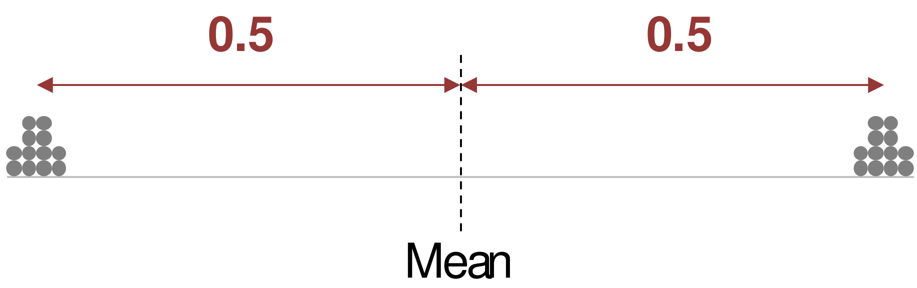
Appendix 11. Methods supplement [posted as supplied by author]

A. For data that can only be whole numbers, some combinations of mean and standard deviation are impossible

The Cole-Dehbi impossibility injunction

New York Heart Association (NYHA) quantifies symptoms in heart failure from class 1 to 4. Individual patients can only have one of those 4 values. By colloquial extension of statistics designed for continuous variables, authors sometimes report mean and SD for NYHA. But for any reported mean NYHA, there is a lower limit on the possible values of SD.

If the mean NYHA is non-integer such as 2.5, the smallest SD occurs with half the patients having NYHA 2 and half NYHA 3. (If instead any patients have NYHA 1 or 4, this would only increase the SD.) For a large sample, SD has a very simple definition: the root-mean-square distance between the data and the mean. Since all the data (at 2 or 3) are exactly 0.5 away from the mean (at 2.5), the SD is therefore 0.5, as sketched below (for clarity using just a few dots to represent very many patients)



For smaller samples, a scaling factor of $\sqrt{\frac{N}{N-1}}$ needs to be applied. SD is therefore slightly larger than the simple root-mean-square concept above, by around 5% for a 10-patient study and 1% for a 50-patient study.

The general formula for all sample sizes combines the correction factor and the root-mean-square, as follows:

$$SD = \sqrt{\frac{N}{N-1}} \times \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}}$$

For other non-integer values of mean NYHA between 2 and 3, the NYHA 2 and 3 patients need to be present in appropriate numbers which will be unequal, and a corresponding minimum possible SD can be calculated in the following way. If a proportion H of the patients is in the higher NYHA class and 1-H in the lower NYHA class, the mean NYHA is H above the lower class. The upper class is at a distance 1-H away from the mean, and the lower class at a distance H away. Therefore, the root mean square distance between the data and the mean for a very large sample is $\sqrt{H^2 \times (1-H) + (1-H)^2 \times H}$, which simplifies to $\sqrt{H \times (1-H)}$.

For all sample sizes, the SD is $\sqrt{\frac{N}{N-1}} \times \sqrt{H \times (1-H)}$.

Rounding also has to be taken into account, since a published rounded mean of 2.2 might represent a raw value anywhere from 2.15 to almost 2.25. In algebraic terminology, a mean of 2.2 (i.e. the fractional part is 0.2) might correspond to an H of 0.2, or anywhere from 0.15 to 0.25. From the formula it can be seen that the H value that is furthest away from 0.5 (i.e. closer to 0 or 1) will give the smallest SD value. However the actual possible values of H depend on the number of patients. If there are 50 patients, the smallest unrounded mean that could be rounded to 2.2 is 2.16 (eight 3's and forty-two 2's), so the possible H that gives the smallest SD will be 0.16.

With these considerations a "smallest mathematically possible SD" can be calculated for any stated mean NYHA and sample size, as shown below for some example sample sizes. When SDs are calculated from individual patient data (even if incorrect), this limit will always be satisfied. When incorrect SDs are written down directly without using individual data, however, this limit may be inadvertently breached.

Fractional part of NYHA class (rounded)	**Minimum possible SD (rounded)**		
	n=10	**n=30**	**n≥50**
0.1	0.3	0.3	0.2
0.2	0.4	0.4	0.4
0.3	0.5	0.4	0.4
0.4	0.5	0.5	0.5
0.5	0.5	0.5	0.5
0.6	0.5	0.5	0.5
0.7	0.5	0.4	0.4

0.8	0.4	0.4	0.4
0.9	0.3	0.3	0.2

B. Limit of standard deviation for a given range

The Centrifuge Clamp

From a large population there is a sample of *N* values (x_1, x_2, …, x_N). Let us denote its mean as $\overline{x}$ and its range (greatest value minus smallest value) as R. The standard deviation (SD) is an index of how far the values are away from $\overline{x}$, and is calculated as follows:

$$SD^2 = \frac{N}{N-1} \times \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}$$

For any given range of data, the greatest possible SD occurs when all the values are at the extremes of the range. If the sample size *N* is even, then it is obvious that the maximum SD occurs when half the values are at the top of the range and half at the bottom of the range. The mean is then the midpoint of the range. All the values are a distance $R/2$ from the mean. Therefore the mean of$(x - \overline{x})^2$ is$(R/2)^2$. Applying the formula above, this upper limit on SD is

$$\mathrm{SD} = \sqrt{\frac{N}{N-1}} \times \frac{R}{2}$$

For odd sample sizes *N*, because the values cannot be distributed equally to the two limits, the SD has to be slightly smaller than the above formula.

For any sample size, if the values are distributed in any other way, then the SD must be smaller than the above formula.

Therefore for any dataset of any size *N* with range R, the data cannot be more centrifugal than the limits of the range, and so SD is always restricted by this formula:

$$\mathrm{SD} \leq \sqrt{\frac{N}{N-1}} \times \frac{R}{2}$$

C. Derivation of age of additional patients from simultaneous change in patient number and standard error of the mean age between publications

Dropout divination

If patients drop out during the study, the standard deviation and the sample size *N* both change. If the mean age stays constant, the age of the patients who dropped out can sometimes be calculated. To do this we need to observe that the square of the standard error of the mean (SEM) is

$$SEM^2 = \frac{SD^2}{N}$$

and that the sample variance is defined as

$$SD^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}$$

From these we obtain

$$\sum_{i=1}^{N}(x_i - \overline{x})^2 = SEM^2 \times N \times (N-1)$$

For example, suppose the SEM is initially 5 in 22 patients with a mean age of 55 years and that, after two patients drop out, the mean age remains 55 years and the SEM falls to 2.

By the formula above, the initial sum of squared deviations from the mean is $5^2 \times 22 \times 21 = 11550$. After the two dropouts, the sum of squared deviations from mean is $2^2 \times 20 \times 19 = 1520$. The contribution of the two dropout patients to the sum of squared deviations from the mean must therefore have been $11550 - 1520 = 10030$, i.e. they were very widely spaced in age. If the lower limit of age for the younger patient was 18 years old, the older patient must have an age that is at least x, defined as the solution of $(18 - 55)^2 + (x - 55)^2 = 10030$. From this we can see that the age of the older dropout patient is 148 years.

D. Discrepancies in pre-post differences

Dubious deltas

If temperature rises from 36.9 by 0.2 degrees, then the final value must be 37.0, 37.1 or 37.2. It cannot be 36.9 or lower, nor 37.3 or higher. This is because allowing for rounding the initial value might have been as low as 36.85 and the increment as small as 0.15, giving 37.0. At the other extreme, the initial value could have been almost 36.95 and the increment almost 0.25 which would have given a result that rounded to 37.2.

This reasoning works equally for subtractions.

If measurements are made for a group of patients before and after an intervention, the mean change must agree in this way with the difference between baseline and final means. A raw disagreement of one in the final decimal place is possible but a larger raw disagreement is not possible.

This can be expressed conveniently as follows:

$a+b=c$ must agree to within one unit in the last decimal place