Measures of Variability: Takeaways 🖻

by Dataquest Labs, Inc. - All rights reserved © 2019

Syntax

• Writing a function that returns the range of an array:

```
def find_range(array):
return max(array) - min(array)
```

• Writing a function that returns the mean absolute deviation of an array:

```
def mean_absolute_deviation(array):
    reference_point = sum(array) / len(array)

distances = []
for value in array:
    absolute_distance = abs(value - reference_point)
    distances.append(absolute_distance)
return sum(distances) / len(distances)
```

• Finding the variance of an array:

```
### If the the array is a `Series` object ###
sample_variance = Series.var(ddof = 1)
population_variance = Series.var(ddof = 0)

### If the array is not a `Series` object ###
from numpyimport var
sample_variance = var(a_sample, ddof = 1)
population_variance = var(a_population, ddof = 0)
```

• Finding the standard deviation of an array:

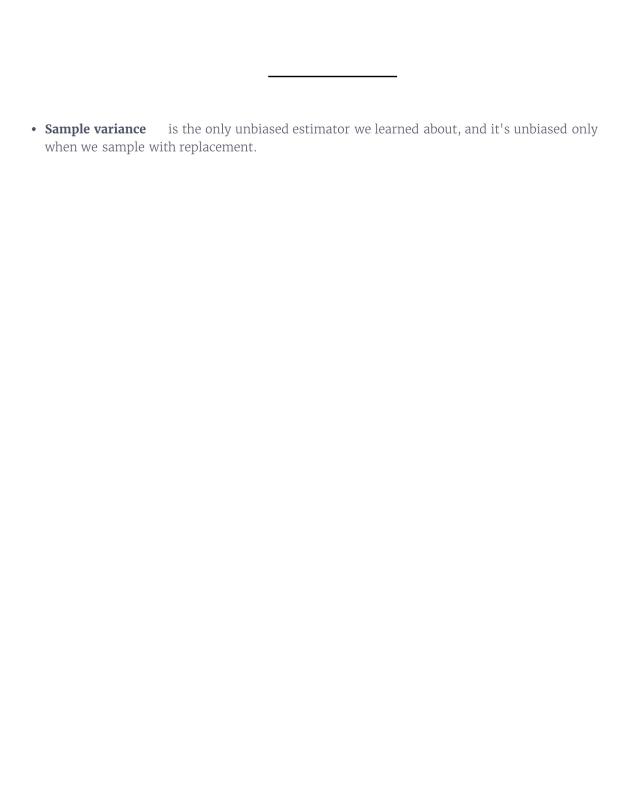
```
### If the array is a `Series` object ###
sample_stdev = Series.std(ddof = 1)
population_stdev = Series.std(ddof = 0)

### If the array is not a `Series` object ###
from numpyimport std
sample_stdev = std(a_sample, ddof = 1)
population_stdev = std(a_population, ddof = 0)
```

Concepts

- There are many ways we can measure the **variability** of a distribution. These are some of the measures we can use:
 - The range.
 - The mean absolute deviation.
 - The variance.
 - The standard deviation.
- Variance and standard deviation are the most used metrics to measure variability. To compute the standard deviation and the variance for a **population**, we can use the formulas:

• To compute the standard deviation and the variance for a **sample**, we need to add the **Bessel's correction** to the formulas above:



Resources

- An intuitive introduction to variance and standard deviation.
- Useful documentation:
 - numpy.var()
 - numpy.std()
 - Series.var()
 - Series.std()



Takeaways by Dataquest Labs, Inc. - All rights reserved $\ensuremath{\text{@}}$ 2019