RESAMPLING

PIKA SILVIANTI

Bootstrap

- Simulasi bootstrap adalah simulasi berbasis resampling data sampel dengan syarat pengembalian pada datanya dalam menyelesaikan statistik ukuran sampel dan harapan sampel tersebut mewakili data populai sebenarnya,
- biasanya ukuran resampling diambil secara ribuan kali agar dapat mewakili data populasinya.
- Metode ini lebih baik untuk ukuran data sampel yang relatif kecil (dalam bukunya Walpole data kecil yaitu n<30)
- Bootsrap diperkenalkan oleh Efron (1979) dan bootstrap lebih efisien secara statistik. Fleksibilitasnya memungkinkan untuk memperkirakan galat baku, nilai-p, selang kepercayaan (CI) dan banyak statistik lainnya tanpa perlu matematika teoretis yang luas (Beasley dan Rodgers, 2009).

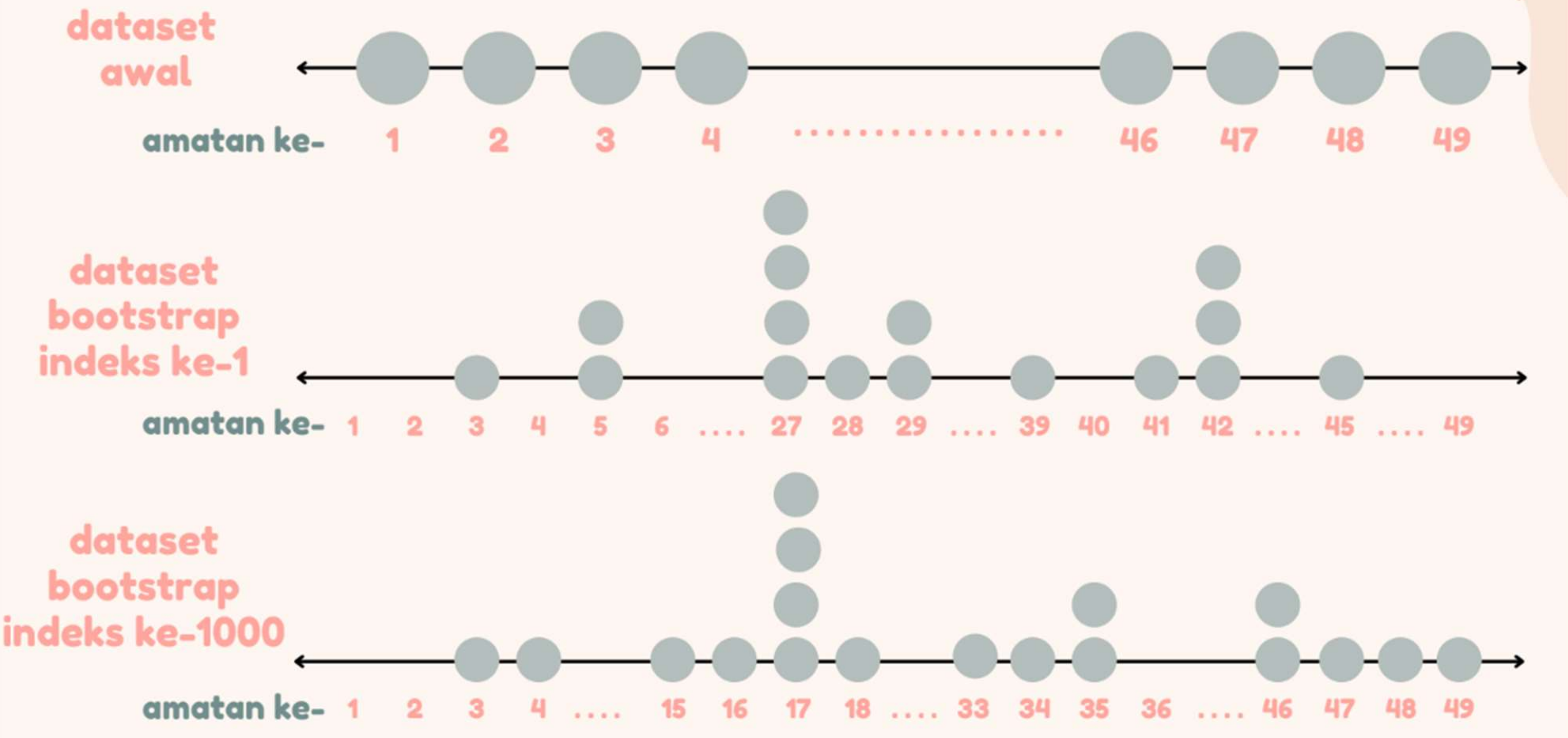
Studi Kasus

Tabel di bawah ini melaporkan jumlah populasi sebanyak 49 amatan dari sepasang kota di AS pada tahun 1920 dan 1930, yang dinotasikan dengan u dan x.

Amatan ke	1	2	3	...	47	48	49
U	138	93	61	...	43	161	36
X	143	104	69	...	50	232	54

Jika (U,X) menotasikan pasangan nilai populasi dari kota yang terpilih secara acak, maka total populasi 1930 merupakan hasil perkalian antara total populasi 1920 dengan rasio dari ekspektasi $\theta = E(X)/E(U)$. Bootstrap resampling akan digunakan untuk menghitung galat baku dari θ.

Ilustrasi


dataset awal
amatan ke- 1 2 3 4 .... 46 47 48 49
dataset bootstrap indeks ke-1
amatan ke- 1 2 3 4 5 6 .... 27 28 29 .... 39 40 41 42 .... 45 49
dataset bootstrap indeks ke-1000
amatan ke- 1 2 3 4 .... 15 16 17 18 .... 33 34 35 36 .... 46 47 48 49

Amatan ke	1	2	3	...	47	48	49
U	138	93	61	...	43	161	36
X	143	104	69	...	50	232	54

dataset awal

thetaBoost= mean(X)/mean(U)

thetaBoost = 1.23902

dataset bootstrap indeks ke-1

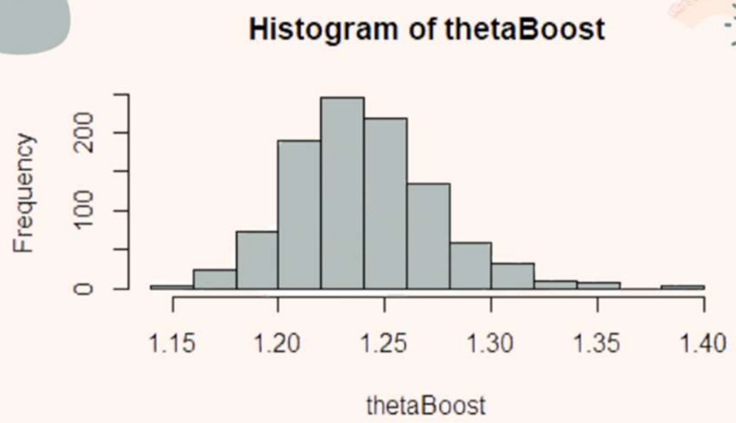
thetaBoost = 1.33271

Amatan ke	3	5	5	...	43	43	45
U	61	48	48	...	40	40	87
X	69	75	75	...	64	64	105

dataset bootstrap indeks ke-1000

thetaBoost = 1.20825

Amatan ke	3	4	6	...	47	48	49
U	61	179	37	...	43	161	36
X	69	260	63	...	50	232	54



$$Var(\hat{\theta}) = 0.0012$$

$$\sqrt{Var(\hat{\theta})} = 0.0347$$

Jackknife

Metode Jackknife menggunakan partisi sistematis dari kumpulan data untuk mengestimasi properti dari estimator dihitung dari sampel penuh. Jackknife dikembangkan oleh Quenouille (1949) dan Tukey (1958) untuk memperkirakan bias dan kesalahan standar (Miller, 1964). Ciri khasnya ialah pengamatan berbeda dikecualikan dalam setiap sampel jackknife (Beasley dan Rodgers, 2009). Quenouille (1949) menyarankan teknik untuk memperkirakan bias dari suatu penduga. Tukey (1958) menciptakan istilah pisau lipat untuk merujuk metode dan berguna dalam mengestimasi varian dari suatu penduga.

Studi Kasus

Tabel di bawah ini melaporkan jumlah populasi sebanyak 49 amatan dari sepasang kota di AS pada tahun 1920 dan 1930, yang dinotasikan dengan u dan x.

Amatan ke	1	2	3	...	47	48	49
U	138	93	61	...	43	161	36
X	143	104	69	...	50	232	54

Jika (U,X) menotasikan pasangan nilai populasi dari kota yang terpilih secara acak, maka total populasi 1930 merupakan hasil perkalian antara total populasi 1920 dengan rasio dari ekspektasi $\boldsymbol{\theta} = E(X)/E(U)$. Jacknife resampling akan digunakan untuk menghitung galat baku dari $\boldsymbol{\theta}$.

Prosedur Jacknife

1. Input Data

2. Bangkitan sampel baru dengan menghilangkan satu amatan pada setiap ulangan. Sehingga banyaknya ulangan yang dapat dilakukan adalah sebanyak n.

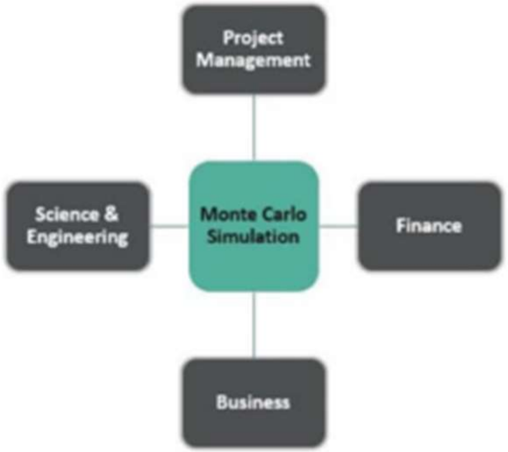
3. Hitung penduga untuk setiap sampel baru

4. Gunakan hasil pada langkah2 untuk menghitung ragam, bisa didekati menggunakan ragam jacknife

Simulasi Monte Carlo

Tipe simulasi probabilistik untuk mencari penyelesaiaan masalah dengan sampling dari proses random.

Monte Carlo Simulation Applica



Simulasi Monte Carlo memanfaatkan informasi mengenai **sebaran data yang diketahui (dihipotesiskan)** dengan pasti.

Simulasi Monte Carlo digunakan untuk melakukan pendugaan peluang (suatu peubah acak) $P(Y>\alpha)$ atau $P(Y<\alpha)$ dengan $Y = g(x)$, $X \sim f(x)$ dan pengujian hipotesis.

Simulasi Monte Carlo

Ilustrasi Simulasi Monte Carlo

Melakukan pengujian hipotesis nilai tengah dengan mengambil 18 contoh acak yang menyebar Poisson (4). Kemudian dilakukan pengulangan dalam pengambilan contoh sebanyak 10000 kali, lalu menghitung proporsi nilai P-value.

Hipotesis

$H0 : \lambda = 4$

$H1 : \lambda \neq 4$

Statistik Uji

```
n = 18
lambda = 4
set.seed(999)
contoh <- rpois(n,lambda)
xbar <- mean(contoh)
stdev <- sd(contoh)
thit <- abs((xbar-4)/(stdev/sqrt(n)))
xbar;stdev;thit
```

```
# [1] 3.777778
# [1] 2.129776
# [1] 0.4426798
```

Titik Kritis

```
k = 10000
set.seed(999)
distribusi <-
matrix(rpois(n*k,lambda),k)
m <- apply(distribusi,1,mean)
s <- apply(distribusi,1,sd)
tdist <- abs((m-4)/(s/sqrt(n)))
y <- ifelse(tdist>thit,1,0)
pvalue <- mean(y)
pvalue
```

```
# [1] 0.6633
```

Kesimpulan

P-value > alpha 5% **(Tak Tolak H0)**
Cukup bukti untuk menyatakan bahwa nilai tengah populasi sama dengan 4 pada taraf nyata 5%