

# LAPORAN UTS

## STATISTIK KOMPUTASI



2023

**Ujian Tengah Semester**

[2141720215]

[Fawaa'el Akbar Firdaus]

[TI-2C]

[ [https://github.com/FaisAkbar/UTS\\_Statistik\\_Komputasi/tree/main](https://github.com/FaisAkbar/UTS_Statistik_Komputasi/tree/main) ]

## Daftar Isi [harus diupdate setiap ada perubahan]

LAPORAN UTS.....	3
------------------	---

SOAL .....	3
------------	---

## LAPORAN UTS

### Soal

1. Berdasarkan “height”,
  - a. Carilah nilai mean, varians, dan simpangan bakunya. (5 poin)

```
a
# mengimport library Pandas
import pandas as pd
# mengimport library Numpy dengan memberikan alias 'np'
import numpy as np
# mengimport library matplotlib.pyplot dengan alias 'plt' yang nantinya akan digunakan untuk melakukan visualisasi data.
import matplotlib.pyplot as plt
# import library Scipy untuk melakukan perhitungan statistik dan matematika.
import scipy

# membaca file CSV dengan nama 'pop1.csv'
# memilih kolom 'height' dari 'data' dan menyimpannya dalam variabel 'table'.
dt = pd.read_csv("/content/drives/MyDrive/Colab Notebooks/pop1.csv")
table = dt["height"]

# menghitung nilai :
# rata-rata dari kolom 'height' yang disimpan dalam variabel 'table' dan menyimpan hasilnya dalam variabel 'mean'.
# variansi dari kolom 'height' yang disimpan dalam variabel 'table' dan menyimpan hasilnya dalam variabel 'variance'.
# standar deviasi dari kolom 'height' yang disimpan dalam variabel 'table' dan menyimpan hasilnya dalam variabel 'std_deviation'.
mean = table.mean()
variance = table.var()
std_deviation = table.std()

# mencetak nilai hasil
print("Mean:", mean)
print("Variance:", variance)
print("Standard Deviation:", std_deviation)
```

Mean: 170.035  
Variance: 126.1588765887659  
Standard Deviation: 11.232046856595902

- b. Carilah Q1, Q2, dan Q3 (kalkulasi interpolasi dengan “midpoint” atau nilai tengah). (5 poin)

```
b
# quartile 1 (Q1) dari data yang terdapat dalam kolom 'height' yang disimpan dalam variabel 'table'
# dengan parameter 0.25 sebagai persentil yang akan dihitung
# method='midpoint' sebagai cara menghitung nilai persentil dan menyimpan hasilnya dalam variabel 'q1'.
# begitu juga dengan q2 dan q3 hanya saja berbeda parameter yang digunakan sebagai persentil
q1 = np.quantile(table, 0.25, method='midpoint')
q2 = np.quantile(table, 0.5, method='midpoint')
q3 = np.quantile(table, 0.75, method='midpoint')

# mencetak hasil Q1, Q2, Q3
print("Q1:", q1)
print("Q2:", q2)
print("Q3:", q3)
```

Q1: 162.0  
Q2: 170.0  
Q3: 178.0

- c. Apakah terdapat outlier? Jika ada, sebutkan jumlahnya! (5 poin)

```

c
# mencari IQR
IQR = q3 - q1
# mencari batas atas dan batas bawah
lower_bound = q1 - (1.5 * IQR)
upper_bound = q3 - (1.5 * IQR)

# Membuat list kosong
lower = []
upper = []
# menentukan outliers
for x in dt["height"]:
    if x < lower_bound:
        lower.append(x)

for x in dt["height"]:
    if x < upper_bound:
        upper.append(x)
a = min(lower)
b = max(upper)

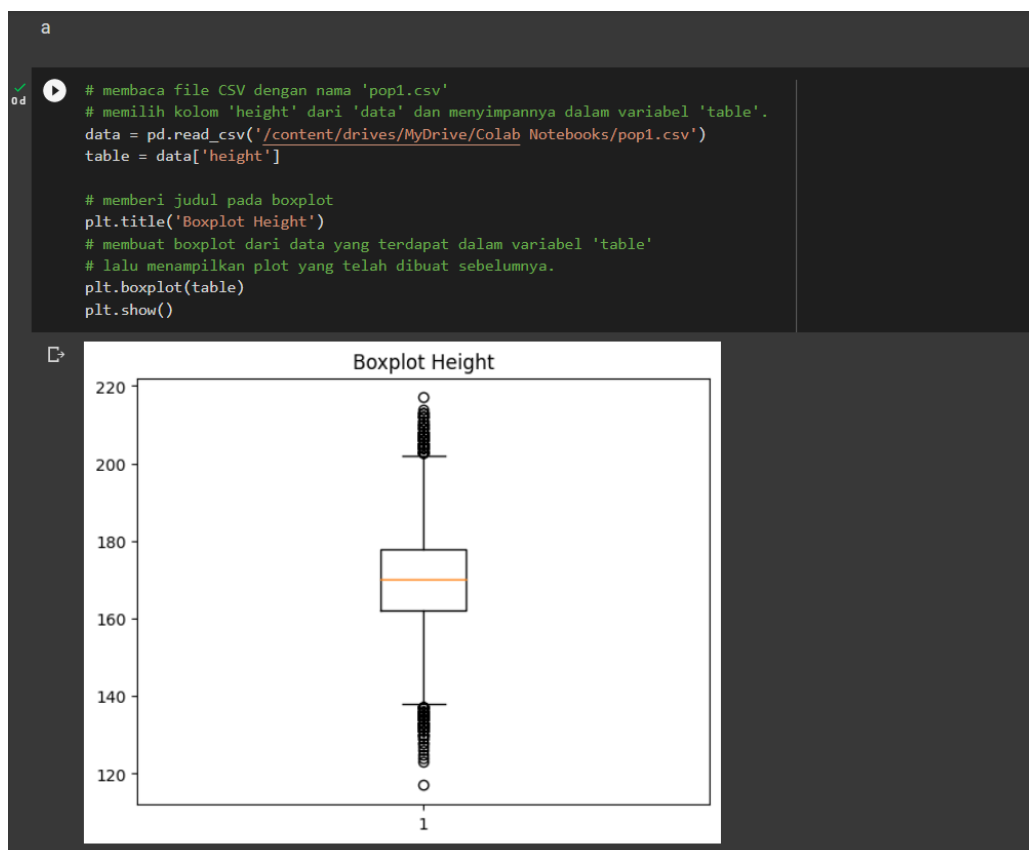
# mencetak hasil
print("IQR: ", IQR)
print("Batas bawah: ", lower_bound)
print("Batas atas: ", upper_bound)
print("Outlier bawah: ", a)
print("Outlier atas: ", b)

IQR: 16.0
Batas bawah: 138.0
Batas atas: 154.0
Outlier bawah: 117
Outlier atas: 153

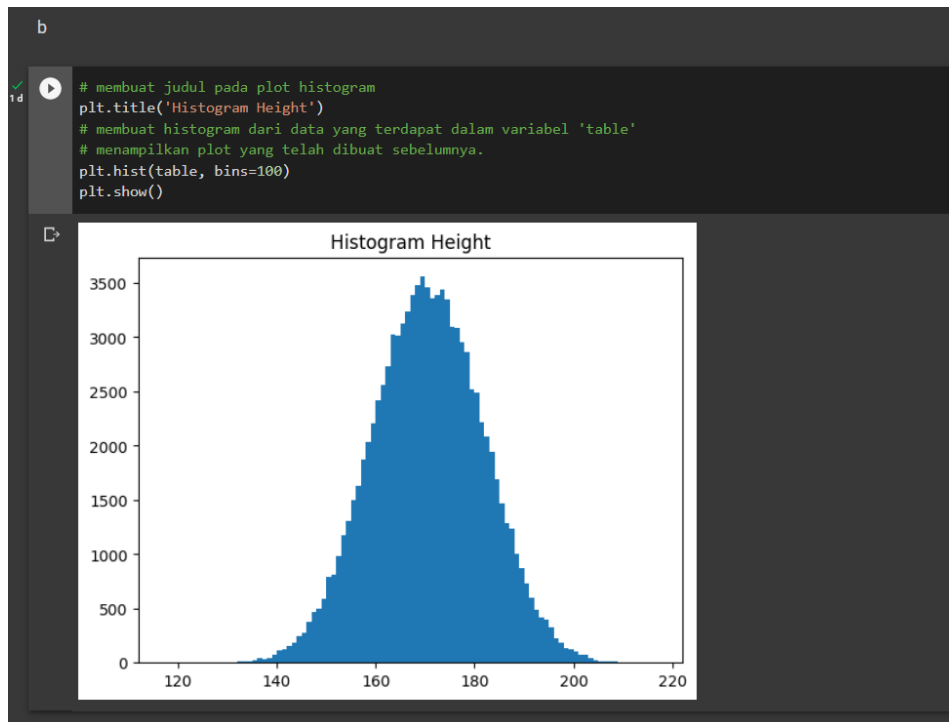
```

2. Berdasarkan data, buatlah,

- a. Boxplot dari “height” (jenis interpolasi bebas, dapat menggunakan linier atau nilai tengah). (10 poin)



- b. Histogram dari “height” dengan jumlah kelompok (bins) adalah 100. (10 poin)



3. Berdasarkan nilai mean dan simpangan baku pada soal nomor 1, tentukan

- a. Berapa nilai z (z-score) untuk orang dengan tinggi hingga 175cm? (10 poin)

a

```
# import fungsi norm dari library SciPy yang berfungsi untuk menghitung distribusi normal
from scipy.stats import norm

# mengisi variable mean
mean = 170.035
# mengisi variable std
std = 11.232

# Menghitung nilai z-score
z_score = (175 - mean) / std
# mencetak nilai z-score dengan format bilangan pecahan 2 angka di belakang koma
print("Z-score untuk tinggi badan > 175 cm : ", "%.2f" % z_score)
```

Z-score untuk tinggi badan > 175 cm : 0.44

+ Kode + Teks

- b. Berapa peluang orang dengan tinggi lebih dari 175cm? (10 poin)

b

```
#menentukan nilai batasan untuk tinggi badan
n = 175

#membuat filter pada kolom height
dt_filter = table[table > n]
#menghitung peluang
peluang = len(dt_filter) / len(table)

# mencetak hasil
print("Total orang dengan tinggi > 175cm : ", len(dt_filter))
print("peluang orang dengan tinggi > 175 : ", peluang)
```

Total orang dengan tinggi > 175cm : 31733  
peluang orang dengan tinggi > 175 : 0.31733

Catatan:

- Gunakan 3 desimal tanpa pembulatan untuk mean dan simpangan baku.
- Gunakan 2 desimal tanpa pembulatan untuk nilai z.
- Gunakan seluruh nilai desimal pada peluang tanpa pembulatan.

4. Tentukan,

a. Peluang dari “MALE” dan “FEMALE”. (5 poin)

```
a
# membaca file CSV 'pop1.csv'
dt = pd.read_csv('/content/drives/MyDrive/Colab Notebooks/pop1.csv')
# menghitung jumlah baris dalam DataFrame 'df' yang memiliki nilai kolom 'sex' bernilai 'MALE'
jumlah_male = dt[dt['sex'] == 'MALE'].count()[0]
# menghitung jumlah baris dalam DataFrame 'df' yang memiliki nilai kolom 'sex' bernilai 'FEMALE'
jumlah_female = dt[dt['sex'] == 'FEMALE'].count()[0]
# menghitung jumlah baris dalam DataFrame 'df'
jumlah_total = dt.count()[0]

# menghitung peluang data yang memiliki nilai kolom 'sex' sama dengan 'MALE' dan 'FEMALE'
peluang_male = jumlah_male / jumlah_total
peluang_female = jumlah_female / jumlah_total

# mencetak hasil
print('Total Data: ', jumlah_total)
print('Total MALE: ', jumlah_male)
print('Peluang MALE:', peluang_male)
print('Total FEMALE: ', jumlah_female)
print('Peluang FEMALE:', peluang_female)
```

Total Data: 100000  
 Total MALE: 51112  
 Peluang MALE: 0.51112  
 Total FEMALE: 48888  
 Peluang FEMALE: 0.48888

b. Jika kita memilih orang secara acak sebanyak 10 kali, berapa peluang terpilihnya 3 laki-laki? (20 poin)

```
b

Peluang terpilihnya 3 laki-laki dari 10 orang yang dipilih secara acak dapat dihitung dengan menggunakan rumus peluang binomial :
• n = 10 (jumlah percobaan)
• k = 3 (jumlah keberhasilan)
• p = 0.5 (peluang keberhasilan dalam satu percobaan)
• q = 0.5 (peluang kegagalan dalam satu percobaan)

Maka peluang terpilihnya 3 laki-laki dari 10 orang yang dipilih secara acak adalah :
menggunakan rumus :
•  $P(X = 3) = C(n,k) * p^k * (1-p)^{n-k}$ 

sehingga, menghasilkan:
 $P(X = 3) = C(10,3) * (0,5)^3 * (0,5)^{10-3}$ 
 $P(X = 3) = C(10,3) * (0,5)^3 * (0,5)^7$ 
 $P(X = 3) = 120 * 0,125 * 0,0078125$ 
 $P(X = 3) = 0,1171875$ 

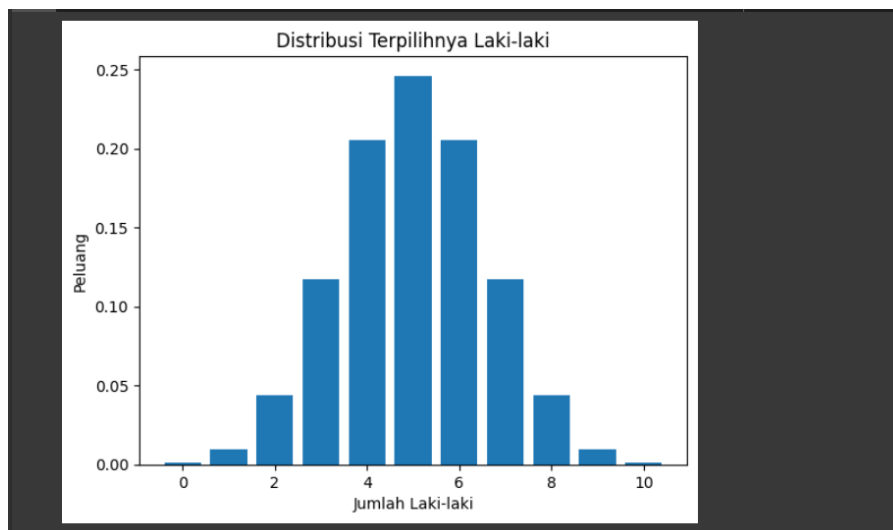
Sehingga peluang terpilihnya 3 laki-laki dari 10 orang yang dipilih secara acak adalah sekitar 0.117 atau 11.7%.
```

- c. Visualisasikan dengan menggunakan diagram batang, distribusi terpilihnya laki-laki mulai dari 0 hingga 10. (20 poin)

```
c
# Mendefinisikan nilai n sebagai 10, yang merupakan jumlah percobaan
n = 10
# Mendefinisikan nilai p sebagai 0,5, yang merupakan probabilitas keberhasilan pada setiap percobaan
p = 0.5

# Membuat list x yang berisi bilangan bulat dari 0 hingga n+1
x = range(n+1)
# Membuat list y yang berisi peluang kejadian dari setiap kemungkinan jumlah laki-laki yang terpilih dalam n percobaan
y = [scipy.special.comb(n, i) * p**i * (1-p)**(n-i) for i in x]

# Membuat bar chart dengan sumbu x berisi list x dan sumbu y berisi list y.
plt.bar(x, y)
# Memberikan label pada sumbu x
plt.xlabel('Jumlah Laki-laki')
# Memberikan label pada sumbu y
plt.ylabel('Peluang')
# Memberikan judul pada grafik
plt.title('Distribusi Terpilihnya Laki-laki')
# Menampilkan grafik
plt.show()
```



Catatan: Gunakan seluruh nilai desimal yang ada