

Praktikum 3: Memeriksa Sebaran Data

Tujuan:

1. Memahami pentingnya memeriksa sebaran data dalam analisis statistik.
2. Mengenal berbagai metode visualisasi untuk memeriksa sebaran data.
3. Mampu melakukan uji statistik untuk menentukan apakah data mengikuti distribusi normal.

Peralatan dan Perangkat Lunak yang Digunakan:

- Laptop/Komputer
- Software Python (Anaconda/Google Colab) dengan library: **numpy**, **matplotlib**, **seaborn**, dan **scipy**.
- Data contoh (dataset yang akan digunakan bisa berupa data numerik sederhana, misalnya pengukuran berat badan, tinggi badan, dll.).

Langkah-langkah Praktikum:

1. Pendahuluan:

- Baca kembali konsep dasar tentang distribusi data dan jenis-jenis distribusi yang umum, seperti distribusi normal, miring ke kiri, miring ke kanan, dll.
- Diskusikan pentingnya memeriksa sebaran data sebelum melakukan analisis statistik.

2. Persiapan Data:

- Gunakan dataset yang disediakan atau buat data acak dengan menggunakan Python.
- Contoh membuat data acak dengan Python:

```
import numpy as np

# Membuat data acak mengikuti distribusi normal
data = np.random.normal(loc=50, scale=10, size=100)
```

```
data
```

3. Visualisasi Data:

- **Histogram:** untuk melihat distribusi data.

```
import matplotlib.pyplot as plt

plt.hist(data, bins=20, edgecolor='black')
plt.title('Histogram Data')
plt.xlabel('Nilai')
plt.ylabel('Frekuensi')
plt.show()
```

- **Box Plot:** untuk mengidentifikasi outlier.

```
plt.boxplot(data)
plt.title('Box Plot Data')
plt.ylabel('Nilai')
plt.show()
```

- **Q-Q Plot (Quantile-Quantile Plot):** untuk memeriksa kesesuaian data dengan distribusi normal.

```
import scipy.stats as stats

stats.probplot(data, dist="norm", plot=plt)
plt.title('Q-Q Plot')
plt.show()
```

4. Uji Kenormalan:

- **Shapiro-Wilk Test** untuk memeriksa kenormalan data.

```
stat, p_value = stats.shapiro(data)
print('Shapiro-Wilk Test Statistic:', stat)
print('p-Value:', p_value)
```

- Jika nilai $p > 0.05$, data dianggap mengikuti distribusi normal.
- **Anderson-Darling Test:** Uji alternatif untuk memeriksa kenormalan data.

```
result = stats.anderson(data, dist='norm')
print('Statistic:', result.statistic)
for i, (cv, sig_level) in enumerate(zip(result.critical_values, result.significance_level)):
    print(f'Level signifikansi {sig_level}%: Nilai kritis = {cv}')
```

Note:

Shapiro-Wilk Test Statistic dan **p-Value** dalam uji Shapiro-Wilk untuk memeriksa kenormalan data:

1. Shapiro-Wilk Test Statistic

- **Definisi:** Shapiro-Wilk Test Statistic adalah nilai statistik yang dihasilkan dari uji Shapiro-

Wilk untuk menilai apakah suatu kumpulan data mengikuti distribusi normal.

- **Interpretasi:**

- Nilai statistik ini berada dalam rentang antara 0 dan 1.
- Semakin mendekati 1, semakin kuat bukti bahwa data mengikuti distribusi normal. Sebaliknya, semakin jauh dari 1, semakin besar kemungkinan data tidak mengikuti distribusi normal.
- Shapiro-Wilk Test Statistic menilai kesesuaian antara data yang diuji dengan distribusi normal berdasarkan korelasi antara nilai data yang diurutkan dan nilai yang diharapkan dari distribusi normal.

2. p-Value

- **Definisi:** p-Value adalah probabilitas yang menunjukkan kemungkinan mendapatkan hasil uji yang sama ekstremnya atau lebih ekstrem daripada hasil yang diperoleh, jika hipotesis nol benar. Dalam konteks uji Shapiro-Wilk, hipotesis nol menyatakan bahwa data berasal dari distribusi normal.
- **Interpretasi:**
 - **Nilai $p > 0.05$:** Tidak ada cukup bukti untuk menolak hipotesis nol. Ini berarti data dapat dianggap mengikuti distribusi normal. Tingkat signifikansi 0.05 sering digunakan sebagai batas untuk memutuskan apakah distribusi data dianggap normal.
 - **Nilai $p \leq 0.05$:** Ada cukup bukti untuk menolak hipotesis nol. Ini menunjukkan bahwa data tidak mengikuti distribusi normal.
- **Signifikansi Nilai p:** Semakin kecil nilai p, semakin kuat bukti bahwa data tidak mengikuti distribusi normal. Sebaliknya, nilai p yang lebih besar menunjukkan bahwa data memiliki kesesuaian yang lebih baik dengan distribusi normal.

Ringkasan:

- Uji Shapiro-Wilk digunakan untuk memeriksa apakah data mengikuti distribusi normal.
- **Shapiro-Wilk Test Statistic** memberikan ukuran tentang sejauh mana data mendekati distribusi normal.
- **p-Value** membantu memutuskan apakah hasil uji tersebut cukup signifikan untuk

menyimpulkan bahwa data tidak normal.

5. Diskusi dan Analisis:

- Interpretasikan hasil visualisasi dan uji statistik.
- Diskusikan apakah data memenuhi asumsi kenormalan atau tidak.
- Jika tidak normal, diskusikan metode alternatif untuk analisis lebih lanjut.

Tugas Praktikum:

1. Lakukan analisis sebaran data pada dataset lain (misalnya, data dari tugas sebelumnya).
2. Bandingkan hasil visualisasi dan uji statistik untuk dua jenis dataset yang berbeda.
3. Tuliskan laporan mengenai hasil pemeriksaan sebaran data beserta kesimpulannya.

Kesimpulan:

- Praktikum ini membantu memahami pentingnya memeriksa sebaran data dan memberikan gambaran tentang metode yang bisa digunakan.
- Mahasiswa diharapkan mampu mengevaluasi hasil analisis dan membuat keputusan yang tepat berdasarkan hasil pemeriksaan sebaran data.