

Pemrograman Berbasis Fungsi dalam Analisis Varians dan Standar Deviasi terhadap Rata-rata Jam Tidur Harian Mahasiswa Sains Data ITERA 2022

**Elok Fiola¹, Jihan Putri Yani², Rani Puspita Sari³, Diana Syafithri⁴,
Hermawan Manurung⁵, Vira Putri Maharani⁶**

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Jalan Terusan Ryacudu, Way Huwi, Kecamatan Jati Agung, Kabupaten Lampung Selatan, Lampung 35365

Pendahuluan

Analisis deskriptif menjadi salah satu bentuk analisis data penelitian yang digunakan untuk menguji generalisasi hasil penelitian berdasarkan satu sampel. Menurut Hasan, statistik deskriptif menjelaskan atau memberikan informasi tentang data, keadaan, atau fenomena tertentu. Dengan kata lain, statistik deskriptif berfungsi untuk menjelaskan situasi, gejala, atau masalah yang diamati. Tujuan dari pengujian statistik deskriptif untuk memberikan gambaran tentang variabel yang akan diteliti (Hasan, 2004).

Pemrograman Berbasis Fungsi (PBF) dapat digunakan untuk melakukan analisis deskriptif dengan memanfaatkan rumus matematika. Varians dan standar deviasi (simpangan baku) merupakan metrik yang sering digunakan untuk mengukur keragaman dalam data statistik. Penggunaan varians dan standar deviasi dalam konteks rata-rata jam tidur harian memungkinkan identifikasi pola tidur yang tidak biasa atau aneh. Studi yang dilakukan oleh (National Sleep Foundation, 2020) menunjukkan bahwa analisis standar deviasi dalam jam tidur harian dapat membantu mengidentifikasi individu yang berisiko mengalami gangguan tidur atau masalah kesehatan yang terkait dengan tidur. Harapan dari keberadaan program ini adalah dapat digunakan dalam menangani kasus untuk meningkatkan efisiensi dalam proses analisis deskriptif data numerik.

Metode

Metode yang digunakan dalam analisis ini mengadaptasi pendekatan Pemrograman Berbasis Fungsi (PBF) yang menggabungkan pendekatan kualitatif melalui studi pustaka dan pendekatan kuantitatif melalui survei untuk memperoleh dataset. Subjek penelitian ini adalah Mahasiswa Sains Data ITERA 2022. Studi pustaka dilakukan dengan mengumpulkan dan menganalisis artikel-artikel yang relevan, dengan salah satu referensi utama adalah artikel karya (Febriani, 2022) yang digunakan sebagai panduan dalam menentukan dataset dan teori dasar yang digunakan. Dalam analisis data, digunakan pendekatan PBF dengan Bahasa Pemrograman Python untuk merancang struktur data, struktur pemrograman, dan kode analisis data deskriptif, termasuk perhitungan varians dan standar deviasi. Penyusunan kode program melibatkan penggunaan berbagai fungsi, baik *built-in* seperti *sum()*, *len()*, *print()*, *input()*, *FileNotFoundError()*, maupun *user-defined* seperti *mean()*, *variance()*, *standard_deviation()*, *read_csv_data()*. Selain itu, modul Python seperti *pandas* juga digunakan untuk membaca data dari file CSV dan memanipulasi *DataFrame*.

Sebelum melakukan perhitungan varians dan standar deviasi, diperlukan nilai rata-rata yang dapat dihitung menggunakan rumus berikut:

$$\bar{x} = \frac{\sum x_i}{n} \quad (1)$$

Dalam menghitung varians, digunakan rumus berikut:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad (2)$$

Sementara untuk menghitung standar deviasi, digunakan rumus:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (3)$$

Dalam rumus-rumus tersebut, \bar{x} adalah nilai rata-rata dari data, x_i merepresentasikan nilai data ke- i , sedangkan n adalah banyak data.

Varians dihitung untuk mengetahui seberapa jauh titik data tersebar dari nilai rata-ratanya. Standar deviasi merupakan akar kuadrat dari varians yang dapat memberikan ukuran sebaran data yang relatif terhadap nilai rata-ratanya (Febriani, 2022).

Pembahasan

Program ini menggunakan fungsi sebagai objek untuk menghitung nilai varians dan standar deviasi. Metode yang digunakan adalah perhitungan statistik deskriptif.

```
import pandas as pd
```

Gambar 1. Import modul pandas

Untuk menjalankan keseluruhan kode program, kita perlu mengimport modul ke dalam program. Modul yang relevan digunakan pada program ini merupakan modul pandas. Pandas adalah *library* yang berfungsi untuk menganalisis dan memanipulasi data dalam bahasa pemrograman Python. Kita juga perlu memberikan nama singkat 'pd'.

```
import pandas as pd

def mean(data):
    return sum(data) / len(data)

def variance(data):
    mu = mean(data)
    return sum((x - mu) ** 2 for x in data) / (len(data) - 1)

def standard_deviation(data):
    return variance(data) ** 0.5

def clean_data(df):
    # Lakukan pembersihan data di sini sesuai kebutuhan
    df.dropna(inplace=True) # Menghapus baris dengan nilai kosong
    return df
```

Gambar 2. Deklarasi fungsi

Kemudian, terdapat 4 fungsi yang didefinisikan yaitu, fungsi *mean(data)*, *variance(data)*, *standart_deviation(data)*, dan *clean_data(df)*. Fungsi *mean(data)* digunakan untuk menghitung rata-rata dari data yang diberikan, sementara itu fungsi *variance(data)* menghitung variansi dari data tersebut. *Standar_deviation(data)* berfungsi untuk menghitung standar deviasi dari dataset yang digunakan, dan yang terakhir merupakan fungsi *clean_data(df)*, kode tersebut berfungsi untuk membersihkan data dari nilai-nilai yang kosong dengan cara menghapus baris-baris yang memiliki nilai 'NaN' atau kosong, dan DataFrame akan mengembalikan nilai data yang telah dibersihkan.

```
def read_csv_data(filename):
    try:
        df = pd.read_csv(filename)
        return df, filename
    except FileNotFoundError:
        return None, "File tidak ditemukan. Pastikan Anda memasukkan nama file yang benar.", None
```

Gambar 3. Membaca file data

Kemudian terdapat fungsi *read_csv_data(filename)* merupakan sebuah fungsi pada python yang bertujuan untuk membaca data dari sebuah file CSV, jika file tidak ditemukan, maka fungsi tersebut akan mengembalikan pesan kesalahan, dan jika berhasil fungsi akan mengembalikan DataFrame(df) yang berisi data pada file CSV, dengan demikian dapat mengantisipasi kemungkinan bahwa file tidak dapat ditemukan.

```
# Penggunaan:
print("Silakan unggah file Anda dalam bentuk file.csv")
filename = input("Masukkan nama file CSV Anda: ")
```

Gambar 4. Perintah unggah file data

Pengguna diminta untuk upload file dalam bentuk CSV, dan jika program telah membaca file CSV, maka digunakan fungsi input untuk memasukan nama file CSV yang akan disimpan dalam variabel ('filename').

```
df, error_message, file_name = read_csv_data(filename)
if error_message:
    print(error_message)
else:
    print("File CSV '{}' berhasil dibaca.".format(file_name))
    column_name = input("Masukkan nama label kolom yang ingin dianalisis: ")
    if column_name in df.columns:
        # Membersihkan data sebelum melakukan perhitungan
        df_cleaned = clean_data(df)
```

Gambar 5. Membaca file data

Fungsi 'read_csv_data' dengan parameter 'filename' yang mendeklarasikan 3 variabel (df, error_message, dan file_name) dipanggil untuk membaca data dari file CSV yang diberikan. Kemudian terdapat kondisi 'if' yang memeriksa ada atau tidaknya error message, jika ada error message maka error message tersebut akan dicetak menggunakan fungsi 'print()'. Kondisi 'else' akan di eksekusi ketika tidak ada error message. Selanjutnya program akan meminta pengguna untuk memasukkan nama kolom yang akan di analisis dengan menggunakan fungsi 'input()' yang akan disimpan dalam variabel 'column_name'. Jika nama kolom yang diberikan pengguna terdapat dalam data frame maka akan dilakukan data cleaning. Data yang sudah melalui proses cleaning akan disimpan dalam variabel 'df_cleaned'.

```
if len(df) != len(df_cleaned):
    print("Data Anda telah dibersihkan.")
else:
    print("Data Anda lengkap.")

data = df_cleaned[column_name].tolist()
banyak_data = len(data)
print("\nDataset untuk kolom '{}':".format(column_name), data)
```

Gambar 6. Pembersihan data

Setelah tahap pembersihan data, dilakukan verifikasi untuk memeriksa apakah ada baris yang telah dihapus selama proses tersebut. Kemudian, data dari kolom yang akan dianalisis diekstraksi dari DataFrame yang telah dipersiapkan dan diubah menjadi sebuah daftar. Panjang data tersebut dihitung dan disimpan dalam variabel 'banyak_data'. Selanjutnya, informasi mengenai data yang akan dianalisis dicetak untuk memberikan gambaran singkat kepada pengguna.

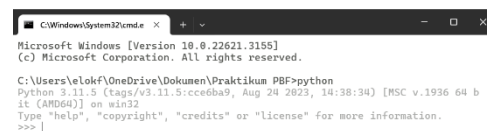
```
result_table = [{"Nama File", file_name},
                 ["Banyak Data", banyak_data],
                 ["Varians", variance(data)],
                 ["Standar Deviasi", standard_deviation(data)]]
print("\nHasil Perhitungan")
for row in result_table:
    print(row[0], ":", row[1])
else:
    print("Label kolom '{}' tidak ditemukan dalam dataset.".format(column_name))
```

Gambar 7. Menampilkan hasil perhitungan

Kode tersebut dijalankan dengan bertujuan untuk membuat sebuah tabel hasil perhitungan statistik berdasarkan data yang telah diproses oleh program sebelumnya. Pertama terdapat 'result_table' yang dibuat menggunakan sebuah list dari list, dimana setiap sublist berisi pasangan label dan nilai yang akan dicetak pada tabel. Label tersebut meliputi nama file, jumlah data, varians, dan standar deviasi dari data yang telah diolah. Kemudian terdapat "Hasil Perhitungan" dicetak untuk menjelaskan user bahwa tabel akan ditampilkan. Jika dataset tidak memiliki label kolom yang sesuai yang dibutuhkan maka pesan kesalahan akan dicetak. Kode diatas membantu menyusun dan menghasilkan hasil perhitungan statistik kepada user.

Kode lengkap dapat diakses pada tautan GitHub berikut: [GitHub Kelompok3 RC Praktikum PBF](#)

Kode tersebut dapat dijalankan pada terminal. Berikut ketika kode dijalankan pada terminal Windows melalui command prompt.



Gambar 8. Input lokasi path folder data

Menjalankan kode program di terminal command prompt diawali dengan menginputkan lokasi path folder keberadaan file program dengan format 'path folder file > python'.

Akan muncul sebuah perintah input, masukan 'import <filename> as fungsi'. Pada kasus ini kami menggunakan file bernama 'kelompok3modul3.py'.

```
>>> import kelompok3modul3 as fungsi
Silakan unggah file Anda dalam bentuk file.csv
Masukkan nama file CSV Anda: Rata-rata Jam Tidur Harian SD2022.csv
File CSV 'Rata-rata Jam Tidur Harian SD2022.csv' berhasil dibaca.
Masukkan nama label kolom yang ingin dianalisis: Rata-rata Jam Tidur
Data Anda lengkap.
```

Gambar 9. Interface pengguna menginput data

Setelah import file berhasil, user akan diminta untuk menginputkan nama file CSV yang berisi dataset. Pada kasus ini kami menggunakan dataset primer 'Rata-rata Jam Tidur Harian SD2022.csv' atau

dapat diakses pada [Dataset Rata rata Jam Tidur Harian S D2022](#). Jika file dataset berhasil terbaca maka program akan memunculkan keterangan “file berhasil dibaca’.dibaca”. Kemudian pengguna akan menginputkan label kolom dari dataset yang ingin dihitung nilai varians dan standar deviasinya. Label kolom harus berisikan data numerik. Program akan mencetak keterangan “Data Anda Lengkap” jika kolom data yang dipilih tidak mengandung nilai null atau NaN.

```
Dataset untuk kolom 'Rata-rata Jam Tidur': [9.0, 6.0, 4.0, 6.0, 5.0, 6.0, 5.0, 5.0, 8.0, 6.0, 5.0, 6.0, 5.5, 6.0, 6.5, 5.0, 5.0, 6.0, 6.0, 7.0, 7.0, 6.0, 6.0, 6.0, 5.5, 6.0, 5.0, 6.0, 5.5]
Hasil Perhitungan
Nama File : Rata-rata Jam Tidur Harian SD2022.csv
Banyak Data : 30
Varians : 0.9206896551724146
Standar Deviasi : 0.9595257449242384
```

Gambar 10. Menampilkan hasil perhitungan

Program akan menampilkan dataset pada kolom yang dipilih dalam bentuk list. Kemudian program akan menampilkan hasil perhitungan varians dan standar defiasi dari data beserta informasi mengenai dataset.

Berikut merupakan 30 data Rata-rata Jam Tidur Harian Mahasiswa Sains Data Angkatan 2022 ITERA:

```
9 6 4 6 5 6 5 5 8 6
5 6 5.5 6 6.5 5 5 6 6 7
7 6 6 6 6 5.5 6 5 6 5.5
```

Berikut deskripsi data Rata-rata Jam Tidur Harian Mahasiswa Sains Data Angkatan 2022 ITERA yang dihasilkan dari PBF:

Deskripsi	Banyak Data	Varians	Standar Deviasi
Jumlah	30	0.921	0.960

Tabel 1. Deskripsi data Rata-rata Jam Tidur Harian Mahasiswa Sains Data Angkatan 2022 ITERA

Simpulan

Program menggunakan pendekatan Pemrograman Berbasis Fungsi (BPF) untuk menghitung varians dan standar deviasi dengan mengintegrasikan pendekatan kualitatif dan kuantitatif. Penulisan kode program melibatkan

pemanfaatan berbagai fungsi dalam Bahasa pemrograman Python. Program ini dapat dijalankan baik pada Jupyter maupun Python. Meskipun menjalankan program pada terminal lebih simpel karena tidak memerlukan pengunggahan file data terlebih dahulu, keduanya menghasilkan hasil varians dan standar deviasi yang serupa.

Data yang digunakan dalam eksperimen ini merupakan data utama dari survei tentang Rata-rata Jam Tidur Harian Mahasiswa Sains Data ITERA 2022. Program berhasil membaca file CSV dengan nama ‘Rata-rata Jam Tidur Harian SD2022.csv’ dan menganalisis kolom ‘Rata-rata Jam Tidur’. Setelah membersihkan data dari nilai-nilai yang kosong, langkah berikutnya adalah melakukan perhitungan varians dan standar deviasi dari dataset. Hasil perhitungan menunjukkan bahwa varians dan standar deviasi dari data Rata-rata Jam Tidur Harian Mahasiswa Sains Data ITERA 2022 sekitar 0.9206896551724146 dan 0.9595257449242384. Total terdapat 30 data lengkap dalam dataset yang diberikan, menandakan bahwa tidak ada nilai kosong dalam analisis tersebut.

Program ini berhasil membaca, membersihkan, dan menganalisis data deskriptif mengenai rata-rata jam tidur mahasiswa Sains Data ITERA 2022. Program ini dapat dimanfaatkan untuk melakukan analisis terhadap data numerik lainnya.

Referensi

- Febriani, S. (2022). Analisis Deskriptif Standar Deviasi. *Jurnal Pendidikan Tambusai*, 910-913.
- Hasan, I. (2004). *Analisa Data Penelitian dengan Statistik*. Jakarta: PT Bumi Aksara.
- National Sleep Foundation. (2020, October 1). Retrieved from News & Articles: <https://www.thensf.org/how-many-hours-of-sleep-do-you-really-need/>