

Traditional One Pass s^2 (in slides)

$$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad (1)$$

$$= \frac{1}{n} \{ \sum x_i^2 - \sum 2x_i \bar{x} + \sum \bar{x}^2 \} \quad (2)$$

$$= \frac{1}{n} \sum x_i^2 - 2\bar{x} \left(\frac{1}{n} \sum x_i \right) + \frac{1}{n} \sum \bar{x}^2 \quad (3)$$

$$= \frac{1}{n} \sum x_i^2 - 2\bar{x}^2 + \bar{x}^2 \quad (4)$$

$$= \frac{1}{n} \sum x_i^2 - \bar{x}^2 \quad (5)$$

Welford One Pass $\bar{x}_i$ and v_i

Definitions to begin with (**not the Welford equations**):

$$\bar{x}_i = \frac{1}{i} (x_1 + x_2 + \dots + x_i) \quad (6)$$

$$Var_i = \left(\frac{1}{i} \sum x_i^2 \right) - \bar{x}_i^2 \quad (7)$$

$$v_i = i \cdot Var_i = (x_1^2 + x_2^2 + \dots + x_i^2) - i\bar{x}_i^2 \quad (8)$$

Equation (6) is the arithmetic mean of sample $\mathcal{X} = \{x_1, x_2, \dots, x_i\}$, equation (7) is the (algebraically correct but numerically challenged) one-pass variance, and equation (8) is $\mathcal{X}$'s **variance** multiplied by i . Equation (8) is simply the **expanded notation** of one pass difference of squares (5).

Welford One Pass $\bar{x}_i$ Derivation

Begin with eqn 6...

$$\bar{x}_i = \frac{1}{i} (x_1 + x_2 + \dots + x_i)$$

... and recognize $\bar{x}_{i-1}$ inside of it...

$$i\bar{x}_i = \underbrace{x_1 + x_2 + \dots + x_{i-1}}_{(i-1)\bar{x}_{i-1}} + x_i$$

$$i\bar{x}_i = (i-1)\bar{x}_{i-1} + x_i$$

$$i\bar{x}_i = i\bar{x}_{i-1} - \bar{x}_{i-1} + x_i$$

... collect terms without the i factor:

$$i\bar{x}_i = i\bar{x}_{i-1} + (x_i - \bar{x}_{i-1})$$

... solve for $\bar{x}_i$

$$\bar{x}_i = \bar{x}_{i-1} + \frac{1}{i} (x_i - \bar{x}_{i-1}) \quad (9)$$

Observe that with $x_0 \equiv 0$ equation 9 at x_1 is $\bar{x}_1 = 0 + \frac{1}{1}(x_1 - 0) = x_1$.

Welford One Pass v_i Derivation

Begin with equation 8...

$$\begin{aligned} v_i &= i \cdot \text{Var}_i = x_1^2 + x_2^2 + \dots + x_i^2 - i\bar{x}_i^2 \\ v_i &= \underbrace{x_1^2 + x_2^2 + \dots + x_{i-1}^2}_{\text{will become } v_{i-1} \dots} + x_i^2 - i\bar{x}_i^2 \end{aligned} \quad (10)$$

and (like Welford $\bar{x}_i$ derivation) think of $x_1^2 + x_2^2 + \dots + x_{i-1}^2$ in terms of v_{i-1} :

$$\begin{aligned} v_{i-1} &= x_1^2 + x_2^2 + \dots + x_{i-2}^2 + x_{i-1}^2 - \underbrace{(i-1)\bar{x}_{i-1}^2}_{\text{move to lhs}} \\ v_{i-1} + (i-1)\bar{x}_{i-1}^2 &= x_1^2 + x_2^2 + \dots + x_{i-1}^2 \\ v_i &= \underbrace{x_1^2 + x_2^2 + \dots + x_{i-1}^2}_{\text{substitute (11) into (10) for } x_1^2 \text{ through } x_{i-1}^2} + x_i^2 - i\bar{x}_i^2 \end{aligned} \quad (11)$$

distribute...

$$v_i = v_{i-1} + i\bar{x}_{i-1}^2 - \bar{x}_{i-1}^2 + x_i^2 - i\bar{x}_i^2$$

factor i out of terms...

$$v_i = v_{i-1} - i(\bar{x}_i^2 - \bar{x}_{i-1}^2) + (x_i^2 - \bar{x}_{i-1}^2)$$

we have two separate differences of squares, factor!...

$$v_i = v_{i-1} - i \underbrace{(\bar{x}_i - \bar{x}_{i-1})}_{\text{ugh}} (\bar{x}_i + \bar{x}_{i-1}) + (x_i - \bar{x}_{i-1})(x_i + \bar{x}_{i-1}) \quad (12)$$

Equation (12) would be factorable if we had x_i instead of $\bar{x}_i$. Let's fix that now: recall Welford equation 9, solved for $(\bar{x}_i - \bar{x}_{i-1})$:

$$\begin{aligned} \bar{x}_i &= \bar{x}_{i-1} + \frac{1}{i}(x_i - \bar{x}_{i-1}) \\ (\bar{x}_i - \bar{x}_{i-1}) &= \frac{1}{i}(x_i - \bar{x}_{i-1}) \end{aligned} \quad (13)$$

Substitute (13) into (12), the $\frac{1}{i}$ cancel...

$$v_i = v_{i-1} - i \left\{ \frac{1}{i}(x_i - \bar{x}_{i-1}) \right\} (\bar{x}_i + \bar{x}_{i-1}) + (x_i - \bar{x}_{i-1})(x_i + \bar{x}_{i-1}) \quad (14)$$

Why did we do this? To get two $(x_i - \bar{x}_{i-1})$ for factoring! Rearranging terms to the right of v_{i-1} in positive first order...

$$\begin{aligned} v_i &= v_{i-1} + (x_i - \bar{x}_{i-1})(x_i + \bar{x}_{i-1}) - (x_i - \bar{x}_{i-1})(\bar{x}_i + \bar{x}_{i-1}) \\ v_i &= v_{i-1} + (x_i - \bar{x}_{i-1}) \{ (x_i + \bar{x}_{i-1}) - (\bar{x}_i + \bar{x}_{i-1}) \} \\ v_i &= v_{i-1} + (x_i - \bar{x}_{i-1})(x_i - \bar{x}_i) \end{aligned}$$

Aside from x_i , we want all the RHS terms to be from the previous iteration $(i-1)$, so use Welford's equation (9) to replace $\bar{x}_i$...

$$v_i = v_{i-1} + (x_i - \bar{x}_{i-1}) \left(x_i - \left\{ \bar{x}_{i-1} + \frac{1}{i} (x_i - \bar{x}_{i-1}) \right\} \right)$$

$$v_i = v_{i-1} + (x_i - \bar{x}_{i-1}) \left\{ (x_i - \bar{x}_{i-1}) - \frac{1}{i} (x_i - \bar{x}_{i-1}) \right\}$$

Notice the pattern $\alpha = x_i - \bar{x}_{i-1}$ in the RHS expression ...

$$\alpha \left(\alpha - \frac{\alpha}{i} \right) = \alpha^2 \left(\frac{i-1}{i} \right)$$

and we have Welford's equation for $v_i = i \cdot \text{Var}_i$:

$$v_i = v_{i-1} + \frac{i-1}{i} (x_i - \bar{x}_{i-1})^2 \tag{15}$$

Observe that (with $x_0 \equiv 0$ and $v_0 \equiv 0$) at $i = 1$ (the first datapoint) $v_1 = 0 + \frac{1-1}{1} (x_1 - 0)^2 = 0$.