Population Variance and Sample Variance

- https://en.wikipedia.org/wiki/Variance#Population_variance_and_sample_variance

Most simply, the sample variance is computed as an average of squared deviations about the (sample) mean, by dividing by n. However, using values other than n improves the estimator in various ways. Four common values for the denominator are n, n − 1, n + 1, and n − 1.5: n is the simplest (population variance of the sample), n − 1 eliminates bias, n + 1 minimizes mean squared error for the normal distribution, and n − 1.5 mostly eliminates bias in unbiased estimation of standard deviation for the normal distribution.

```
n  for population variance
n-1 for unbiased estimator of sample variance
n+1 minimizes MSE for the normal distribution
n- 1.5 mostly eliminates bias in unbiased estimation of standard deviation
```

Biased and unbiased sample standard deviation

- `numpy` uses `ddof=0` by default and gives biased estimator for `np.std, np.var, np.nanstd`
- `pandas` uses `ddof=1` by default and gives unbiased estimator for `ser.std, ser.var`
- `scipy.stats.sem` gives standard error of mean `s/sqrt(n)` and uses `ddof=1` and gives unbiased estimator. We can use sem to calculate confidence interval. `stats.t.interval(alpha=1-alpha, df=degreeoffreedom, loc=mean,scale=sem)` (there is no x variable, only alpha mean etc).