach eine Stabilisierung der Parameter zu erwarten ist. Auch könnte getestet werden, inwieweit eine unterschiedliche Reihenfolge bei der Eingabe derselben Stichprobenwerte das Ergebnis beeinflusst.

Das Programm kann auch bei der Berechnung von Konfidenzintervallen nach der Eingabe von z.B. 10, 20, 30,... Stichprobenwerten eingesetzt werden. Die Auswirkung eines größer werdenden Stichprobenumfangs auf das Konfidenzintervall wird offensichtlich.

Programmausdruck:

```

10 REM Programm von John S. Croucher
20 REM Macquarie University, North Ryde, N.S.W., Australia
30 PRINT "Anzahl der hoechstens "
40 PRINT "auszuwertenden Stichprobenwerte"
50 INPUT N: PRINT
60 DIM X(N): REM Stichprobenwerte
70 DIM M(N): REM Mittelwerte
80 DIM S(N): REM Standardabweichung
90 DIM PM(N): REM Proz.Abw.vom letzten Mittelwert
100 DIM PS(N): REM Proz.Abw.von letzter Standardabw.
110 DIM V(N): REM Varianz
120 FOR I = 1 TO N
130 PRINT "EINGABE DES ";I;"-TEN"
140 PRINT "STICHPROBENWERTE:";
150 REM Stichprobenwert 9999 beendet das Programm
160 INPUT X(I)
170 IF X(I) = 9999 THEN 330
180 IF I < > 1 THEN 200
190 M(I) = X(I):S(I) = 0:PM(I) = 0: GOTO 270
200 M(I) = M(I - 1) + (X(I) - M(I - 1)) / I
210 PM(I) = 100 * (M(I) - M(I - 1)) / M(I - 1)
220 V(I) = (1 - 1 / (I - 1)) * V(I - 1) + 1 * ((M(I) - M(I - 1)) ^ 2)
230 S(I) = SQR (V(I))
240 PS(2) = 100
250 IF I = 2 THEN 270
260 PS(I) = 100 * (S(I) - S(I - 1)) / S(I - 1)
270 PRINT "NEUER MITTELWERT=";M(I); TAB( 30);
280 PRINT "PROZ.ABW.VOM LETZTEN MW=";PM(I)
290 PRINT "NEUE STANDARDABW.=";S(I); TAB( 30);
300 PRINT "PROZ.ABW.VON LETZTER ST.ABW.=";PS(I)
310 PRINT
320 NEXT I
330 PRINT "ENDE"

```

Beispiel für einen Programmablauf:

Obwohl das Programm natürlich für große Zahlen von Stichprobenwerten gedacht ist, soll an dieser Stelle der Programmablauf nach Eingabe von nur vier Werten (73,0 / 62,7 / 59,3 / 68,2) vorgestellt werden.

Die Frage nach der "Anzahl der höchstens auszuwertenden Stichprobenwerte" kann mit jeder beliebigen Zahl beantwortet werden, die voraussichtlich nicht überschritten wird. In unserem Beispiel wird die Zahl 20 eingegeben, obwohl von nur 4 Stichprobenwerten ausgegangen wird. So können, falls gewünscht, bis zu 16 weitere Stichprobenwerte eingegeben werden. Zur Beendigung des Programms wird hier als fünfter Wert 9999 eingegeben.

Anzahl der hoechstens  
auszuwertenden Stichprobenwerte  
?20

EINGABE DES 1-TEN	
STICHPROBENWERTE:773,0	
NEUER MITTELWERT=73	PROZ.ABW.VOM LETZTEN MW=0
NEUE STANDARDABW.=0	PROZ.ABW.VON LETZTER ST.ABW.=0

EINGABE DES 2-TEN	
STICHPROBENWERTE:62,7	
NEUER MITTELWERT=67,95	PROZ.ABW.VOM LETZTEN MW=-7,05479453
NEUE STANDARDABW.=7,28319986	PROZ.ABW.VON LETZTER ST.ABW.=100

EINGABE DES 3-TEN	
STICHPROBENWERTE:59,3	
NEUER MITTELWERT=65	PROZ.ABW.VOM LETZTEN MW=-4,20044215
NEUE STANDARDABW.=7,13372274	PROZ.ABW.VON LETZTER ST.ABW.=-2,05235513

EINGABE DES 4-TEN	
STICHPROBENWERTE:68,2	
NEUER MITTELWERT=65,8	PROZ.ABW.VOM LETZTEN MW=1,23076925
NEUE STANDARDABW.=6,04041942	PROZ.ABW.VON LETZTER ST.ABW.=-15,3258453

EINGABE DES 5-TEN	
STICHPROBENWERTE:9999	
ENDE	

Anmerkung:

Während im Originalbeitrag die Ergebnisse dieses Probelaufes (nicht einheitlich) gerundet sind, werden hier genau die Werte abgedruckt, die vorliegendes Programm auf dem APPLE IIe liefert. Geeignete Rundungen müssen im Programm eingeplant werden.

LITERATUR

- [1] ANDERSON, C. and BARNETT, V. (1983): INSTEP with the microcomputer. Teaching Statistics, Vol.5, Nr. 1,2-7
- [2] SEARLE, S.R. (1983): The recurrence formulae for means and variances. Teaching Statistics, Vol.5, Nr. 1,7-10.
- [3] SEARLE, S.R. (1983): Übersetzung von [2]. In: Stochastik in der Schule (1983), Nr. 3,5-9.