

**LAPORAN PRAKTIKUM**  
**MATA KULIAH DATA SCIENCE**  
**PERTEMUAN 6**  
**STATISTIKA DESKRIPTIF DAN EKSPLORASI DATA**  
**(Ukuran Pemusatan, Penyebaran, dan Korelasi)**

**Nama :** Peno  
**NIM :** 221220095  
**Kelas :** 29 (DS B)  
**Tanggal Praktikum :** 31 Oktober 2025  
**Tanggal Pengumpulan :** 31 Oktober 2025



**PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNIK DAN ILMU  
KOMPUTER UNIVERSITAS MUHAMMADIYAH PONTIANAK  
2025**

## 1. Pendahuluan

### 1.1 Tujuan Praktikum

Setelah melakukan praktikum ini, mahasiswa diharapkan dapat:

1. Memahami konsep ukuran pemusatan data (*Mean, Median, Mode*) dan kapan penggunaannya yang tepat.
2. Menganalisis ukuran penyebaran data (*Variance, Standard Deviation*) untuk melihat distribusi data.
3. Mengidentifikasi hubungan antar variabel menggunakan teknik korelasi (*Correlation & Covariance*).
4. Mengenali fenomena anomali statistik seperti *Simpson's Paradox*.

### 1.2 Teori Singkat

- **Mean vs Median:** Mean adalah rata-rata aritmatika, sedangkan Median adalah nilai tengah setelah data diurutkan. Median lebih tangguh (*robust*) terhadap *outlier* (penciran) dibandingkan Mean.
- **Variance & Standar Deviasi:** Mengukur seberapa jauh data tersebar dari rata-ratanya. Semakin besar nilainya, semakin bervariasi datanya.
- **Korelasi:** Mengukur kekuatan dan arah hubungan linier antara dua variabel (range -1 hingga +1).
- **Simpson's Paradox:** Fenomena di mana tren yang muncul pada beberapa kelompok data menghilang atau berbalik arah ketika kelompok-kelompok tersebut digabungkan.

### 1.3 Tools yang Digunakan

- Google Colab / Jupyter Notebook
- Python 3.x
- Library: pandas, matplotlib, seaborn, numpy

## 2. Pelaksanaan Praktikum

### 2.1 Bagian 1: Ukuran Pemusatan (Mean vs Median)

Pada bagian ini, dilakukan analisis terhadap data gaji (*Salary*) karyawan untuk melihat dampak *outlier* terhadap Mean dan Median.

#### 2.1.1 Hasil Visualisasi

(Tempelkan gambar median\_example.png di sini)

Gambar 1. Perbandingan Mean dan Median pada Data Gaji

#### 2.1.2 Analisis

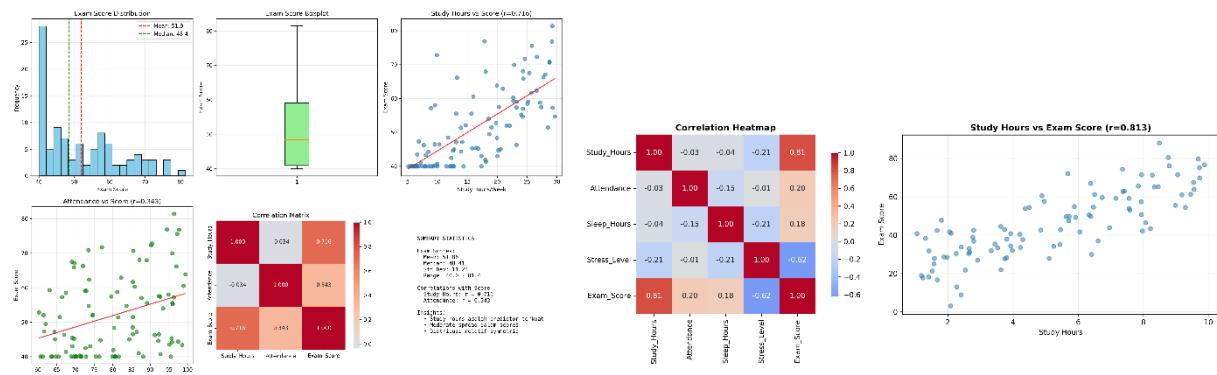
Berdasarkan grafik di atas:

- Terdapat 5 data gaji karyawan. Sebagian besar berada di kisaran 30-45 juta/tahun, namun ada satu karyawan (Employee 4) dengan gaji sangat tinggi yaitu 200 juta/tahun (*outlier*).
- **Mean (Rata-rata):** Tertarik naik hingga **70 juta/tahun** karena pengaruh nilai 200 juta tersebut. Ini tidak merepresentasikan mayoritas karyawan.
- **Median (Nilai Tengah):** Berada di angka **40 juta/tahun**. Nilai ini lebih mewakili kondisi gaji mayoritas karyawan karena tidak terpengaruh oleh nilai ekstrem.
- **Boxplot:** Menunjukkan adanya *outlier* (lingkaran kecil di atas) yang mengonfirmasi ketimpangan distribusi data.

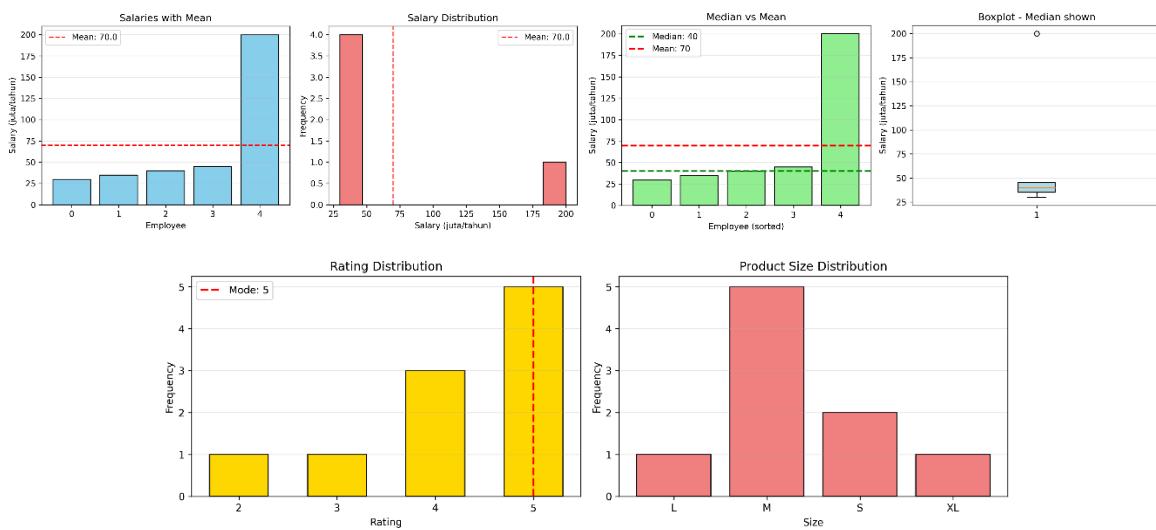
## 2.2 Bagian 2: Ukuran Penyebaran & Korelasi

Bagian ini menganalisis hubungan antara Jam Belajar (*Study Hours*) dengan Nilai Ujian (*Exam Score*).

### 2.2.1 Hasil Visualisasi



Gambar 2. Heatmap Korelasi dan Scatter Plot Study Hours vs Exam Score



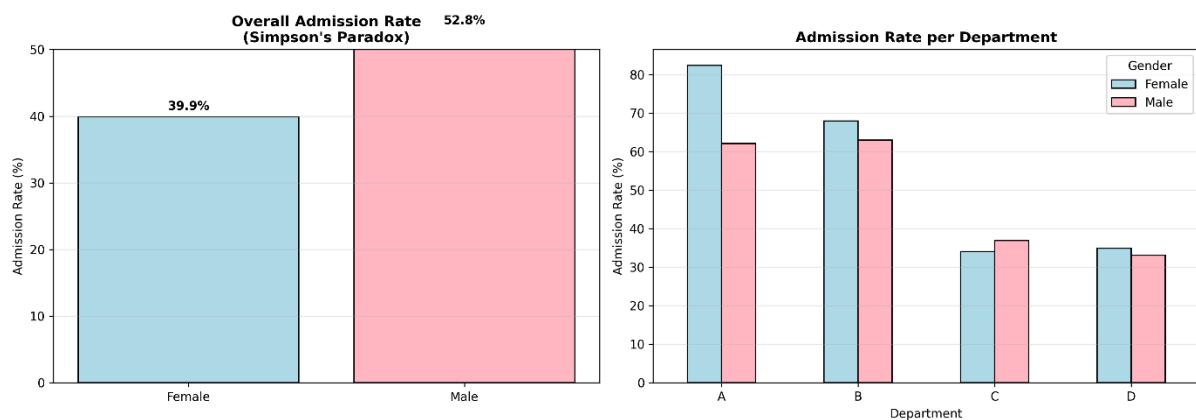
### 2.2.2 Analisis

- **Korelasi Positif:** Terlihat korelasi kuat ( $r = 0.81$  atau mendekati 1) antara Study\_Hours dan Exam\_Score. Artinya, semakin lama durasi belajar, nilai ujian cenderung semakin tinggi.
- **Korelasi Negatif:** Terdapat korelasi negatif antara Stress\_Level dan Exam\_Score ( $r = -0.62$ ). Artinya, tingkat stres yang tinggi justru berkaitan dengan penurunan nilai ujian.
- **Distribusi:** Histogram menunjukkan sebaran nilai ujian mahasiswa, dengan garis putus-putus merah menandakan Mean dan hijau menandakan Median.

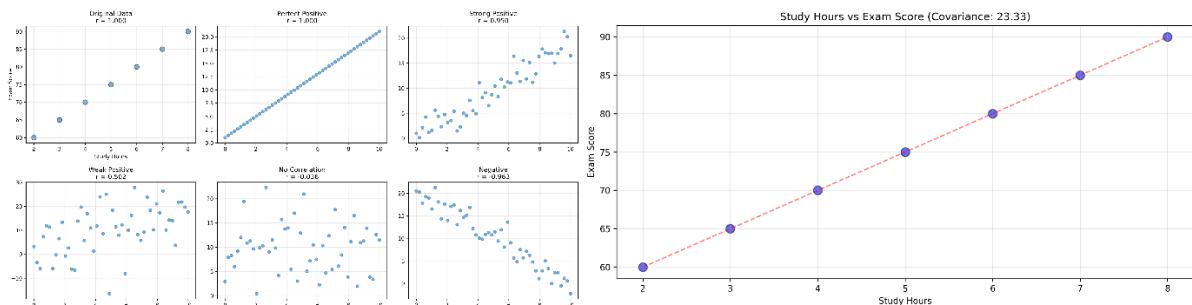
### 2.3 Bagian 3: Fenomena Simpson's Paradox

Menganalisis data tingkat penerimaan (*Admission Rate*) berdasarkan Gender.

#### 2.3.1 Hasil Visualisasi



Gambar 3. Fenomena Simpson's Paradox pada Data Penerimaan



#### 2.3.2 Analisis

- **Secara Agregat (Overall):** Tingkat penerimaan Laki-laki (Male) terlihat lebih tinggi (52.8%) dibandingkan Perempuan (Female) yang hanya 39.9%. Sekilas ini terlihat bias.
- **Per Departemen:** Namun, ketika data dipecah per Departemen (A, B, C, D), terlihat bahwa Perempuan justru memiliki tingkat penerimaan yang **lebih tinggi atau setara** di hampir setiap departemen dibandingkan Laki-laki.

- **Kesimpulan:** Ini adalah *Simpson's Paradox*. Perbedaan rata-rata gabungan disebabkan oleh proporsi pelamar perempuan yang lebih banyak mendaftar ke departemen yang selektif (sulit ditembus), bukan karena diskriminasi di setiap departemen.

### 3. Hasil dan Pembahasan

#### 3.1 Ringkasan Temuan

Konsep	Temuan Utama pada Data
Central Tendency	Median lebih akurat daripada Mean untuk data gaji yang memiliki <i>outlier</i> ekstrem.
Korelasi	Jam belajar memiliki dampak positif terbesar terhadap nilai ujian, sementara stres berdampak negatif.
Paradoks Data	Kesimpulan dari data gabungan (agregat) bisa menyesatkan jika tidak melihat detail sub-kelompoknya ( <i>Simpson's Paradox</i> ).

#### 3.2 Pembahasan

Dalam pengolahan data praktikum ini, terbukti bahwa mengandalkan satu metrik saja (seperti Mean) bisa berbahaya. Pada kasus gaji, penggunaan Mean seolah-olah menunjukkan kesejahteraan karyawan tinggi (70 juta), padahal kenyataannya mayoritas hanya menerima 40 juta. Oleh karena itu, visualisasi seperti *Boxplot* dan *Histogram* sangat penting untuk mendeteksi distribusi data sebelum mengambil keputusan. Selain itu, *Heatmap* sangat membantu dalam menyeleksi fitur mana yang paling berpengaruh terhadap target (nilai ujian).

### 4. Kesimpulan

Berdasarkan praktikum statistika deskriptif dan eksplorasi data ini, dapat disimpulkan:

1. **Median** adalah ukuran pemusatan yang lebih *robust* (tahan) terhadap pencilan dibandingkan Mean.
2. **Korelasi** membantu mengidentifikasi variabel yang saling berhubungan, namun korelasi tidak selalu menunjukkan sebab-akibat (*causality*).
3. **Visualisasi Data** (seperti Scatter Plot dan Heatmap) mutlak diperlukan untuk memahami pola yang tidak terlihat hanya dengan melihat tabel angka.
4. Penting untuk mewaspadai **Simpson's Paradox**, di mana tren global bisa bertentangan dengan tren lokal, sehingga analisis harus dilakukan secara mendalam (granular).

### 5. Daftar Pustaka

1. Modul Praktikum Data Science, Program Studi Teknik Informatika, Universitas Muhammadiyah Pontianak, 2025.<sup>2222</sup>

+1

2. Pandas Documentation, "Descriptive Statistics", <https://pandas.pydata.org/docs/>

3. Matplotlib Documentation, "Visualizing Data Distributions", <https://matplotlib.org/>