

## 9: Resampling Methods (The Bootstrap)

```
$ echo "Data Science Institute"
```

## Activity (15 minutes)

Watch this video about Bootstrapping: <https://www.youtube.com/watch?v=uGsf3spCM3Y>

# The Bootstrap

Suppose we wish to find the average of the population of Toronto  $\mu$  and we have a sample of size  $n$ . We can find the mean of the sample  $\mu_s$  but this does not give any indication for how this compares to the true population mean  $\mu$ .

# The Bootstrap

◆ *The bootstrap can be used to quantify the uncertainty of an estimate* ◆ in the following way:

- Randomly sample  $n$  observations from the original sample to acquire a new sample of the same size (repeat observations are allowed).
- Compute the desired statistic (i.e. average age) of this new sample.
- Repeat steps 1-2 many times.
- Compute the standard error (SE) of the estimates.

This method is able to give us an estimate of the variability associated with our sample mean  $\mu_s$ .

# Breakout Room

When should you use Bootstrapping over Cross Validation methods?

# Exercises: The Bootstrap

Open the The Bootstrap Jupyter Notebook file.

- Go over the "The Bootstrap" section together as a class.

# References

Chapter 5 of the ISLP book:

James, Gareth, et al. "Resampling Methods." An Introduction to Statistical Learning: with Applications in Python, Springer, 2023.