Measures of Variability: Takeaways

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 numpy import 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 numpy import 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 σ and the variance σ^2 for a **population**, we can use the formulas:

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

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

- To compute the standard deviation s and the variance s^2 for a **sample**, we need to add the **Bessel's correction** to the formulas above:

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

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

- **Sample variance** s^2 is the only unbiased estimator we learned about, and it's unbiased only when we sample with replacement.

Resources

- [An intuitive introduction to variance and standard deviation](https://).

- Useful documentation:
  - `numpy.var()`
  - `numpy.std()`
  - `Series.var()`
  - `Series.std()`