STAT503 Metode Statistik

Ujian Akhir Semester (UAS): Alternatif Penyelesaian

Yosep Dwi Kristanto

2024-01-03

1 Motivasi Belajar

Seorang guru ingin mengetahui apakah secara rata-rata motivasi belajar peserta didiknya meningkat setelah mereka mengikuti pembelajarannya. Untuk itu, guru tersebut menghitung selisih motivasi belajar setiap peserta didik dengan mengurangi motivasi awal dengan motivasi akhirnya. Statistik deskriptif selisih motivasi tersebut, yaitu `d`, disajikan pada Tabel 1.

Tabel 1: Statistik deskriptif selisih motivasi belajar peserta didik

Variabel	n	$\bar{x}$	s
`d`	58	-1,1	5,01

Berdasarkan Tabel 1, guru tersebut menduga bahwa terjadi peningkatan motivasi belajar peserta didiknya.

1. Untuk menguji klaim guru tersebut, tuliskan hipotesis nol dan alternatifnya.
2. Anggap bahwa sampel guru tersebut bebas. Karena sampelnya besar, kita langsung dapat menggunakan uji t. Dengan menggunakan uji ini, berapakah nilai-P yang didapatkan?
3. Tuliskan kesimpulan dan interpretasi uji t tersebut!

Alternatif Jawaban

1. Hipotesis nolnya adalah H_0: Tidak terjadi peningkatan motivasi belajar, yaitu $\mu_d = 0$. Hipotesis alternatifnya adalah terjadi peningkatan motivasi belajar, yaitu $\mu_d < 0$.
2. Untuk menentukan nilai-P, pertama kita hitung statistik ujinya dengan kode berikut.

```r
n_motiv <- 58                                          # (1)
rerata_d_motiv <- -1.1
s_d_motiv <- 5.01
```

```r
SE_d_motiv <- s_d_motiv / sqrt(n_motiv)              # ②
stat_uji_t_motiv <- (rerata_d_motiv - 0) / SE_d_motiv # ③
print(stat_uji_t_motiv)                               # ④
```

① Membuat variabel-variabel `n_motiv`, `rerata_d_motiv`, dan `s_d_motiv` yang secara berturut-turut untuk menyimpan nilai n, $\bar{x}$, dan s pada Tabel 1.
② Menghitung SE dari variabel `d` dengan rumus $SE = \frac{s}{\sqrt{n}}$.
③ Menghitung statistik uji dengan rumus $t = \frac{\bar{x}-\mu}{SE}$. Karena hipotesis nol diasumsikan benar, μ bernilai 0.
④ Cetak hasil perhitungan statistik uji t.

```
[1] -1.672126
```

Kita gunakan distribusi-t dengan $\text{df} = n-1 = 58-1 = 57$. Karena uji hipotesis yang kita lakukan adalah uji ekor kiri, nilai-P uji ini dapat ditentukan sebagai berikut.

```r
df_motiv <- n_motiv - 1                 # ①
P_motiv <- pt(stat_uji_t_motiv,         # ②
   df = df_motiv,
   lower.tail = TRUE)
print(P_motiv)                          # ③
```

① Menghitung derajat bebas dan menyimpan nilainya sebagai `df_motiv`.
② Menentukan nilai-P, yaitu luas daerah di sebelah kiri (`lower.tail = TRUE`) t yang senilai `stat_uji_t_motiv` pada distribusi-t yang memiliki derajat bebas `df_motiv`.
③ Tampilkan hasil perhitungan nilai-P.

```
[1] 0.0499904
```

3. Dengan tingkat signifikansi $\alpha = 0{,}05$, kita tolak hipotesis nol. Dengan demikian, terdapat cukup bukti untuk mendukung dugaan guru tersebut bahwa terjadi peningkatan motivasi belajar peserta didiknya.

2 Tinggi Badan Pemain Sepak Bola

Seorang penikmat sepak bola menduga bahwa pemain sepak bola pada posisi penyerang, gelandang, pemain belakang, dan penjaga gawang memiliki tinggi badan yang reratanya sama. (a) Ujilah dugaan tersebut dengan data `sampel_pemain_bola`. Sampel tersebut merupakan sampel acak sederhana. (b) Jika dugaan tersebut ditolak, carilah rerata mana yang berbeda.

```r
sampel_pemain_bola <- tribble(
  ~posisi,              ~tinggi_badan,
  "Penjaga Gawang",     188,
  "Penjaga Gawang",     190,
  "Penjaga Gawang",     190,
  "Penjaga Gawang",     190,
```

```
    "Penjaga Gawang",   185,
    "Penjaga Gawang",   191,
    "Penjaga Gawang",   188,
    "Penjaga Gawang",   196,
    "Pemain Belakang",    190,
    "Pemain Belakang",    185,
    "Pemain Belakang",    189,
    "Pemain Belakang",    172,
    "Pemain Belakang",    184,
    "Pemain Belakang",    184,
    "Pemain Belakang",    191,
    "Pemain Belakang",    186,
    "Pemain Belakang",    179,
    "Pemain Belakang",    186,
    "Gelandang",          183,
    "Gelandang",        183,
    "Gelandang",        173,
    "Gelandang",        174,
    "Gelandang",        174,
    "Gelandang",        181,
    "Gelandang",        181,
    "Gelandang",        173,
    "Gelandang",        186,
    "Gelandang",        176,
    "Penyerang",        187,
    "Penyerang",        181,
    "Penyerang",        181,
    "Penyerang",        181,
    "Penyerang",        189,
    "Penyerang",        189,
    "Penyerang",        181,
    "Penyerang",        184,
    "Penyerang",        172
  )
```

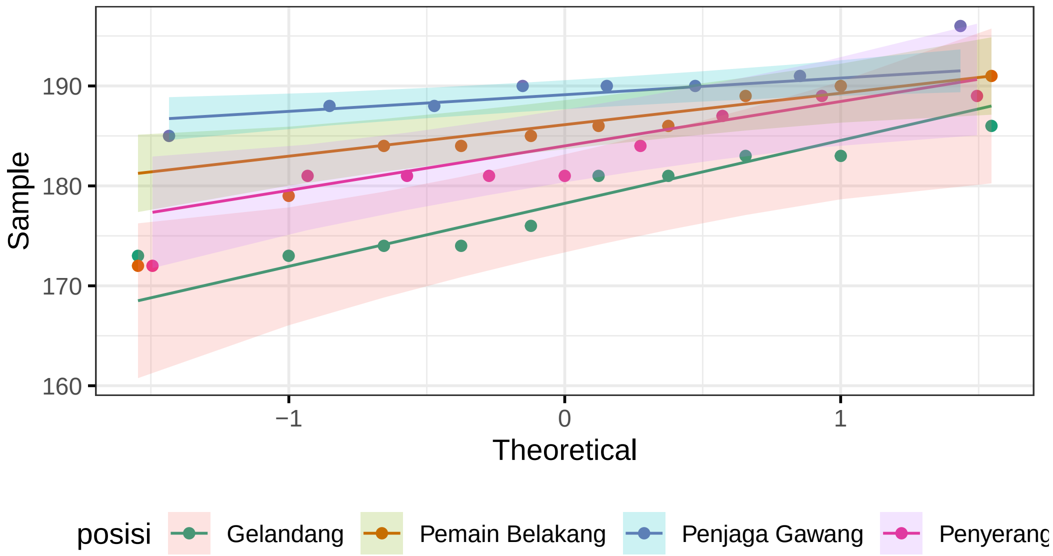
2.1 Alternatif Jawaban

2.1.1 Pemeriksaan Kondisi

Seperti yang dinyatakan di soal, sampel tersebut merupakan sampel acak sederhana. Dengan demikian, kondisi bebas terpenuhi. Kondisi berikutnya yang perlu diuji adalah normalitas data sampel pada tiap-tiap kategorinya. Untuk mengujinya, kita dapat menggunakan diagram peluang normal. Diagram ini ditunjukkan pada Gambar 1.

```
sampel_pemain_bola %>%                                    ①
  ggqqplot(x = "tinggi_badan",                            ②
           color = "posisi",
           ggtheme = theme_bw()) +
  scale_color_brewer(palette = "Dark2") +                 ③
  theme(legend.position = "bottom")                       ④
```

① Memanggil data `sampel_pemain_bola`, dan kemudian,
② berdasarkan data tersebut dibuat diagram peluang normal terhadap variabel `tinggi_badan` untuk setiap `posisi`-nya.
③ Menggunakan palet warna `Dark2`.
④ Meletakkan legenda di bawah.



Gambar 1: Diagram peluang normal `tinggi_badan` pemain sepak bola pada tiap-tiap `posisi`-nya.

Berdasarkan Gambar 1, kita dapat mengasumsikan bahwa `tinggi_badan` tersebut berasal dari populasi yang berdistribusi normal. Asumsi ini tidak ditolak dengan uji Shapiro-Wilk yang hasilnya disajikan pada Tabel 2 karena uji tersebut mendapatkan nilai-P yang cukup besar pada setiap `posisi`-nya.

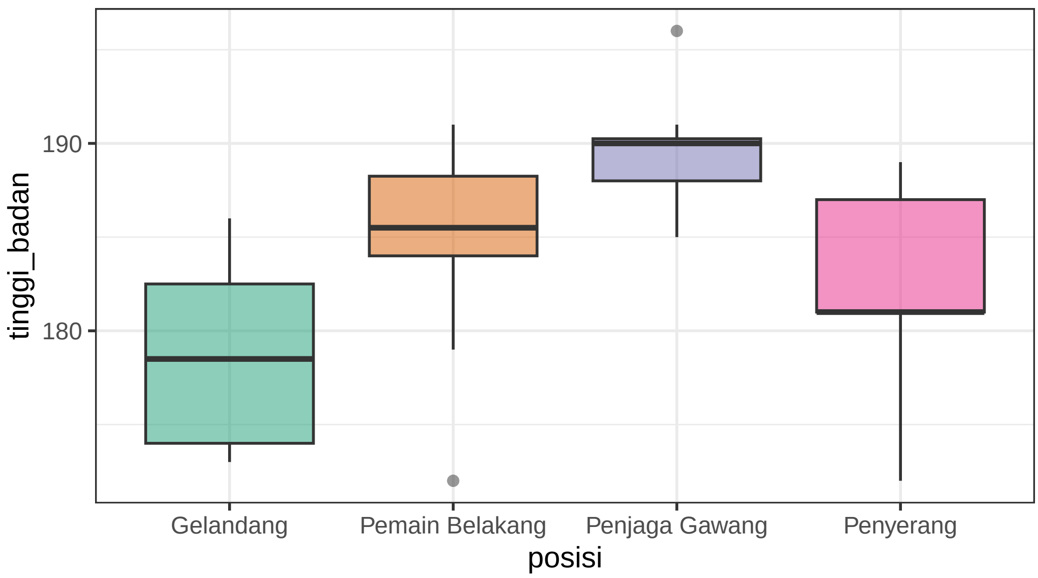
```
sampel_pemain_bola %>%                                    ①
  group_by(posisi) %>%                                    ②
  shapiro_test(tinggi_badan) %>%                          ③
  kbl(linesep = "",                                       ④
      booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped",              ⑤
                                  "condensed"),
                full_width = FALSE)
```

Tabel 2: Hasil normalitas dengan uji Shapiro-Wilk.

posisi	variable	statistic	p
Gelandang	tinggi_badan	0.8720342	0.1055679
Pemain Belakang	tinggi_badan	0.8863168	0.1540703
Penjaga Gawang	tinggi_badan	0.9104863	0.3575007
Penyerang	tinggi_badan	0.8784084	0.1510723

Selanjutnya kita lihat variansi `tinggi_badan` pada tiap-tiap `posisi`-nya. Hal ini dapat kita lakukan dengan membuat boxplot.

```
sampel_pemain_bola %>%
  ggplot(aes(x = posisi,
             y = tinggi_badan,
             fill = posisi)) +
  geom_boxplot(alpha = .5) +
  theme_bw() +
  scale_fill_brewer(palette = "Dark2") +
  theme(legend.position = "none")
```



Gambar 2: Distribusi `tinggi_badan` pemain sepak bola pada tiap-tiap `posisi`-nya.

Berdasarkan boxplot yang ditampilkan pada Gambar 2, kita melihat variansinya agak berbeda. Kita pastikan hal ini dengan menggunakan uji Levene.

```
sampel_pemain_bola %>%
  levene_test(tinggi_badan ~ posisi) %>%
  kbl(linesep = "",
      booktabs = TRUE) %>%
```

Tabel 3: Hasil uji Levene terhadap `tinggi_badan` pemain sepak bola pada tiap-tiap `posisi`-nya.

df1	df2	statistic	p
3	33	0.9050239	0.4491482

```
  kable_styling(latex_options = c("striped",
                                  "condensed"),
                full_width = FALSE)
```

Berdasarkan hasil uji Levene yang disajikan pada Tabel 3, kita gagal menolak asumsi homogenitas variansi. Dengan demikian, kita dapat mengasumsikan bahwa variansi `tinggi_badan` pemain sepak bola pada setiap `posisi`-nya homogen.

Kita telah menunjukkan bahwa data `sampel_pemain_bola` memenuhi kondisi-kondisi ANOVA. Untuk itu, selanjutnya kita lakukan uji tersebut.

2.1.2 Hipotesis Nol dan Alternatifnya

Hipotesis nol dan hipotesis alternatif uji ANOVA yang akan kita lakukan adalah sebagai berikut.

- H_0: Rerata `tinggi_badan` pemain sepak bola pada setiap `posisi`-nya sama, yaitu $\mu_1 = \mu_2 = \mu_3 = \mu_4$ dengan μ_1, μ_2, μ_3, dan μ_4 secara berturut-turut adalah rerata `tinggi_badan` pemain sepak bola pada `posisi` penjaga gawang, pemain belakang, gelandang, dan penyerang.
- H_A: Terdapat paling tidak sepasang rerata yang berbeda.

2.1.3 Tingkat Signifikansi

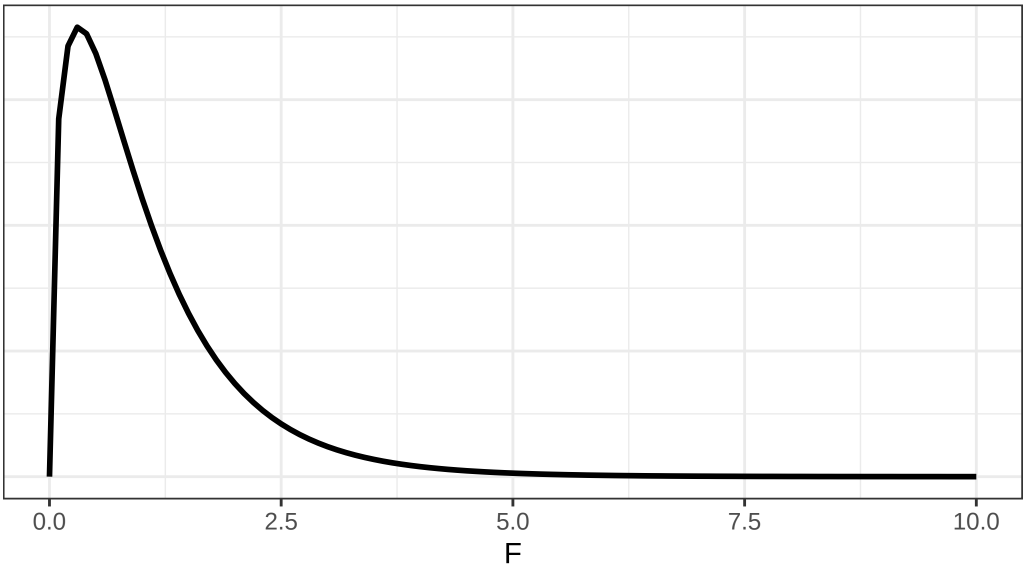
Kita gunakan tingkat signifikansi $\alpha = 0{,}05$.

2.1.4 Variabilitas Statistik

Data sampel yang kita miliki, yaitu `sampel_pemain_bola`, berukuran $n = 37$ dan memiliki $k = 4$ kategori. Dengan demikian, statistik F sampel-sampelnya mengikuti distribusi-F dengan derajat bebas $\mathrm{df}_1 = 4 - 1 = 3$ dan $\mathrm{df}_2 = 37 - 4 = 33$. Distribusi F tersebut ditunjukkan pada Gambar 3.

```
k_pemain_bola <- sampel_pemain_bola %>%
  distinct(posisi) %>%
  nrow()
n_pemain_bola <- sampel_pemain_bola %>%
  nrow()
df1_pemain_bola <- k_pemain_bola - 1
df2_pemain_bola <- n_pemain_bola - k_pemain_bola
```

```r
data.frame(x = c(0, 10)) %>%
  ggplot(aes(x = x)) +
  geom_function(fun = df,
                args = list(df1 = df1_pemain_bola,
                            df2 = df2_pemain_bola),
                linewidth = 1) +
  theme_bw() +
  theme(axis.title.y = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank()) +
  labs(x = "F")
```



Gambar 3: Distribusi-F dengan $\mathrm{df}_1 = 3$ dan $\mathrm{df}_2 = 33$.

2.1.5 Statistik Uji dan Nilai-*P*

Kita selanjutnya hitung statistik uji F dan nilai-P untuk data `sampel_pemain_bola`. Hasil disajikan pada

```r
hasil_anova_pemain_bola <- sampel_pemain_bola %>%
  anova_test(formula = tinggi_badan ~ posisi)
hasil_anova_pemain_bola %>%
  kbl(linesep = "",
      booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped",
                                  "condensed"),
                full_width = FALSE)
```

Berdasarkan Tabel 4, kita dapatkan F yang kurang lebih sebesar 8.142 dan nilai-P yang kurang lebih senilai 3.4×10^{-4}.

Tabel 4: Statistik uji F dan nilai-P dari uji ANOVA

Effect	DFn	DFd	F	p	p<.05	ges
posisi	3	33	8.142	0.00034	*	0.425

Tabel 5: Hasil uji perbandingan berganda Tukey terhadap data `sampel_pemain_bola`.

group1	group2	estimate	conf.low	conf.high	p.adj
Gelandang	Pemain Belakang	6.200000	0.2593495	12.1406505	0.038100
Gelandang	Penjaga Gawang	11.350000	5.0489886	17.6510114	0.000152
Gelandang	Penyerang	4.377778	-1.7256604	10.4812159	0.231000
Pemain Belakang	Penjaga Gawang	5.150000	-1.1510114	11.4510114	0.141000
Pemain Belakang	Penyerang	-1.822222	-7.9256604	4.2812159	0.850000
Penjaga Gawang	Penyerang	-6.972222	-13.4269394	-0.5175051	0.030200

2.1.6 Kesimpulan dan Interpretasi

Karena kita mendapatkan nilai-P yang kurang dari $\alpha = 0{,}05$, kita tolak hipotesis nol. Dengan demikian, terdapat cukup bukti untuk menolak dugaan penikmat sepak bola tersebut.

2.1.7 Uji Hipotesis Pasca-ANOVA

Sekarang kita akan mengidentifikasi rerata `posisi` mana saja yang berbeda. Untuk melakukannya, kita gunakan uji perbandingan berganda Tukey. Hasil uji ini disajikan pada Tabel 5.

```
hasil_tukey_pemain_bola <- sampel_pemain_bola %>%
  tukey_hsd(formula = tinggi_badan ~ posisi)
hasil_tukey_pemain_bola %>%
  select(-term, -null.value, -p.adj.signif) %>%
  kbl(linesep = "",
      booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped",
                                  "condensed"))
```

Berdasarkan Tabel 5, kita mendapatkan tiga pasang rerata `tinggi_badan` yang perbedaannya signifikan. Ketiga pasang tersebut adalah gelandang dan pemain belakang, gelandang dan penjaga gawang, serta penjaga gawang dan penyerang.