Bessel's Correction: Effects of (n-1) as the denominator in Standard deviation

By
Saradindu Sengupta

ML Engineer @Nunam

Mean, Variance and Standard Deviation

Sample Variance

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

Population Variance

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

Mean

$$\mu = \frac{\sum_{i=1}^{n} (x_i)}{n}$$

Sample Standard Deviation

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

Population Standard Deviation

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

Why (n-1)

Assume n number of samples are drawn from a population N with mean μ and variance s^2.

The sample standard deviation and variance would contain a bias in all the scenarios because sample mean(x) would be used to calculate sample variance and standard deviation instead of population mean (μ)

Any value in the sample n is going to be closer to sample mean(x) population mean(μ). The sum of squares in variance and standard deviation would be large with population mean(μ) than sample mean(x)

What to do

$$E\left[s^2\right]=\sigma^2+\mu^2-\left(\frac{\sigma^2}{N}+\mu^2\right)=\sigma^2\left(1-\frac{1}{N}\right)$$

To get an unbiased estimator, multiplying both sides by constant (N-1)/N would give

$$S^2=\frac{\sum_{i=1}^{N}\left(x_i-\mu\right)^2}{N-1} \qquad s=\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-\mu\right)^2}{N-1}}$$

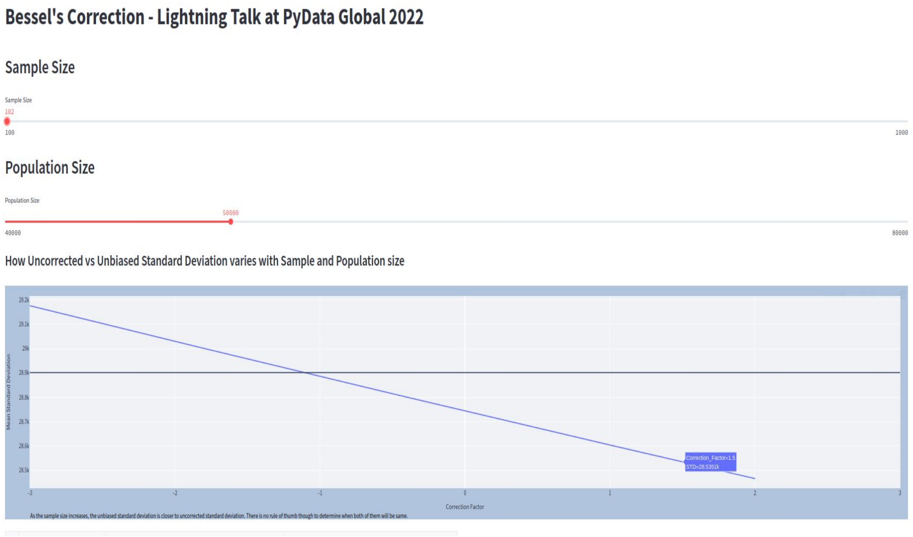
When To use

In cases where the sample size is small compared to the population, because the sum of square deviation is going to be severe

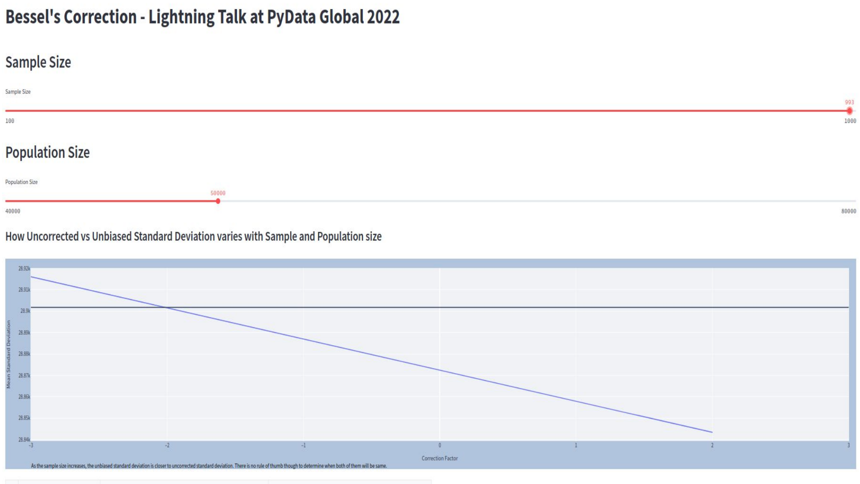
Demo

Link: https://pydata-global-2022-lightning-talk.streamlit.app/

Low Sample Size



High Sample Size



Thank You

/in/saradindusengupta

@iamsaradindu

/saradindusengupta