

Modul 5 Komputasi Statistik

Resampling Methods: Jackknife

Pendahuluan

Dalam statistika, dapat digunakan sampel untuk dapat memberikan kesimpulan mengenai kondisi dengan keuntungan lebih menghemat waktu, dan biaya. Namun, karena sampel yang diambil tidaklah mencakup semua populasi, maka terdapat standar deviasi, standar galat, dan selang kepercayaan dari setiap ekspresi statistik. Jika kita mengambil sampel yang berbeda, maka akan menghasilkan statistik deskriptif (mean, median, dll) yang berbeda pula. Karena standar galat dihitung berdasarkan sampel, estimasi dapat menjadi bias terhadap sampel.

Teknik Resampling adalah melakukan pengambilan subsample (boleh data berulang) dari sampel kemudian dihitung setiap statistik deskriptifnya. Pengambilan sampel berulang kali akan mendapatkan distribusi dari nilai statistik deskriptif. Distribusi ini yang dapat memberikan besaran empirik atau akurasi mengenai beberapa tes statistik. Teknik Resampling memiliki beberapa keunggulan, diantaranya:

1. Lebih sedikit asumsi

Contoh: Dimana metode resampling tidak mengharuskan distribusi normal dan ukuran sampel yang besar.

2. Akurasi yang lebih besar. Metode bootstrap pada praktiknya menghasilkan hasil yang lebih akurat.
3. Bersifat umum. Metode resampling sangat mirip untuk berbagai statistic dan tidak memerlukan formula baru untuk setiap statistik.

Beberapa teknik resampling diantaranya: **Teknik Bootstrap** dan **Teknik Jackknife**. Teknik ini dapat digunakan ketika:

1. Menurunkan model statistik.
2. Model statistik yang terlalu rumit
3. Model yang mungkin tidak cukup valid
4. Ukuran presisi yang akurat dalam model statistik hanya dimungkinkan dengan besar n

Teknik Resampling: Jackknife

Metode Jackknife adalah metode resampling dengan menghapus satu data dari sampel. Kemudian dari sampel yang sudah dihapus satu data, dihitung mean dan median. Jika sampel memiliki n data, artinya akan terdapat distribusi mean atau median sebanyak n , hasil dari melakukan penghitungan mean atau median dari sampel yang sudah dihapus satu data sebanyak n kali.

Mengenerate Data

Fungsi **rnorm** di R digunakan untuk menghasilkan data acak yang terdistribusi normal (Gaussian). Ini adalah salah satu fungsi dasar dalam statistik untuk simulasi dan analisis data. Contoh dalam R:

```
rnorm(20, mean = 60, sd = 2.5)
```

Sampling Data

Sampling adalah teknik penting dalam analisis data untuk memperoleh subset dari data atau populasi untuk analisis lebih lanjut. Beberapa teknik sampling yang bisa digunakan diantaranya:

1. Sampling acak (random sampling)
2. Sampling Berstrata
3. Sampling Klaster
4. Sampling Sistematis
5. Sampling bertahap
6. Sampling Tidak Acak
7. Sampling Snowball

Contoh penggunaan dalam R:

```
# Vektor data
```

```
data <- 1:10
```

```
# Sampling acak tanpa pengulangan
```

```

set.seed(123)

sample_data <- sample(data, size = 5)

print(sample_data)

# Sampling acak dengan pengulangan

sample_data_replacement <- sample(data, size = 5, replace = TRUE)

print(sample_data_replacement)

```

set.seed() adalah fungsi dalam R yang digunakan untuk mengatur nilai awal (seed) dari generator angka acak. Ini memastikan bahwa setiap kali Anda menjalankan kode yang melibatkan elemen acak, hasil yang dihasilkan akan konsisten dan dapat direproduksi.

Menghapus Data

Langkah-langkah yang dilakukan:

1. Tinjau data $\theta: \{x_1, x_2, x_3, \dots, x_n\}$
2. Misalkan θ_i adalah data yang diperoleh dengan menghapus x_i .
3. Dalam R, kita dapat memperoleh sebuah data hasil dari penghapusan pada indeks tertentu.

Contoh dalam R:

```

nilai<-c(7,8,9,10)

nilai[-1]    #menghilangkan data ke-1

nilai[-2]    #menghilangkan data ke-2

```

Contoh Kasus 1:

Misalkan terdapat data={10,20,30,40,50}. Gunakan teknik Jackknife untuk menghitung rata-rata (mean) dari data ini setelah menghilangkan satu observasi pada satu waktu, dan kemudian menghitung estimasi bias dan variabilitas dari estimasi tersebut.

```

# Data awal

data <- c(10, 20, 30, 40, 50)

# Fungsi untuk menghitung rata-rata Jackknife

```

```

jackknife_mean <- function(data) {
  n <- length(data)
  mean_original <- mean(data)
  mean_jackknife <- numeric(n)

  for (i in 1:n) {
    data_jack <- data[-i]
    mean_jackknife[i] <- mean(data_jack)
  }

  # Estimasi Jackknife untuk mean
  mean_jackknife_estimate <- mean(mean_jackknife)
  bias <- mean_jackknife_estimate - mean_original
  variance <- sum((mean_jackknife - mean_jackknife_estimate)^2) / n

  list(mean_jackknife = mean_jackknife_estimate, bias = bias, variance = variance)
}

# Hitung estimasi Jackknife untuk mean
result_mean <- jackknife_mean(data)
result_mean

```

Contoh Kasus 2:

Anda sedang mengevaluasi waktu yang dibutuhkan untuk menyelesaikan proyek dari lima tim berbeda dalam sebuah perusahaan. Waktu penyelesaian (dalam hari) yang dicatat adalah:

$$\{12,15,20,18,14\}$$

Gunakan Teknik Jackknife untuk mengetahui seberapa stabil median waktu penyelesaian jika satu tim dihilangkan dari perhitungan!

```

# Data awal
data <- c(12,15,20,18,14)

# Fungsi untuk menghitung median Jackknife
jackknife_median <- function(data) {

```

```
n <- length(data)
median_original <- median(data)
median_jackknife <- numeric(n)

for (i in 1:n) {
  data_jack <- data[-i]
  median_jackknife[i] <- median(data_jack)
}

# Estimasi Jackknife untuk median
median_jackknife_estimate <- median(median_jackknife)
bias <- median_jackknife_estimate - median_original
variance <- sum((median_jackknife - median_jackknife_estimate)^2) / n

list(median_jackknife = median_jackknife_estimate, bias = bias, variance = variance)
}

# Hitung estimasi Jackknife untuk median
result_median <- jackknife_median(data)
result_median
```