

UNIVERSITAS TERBUKA

Tugas ke-1
Pengantar Sains Data
STDA4101

Nama : MOH EVAN ALSADIK
NIM : 053779188
Kelas : STDA4101.1
UPBJJ-UT : Bandung

Dosen Pengampu Mata Kuliah:
Dimas Agung Prasetyo, S.Kom., M.S. | 01001664

Sains Data
Fakultas Teknologi dan Sains
2024

Soal & Pembahasan

Berdasarkan data excel terlampir, lakukan analisis statistik deskriptif dengan cara menentukan:

- 1. Ukuran pemusatan data: Mean, Median, dan Mode**
- 2. Ukuran penyebaran data: Range, Variance, dan Standard Deviation**
- 3. Bentuk distribusi data: Symmetry, Skewness, dan Kurtosis.**
- 4. Buatlah visualisasi data univariat menggunakan Quartile Plot (box plot) dan berikan ulasan Anda tentang plot tersebut.**

Tugas 1 SDTA4101

November 4, 2024

Nama: Moh Evan Alsadik | Nim: 053779188 | STDA4101

```
[78]: import pandas as pd
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns

file_tugas = "Tugas 1 STDA4101-2024.1.xlsx"
tugas_df = pd.read_excel(file_tugas)
# tugas_df.describe()

umur = tugas_df["AGE"]
print("\n")
```

0.1 Berdasarkan data excel terlampir, lakukan analisis statistik deskriptif dengan cara menentukan:

0.1.1 1. Ukuran pemusatan data: Mean, Median, dan Mode

Ukuran Pemusatan Ukuran pemusatan atau *central tendency* merupakan nilai yang dipandang dapat memberikan gambaran mengenai tata letak suatu data. Utamanya digunakan dalam mencari titik fokus atau nilai tengah yang mewakili sekumpulan data.

```
[39]: mean = umur.mean()
median = umur.median()
modus = umur.mode().tolist() # Jadikan list kalau-kalau modusnya lebih dari
↪ satu.

# Menampilkan hasil.
print("Mean:", mean) # jumlah seluruh nilai dalam dataset dibagi jumlah data.
print("Median:", median) # nilai tengah dari suatu datadet yang telah diurutkan.
print("Modus:", modus) # nilai dengan frekuensi paling banyak dalam dataset.
print("\n")
```

Mean: 45.82

Median: 47.5

Modus: [48, 61]

0.1.2 2. Ukuran penyebaran data: Range, Variance, dan Standard Deviation

Ukuran Penyebaran Ukuran yang menyatakan seberapa besar nilai pada data memiliki perbedaan atau variasi dengan nilai ukuran pusatnya, atau seberapa besar penyimpangan nilai-nilai data dengan nilai pusatnya disebut ukuran pemusatan. Ukuran penyebaran dapat memberikan informasi mengenai apakah data cenderung berkumpul di sekitaran pusat atau tersebar luas.

```
[41]: data_range = umur.max() - umur.min()
      variance = umur.var()
      std_dev = umur.std()

      # Menampilkan hasil
      print(f"Range: {data_range}") # selisih antara nilai maksimum dan minimum dalam
      ↳ dataset.
      print(f"Variance: {variance:.2f}") # nilai yang memperlihatkan seberapa jauh
      ↳ nilai-nilai dalam data menyebar dari nilai pusatnya.
      print(f"Standard Deviation: {std_dev:.2f}") # menunjukkan seberapa jauh
      ↳ rata-rata penyebaran data dari nilai rata-rata (mean) secara umum.
      print("\n")
```

Range: 53

Variance: 187.38

Standard Deviation: 13.69

0.1.3 3. Bentuk distribusi data: Symmetry, Skewness, dan Kurtosis.

```
[96]: skewness = stats.skew(umur)
      kurtosis = stats.kurtosis(umur)

      print("Symmetry:")
      print(f"\tApakah mean ({mean}) = median ({median}) = modus pertama ({modus[0]}):
      ↳ {mean == median == modus[0]}")
      print(f"\tApakah mean ({mean}) = median ({median}) = modus kedua ({modus[1]}):
      ↳ {mean == median == modus[1]}")
      print(f"Skewness: {skewness:.2f}") # Kemiringan distribusi data.
      print(f"Kurtosis: {kurtosis:.2f}") # Keruncingan kurva.
```

Symmetry:

Apakah mean (45.82) = median (47.5) = modus pertama (48): False

Apakah mean (45.82) = median (47.5) = modus kedua (61): False

Skewness: -0.09

Kurtosis: -1.01

Uraian Symmetry: Data bisa dikatakan simetris jika memiliki distribusi yang sama di kedua sisi median. Dengan kata lain, jika membagi data di median (50:50), jumlah data di sebelah kiri dan kanan harus hampir sama. Tapi untuk sedikit kemudahan, gunakan nilai ukuran pemusatan sebelumnya saja. - Jika nilai mean = median = modus: distribusinya dianggap simetris, - Jika nilai mean < median: distribusi miring ke kiri (skewed left), dan - Jika nilai mean > median: distribusi miring ke kanan (skewed right).

Dalam kasus ini, kondisi nomor 1 sudah jelas tidak terpenuhi, berarti distribusinya **asimetri**. Ciri distribusi data asimetri adalah nilai mean, median, dan modusnya tidak sama. Lalu, kondisi nomor 2 terpenuhi yaitu mean < median ($45.82 < 47.5$) maka itu berarti distribusi datanya miring ke kiri (skewed left). Hal ini akan dipertegas di bagian selanjutnya.

Skewness: Kemencengan atau kecondongan adalah tingkat ketidaksimetrisan dari sebuah distribusi. Skewness diartikan sebagai kemiringan distribusi data dan nilainya berguna untuk mengukur sejauh mana data miring ke satu sisi. - Jika nilai skewness = 0: distribusi simetris, - Jika nilai skewness > 0: distribusi miring ke kanan (lebih banyak nilai tinggi), dan - Jika nilai skewness < 0: distribusi miring ke kiri (lebih banyak nilai rendah).

Hasil atau nilai skewness yang saya dapatkan dari data ini adalah -0.09 yang mana nilai ini kurang dari nol (0) maka bentuk distribusi datanya **sedikit miring ke sebelah kiri**.

Kurtosis: Ketinggian kurva atau keruncingan adalah tingkat kepuncakan maupun kedataran dari sebuah distribusi data yang biasanya diambil secara relatif terhadap suatu distribusi normal. - Jika nilai kurtosis = 0: (dalam beberapa definisi) ini berarti kurtosis normal. - Jika nilai kurtosis > 0: puncak distribusi sangat tajam, memiliki ekor yang lebih berat serta menunjukkan bahwa ada lebih banyak nilai ekstrem (outlier). - Jika nilai kurtosis < 0: puncak distribusi lebih datar dan juga menunjukkan bahwa datanya lebih tersebar.

Terakhir, nilai kurtosis yang muncul adalah -1,01 yang menunjukkan bahwa **distribusi datanya lebih datar dan menyebar**. Hal ini karena angka $-1,01 < 0$ sesuai kondisi ketiga pada uraian sebelumnya.

0.1.4 4. Buatlah visualisasi data univariat menggunakan Quartile Plot (box plot) dan berikan ulasan Anda tentang plot tersebut.

Visualisasi data

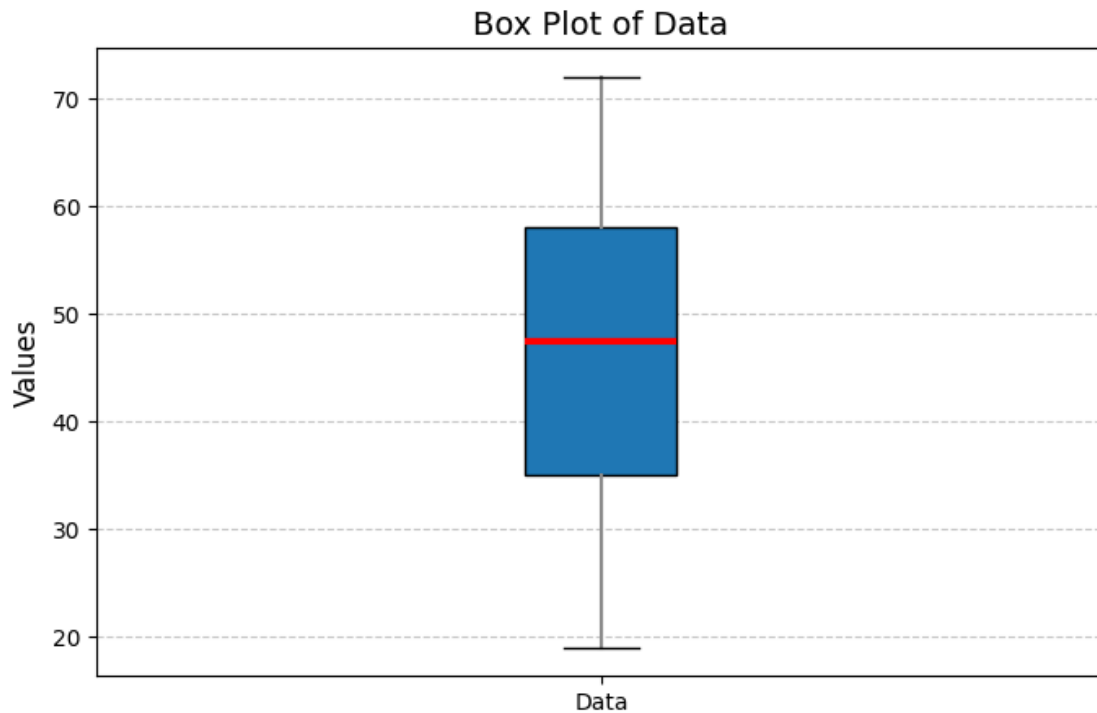
```
[121]: # Visualisasi Box Plot
plt.figure(figsize=(8, 5)) # Ukuran Gambar
box = plt.boxplot(data, patch_artist=True)

# Whiskers dan Garis Median
for whisker in box['whiskers']:
    whisker.set(color='#8B8B8B', linewidth=1.5)
for median in box['medians']:
    median.set(color='red', linewidth=3)

# Judul dan Label
plt.title('Box Plot of Data', fontsize=14)
plt.ylabel('Values', fontsize=12) # Label pada sumbu y
plt.xticks([1], ['Data']) # Label pada sumbu X
```

```
# Grid
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Tampilkan plot
plt.show()
```



Ulasan saya mengenai plot diatas: Grafik diatas menggunakan *box plot* untuk menggambarkan distribusi data umur. Persegi panjang yang berdiri vertikal menunjukkan jangkauan interkuartil (IQR) dalam mengukur sebaran data dengan mengabaikan nilai ekstrim. Nilai jangkauan interkuartil didapatkan dari selisih antara kuartil atas (Q3) dan kuartil bawah (Q1). Lalu, Garis horizontal yang berada di dalam box mewakili nilai tengah atau median (Q2) dari data. Garis Vertikal yang keluar kotak-persegi disebut *Whiskers*. Pada tiap ujung *whiskers* terdapat garis vertikal yang menunjukkan nilai data terkecil (dibawah) dan nilai data terbesar (diatas). Dan titik diluar *Whiskers* merupakan outlier, yaitu nilai pencilan yang letaknya jauh dari titik atau nilai pusat data.

Berdasarkan grafik yang telah ditampilkan, terlihat data terkecilnya adalah 19 (umur 19 tahun) sedangkan data terbesarnya adalah 72 (umur 72 tahun). Grafik ini juga tidak menunjukkan adanya outlier, sehingga data berada dalam rentang normal. Hal ini dibuktikan pada grafik diatas yang tidak memiliki bintang-bintang atau titik-titik diluar *whiskers*.

Selanjutnya grafik ini menunjukkan bahwa 25% data memiliki nilai kurang dari 35 (tahun), 50% dari data nilainya kurang dari 47 (tahun), dan 75% data nilainya kurang dari 58 (tahun), serta

sisanya kurang dari 72 (tahun).

Dengan demikian, grafik box plot ini mencerminkan distribusi data umur yang cukup merata tanpa variasi yang berlebihan.

0.1.5 Sumber Referensi

Sutikno., & Ratnaningsih, D. J. (2025). Metode Statistika I. Modul 01 & 02. Tangerang, Banten. Universitas Terbuka.

<https://binus.ac.id/malang/2022/04/ukuran-pemusatan-data/#:~:text=1.-,Mean,digunakan%20dalam%20banyak%20aplikasi%20statistik.>

<https://www.gramedia.com/literasi/aturan-distribusi-normal/>

<https://accounting.binus.ac.id/2021/08/12/memahami-nilai-skewness-ukuran-kemiringan-dalam-statistik-deskriptif/>

<https://binus.ac.id/malang/2022/04/ukuran-penyebaran-data/#:~:text=Interquartile%20Range%20%E2%80%93,Definisi,terhadap%20outlier%20dibandingkan%20dengan%20rentang.>