
Subject: Re: VARIANCE in IDL
Posted by [Martin Schultz](#) on Wed, 24 Feb 1999 08:00:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Justin Ashmall wrote:

```
>  
> Dear All,  
>  
> [...]  
>  
> VARIANCE = { SUM [ (x - mean_x)^2 ] } / N  
>  
> [vs.]  
>  
> VARIANCE = { SUM [ (x - mean_x)^2 ] } / N-1  
>
```

as N approaches infinity, the tiny 1 doesn't matter. For samples with low N, it is always questionable what you can learn from the variance. In these cases it is often preferable to use the absolute difference from the mean as a measure for the scatter of the data.

Example:

```
a = [ -1., 1., 2., 8. ]
```

```
b = [ -1., 1., 4., 6. ]
```

	A	B
MEAN=	2.50000	2.50000
VAR(N)=	11.2500	7.25000
VAR(N-1)=	15.0000	9.66667
ABS.DIFF=	2.75000	2.50000

While the variance differs by more than 50% between the two cases, the absolute difference is only 10% which I would consider a fairer result. Because it is a quadratic measure, the variance is very sensitive to outliers, and these have a much greater influence on results for small samples.

Martin.

PS: this is a one-liner to compute the absolute difference:

```
ma = mean(a) & adiff = total(abs(a-ma))/n_elements(a)
```

--

Dr. Martin Schultz

Department for Engineering&Applied Sciences, Harvard University
109 Pierce Hall, 29 Oxford St., Cambridge, MA-02138, USA

phone: (617)-496-8318

fax : (617)-495-4551

e-mail: mgs@io.harvard.edu

Internet-homepage: <http://www-as.harvard.edu/people/staff/mgs/>
