

STAT503 Metode Statistik

Ujian Tengah Semester (UTS): Landasan Inferensi

Yosep Dwi Kristanto

```
library(tidyverse)
library(infer)
library(knitr)
library(kableExtra)
```

①
②
③
④

- ① Metapaket yang memuat paket `{dplyr}` dan `{ggplot2}` yang dalam dokumen ini secara berturut-turut digunakan untuk mengolah data dan membuat diagram-diagram statistik.
- ② Paket yang digunakan untuk melakukan inferensi dengan metode simulasi.
- ③ Dalam dokumen ini, paket `{knitr}` digunakan untuk membuat tabel sederhana.
- ④ Paket yang digunakan untuk mengatur tampilan tabel.

1 Soal: Video di Luar Ruangan

Misalkan kalian ingin mengestimasi proporsi video-video Youtube yang pengambilan gambarnya (sebagian atau keseluruhan) dilakukan di luar ruangan. Kalian memilih sampel acak yang terdiri dari 154 video Youtube dan kalian memperoleh 45 video yang pengambilan gambarnya dilakukan di luar ruangan.

Buatlah sebuah selang kepercayaan 95% untuk mengestimasi persentase video-video Youtube yang pengambilan gambarnya dilakukan di luar ruangan!

2 Jawaban

Permasalahan yang diberikan akan diselesaikan melalui lima tahapan. Tahapan-tahapan tersebut dipaparkan sebagai berikut.

2.1 Mengklarifikasi Pertanyaan

Pertanyaan berikut ini akan dijawab dalam artikel ini.

Berapakah persentase video-video di Youtube yang pengambilan gambarnya dilakukan di luar ruangan?

Tabel 1: Beberapa baris dalam data `sampel_videoYt`.

id	LuarRuangan
1	Ya
2	Ya
3	Ya
4	Ya
5	Ya
150	Tidak
151	Tidak
152	Tidak
153	Tidak
154	Tidak

2.2 Mempersiapkan dan Mengeksplorasi Data

Berdasarkan informasi dalam soal, kita buat data `sampel_videoYt` dengan menggunakan kode berikut.

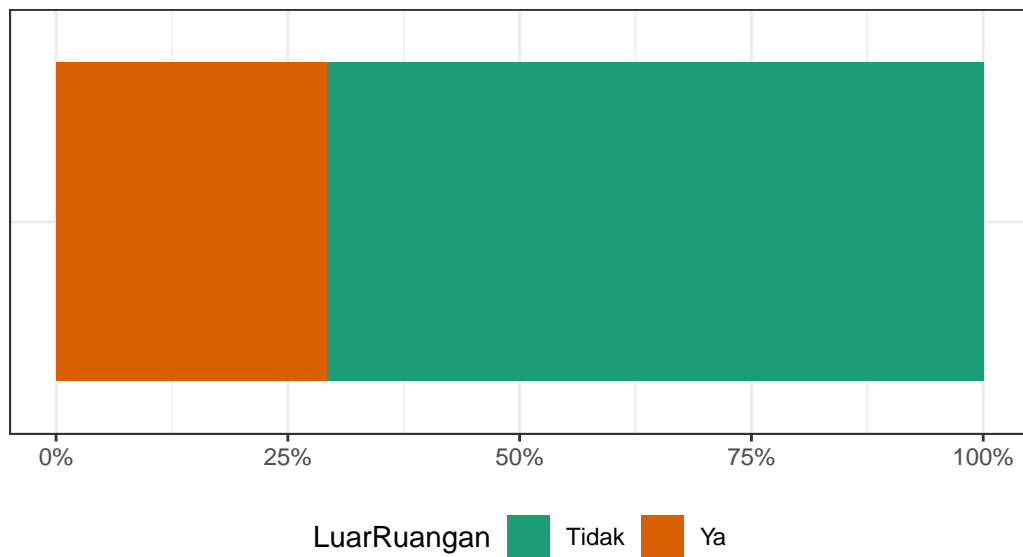
```
sampel_videoYt <- data.frame(
  id = 1:154,
  LuarRuangan = c(rep("Ya", 45),
                   rep("Tidak", 109))
)
```

- ① Membuat objek `sampel_videoYt` yang merupakan dataframe,
- ② yang memiliki variabel pertama bernama `id` yang isinya bilangan bulat terurut dari 1 sampai 154,
- ③ dan variabel kedua bernama `LuarRuangan` yang berisi “Ya” sebanyak 45 dan “Tidak” sebanyak 109.

Tabel 1 menyajikan beberapa baris (lima baris paling atas dan paling bawah) dalam data `sampel_videoYt`.

```
sampel_videoYt %>%
  slice(c(1:5, 150:154)) %>%
  kbl(linesep = "",
      booktabs = TRUE) %>%
  kable_styling(bootstrap_options = c("striped",
                                       "condensed"),
               latex_options = c("striped",
                                 "condensed"),
               full_width = FALSE)
```

Visualisasi yang menunjukkan rasio video-video Youtube yang gambarnya diambil di luar ruangan ditunjukkan pada Gambar 1.



Gambar 1: Visualisasi rasio video-video Youtube yang pengambilan gambarnya di luar ruangan atau tidak.

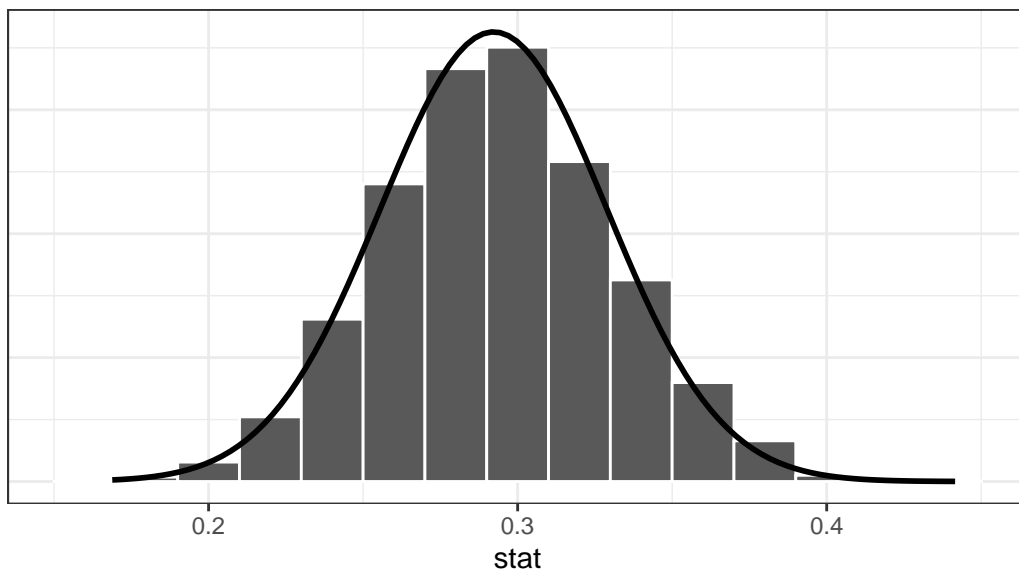
2.3 Memodelkan Variabilitas Statistik

Untuk memodelkan variabilitas statistik-statistiknya dengan menggunakan metode bootstrap, pertama kita buat distribusi bootstrapnya dengan melakukan 10000 simulasi.

Gambar 2 menunjukkan visualisasi distribusi bootstrap dan distribusi sampling dari model matematisnya.

```
mu <- 45/154
SE <- sqrt(mu*(1-mu)/154)

dist_boot_videoYt %>%
  ggplot(aes(x = stat)) +
  geom_histogram(aes(y = after_stat(density)),
    binwidth = 0.02,
    col = "white") +
  stat_function(fun = dnorm,
    args = list(mean = mu, sd = SE),
    linewidth = 1) +
  theme_bw() +
  theme(axis.title.y = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank())
```



Gambar 2: Visualisasi distribusi sampling hasil simulasi bootstrap dan model matematis kurva normal

2.4 Membuat Selang Kepercayaan

Selang kepercayaannya akan dibuat dengan menggunakan dua metode, yaitu metode bootstrap dan model matematis yang didasarkan pada distribusi normal.

2.4.1 Metode Bootstrap

Ujung-ujung selang kepercayaan persentil 95% dapat ditentukan dengan kode berikut.

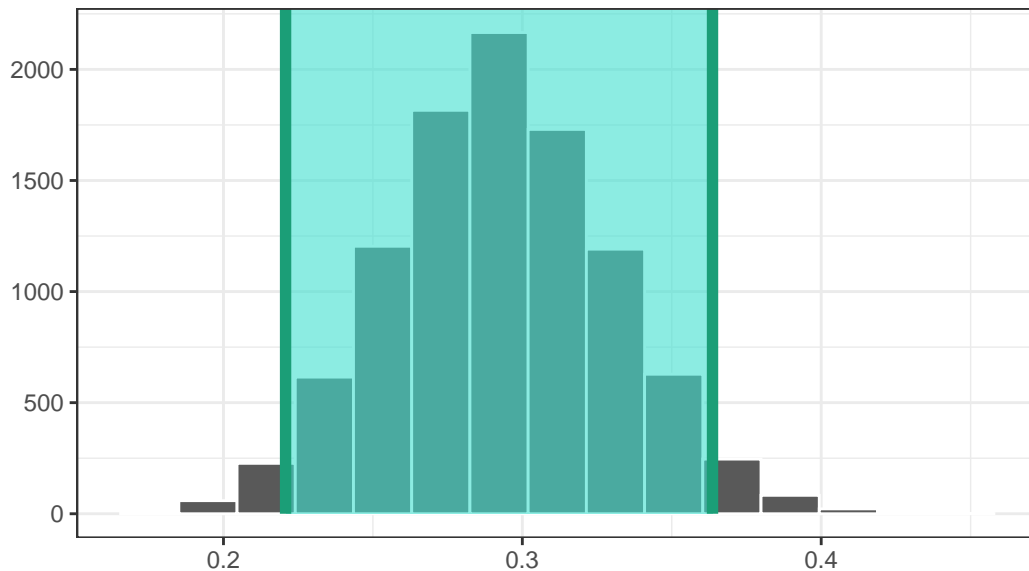
```
sk_persentil <- dist_boot_videoYt %>%
  get_ci(level = .95)
sk_bawah <- as.numeric(sk_persentil[1, 1])
sk_atas <- as.numeric(sk_persentil[1, 2])
print(sk_persentil)
```

```
# A tibble: 1 x 2
  lower_ci upper_ci
  <dbl>     <dbl>
1    0.221    0.364
```

Jadi, selang kepercayaan 95% yang kita peroleh dari metode bootstrap adalah (22.08%, 36.36%). Selang kepercayaan tersebut divisualisasikan pada Gambar 3.

```
visualise(dist_boot_videoYt) +
  shade_confidence_interval(endpoints = sk_persentil,
                           color = "#1b9e77") +
  theme_bw() +
```

```
theme(title = element_blank())
```



Gambar 3: Visualisasi selang kepercayaan persentil 95% yang diperoleh dari metode bootstrap.

2.4.2 Metode Kurva Normal

Ujung-ujung selang kepercayaan 95% berdasarkan model matematis distribusi normal dapat ditentukan dengan kode berikut.

```
z_bintang <- qnorm(0.975, mean = 0,
                  sd = 1, lower.tail = TRUE)
pias_galat <- z_bintang * SE
sk_bawah_matematis <- mu - pias_galat
sk_atas_matematis <- mu + pias_galat
sk_matematis <- tibble(
  batas_bawah = sk_bawah_matematis,
  batas_atas = sk_atas_matematis)
print(sk_matematis)
```

```
# A tibble: 1 x 2
  batas_bawah batas_atas
    <dbl>      <dbl>
1    0.220      0.364
```

Dengan demikian, selang kepercayaan yang kita peroleh adalah (22.04%, 36.4%).

2.5 Membuat Kesimpulan

Jadi, kita 95% yakin bahwa persentase video-video Youtube yang pengambilan gambarnya dilakukan di luar ruangan adalah (22.08%, 36.36%) berdasarkan model bootstrap atau (22.04%, 36.4%) berdasarkan model matematis distribusi normal.