

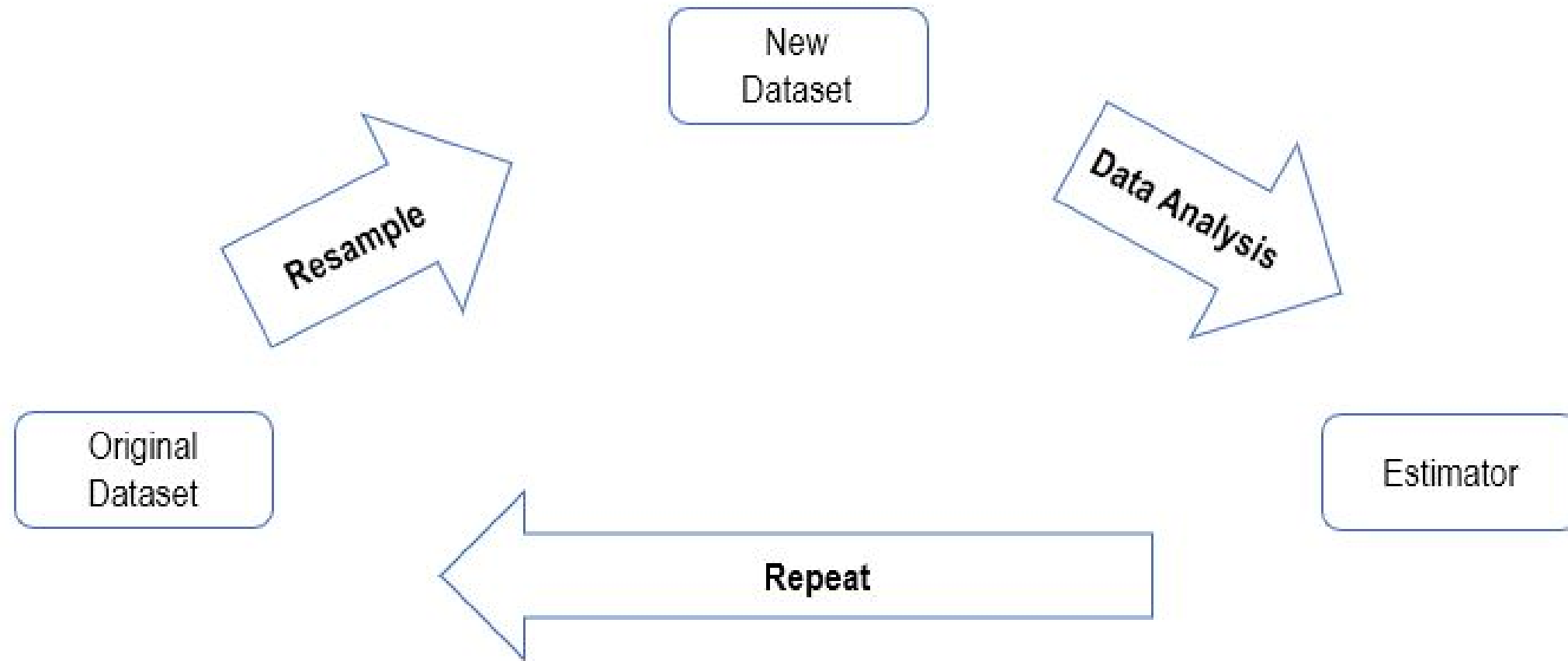
Introduction to resampling methods

STATISTICAL SIMULATION IN PYTHON



Tushar Shanker
Data Scientist

Resampling workflow



Why resample?

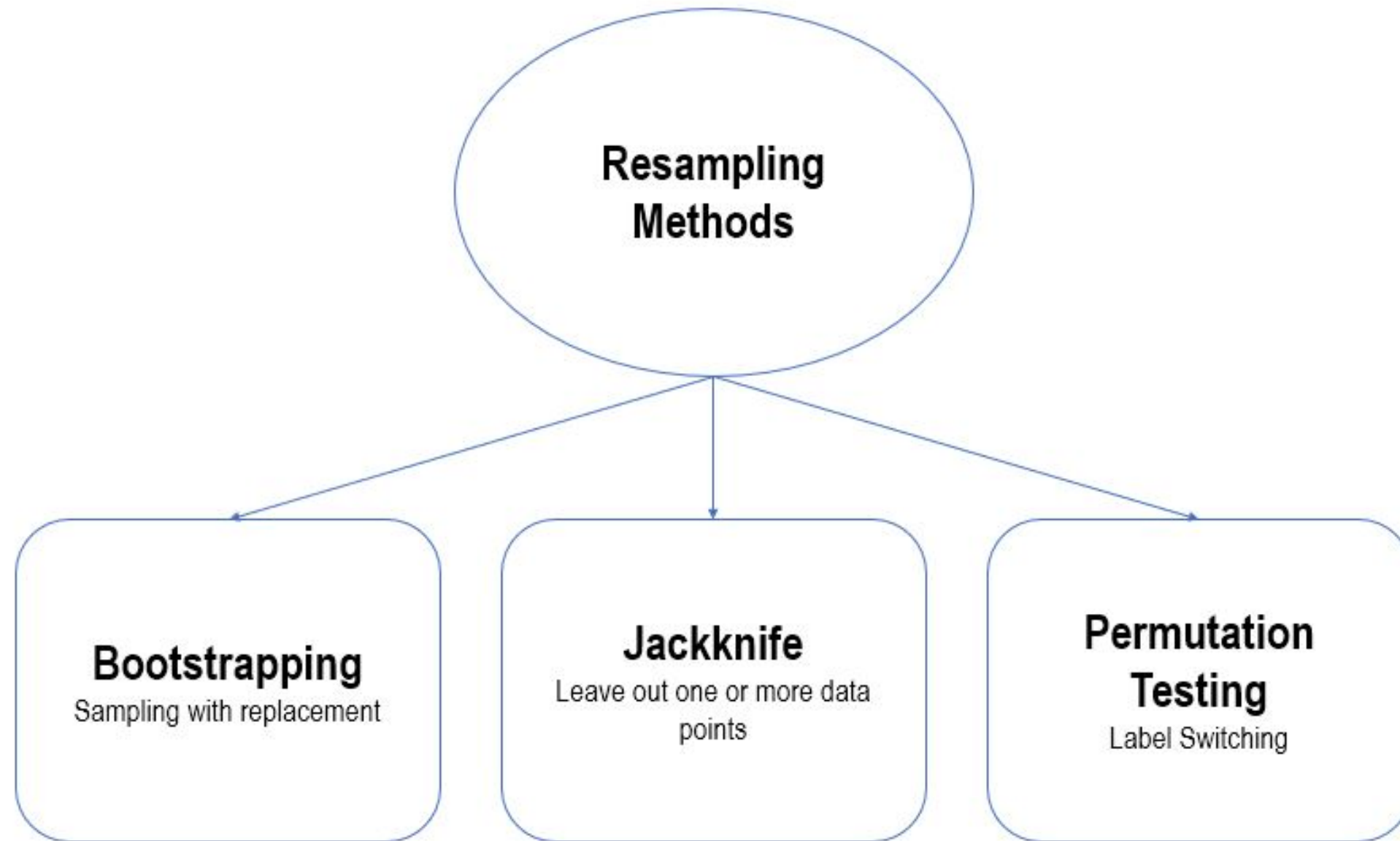
Advantages

- Simple implementation procedure.
- Applicable to complex estimators.
- No strict assumptions.

Drawbacks

- Computationally expensive.

Types of resampling methods



Let's practice!

STATISTICAL SIMULATION IN PYTHON

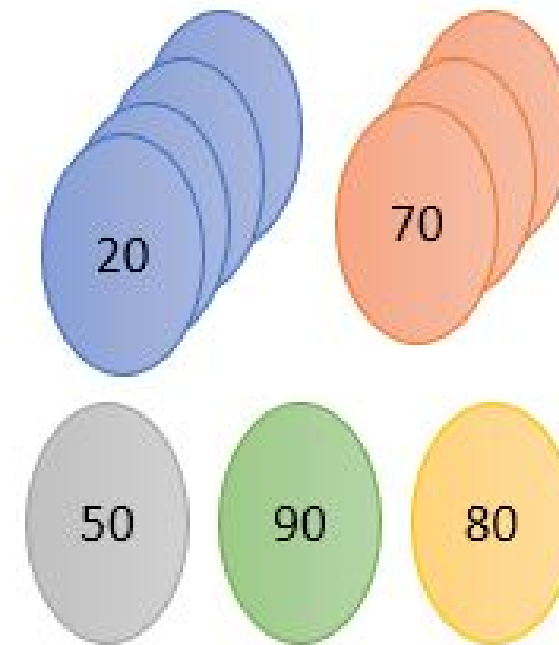
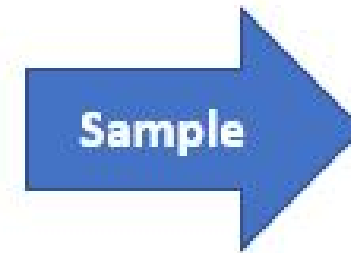
Bootstrapping

STATISTICAL SIMULATION IN PYTHON

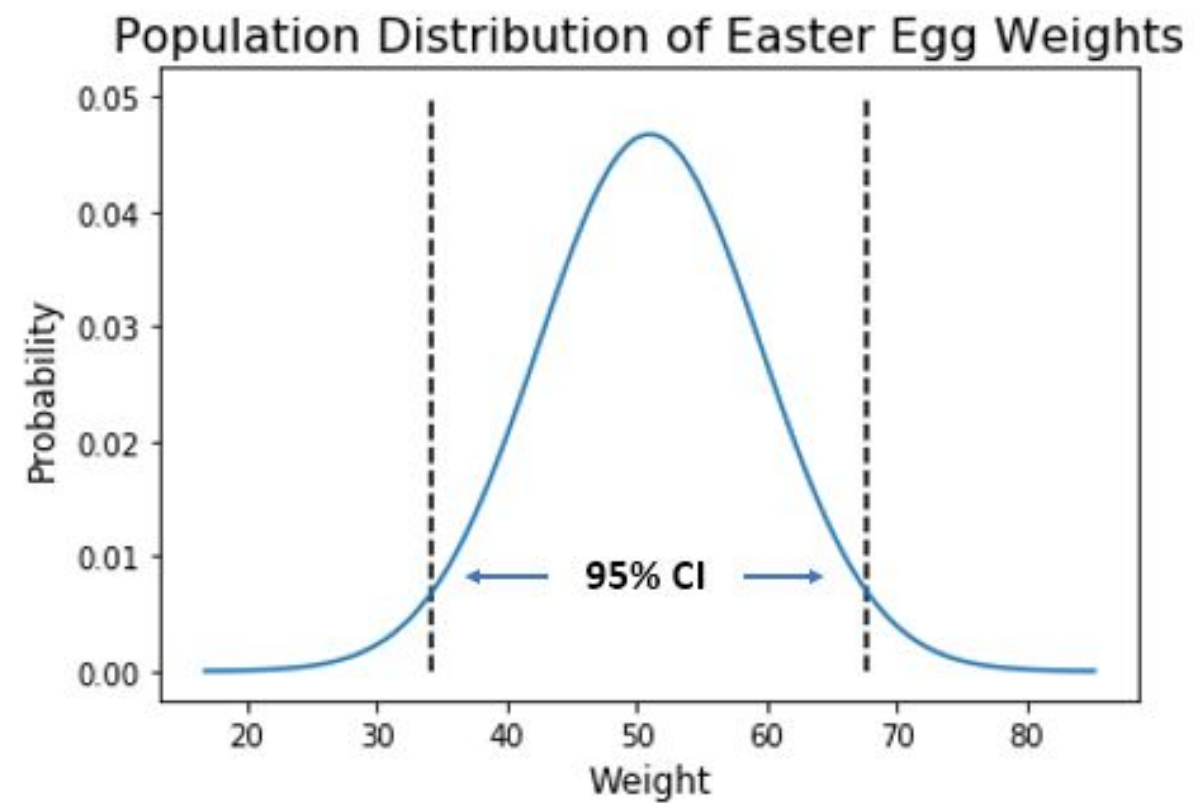
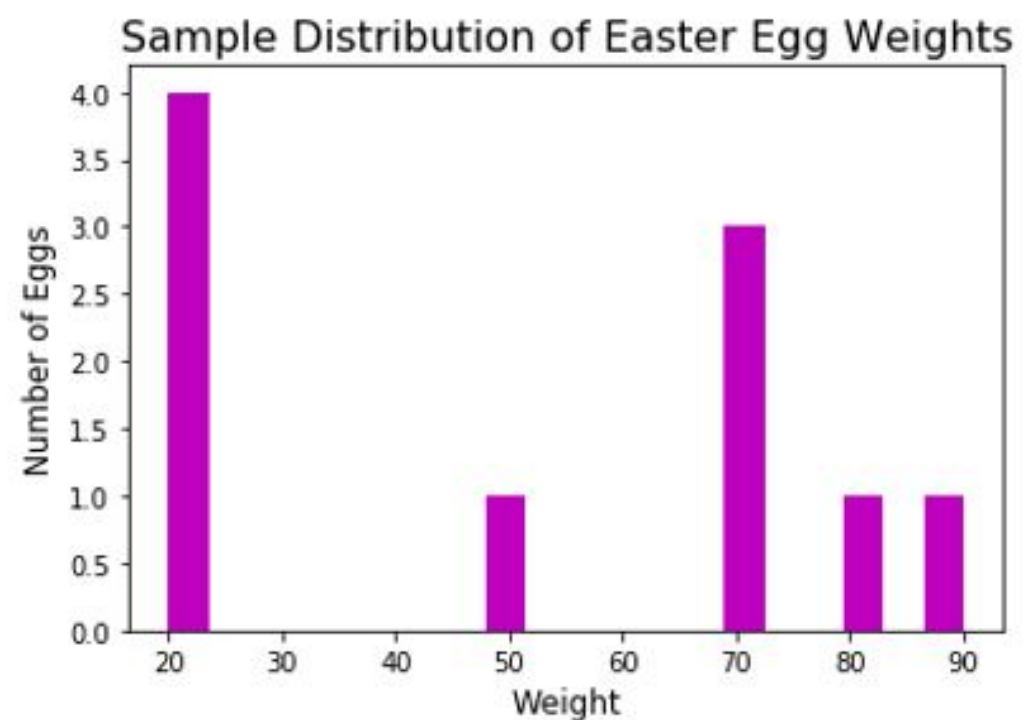


Tushar Shanker
Data Scientist

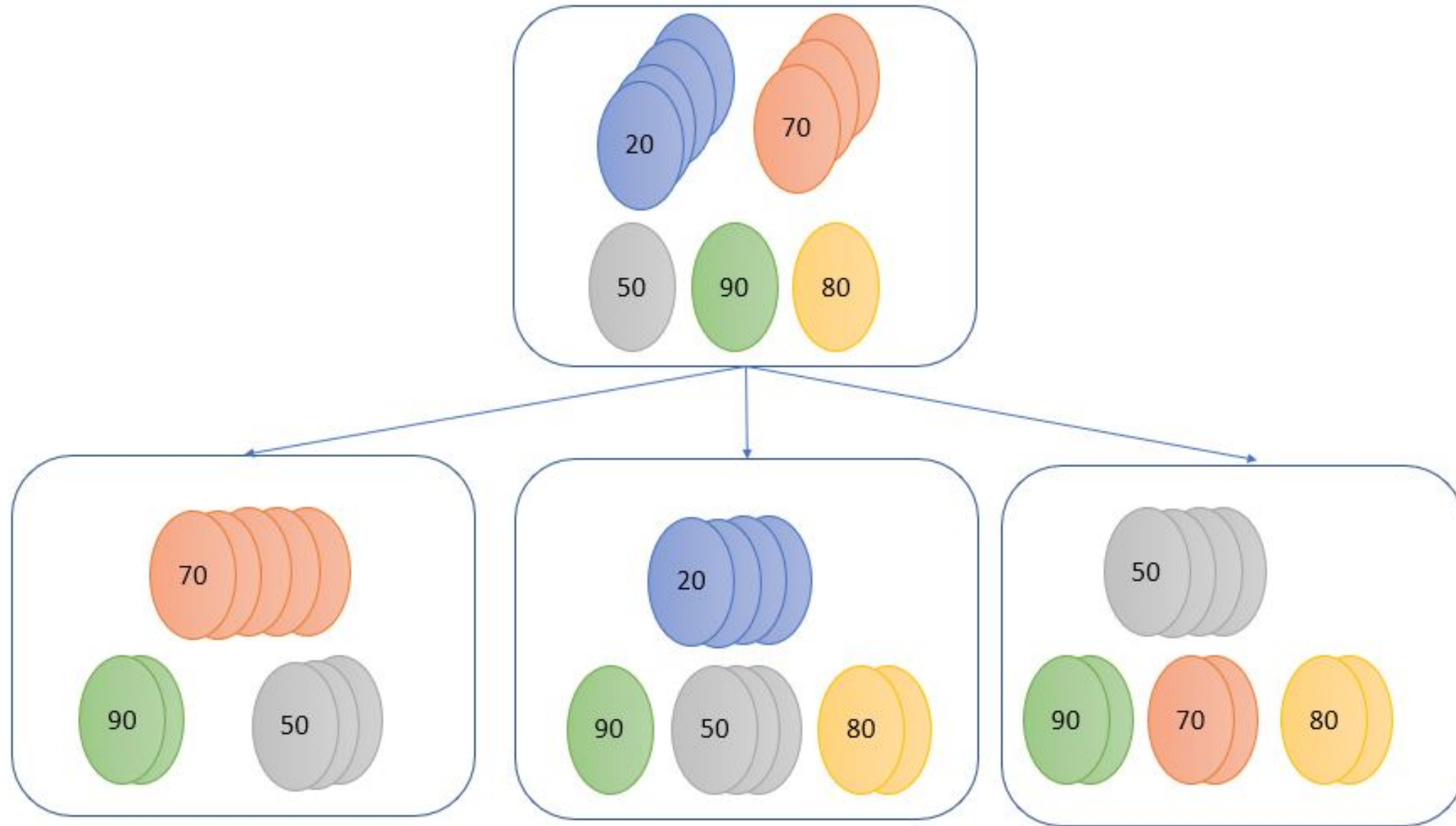
Easter eggs



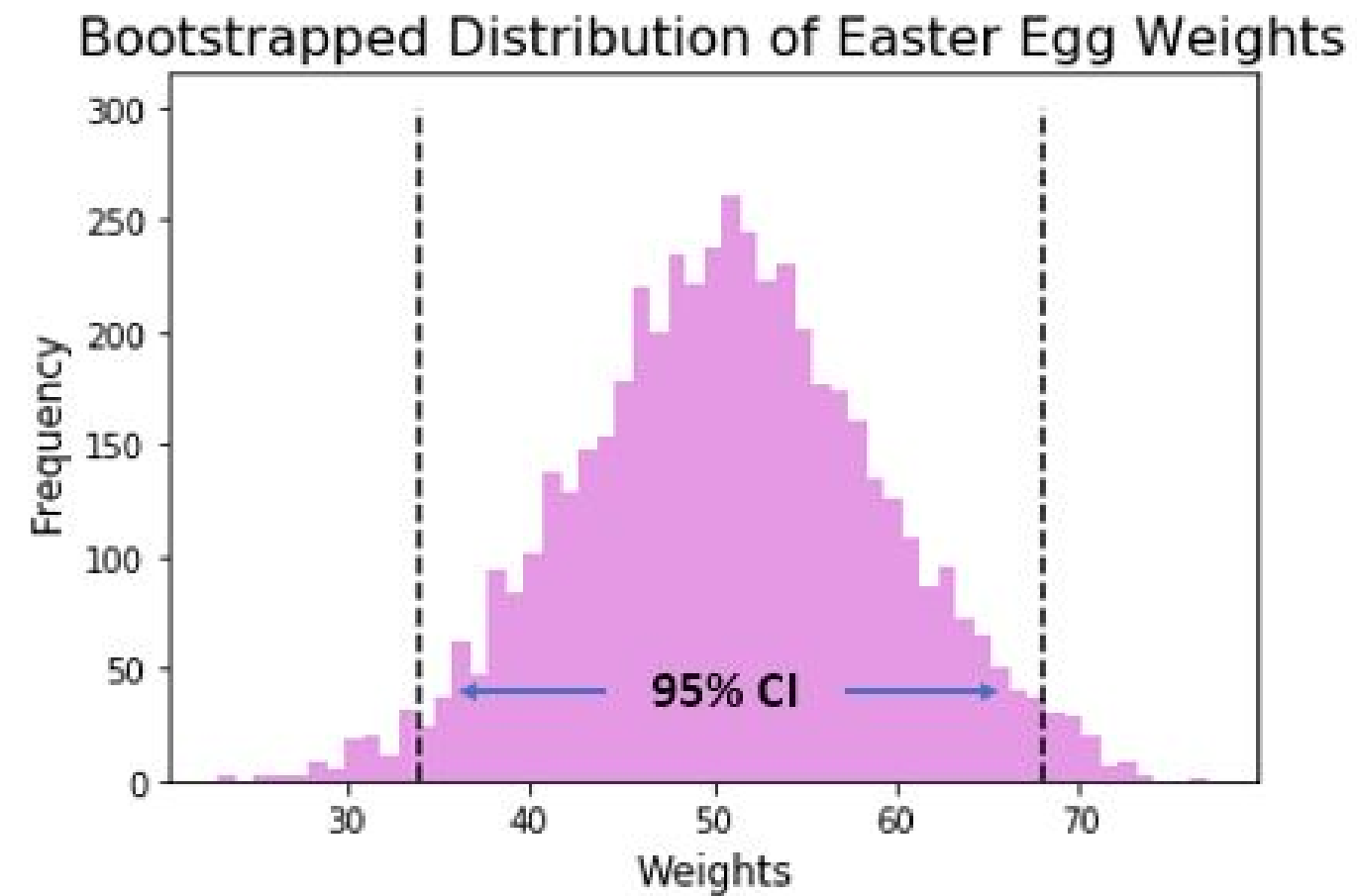
Easter eggs



Bootstrapping Easter eggs



Bootstrapped distribution



Bootstrap - Good to know

- Run at least 5-10k iterations.
- Expect an approximate answer.
- Consider bias correction.

Let's practice!

STATISTICAL SIMULATION IN PYTHON

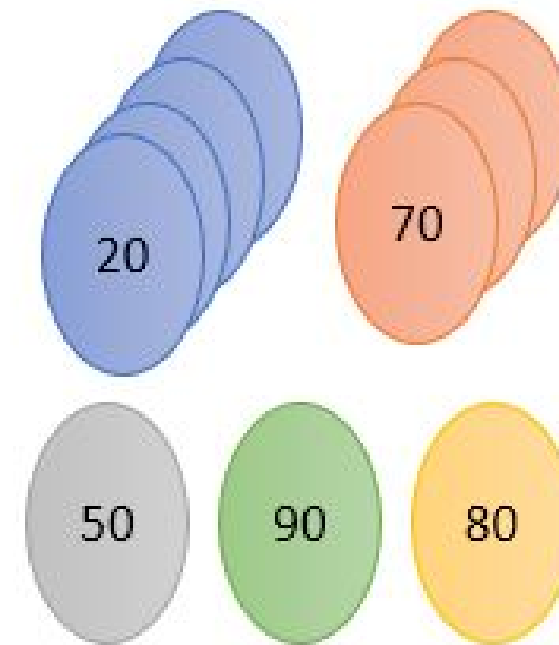
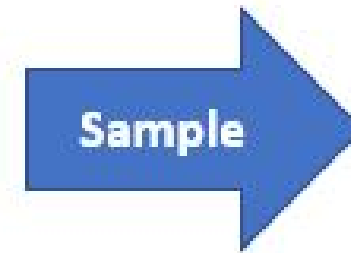
Jackknife resampling

STATISTICAL SIMULATION IN PYTHON

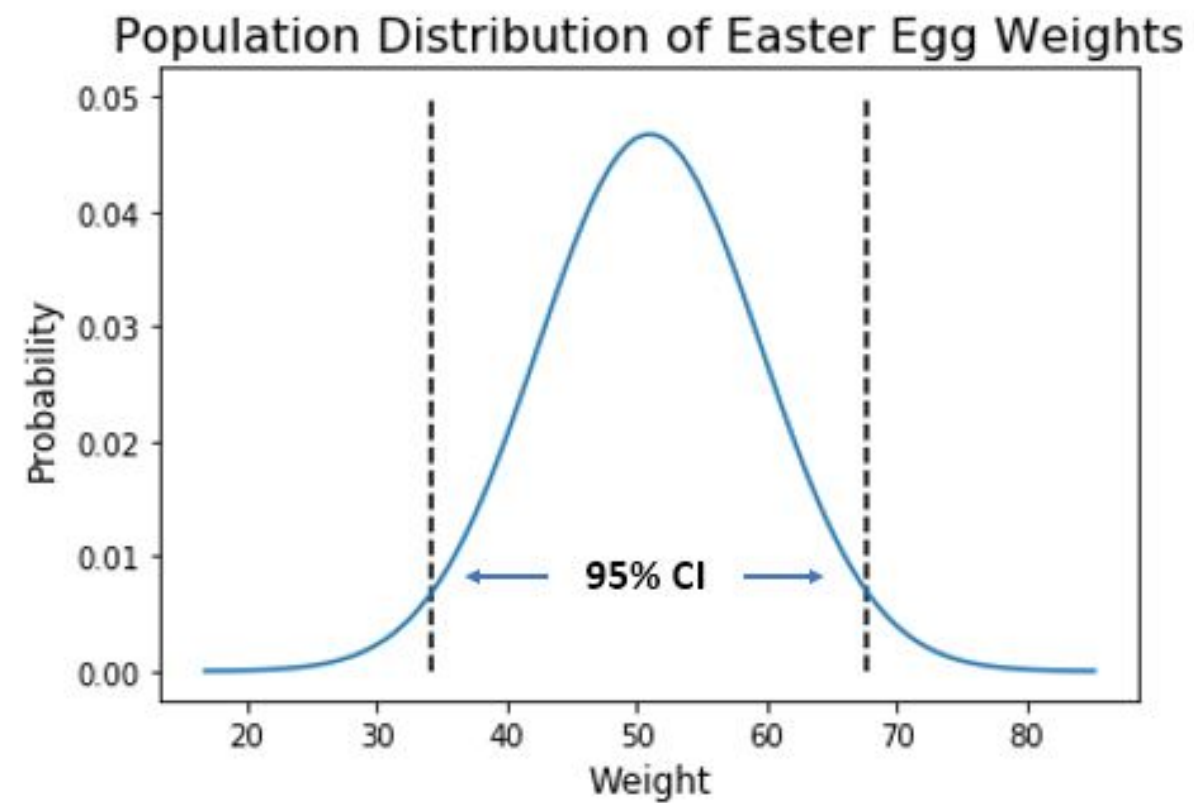
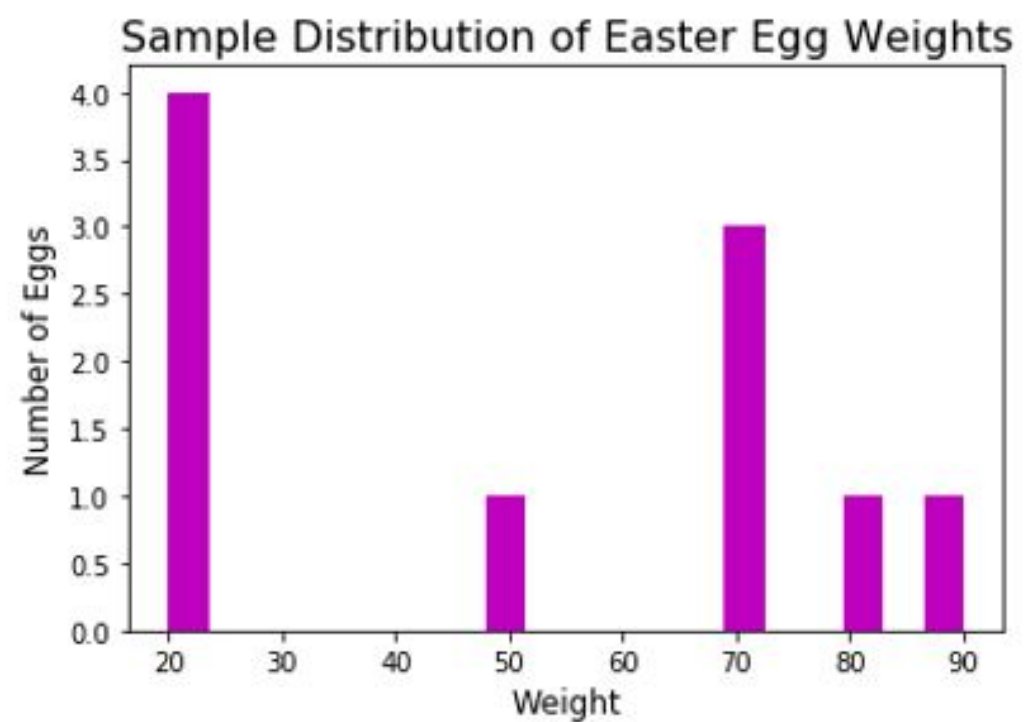


Tushar Shanker
Data Scientist

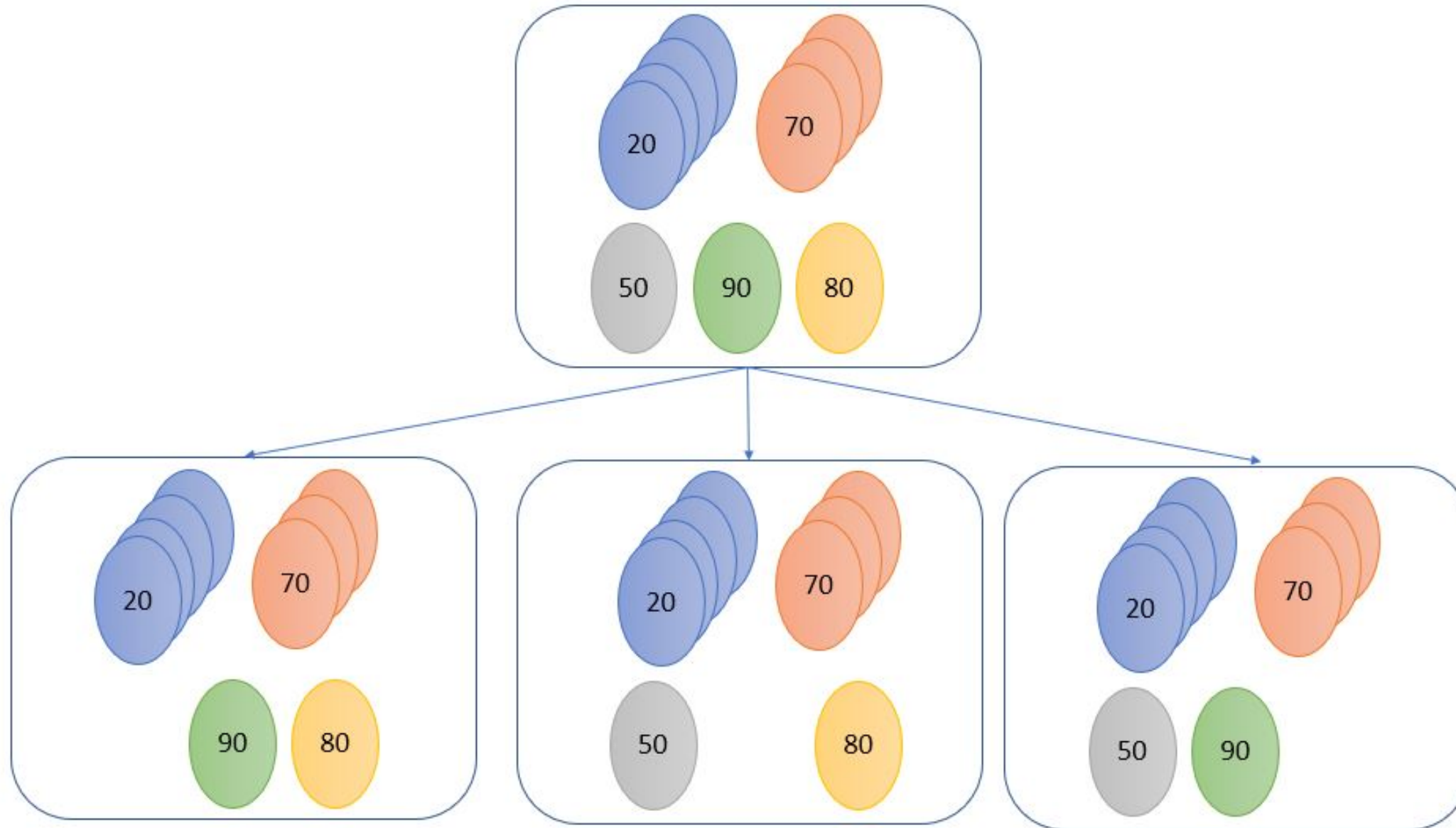
Easter eggs



Easter eggs



Jackknifing Easter eggs



Jackknife estimate

Jackknife Estimate

$$\hat{\theta}_{jackknife} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i$$

$\hat{\theta}_{jackknife}$: Jackknife Estimate, $\hat{\theta}_i$: Estimate for each Jackknife Sample

Variance of Jackknife Estimate

$$Var(\hat{\theta}_{jackknife}) = \frac{n-1}{n} \sum (\hat{\theta}_i - \hat{\theta}_{jackknife})^2$$

$\hat{\theta}_{jackknife}$: Jackknife Estimate $\hat{\theta}_i$: Estimate for each Jackknife Sample

Useful When underlying Distribution of the data is unknown

Jackknife vs Bootstrap

Jackknife

- Mean Weight = 51g
- 95% CI = [33.36g, 68.64g]

Leave one out resampling

Bootstrap

- Mean Weight = 50.8g
- 95% CI = [35g, 67.03g]

Sampling With replacement

Let's practice!

STATISTICAL SIMULATION IN PYTHON

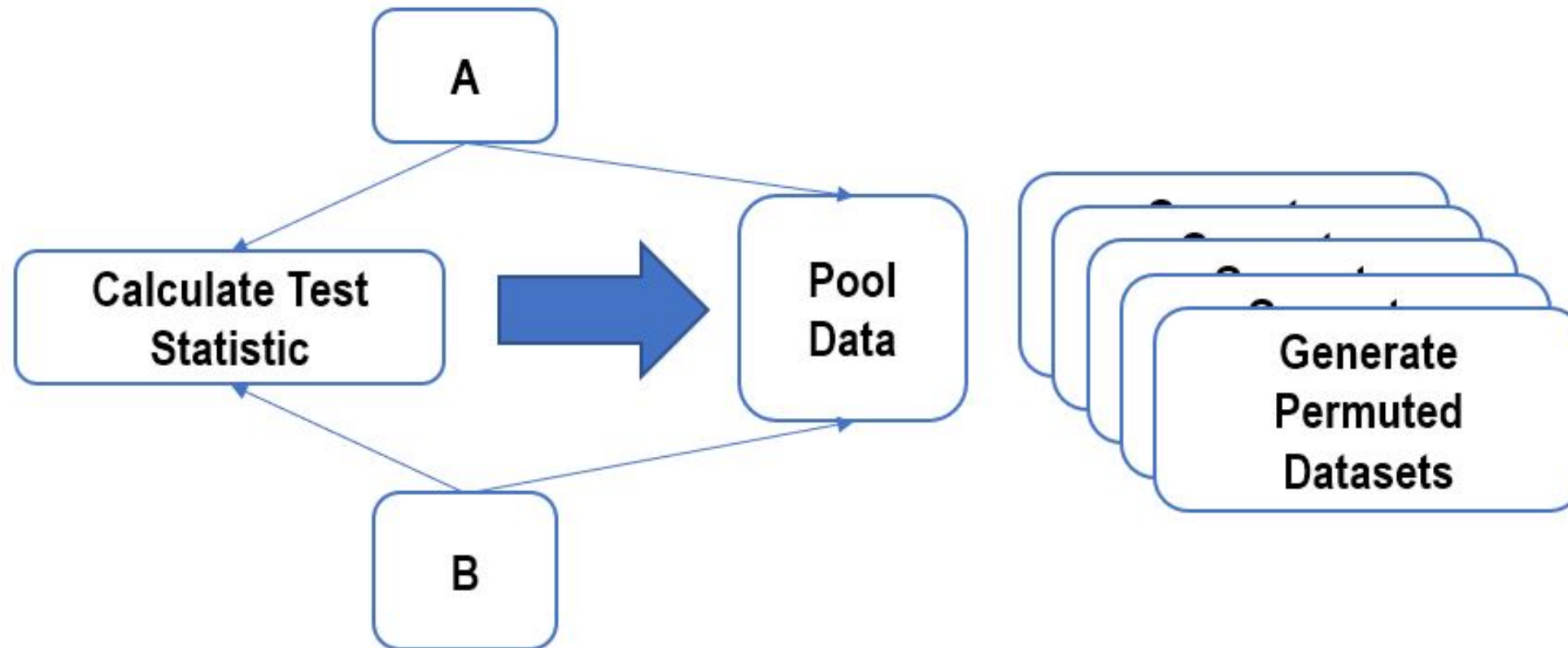
Permutation testing

STATISTICAL SIMULATION IN PYTHON

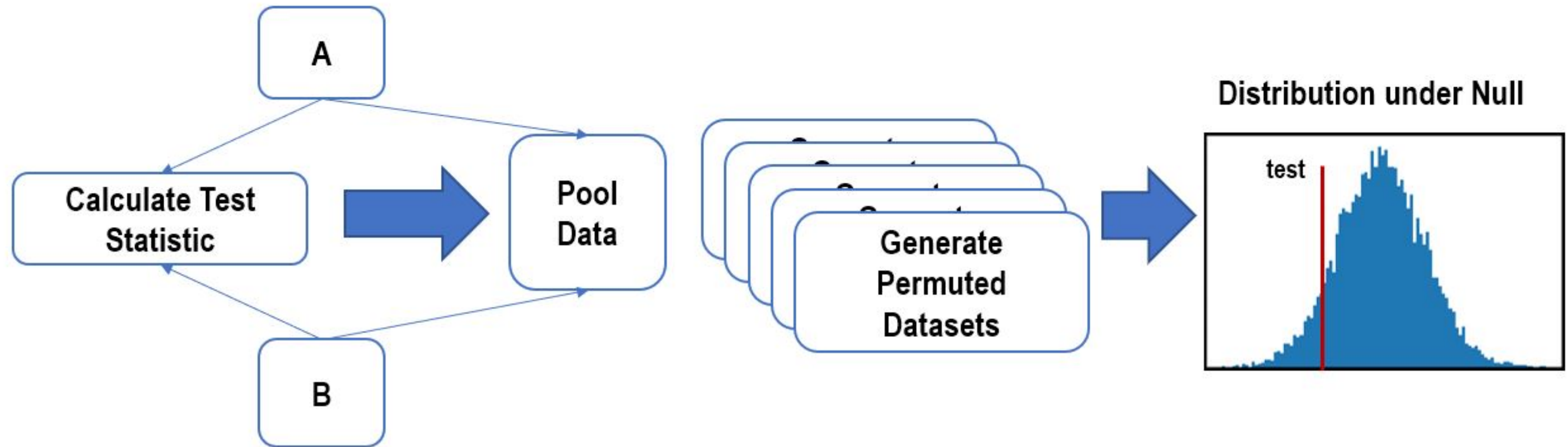


Tushar Shanker
Data Scientist

Steps involved



Steps involved



Discussion

Advantages

- Very flexible
- No strict assumptions
- Widely applicable

Drawbacks

- Computationally Expensive
- Custom coding required

Donation website

Donation Website Design Comparison



Let's practice!

STATISTICAL SIMULATION IN PYTHON