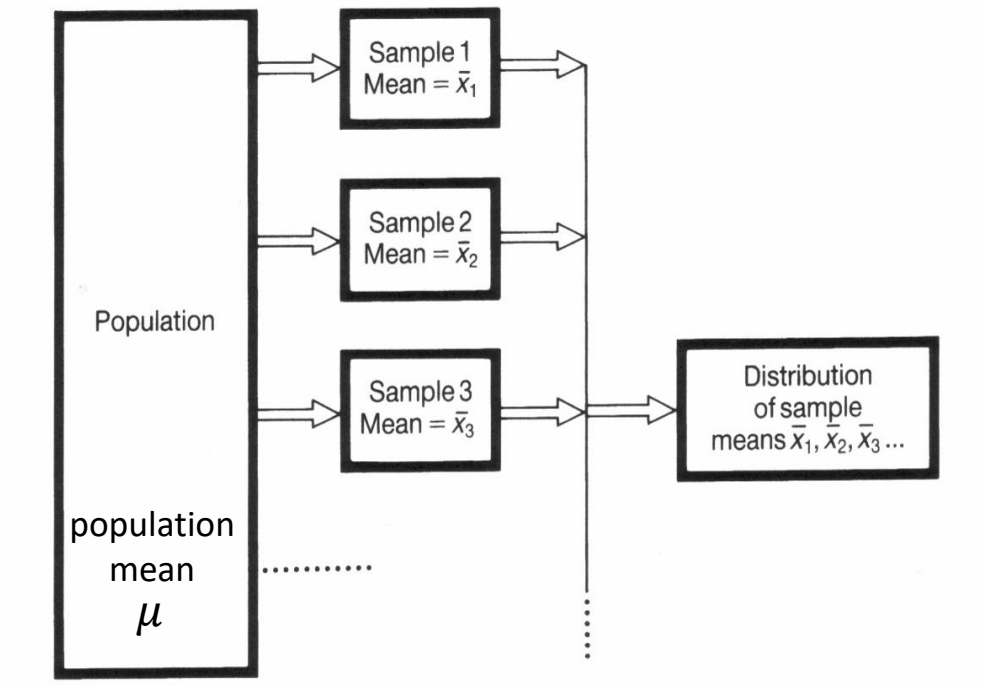
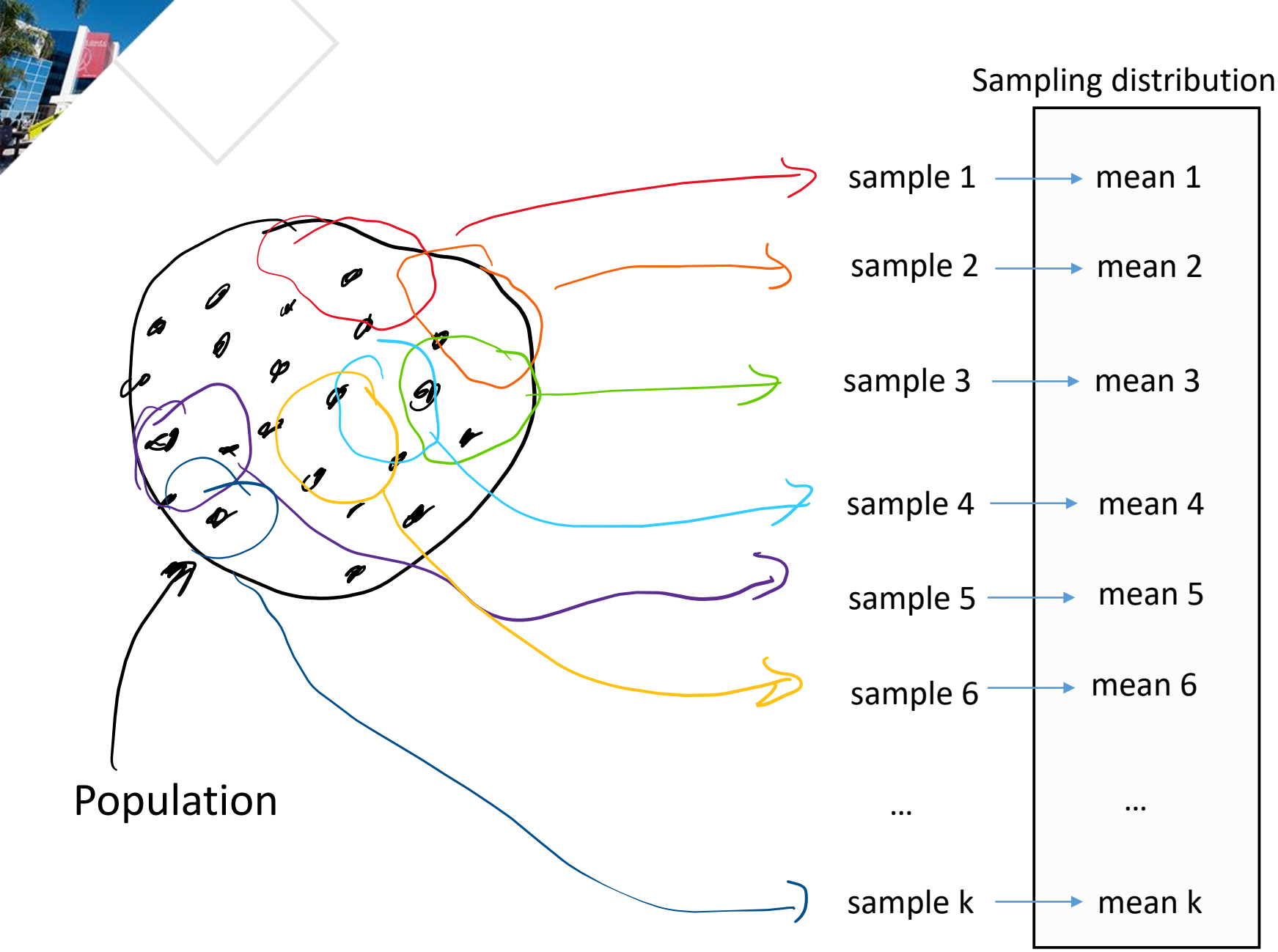
SAMPLING DISTRIBUTION

SAMPLING FLUCTUATIONS

- For every sample that I extract, I will obtain different statistics and estimates.
- The variation is described by a **sampling distribution**.




Sampling distribution
sample 1
mean 1
sample 2
mean 2
sample 3
mean 3
sample 4
mean 4
sample 5
mean 5
sample 6
mean 6
...
...
sample k
mean k
Population

SAMPLING DISTRIBUTIONS: USEFUL FACTS

Consider an univariate population with mean **μ** and standard deviation **σ**. Let's take samples of size **n**. Then:

- If a population is normally distributed, the sampling distribution of the mean is also normally distributed.
- With large samples (more than 30) the sampling distribution of the mean is approximately normally distributed regardless of the population distribution.
- In the two previous cases, the sampling distribution of the sample means has a mean **μ** and standard deviation $\frac{\sigma}{\sqrt{n}}$

EXAMPLE ON PAPER

- Consider a dataset with all the weekly hours of Netflix consumption on European families. Suppose that this dataset has an average viewing time of 6 hours and a standard deviation of 1,2 hours.

- We collect many subsamples of 100 households each and compute the average viewing time for each of the samples. **What is the average of these subsamples? And the standard error?**

Solution:

Mean of the samples must also be (close to) 6 hours

Standard error must be approximately $\frac{1,2}{\sqrt{100}} = 0,12$ hours

Python practice

Let’s sample randomly some known datasets and distributions, to see how the sample mean is actually a random variable.