
Subject: VARIANCE in IDL

Posted by [ashmall](#) on Tue, 23 Feb 1999 08:00:00 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear All,

I have a question regarding the variance as calculated by IDL - I expect to get thoroughly flamed by some statistician types but I'm keen to know if I'm wrong!

I always thought the definition of variance was the mean of the squares of the differences from the mean, i.e.:

$$\text{VARIANCE} = \{ \text{SUM} [(x - \text{mean}_x)^2] \} / N$$

and this is what I *thought* I was getting from IDL - it wasn't until I was testing a prog to calculate the means and variances of rows and columns of an array that I spotted that IDL's variance has N-1 as the denominator:

$$\text{VARIANCE} = \{ \text{SUM} [(x - \text{mean}_x)^2] \} / N-1$$

Now I realise the latter (let's call it $\text{Var}(n-1)$) is the best estimate of the variance of the overall population, if my data is a sample from that population, but that's not what I want (or expect) from the variance function.

More worrying is the fact that this isn't mentioned in any way in the on-line help for the VARIANCE function (although the equation does appear in the help on the MOMENT function). Perhaps a keyword to the function would be in order so you could select if you wanted "population estimate" or "sample" variance at the very least.

A simple example is given calculating $\text{Var}(n)$ and $\text{Var}(n-1)$ on the numbers 1,2,3,4,5. The mean is obviously 3 but I would say the variance is 2.0 ($\text{Var}(n)$), not 2.5 as given by IDL ($\text{Var}(n-1)$).

I'd be interested to hear if my definition of variance is correct and whether other people made the same assumption regarding variance as myself. Incidentally, I use IDL 5.1.1.

Thanks,

Justin
