Mittelwert und Stichproben-Varianz

Aus einer Datenquelle stammen eindimensionalen Messwerte *x[i], i=1, 2, 3, ..., N*

der Mittelwert: *m = Summe **(** x[i] **)** / N*
die Stichprobenvarianz *v = Summe **(** (x[i] - m)^2 **)** / (N-1)*

Die Stichprobenvarianz oder alternativ die Streuung (= Wurzel aus der Stichprobenvarianz) misst die durchschnittliche Abweichung vom Mittelwert. Wurden alle Stichproben **gespeichert**, kann man mit diesen Formeln **zuerst** den Mittelwert *m* und mit diesem **dann** *v* berechnen.

Problem: Wie berechnet man diese Kennzahlen OHNE alle Messwerte zu speichern? Beim Mittelwert *m* ist es noch einfach. Hier genügt die akkumulierte Summe der Messungen, die man am Ende durch die Stichprobenanzahl N dividieren kann. Zur nachträglichen Berechnung der Strichprobenvarianz eignet sich obige Formel nicht, da man den Mittelwert schon am Anfang kennen müsste. Man kann aber die Berechnungsformel algebraisch umformen zu

*v = [Summe **(** x[i]^2 **)** - N * m^2] / (N-1)*

In dieser Formel kommt nur mehr die akkumulierte Quadratsumme der Werte vor!

Praktische Durchführung:
Man berechnet laufend die Partialssummen der *x[i]* und der *x[i]*^{2} aller Messwerte:

```
int n = 0;                          // Zählvariable für die Anzahl der Werte
double sum = 0., sum2 = 0.;         // zur Berechnung der 2 Summen

while  (es gibt einen neuen Wert x)
{      ++n;                 // die neue Stichprobe zählen
       sum += x;            // Summe updaten
       sum2 += x*x;         // Quadrat-Summe updaten
}

// hier enthält  n die Anzahl der gelesenen Zahlen
if (n > 0)         // gabs überhaupt Werte ?
      m = sum/n;

if (n > 1)         // mindestens  2?
      v = ( sum2 - n * m * m ) / (n-1);
```