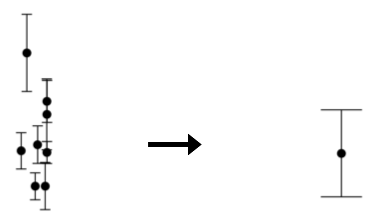
Determining the total standard deviation of a quantity measured on multiple instances



The mean value, $\bar{X}$, and standard deviation, $\sigma_{\bar{X}}$, of a quantity can be determined by making multiple measurements on multiple instances (*i.e.* mean values $\langle x \rangle_p$ and associated standard deviations σ_p). The correct way to determine the total standard deviation is:

$$\sigma_{\bar{X}}^2 = \left(\frac{P(N-1)}{PN-1}\right)\frac{1}{P}\sum_{p=1}^{P}\sigma_p^2 + \left(\frac{PN}{PN-1}\right)\left[\frac{1}{P}\sum_{p=1}^{P}\langle x \rangle_p^2 - \frac{1}{P^2}\left(\sum_{p=1}^{P}\langle x \rangle_p\right)^2\right] \tag{0.1}$$

where there are N measurements on P instances. Or for large NP, this formula is approximately

$$\boxed{\sigma_{\bar{X}}^2 \approx \langle \sigma_p^2 \rangle + \left(\text{std}\left[\langle x \rangle_p\right]\right)^2} \tag{0.2}$$

Derivation

Start with K total data points measuring the same quantity.

$$x_1, x_2, x_3, \ldots, x_K$$

Now partition the data into P sets of N measurements (so, $K = n_1 + n_2 + \cdots + n_N$)

$$x_1, x_2, x_3, \ldots, x_N \in \mathbf{p}_1$$

$$x_{N+1}, x_{N+2}, \ldots, x_{2N} \in \mathbf{p}_2$$

$$x_{2N+1}, x_{2N+2}, \ldots, x_{3N} \in \mathbf{p}_3$$

$$\ldots$$

Now relabel the data points to reflect this new partitioning.

$$x_{n,p}$$

where the first of the indices refers to the measurement number within the set and the second one refers to the set number (*i.e.* point n from set p) thus, $K = NP$. From these partitioned sets compute the average and variance

$$\langle x \rangle_p = \frac{1}{N}\sum_{n=1}^{N} x_{n,p} \tag{0.3}$$

$$\sigma_p^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(x_{n,p} - \langle x \rangle_p\right)^2 \tag{0.4}$$

The average of all K data points is:

$$\bar{X} = \frac{1}{K}\sum_{k=1}^{K} x_k = \frac{1}{P}\sum_{p=1}^{P}\langle x\rangle_{p=} = \frac{1}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p} \tag{0.5}$$

And the standard deviation of all K data points is:

$$\sigma_{\bar{X}}^2 = \frac{1}{K-1}\sum_{k=1}^{K}\left(x_k - \frac{1}{K}\sum_{k=1}^{K} x_k\right)^2 = \frac{1}{K-1}\sum_{k=1}^{K}\left(x_k - \frac{1}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p}\right)^2 \tag{0.6}$$

Since $K = NP$ and averaging is a linear operation:

$$\sigma_{\bar{X}}^2 = \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N}\left(x_{n,p} - \frac{1}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p}\right)^2 \tag{0.7}$$

1 Deriving the exact formula

multiply out equation 0.7

$$\sigma_{\bar{X}}^2 = \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N}\left[x_{n,p}^2 - 2\frac{x_{n,p}}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p} + \left(\frac{1}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p}\right)^2\right] \tag{1.1}$$

substitute eq 0.5 in the right most term

$$= \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N}\left[x_{n,p}^2 - 2\frac{x_{n,p}}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p} + \left(\frac{1}{P}\sum_{p=1}^{P}\langle x\rangle_p\right)^2\right] \tag{1.2}$$

regroup/distribute

$$= \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N}\left[x_{n,p}^2 - 2\frac{x_{n,p}}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p}\right] + \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N}\left(\frac{1}{P}\sum_{p=1}^{P}\langle x\rangle_p\right)^2 \tag{1.3}$$

that last term is really $\sum_p\sum_n \bar{X}^2$ which is the same as $\bar{X}^2\sum_p\sum_n = \bar{X}^2 PN$

$$= \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N}\left[x_{n,p}^2 - 2\frac{x_{n,p}}{NP}\sum_{p=1}^{P}\sum_{n=1}^{N} x_{n,p}\right] + \frac{NP}{NP-1}\bar{X}^2 \tag{1.4}$$

distribute, regroup the first term and substitute 0.5 in the middle term

$$= \frac{1}{NP-1}\sum_{p=1}^{P} N\left[\frac{1}{N}\sum_{n=1}^{N} x_{n,p}^2\right] - \frac{1}{NP-1}\sum_{p=1}^{P}\sum_{n=1}^{N} 2x_{n,p}\bar{X} + \frac{NP}{NP-1}\bar{X}^2 \tag{1.5}$$

pull out the $\bar{X}$ of the sum in the middle term then substitute 0.5

$$= \frac{1}{NP-1}\sum_{p=1}^{P} N\left[\frac{1}{N}\sum_{n=1}^{N} x_{n,p}^2\right] - \frac{2NP\bar{X}^2}{NP-1} + \frac{NP}{NP-1}\bar{X}^2 \tag{1.6}$$

clean up the last two terms

$$= \frac{1}{NP-1}\sum_{p=1}^{P} N\left[\frac{1}{N}\sum_{n=1}^{N} x_{n,p}^2\right] - \frac{NP}{NP-1}\bar{X}^2 \tag{1.7}$$

Recall that,

$$\sigma^2 = \frac{1}{M-1}\sum_{m=1}^{M} x_m^2 - \frac{M}{M-1}\left(\frac{1}{M}\sum_{m=1}^{M} x_m\right)^2 \tag{1.8}$$

or,

$$\frac{1}{M}\sum_{m=1}^{M} x_m^2 = \frac{M-1}{M}\sigma^2 + \left(\frac{1}{M}\sum_{m=1}^{M} x_m\right)^2$$

substitute this relationship into 1.7,

$$\sigma_{\bar{X}}^2 = \frac{1}{NP-1}\sum_{p=1}^{P} N\left[\frac{N-1}{N}\sigma_p^2 + \left(\frac{1}{N}\sum_{n=1}^{N} x_{n,p}\right)^2\right] - \frac{NP}{NP-1}\bar{X}^2 \tag{1.9}$$

distribute the left most sum, and in the second term substitute 0.3

$$= \frac{N-1}{NP-1}\sum_{p=1}^{P}\sigma_p^2 + \frac{1}{NP-1}\sum_{p=1}^{P} N\left\langle x\right\rangle_p^2 - \frac{NP}{NP-1}\bar{X}^2 \tag{1.10}$$

factor out the fraction

$$= \frac{N-1}{NP-1}\sum_{p=1}^{P}\sigma_p^2 + \frac{N}{NP-1}\left[\sum_{p=1}^{P}\left\langle x\right\rangle_p^2 - P\bar{X}^2\right] \tag{1.11}$$

or

$$\boxed{\sigma_{\bar{X}}^2 = \left(\frac{P(N-1)}{NP-1}\right)\frac{1}{P}\sum_{p=1}^{P}\sigma_p^2 + \left(\frac{NP}{NP-1}\right)\left[\frac{1}{P}\sum_{p=1}^{P}\left\langle x\right\rangle_p^2 - \frac{1}{P^2}\left(\sum_{p=1}^{P}\left\langle x\right\rangle_p\right)^2\right]} \tag{1.12}$$

2 Approximation

For $N \to \infty$ we find that

$$\sigma_{\bar{X}}^2 \to \frac{1}{P}\sum_{p=1}^{P}\sigma_p^2 + \left[\frac{1}{P}\sum_{p=1}^{P}\left\langle x\right\rangle_p^2 - \bar{X}^2\right] \tag{2.1}$$

now if we also assume that $P \to \infty$ and recall again relation 1.8

$$= \left\langle\sigma_p^2\right\rangle + \frac{1}{P}\sum_{p=1}^{P}\left(\left\langle x\right\rangle_p - \bar{X}\right)^2 \tag{2.2}$$

So,

$$\sigma_{\bar{X}}^2 \approx \left\langle\sigma_p^2\right\rangle + \left(\mathrm{std}\left[\left\langle x\right\rangle_p\right]\right)^2 \tag{2.3}$$