

STATISTIK KOMPUTASI
UJIAN TENGAH SEMESTER



MUHAMMAD IQBAL MAKMUR AL-MUNIRI

2241720099

TEKNIK INFORMATIKA
TEKNOLOGI INFORMASI

Link Github Pengerjaan: https://github.com/GalaAldebara/StatKom_Semester4.git

A. Bagian 1

Pada UTS ini, kita akan mencoba menganalisis data tinggi masyarakat yang berjumlah 100.000 data. Data dapat diunduh pada laman <https://intip.in/datauts>

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

Jawab:

Kode Program

```
import pandas as pd
import numpy as np

data = pd.read_csv("pop1.csv")

print(data.head())

tinggi = data['height']
mean_tinggi = np.mean(tinggi)
varians_tinggi = np.var(tinggi)
simpangan_baku_tinggi = np.sqrt(varians_tinggi)
print("Mean tinggi:", mean_tinggi)
print("Varians tinggi:", varians_tinggi)
print("Simpangan baku tinggi:", simpangan_baku_tinggi)
```

Output Kode Program

```
[Running] python -u "d:\Polinema\Semester 4\StatKom\UTS.py"
   id  sex  height
0  5696379  FEMALE    182
1  3019088   MALE    168
2  2038883   MALE    172
3  1920587  FEMALE    154
4  6006813   MALE    174
Mean tinggi: 170.035
Varians tinggi: 126.15761500000002
Simpangan baku tinggi: 11.231990696221219
```

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

Jawab:

Kode Program

```
Q1 = np.percentile(tinggi, 25)
Q2 = np.percentile(tinggi, 50)
Q3 = np.percentile(tinggi, 75)
print("Q1:", Q1)
print("Q2 (Median):", Q2)
print("Q3:", Q3)
```

Output Kode Program

```
Q1: 162.0
Q2 (Median): 170.0
Q3: 178.0
```

- c. Apakah terdapat outlier? Jika ada, sebutkan jumlahnya!

Jawab:

Ya terdapat outliers, berikut adalah kode programnya

```
IQR = Q3 - Q1
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR
outliers = data[(tinggi < batas_bawah) | (tinggi > batas_atas)]
jumlah_outliers = len(outliers)
print("Jumlah outliers:", jumlah_outliers)
```

Output Kode Program

```
Q1: 162.0
Q2 (Median): 170.0
Q3: 178.0
Jumlah outliers: 304
```

Hasil telah saya simpan kedalam file excel Bernama “hasil_statistik” yang terdapat pada github saya.

P5							
	A	B	C	D	E	F	G
1	Mean Tinggi	Varians Tinggi	Simpangan Baku Tinggi	Q1	Q2 (Median)	Q3	Jumlah Outliers
2	170,035	126,157615	11,2319907	162	170	178	304
3							

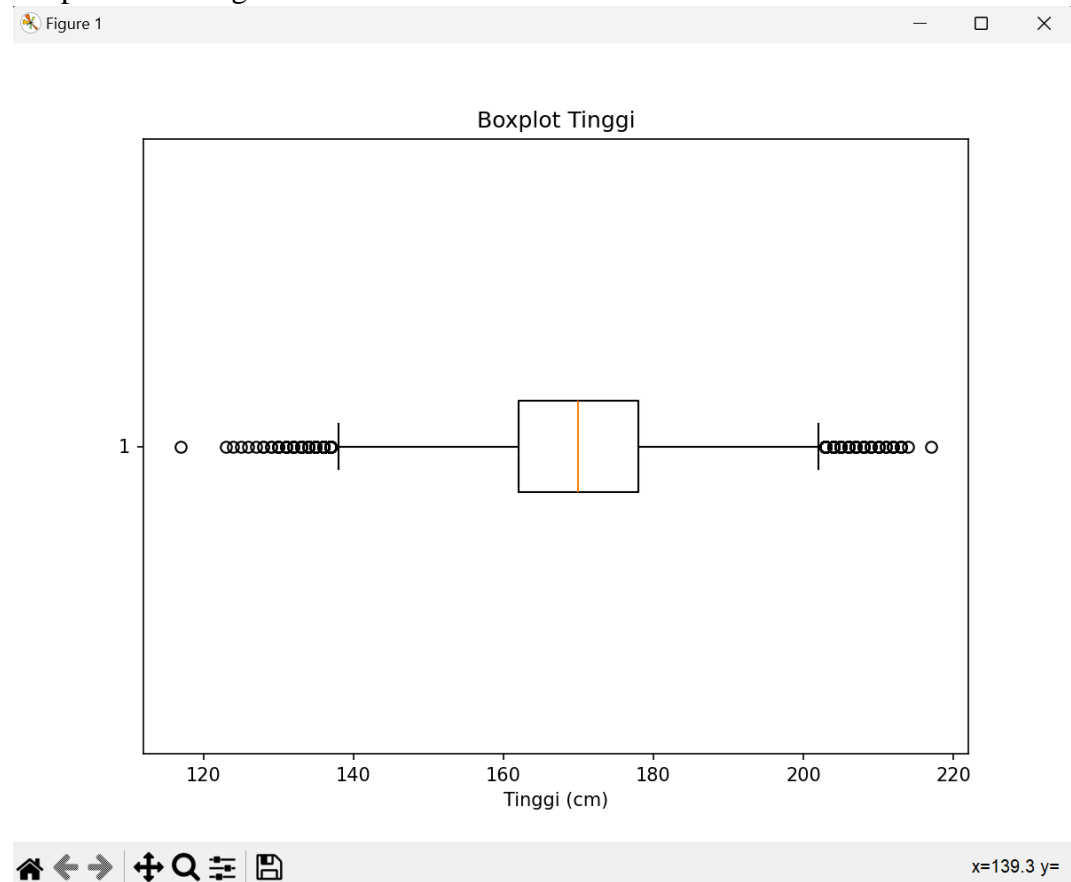
2. Berdasarkan data, buatlah,
- a. Boxplot dari “height” (jenis interpolasi bebas, dapat menggunakan linier atau nilai tengah).

Jawab:

Kode Program

```
plt.figure(figsize=(8, 6))
plt.boxplot(data['height'], vert=False)
plt.title('Boxplot Tinggi')
plt.xlabel('Tinggi (cm)')
plt.show()
```

Output Kode Program



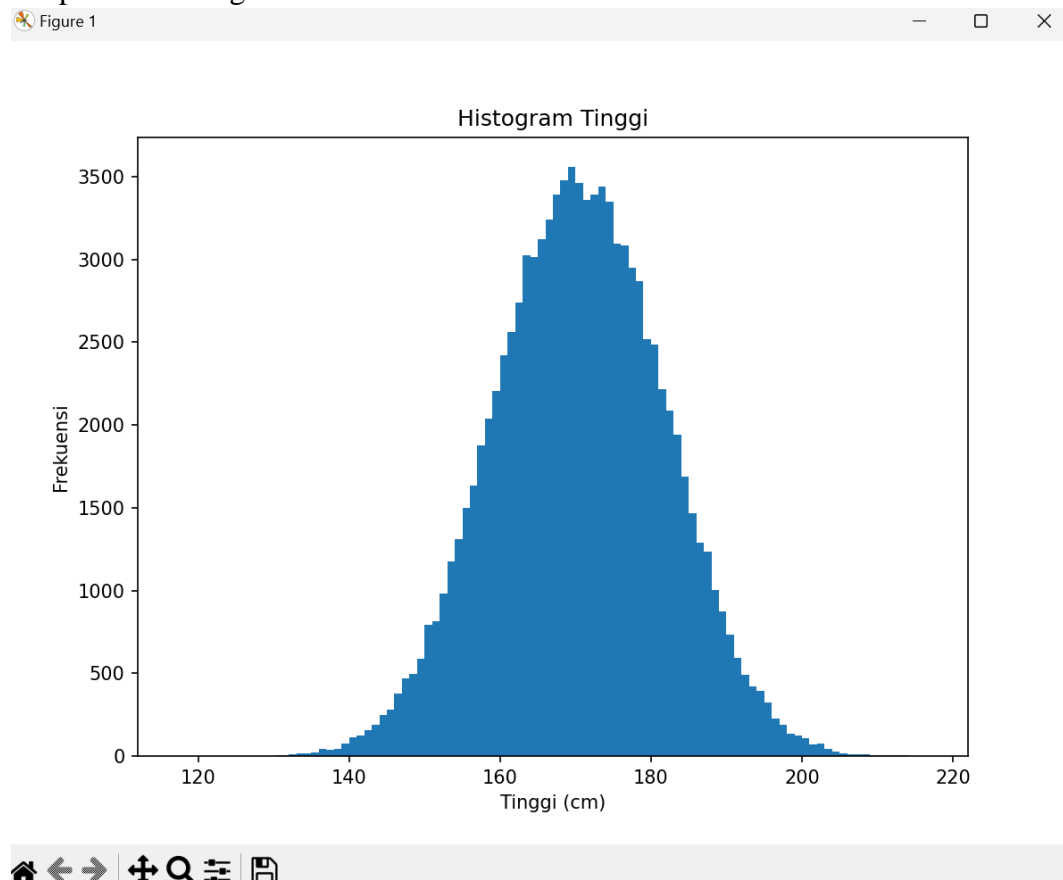
- b. Histogram dari “height” dengan jumlah kelompok (bins) adalah 100.

Jawab:

Kode Program

```
plt.figure(figsize=(8, 6))
plt.hist(data['height'], bins=100)
plt.title('Histogram Tinggi')
plt.xlabel('Tinggi (cm)')
plt.ylabel('Frekuensi')
plt.show()
```

Output Kode Program



3. Berdasarkan nilai mean dan simpangan baku pada soal nomor 1, tentukan
- a. Berapa nilai z (z-score) untuk orang dengan tinggi hingga 175cm?

Jawab:

Kode Program

```
x = 175
mean_tinggi = 170.035
simpangan_baku_tinggi = 11.232
z_score = (x - mean_tinggi) / simpangan_baku_tinggi
print("Nilai z-score untuk tinggi 175cm:", round(z_score, 2))
```

Output Kode Program

```
[Running] python -u "d:\Polinema\Semester 4\StatKom\UTS\No2&3.py"
Nilai z-score untuk tinggi 175cm: 0.44
```

- b. Berapa peluang orang dengan tinggi lebih dari 175cm?

Jawab:

Kode Program

```
peluang_lebih_dari_175 = 1 - norm.cdf(z_score)
print("Peluang orang dengan tinggi lebih dari 175cm:",
round(peluang_lebih_dari_175, 3))
```

Output Kode Program

```
[Running] python -u "d:\Polinema\Semester 4\StatKom\UTS\No2&3.py"
Nilai z-score untuk tinggi 175cm: 0.44
Peluang orang dengan tinggi lebih dari 175cm: 0.329
```

B. Bagian 2

1. Rata-rata konsentrasi kolesterol darah pada populasi dalam jumlah besar pria dewasa usia 50-60 tahun adalah 200mg/dl dengan simpangan baku 20mg/dl. Dianggap pengukuran kolesterol bersebaran normal. Berapakah peluang bahwa 100 pria dari kelompok umum ini akan memiliki kolesterol di bawah 204 ml/dl?

Jawab:

Pertama menghitung z-score untuk nilai ambang kolesterol, lalu menggunakan z-score yang dihitung untuk mencari peluang menggunakan distribusi normal standar.

Kode Program

```
from scipy.stats import norm

x = 204
mean = 200
std_dev = 20
z_score = (x - mean) / std_dev

peluang = norm.cdf(z_score)

print("Peluang bahwa 100 pria memiliki kolesterol di bawah 204 mg/dl:",
peluang)
```

Output Kode Program

```
[Running] python -u "d:\Polinema\Semester 4\StatKom\UTS\Bagian2.py"
Peluang bahwa 100 pria memiliki kolesterol di bawah 204 mg/dl: 0.579259709439103
```

2. Misalkan spesies tanaman tertentu dengan populasi tidak diketahui (sangat banyak) memiliki varians tinggi 16 cm. Diambil sampel sebanyak 25 secara acak kemudian diperoleh rata-rata tingginya adalah 15 cm. Hitunglah rentang nilai rata-rata populasi dari tinggi tanaman tersebut dengan tingkat kepercayaan 95%.

Jawab:

Kode Program

```
import numpy as np
def interval_kepercayaan(rata_rata_sampel, simpangan_baku_populasi,
    ukuran_sampel, tingkat_kepercayaan):
    z = 1.96
    margin_error = z * (simpangan_baku_populasi /
np.sqrt(ukuran_sampel))
    batas_bawah = rata_rata_sampel - margin_error
    batas_atas = rata_rata_sampel + margin_error
    return (batas_bawah, batas_atas)

rata_rata_sampel = 15
simpangan_baku_populasi = np.sqrt(16)
ukuran_sampel = 25
tingkat_kepercayaan = 0.95

interval = interval_kepercayaan(rata_rata_sampel,
    simpangan_baku_populasi, ukuran_sampel, tingkat_kepercayaan)

print("Rentang nilai rata-rata populasi dari tinggi tanaman dengan
    tingkat kepercayaan 95%")
print("Rentang nilai rata-rata populasi:", interval)
```

Output Kode Program

```
[Running] python -u "d:\Polinema\Semester 4\StatKom\UTS\Bagian2No2.py"
Rentang nilai rata-rata populasi dari tinggi tanaman dengan tingkat kepercayaan 95%
Rentang nilai rata-rata populasi: (13.432, 16.568)
```