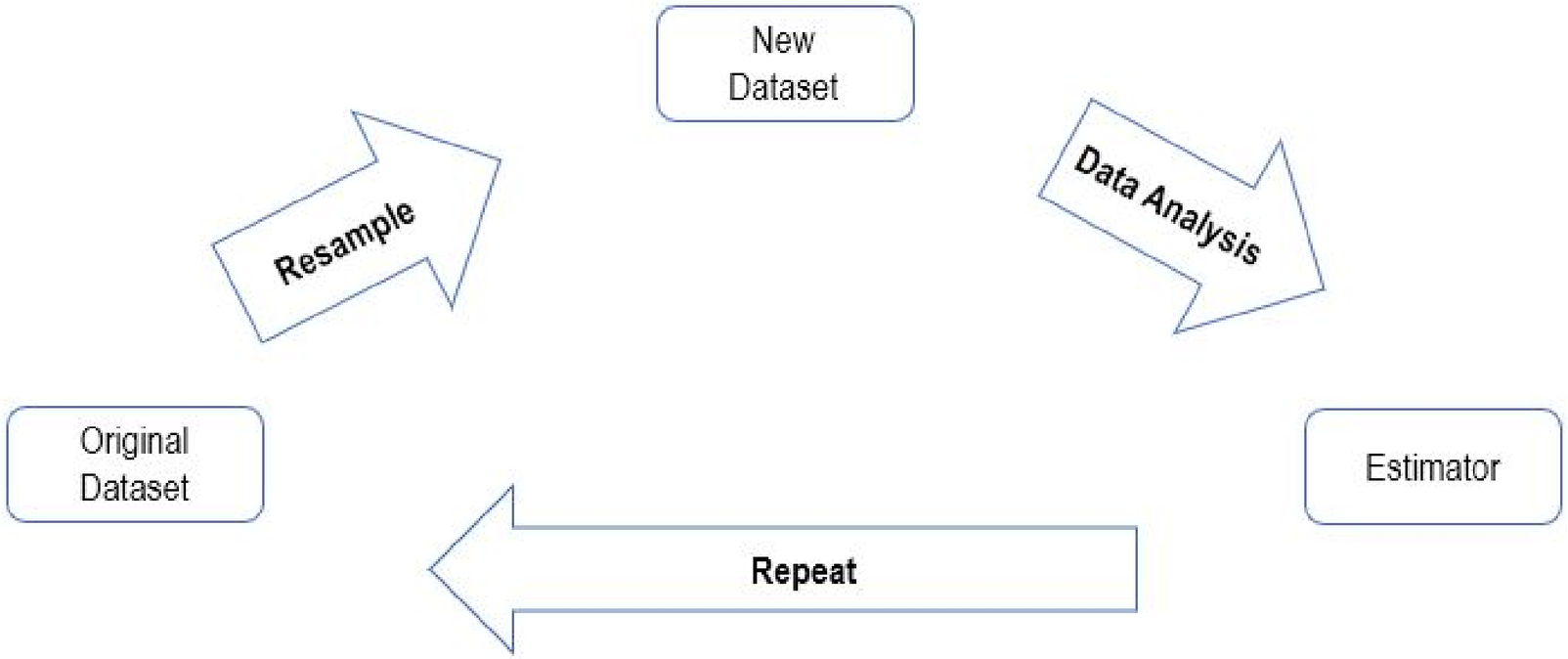
Introduction to resampling methods

STATISTICAL SIMULATION IN PYTHON

Tushar Shanker
Data Scientist

datacamp

Resampling workflow



Why resample?

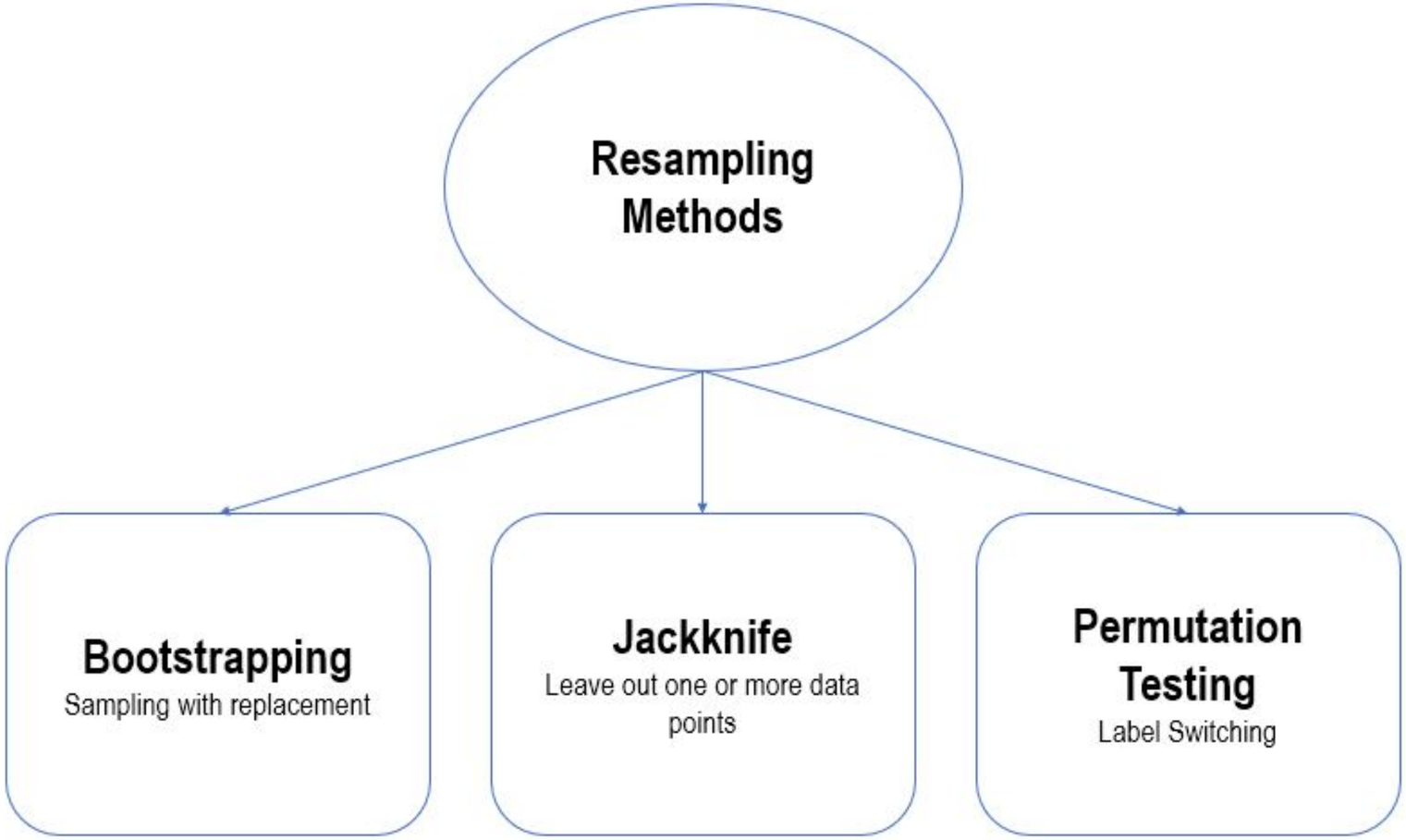
Advantages

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

Drawbacks

- Computationally expensive.

Types of resampling methods



Let's practice!

STATISTICAL SIMULATION IN PYTHON

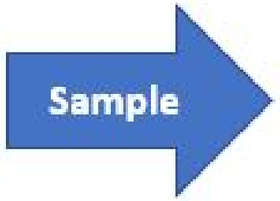
Bootstrapping

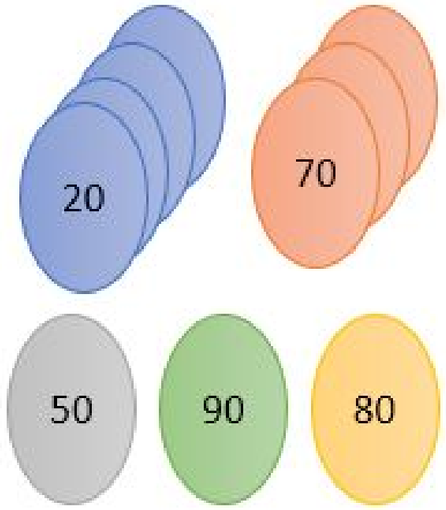
STATISTICAL SIMULATION IN PYTHON

Tushar Shanker
Data Scientist

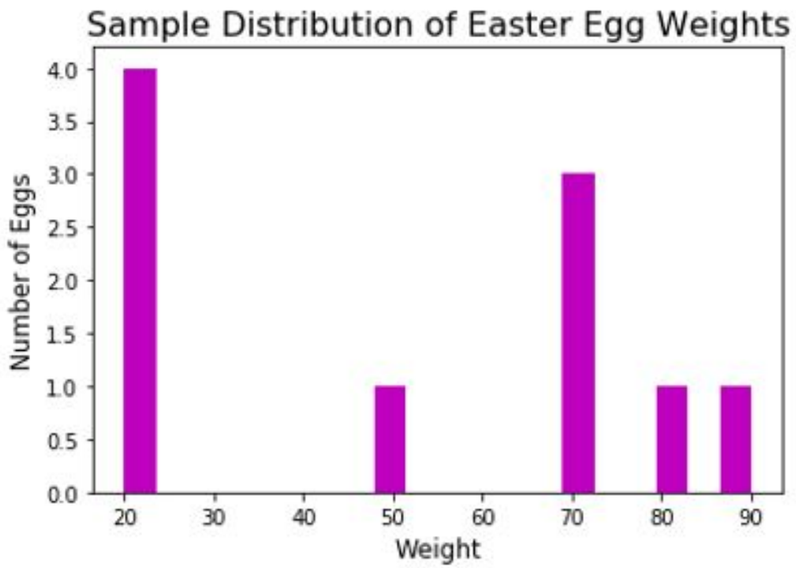
datacamp

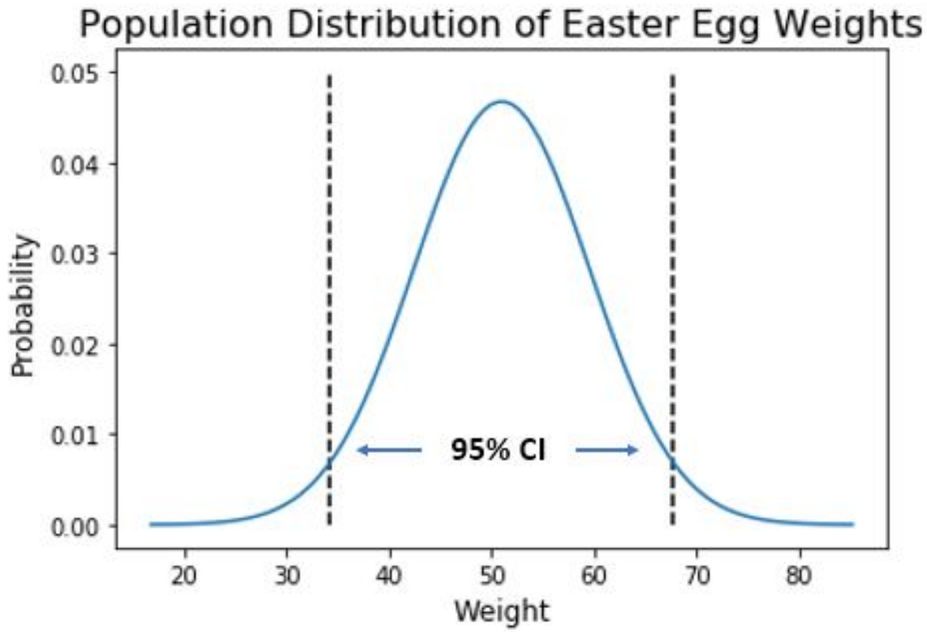
Easter eggs



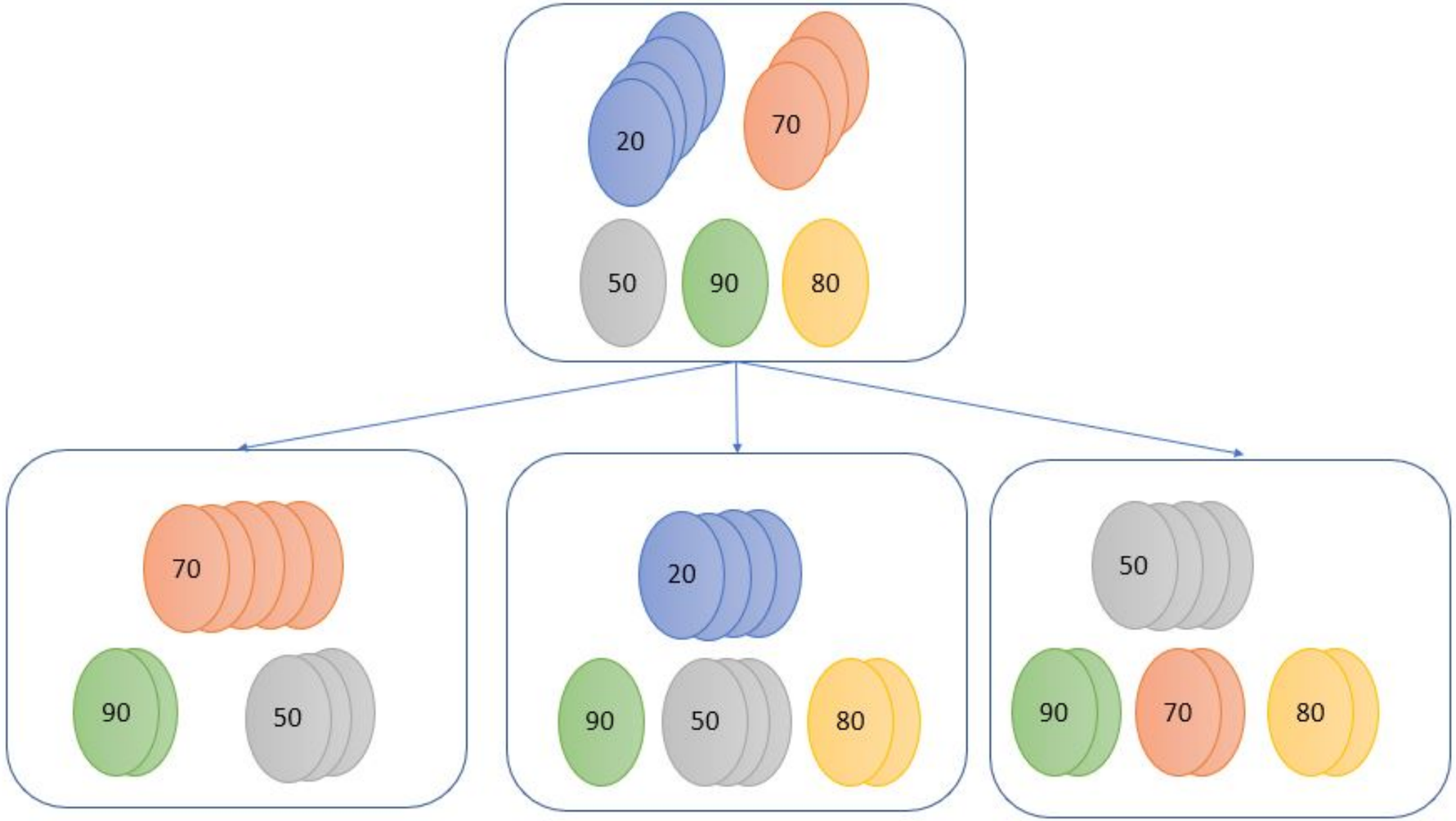


Easter eggs

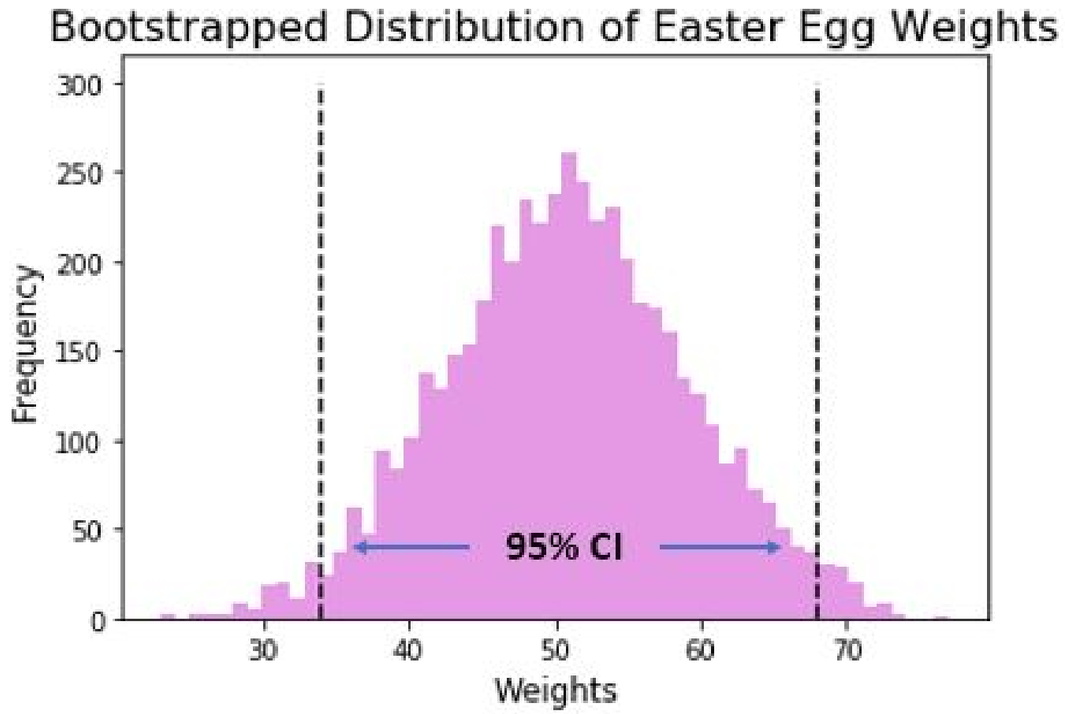

Sample Distribution of Easter Egg Weights
Number of Eggs
Weight
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
20
30
40
50
60
70
80
90


Population Distribution of Easter Egg Weights
Probability
Weight
95% CI
0.00
0.01
0.02
0.03
0.04
0.05
20
30
40
50
60
70
80

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

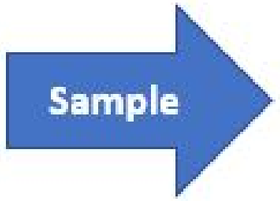
Jackknife resampling

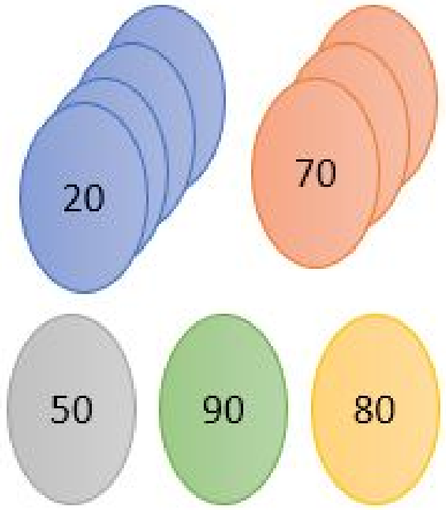
STATISTICAL SIMULATION IN PYTHON

Tushar Shanker
Data Scientist

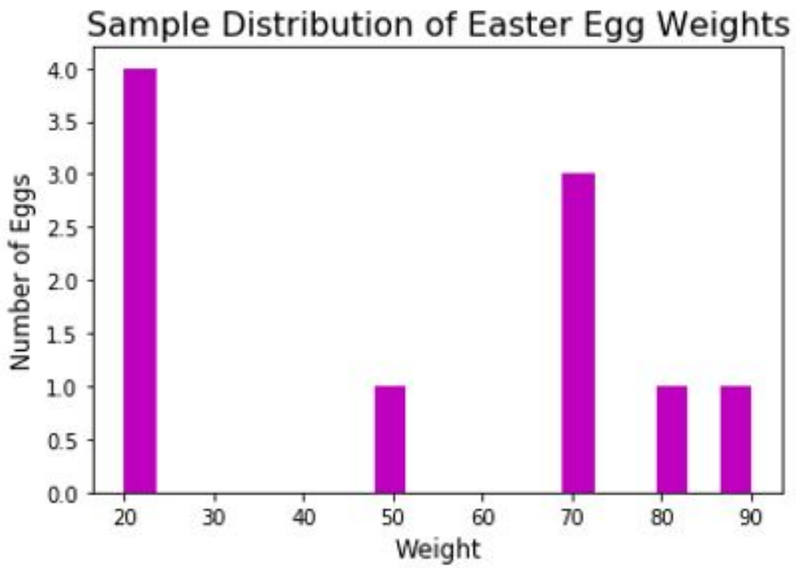
datacamp

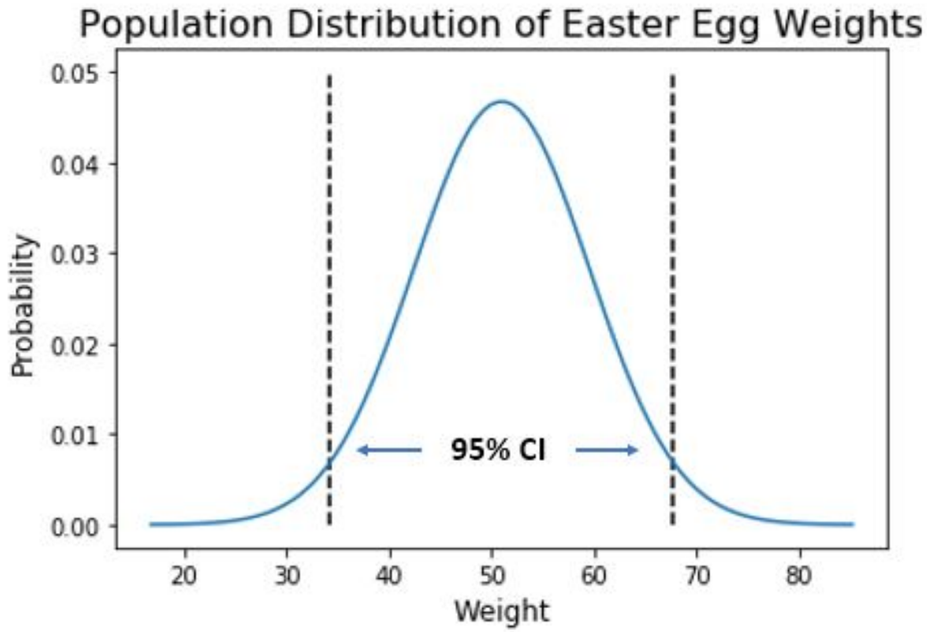
Easter eggs



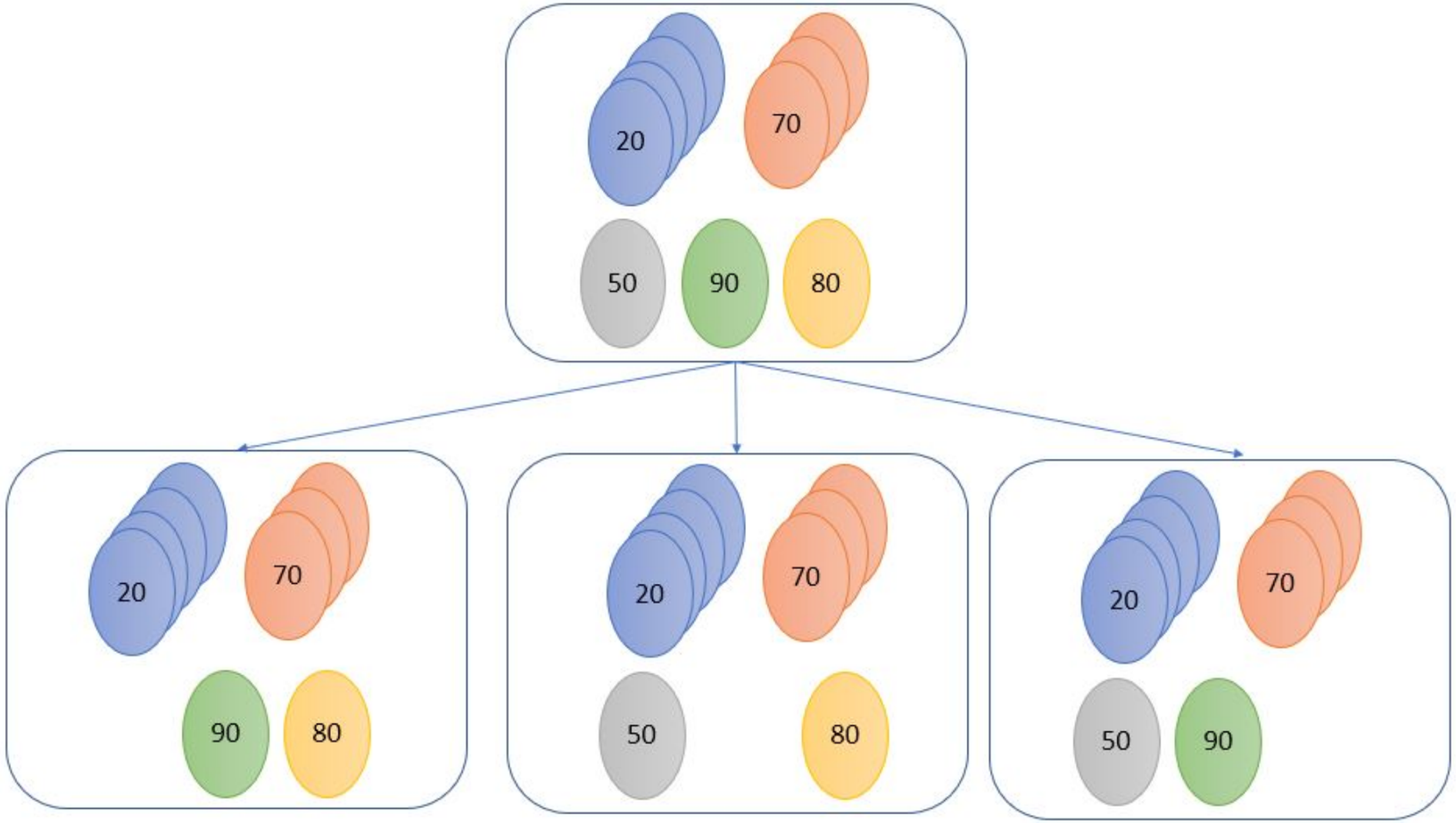


Easter eggs


Sample Distribution of Easter Egg Weights
Number of Eggs
Weight
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
20
30
40
50
60
70
80
90


Population Distribution of Easter Egg Weights
Probability
Weight
95% CI
0.00
0.01
0.02
0.03
0.04
0.05
20
30
40
50
60
70
80

Jackknifing Easter eggs



Jackknife estimate

Jackknife Estimate

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

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

Variance of Jackknife Estimate

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

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

Jackknife vs Bootstrap

Jackknife

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

Bootstrap

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

Let's practice!

STATISTICAL SIMULATION IN PYTHON

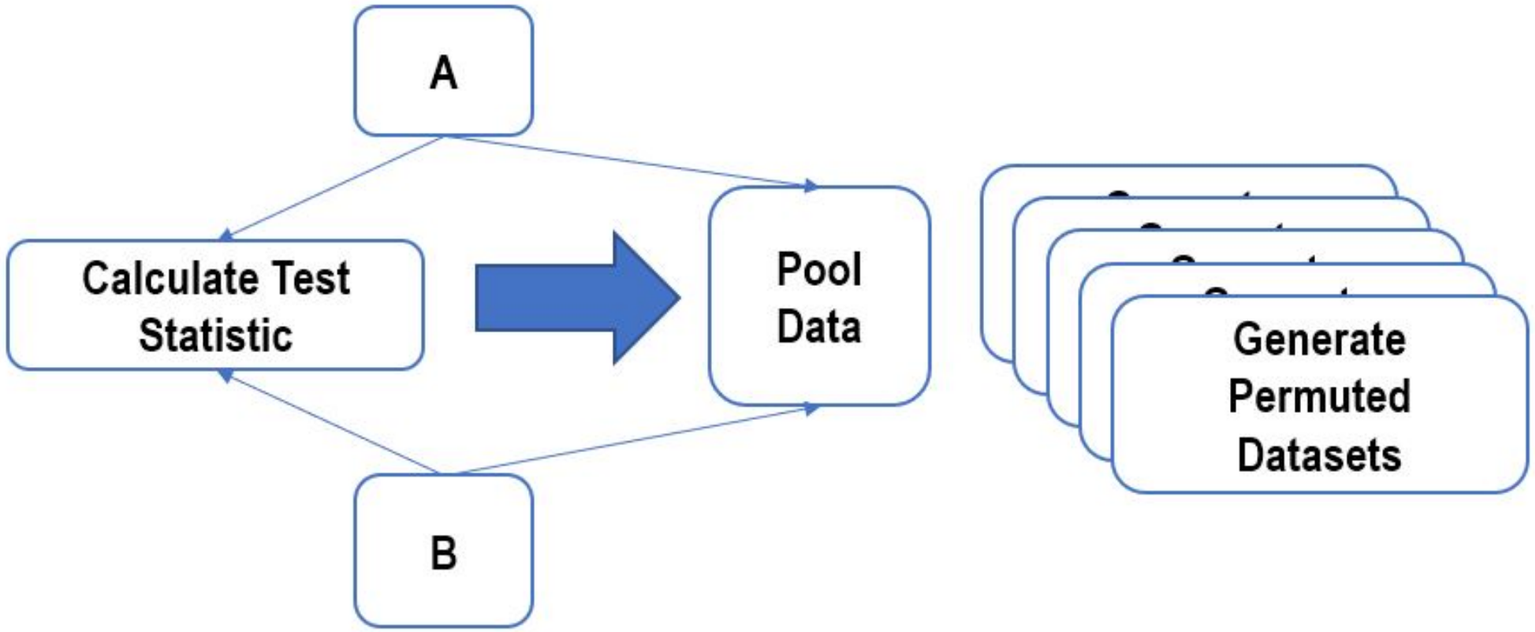
Permutation testing

STATISTICAL SIMULATION IN PYTHON

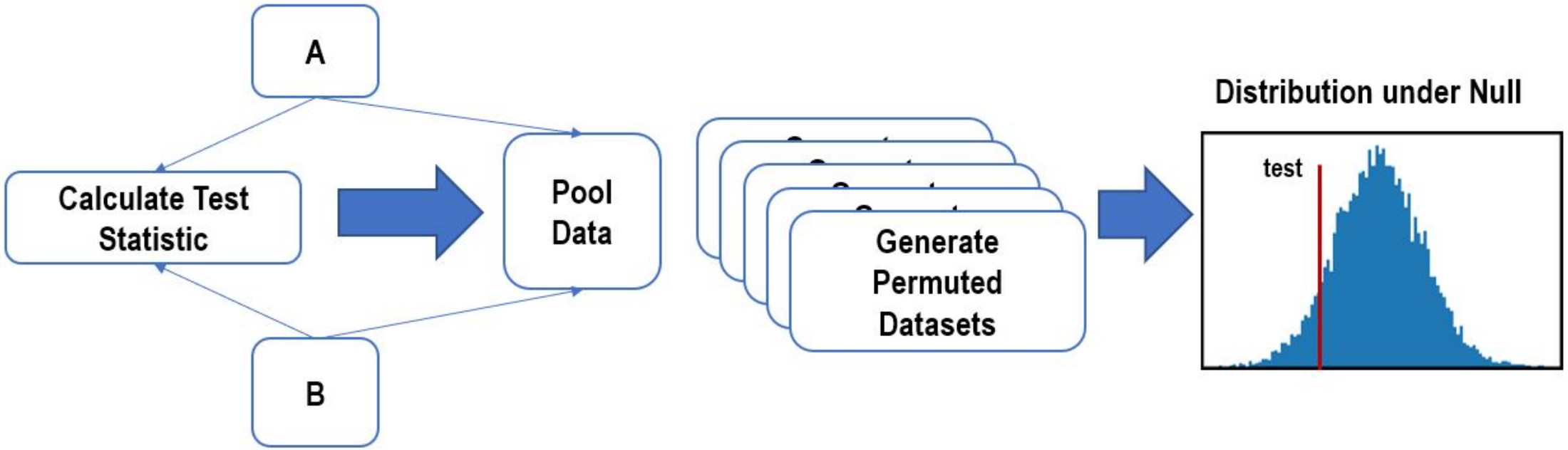
Tushar Shanker
Data Scientist

datacamp

Steps involved



Steps involved



Discussion

Advantages

- Very flexible
- No strict assumptions
- Widely applicable

Drawbacks

- Computationally Expensive
- Custom coding required

Donation website

Let's practice!

STATISTICAL SIMULATION IN PYTHON