

# Re: standard deviation, but without the mean

---

*Source:* <http://sci.tech-archive.net/Archive/sci.stat.math/2006-03/msg00178.html>

---

- *From:* "Ray Koopman" <[koopman@xxxxxx](mailto:koopman@xxxxxx)>
  - *Date:* 10 Mar 2006 16:09:01 -0800
- 

David A. Heiser wrote:

"Ray Koopman" <[koopman@xxxxxx](mailto:koopman@xxxxxx)> wrote in message  
<news:1141669275.577944.112830@xx>

richardstartz@xxxxxxxxxxx wrote:

The standard deviation is the square root of the variance (of course).

There's a standard formula for computing the variance from a running

sum. Suppose Xsum is the sum of the the first n numbers and that X2 is

the sum of the squares. Then

$$\text{var} = X2/n - (Xsum/n)^2$$

Just keep track if X2 and Xsum as you go.

Using running totals can sometimes lead to cancellation errors. See <http://tinyurl.com/d6ax2> for a stable algorithm.

++++  
Actually this is not a correct algorithm, just an approximation.

I've just gone through Welford's algorithm (a one pass calculation) using xnumbers on the NIST data sets and have found Welford to obtain correct values. The errors come in due to the general problem of summation of lists of numbers, which no algorithm can fix. The solution of course is to do it with as many digits as possible, so that the summation errors are not important.

For example, running Welford's on the NIST NumAcc4 data set (1001 values) using a 30 digit exact computation comes out to the theoretical value, with the error in the least 9 digits (21 accurate digits).. By using Kahan's method of summation, this reduces the error to about 7 digits.

Re: standard deviation, but without the mean

If people would make the effort to read Knuth, a lot of the misconceptions would disappear. Welford published in *Technometrics*, 1962, pg 419–420. All this is in Knuth. Apparently everybody else missed it.

David Heiser

The pseudocode that Miller attributes to Jennrich is what [http://en.wikipedia.org/wiki/Algorithms\\_for\\_calculating\\_variance](http://en.wikipedia.org/wiki/Algorithms_for_calculating_variance) calls Algorithm II, saying that it is "due to Knuth, who cites Welford." Coding it in Mathematica as

```
n = m = ss = 0;  
Scan[(d = # - m; m += d/(++n); ss += d*(# - m))&, data];  
{m, Sqrt[ss/(n-1)]}
```

and applying it to the NumAcc4 data gives

```
{10000000.200000001, 0.10000000055890503}
```

for the mean and s.d.

.