

Program Studi Teknik Informatika  
Sekolah Teknik Elektro dan Informatika  
Institut Teknologi Bandung

## Tugas Besar IF2220 Probabilitas dan Statistika

### Penarikan Kesimpulan dan Pengujian Hipotesis

#### Update Spesifikasi Soal

04 april 2023 - Terdapat kesalahan penulisan atritut pada nomor 4e dari "total sulfat dioxide" menjadi "total sulfure dioxide"

#### Tujuan:

- Mahasiswa memahami dan dapat menyelesaikan persoalan distribusi peluang variabel random diskrit dan kontinu, dan
- mahasiswa mampu menyelesaikan persoalan untuk menarik kesimpulan mengenai parameter populasi yang diperoleh dari data hasil eksperimen.
- Mahasiswa mampu menyelesaikan persoalan pengujian hipotesis.

#### Petunjuk pengerjaan tugas:

1. Dikerjakan berkelompok (2 orang) dalam kelas yang sama. Anggota kelompok bebas dan cantumkan nim anggota pada link berikut [Link Anggota](#).
2. Untuk menjawab soal, mahasiswa diharuskan membuat program bahasa **python atau R** yang ditulis pada Jupyter Notebook.
3. Arsip yang dikumpulkan: **File zip** yang berisi file **.ipynb** dan **.pdf** hasil *export* dari notebook dengan nama file **[Kelas]-T1-IF2220-[NIM].zip** dengan NIM adalah NIM terkecil anggota kelompok dan Kelas adalah K01,K02, dan sebagainya. File zip dapat diunggah melalui assignment **edunex**.
4. Tuliskan nomor soal dan keterangan pengerjaan **selengkap mungkin** dengan menggunakan Text di Jupyter Notebook, atau di comment di badan Code.
5. Untuk tes hipotesis, wajib menuliskan ke-6 langkah testing.
6. Batas pengumpulan adalah **18 April 2023**.

#### Enam Langkah Testing:

1. Tentukan Hipotesis nol ( $H_0: \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).
2. Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .
3. Tentukan tingkat signifikan  $\alpha$ .
4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.
5. Hitung nilai uji statistik dari data sample. Hitung *p-value* sesuai dengan uji statistik yang digunakan.
6. Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika *p-value* lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

### Soal Tugas

Diberikan sebuah data anggur.csv yang dapat diakses pada utas berikut: [Dataset Tugas Besar IF2220](#). anggur.csv merupakan data metrik kualitas wine (minuman anggur) yang mengandung 12 kolom sebagai berikut:

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar
5. chlorides
6. free sulfur dioxide
7. total sulfur dioxide
8. density
9. pH
10. sulphates
11. alcohol
12. quality

Kolom 1-11 adalah kolom atribut (non-target), sedangkan kolom 12 adalah kolom target. Anda diminta untuk melakukan analisis statistika sebagai berikut:

1. Menulis deskripsi statistika (*Descriptive Statistics*) dari semua kolom pada data yang bersifat numerik, terdiri dari mean, median, modus, standar deviasi, variansi, range, nilai minimum, maksimum, kuartil, IQR, skewness dan kurtosis. Boleh juga ditambahkan deskripsi lain.
2. Membuat Visualisasi plot distribusi, dalam bentuk histogram dan boxplot untuk setiap kolom numerik. Berikan uraian penjelasan kondisi setiap kolom berdasarkan kedua plot tersebut.

3. Menentukan setiap kolom numerik berdistribusi normal atau tidak. Gunakan normality test yang dikaitkan dengan histogram plot.
4. Melakukan test hipotesis 1 sampel,
  - a. Nilai rata-rata pH di atas 3.29?
  - b. Nilai rata-rata Residual Sugar tidak sama dengan 2.50?
  - c. Nilai rata-rata 150 baris pertama kolom sulphates bukan 0.65?
  - d. Nilai rata-rata total sulfur dioxide di bawah 35?
  - e. Proporsi nilai **total sulfur dioxide** yang lebih dari 40, adalah tidak sama dengan 50% ?
5. Melakukan test hipotesis 2 sampel,
  - a. Data kolom fixed acidity dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata kedua bagian tersebut sama?
  - b. Data kolom chlorides dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata bagian awal lebih besar daripada bagian akhir sebesar 0.001?
  - c. Benarkah rata-rata sampel 25 baris pertama kolom Volatile Acidity sama dengan rata-rata 25 baris pertama kolom Sulphates ?
  - d. Bagian awal kolom residual sugar memiliki variansi yang sama dengan bagian akhirnya?
  - e. Proporsi nilai setengah bagian awal alcohol yang lebih dari 7, adalah lebih besar daripada, proporsi nilai yang sama di setengah bagian akhir alcohol?

#### **Komponen Penilaian:**

- Nomor 1 dan 2 : Kelengkapan jawaban dan ketepatan nilai
- Nomor 3, 4, dan 5 : Kelengkapan jawaban, ketepatan nilai, dan kejelasan metode yang digunakan

#### **Lain-lain:**

1. Keterlambatan pengumpulan akan menyebabkan nilai menjadi nol.
2. Segala bentuk kecurangan akan ditindaklanjuti oleh asisten.
3. Segala pertanyaan hanya dapat ditanyakan melalui pranala [Link QnA](#).

#### **Referensi:**

1. Dokumentasi R - <https://www.rdocumentation.org/>
2. Project Jupyter - <http://jupyter.org/>
3. Pandas - <https://pandas.pydata.org/>

4. Matplotlib - <https://matplotlib.org/>
5. R tutorial - <https://www.tutorialspoint.com/r/index.htm>
6. R In Jupyter Notebook - <https://docs.anaconda.com/navigator/tutorials/r-lang/>
7. R basic plotting - [https://hbctraining.github.io/Intro-to-R/lessons/basic\\_plots\\_in\\_r.html](https://hbctraining.github.io/Intro-to-R/lessons/basic_plots_in_r.html)
8. Walpole, dkk. 2012. Probability and Statistics for Engineers and Scientists: Ninth Edition.