

Why n-1?

Population Variance

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Sample Variance

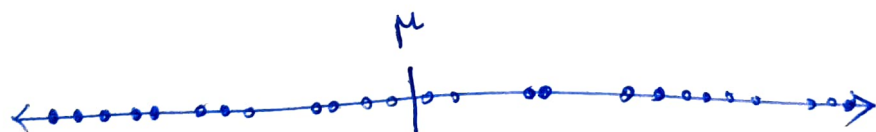
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Reason for n-1:

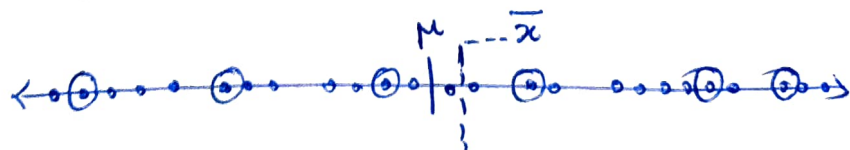
If we divide the numerator by  $n$ , the sample variance value will be very less compared to the population variance, i.e. we are likely to underestimate the true population variance.

Example:

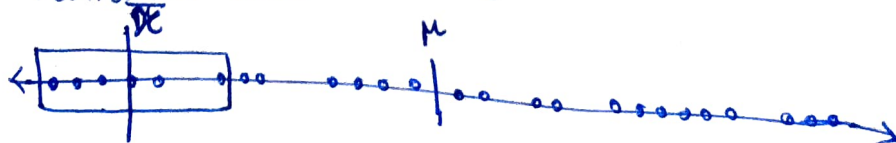
Let us consider an example where  $N=25$



Let us take a sample where  $n=6$



In this case, the sample variance is nearly ~~equal~~ to true variance which is alright. But, in the other case



The sample variance of all the data points will be much far away from the true variance, which is biased.

In order to rectify this, experts have performed various calculations and arrived to conclusion that dividing it by  $n-1$  will make it unbiased.