

**UJIAN AKHIR SEMESTER
METODE SIMULASI RESAMPLING**



IPB University
— Bogor Indonesia —

ANGGA FATHAN ROFIQY

G1401211006

Dosen Pengajar

Dr. Farit Mochamad Afendi S.Si., M.Si.

DEPARTEMEN STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
IPB UNIVERSITY

2024

A. Skenario

Anda bekerja untuk agensi pemasaran yang memiliki spesialisasi dalam periklanan digital. Klien Anda, suatu perusahaan e-commerce populer, ingin menilai efektivitas kampanye iklan mereka baru-baru ini. Klien Anda telah mengumpulkan data tentang interaksi pengguna (rasio klik-impresi) untuk dua desain iklan yang berbeda: "Iklan A" dan "Iklan B".

B. Tujuan

Simulasi ini bertujuan menduga desain iklan mana yang memiliki performa lebih baik dalam hal interaksi pengguna.

C. Data

Data yang diberikan mencakup dua desain iklan dari perusahaan e-commerce: "Iklan A" dengan 100,000 impresi dan 2,500 klik, serta "Iklan B" dengan 80,000 impresi dan 2,000 klik. Perbandingan ini bertujuan untuk mengevaluasi efektivitas kedua desain iklan dalam mendapatkan interaksi pengguna.

D. Soal

1. Pengujian Hipotesis:
 - Nyatakan hipotesis nol dan hipotesis alternatif untuk membandingkan rasio klik impresi Iklan A dan Iklan B.
2. Distribusi Sampling:
 - Jelaskan mengapa kita menggunakan distribusi sampling dalam pengujian hipotesis.
3. Simulasi Monte Carlo:
 - Jelaskan bagaimana Anda akan menggunakan simulasi Monte Carlo untuk memperkirakan p-value untuk membandingkan rasio klik-impresi Iklan A dan Iklan B.
 - Lakukan simulasi dan intepretasikan hasilnya berdasarkan yang anda jelaskan!
4. Resampling Bootstrap:
 - Jelaskan bagaimana resampling bootstrap dapat membantu kita menduga selang kepercayaan untuk perbedaan rasio klik-impresi antara Iklan A dan Iklan B.
 - Lakukan simulasi dan intepretasikan hasilnya berdasarkan yang anda jelaskan!

E. Hasil dan Pembahasan

1. Pengujian Hipotesis

$$H_0: p_A = p_B$$

$$H_1: p_A \neq p_B$$

Keterangan:

- H_0 : Rasio klik-impresi dari Iklan A **sama** dengan rasio klik-impresi dari Iklan B.
- H_1 : Rasio klik-impresi dari Iklan A **berbeda** dengan rasio klik-impresi dari Iklan B.

2. Distribusi Sampling

Dalam pengujian hipotesis, kita menggunakan distribusi sampling untuk memahami bagaimana statistik uji (seperti perbedaan rasio klik-impresi) terdistribusi di bawah hipotesis nol. Ini membantu kita menentukan seberapa ekstrem hasil yang kita amati dibandingkan dengan apa yang diharapkan jika hipotesis nol benar.

Distribusi sampling memungkinkan kita untuk menghitung p-value, yang menunjukkan probabilitas mendapatkan hasil yang setidaknya se-ekstrem hasil yang diobservasi. Dengan demikian, kita dapat membuat keputusan yang objektif tentang apakah ada cukup bukti untuk menolak hipotesis nol atau tidak, memberikan dasar probabilistik yang kuat untuk seluruh proses pengujian hipotesis.

3. Simulasi Monte Carlo

Simulasi Monte Carlo dapat digunakan untuk memperkirakan p-value dengan menjalankan banyak eksperimen simulasi. Kita akan mensimulasikan distribusi dari perbedaan rasio klik-impresi di bawah hipotesis nol dan melihat seberapa sering hasil yang kita amati (atau lebih ekstrem) terjadi.

Langkah-langkah Simulasi Monte Carlo:

1. Menggabungkan Data: Menggabungkan impresi dan klik dari kedua iklan untuk mendapatkan rasio klik-impresi gabungan.
2. Simulasi Klik: Menggunakan distribusi binomial untuk mensimulasikan jumlah klik untuk Iklan A dan B berdasarkan rasio klik-impresi gabungan.
3. Menghitung Rasio Simulasi: Menghitung rasio klik-impresi dari hasil simulasi.
4. Membandingkan Perbedaan: Membandingkan perbedaan rasio klik-impresi antara simulasi dan data observasi.
5. Menghitung p-value: Menghitung proporsi simulasi di mana perbedaan rasio klik-impresi lebih besar atau sama dengan yang diobservasi.

Simulasi dengan program R:

```
set.seed(86)

# Data
impresi_A <- 100000
klik_A <- 2500
impresi_B <- 80000
klik_B <- 2000
```

```

# Rasio klik-impresi
p_A <- klik_A / impresi_A
p_B <- klik_B / impresi_B

# Mengatur jumlah simulasi
n_sim <- 10000

# Data gabungan
impresi_gab <- impresi_A + impresi_B
klik_gab <- klik_A + klik_B
p_gab <- klik_gab / impresi_gab

# Insialisasi Nilai P
p_value_mc <- 0

# Simulasi Monte Carlo
for (i in 1:n_sim) {
  # Simulasikan jumlah klik untuk Iklan A dan Iklan B dari distribusi
  binomial
  klik_A_sim <- rbinom(1, impresi_A, p_gab)
  klik_B_sim <- rbinom(1, impresi_B, p_gab)

  # Hitung rasio klik-impresi simulasi untuk A dan B
  p_A_sim <- klik_A_sim / impresi_A
  p_B_sim <- klik_B_sim / impresi_B

  # Hitung perbedaan absolut antara rasio klik-impresi simulasi dan
  observasi
  p_diff_sim <- abs(p_A_sim - p_B_sim)
  p_diff_obs <- abs(p_A - p_B)

  # Periksa apakah perbedaan rasio simulasi lebih besar atau sama dengan
  # perbedaan rasio observasi (hipotesis alternatif satu arah:  $p_A > p_B$ )
  if (p_A_sim - p_B_sim >= p_A - p_B) {
    p_value_mc <- p_value_mc + 1
  }
}

p_value_mc <- p_value_mc / n_sim

cat("Simulasi Monte Carlo p-value:", p_value_mc, "\n")

# Output:
# Simulasi Monte Carlo p-value: 0.5002

```

Hasil simulasi Monte Carlo menunjukkan bahwa *p-value* adalah sekitar 0.5002. Ini berarti bahwa jika hipotesis nol ($H_0: p_A = p_B$) benar, kemungkinan untuk

mendapatkan perbedaan rasio klik-impresi yang sama atau lebih besar daripada yang diamati adalah sekitar 50.02%.

Karena p-value lebih besar dari alpha 5%, kita tidak dapat menolak hipotesis nol. Dengan kata lain, tidak ada cukup bukti untuk menyatakan bahwa Iklan A memiliki rasio klik-impresi yang lebih tinggi secara signifikan dibandingkan dengan Iklan B.

4. Resampling Bootstrap

Bootstrap adalah teknik statistik yang memungkinkan kita untuk memperkirakan distribusi sampel dari statistik tertentu (misalnya, rata-rata, varians, proporsi) dengan cara mengulang proses sampling dari data asli dengan pengembalian (resampling with replacement). Ini sangat berguna ketika kita tidak dapat membuat asumsi yang kuat tentang distribusi data kita atau ketika kita memiliki ukuran sampel yang kecil.

Kelebihan Bootstrap

1. Non-parametrik: Tidak memerlukan asumsi tentang distribusi data.
2. Akurasi: Dapat memberikan estimasi yang baik untuk statistik sampel.
3. Fleksibilitas: Dapat digunakan untuk berbagai jenis data dan statistik.
4. Selang Kepercayaan: Dapat digunakan untuk menduga selang kepercayaan untuk statistik sampel.

Resampling bootstrap adalah metode yang sangat berguna untuk menduga selang kepercayaan dari statistik sampel. Selang kepercayaan adalah rentang nilai yang digunakan untuk mengestimasi parameter populasi dengan tingkat kepercayaan tertentu (misalnya, 95%). Bootstrap memungkinkan kita untuk mengestimasi selang kepercayaan dengan cara yang fleksibel dan non-parametrik, tanpa asumsi kuat tentang distribusi data.

Simulasi dengan program R:

```
# Jumlah simulasi bootstrap
n_sim <- 100

# Data gabungan
data_gab <- c(
  rep(1, klik_A), rep(0, impresi_A - klik_A),
  rep(1, klik_B), rep(0, impresi_B - klik_B)
)

# Simulasi bootstrap
p_diff_boots <- numeric(n_sim)

for (i in 1:n_sim) {
  # Ambil sampel bootstrap dengan pengembalian dari data gabungan
  sampel_boots <- sample(data_gab, replace = TRUE, size =
length(data_gab))
```

```

# Hitung jumlah klik bootstrap untuk A dan B
klik_A_boots <- sum(sampel_boots[1:impresi_A])
klik_B_boots <- sum(sampel_boots[(impresi_A + 1):(impresi_A +
impresi_B)])

# Hitung rasio klik-impresi bootstrap untuk A dan B
p_A_boots <- klik_A_boots / impresi_A
p_B_boots <- klik_B_boots / impresi_B

# Hitung perbedaan rasio klik-impresi bootstrap
p_diff_boots[i] <- p_A_boots - p_B_boots
}

# Hitung selang kepercayaan 95% dari distribusi bootstrap
ctr_diff_ci <- quantile(p_diff_boots, c(0.025, 0.975))

cat("Selang Kepercayaan 95% Perbedaan Rasio Klik-Impresi (Bootstrap):\n",
    ctr_diff_ci, "\n")

# Output:
# Selang Kepercayaan 95% Perbedaan Rasio Klik-Impresi (Bootstrap):
# -0.001085125 0.001372125

```

Selang kepercayaan 95% ini menunjukkan bahwa kita 95% yakin bahwa perbedaan sebenarnya dalam rasio klik-impresi antara Iklan A dan Iklan B berada dalam rentang -0.001085125 hingga 0.001372125 .

Karena selang kepercayaan mencakup nol (0), ini menunjukkan bahwa tidak ada bukti yang cukup kuat untuk menolak hipotesis nol bahwa tidak ada perbedaan signifikan antara rasio klik-impresi dari kedua iklan. Artinya, perbedaan yang diamati antara rasio klik-impresi dari Iklan A dan Iklan B bisa jadi disebabkan oleh variasi acak dalam sampel.

Dari perspektif praktis, selang kepercayaan yang mencakup nol mengindikasikan bahwa performa kedua iklan tersebut dalam hal interaksi pengguna (rasio klik-impresi) tidak berbeda secara signifikan.

Perusahaan e-commerce, dapat mempertimbangkan faktor lain (misalnya, biaya pembuatan, target demografis, dll.) dalam memutuskan desain iklan mana yang akan dipilih, karena dari sisi efektivitas klik, keduanya tidak menunjukkan perbedaan yang signifikan.

F. Simpulan

Dari hasil pengujian hipotesis menggunakan simulasi Monte Carlo dan resampling bootstrap, dapat disimpulkan bahwa tidak ada bukti yang cukup untuk mendukung adanya perbedaan yang signifikan antara rasio klik-impresi untuk Iklan A dan Iklan B. Perusahaan e-commerce sebaiknya mempertimbangkan faktor-faktor lain atau melakukan lebih banyak eksperimen untuk lebih memahami efektivitas iklan mereka.

G. Lampiran

[Link File Analsis .rmd](#)