MSV 9: The "Do We Divide by n or n – 1?" Sheet

Suppose you are given the population of numbers 1, 3, 4, 6, 8 and 9,
where each number is equally likely to be picked each time.

To find the 'population variance' = σ^2 (= 'msd' (mean square deviation) for MEI)
(= σ_n^2 on some calculators)
= the 'variance of the entire population',
divide by n.

So for this population, $\sigma^2 = \dfrac{\sum x^2 - n\bar{x}^2}{n} = \dfrac{207 - 6(\frac{31}{6})^2}{6} = 7.81$ (3sf).
So σ (= rmsd ('root mean square deviation') for MEI) = $\sqrt{7.805555} = 2.79$ (3sf).

However, if we choose a sample from this population, let's say of size 3,
we now switch to using 'sample variance' = s^2 (= 'variance' for MEI)
(= σ_{n-1}^2 on some calculators),
which means dividing by n – 1.

Suppose our sample is 3, 4 and 8.

So here, $s^2 = \dfrac{\sum x^2 - n\bar{x}^2}{n-1} = \dfrac{89 - 3(\frac{15}{3})^2}{2} = 7.$
So s (= 'standard deviation' for MEI) = $\sqrt{7} = 2.65$ (3sf).

Why do we do use n – 1 here?
Because we would like the expected (average) value of the sample variance
to be equal to the population variance σ^2.

It turns out that the expected (average) value of s^2 for samples size n
(so dividing by n – 1) is exactly σ^2 for the whole population.
So dividing by n rather than n – 1
would on average give an estimate for σ^2 that was too small.

Try the spreadsheet that accompanies this page and try to check the above.

Note: statisticians by convention use
Roman characters (the ones we normally use) when talking about <u>*sample*</u> *statistics*
(for example, s for standard deviation and $\bar{x}$ *for the sample mean),*
and Greek characters when talking about <u>*population*</u> *parameters*
(for example, μ *for the population mean*
and σ *for the population standard deviation.)*

www.making-statistics-vital.co.uk