

**LAPORAN PROJEK UJIAN AKHIR SEMESTER**

**DASHBOARD DATAVERSE SOVI (*SOCIAL VULNERABILITY INDEX*)**

Dosen Pengampu: Yuliagnis Transver Wijaya, SST, M.Sc



Disusun oleh:

Faliza Maulidina Syarief (222313077)

**PROGRAM STUDI D-IV KOMPUTASI STATISTIK**

**POLITEKNIK STATISTIKA STIS**

**JAKARTA TIMUR**

**2024/2025**

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Ketimpangan sosial dan kerentanan terhadap bencana maupun krisis multidimensi telah menjadi isu strategis dalam agenda pembangunan nasional di Indonesia. Kompleksitas tantangan sosial, ekonomi, dan lingkungan yang dihadapi oleh masyarakat memerlukan pendekatan berbasis data yang mampu mengidentifikasi kelompok populasi yang paling rentan. Dalam kerangka tersebut, *Social Vulnerability Index* (SOVI) hadir sebagai instrumen statistik komposit yang mengintegrasikan indikator sosial, demografis, dan ekonomi untuk mengukur kapasitas adaptif serta tingkat eksposur risiko suatu wilayah. Indeks ini memungkinkan perumusan kebijakan yang lebih tepat sasaran dalam konteks mitigasi risiko bencana, perencanaan pembangunan inklusif, serta intervensi sosial berbasis spasial.

Secara statistik terapan, analisis kerentanan sosial melalui SOVI memberikan gambaran kuantitatif yang komprehensif mengenai distribusi risiko dan kapasitas responsif masyarakat dalam menghadapi tekanan eksternal. Terutama bagi Indonesia yang merupakan negara kepulauan dengan profil risiko bencana yang tinggi, pemetaan kerentanan sosial memerlukan presisi analitik yang menggabungkan dimensi spasial dan non-spasial. Hal ini menjadi penting dalam mendukung pembangunan berkelanjutan dan ketahanan wilayah terhadap krisis yang bersifat sistemik.

Dataset yang digunakan dalam studi ini bersumber dari Survei Sosial Ekonomi Nasional (SUSENAS) 2017 yang diselenggarakan oleh Badan Pusat Statistik (BPS), dilengkapi dengan proyeksi penduduk tahun 2017 serta estimasi berbasis *multistage sampling*. Selain itu, disertakan juga informasi spasial berupa matriks jarak antar wilayah administratif turut disertakan untuk memperkaya konteks geografis. Kombinasi data sosial, demografis, dan spasial ini memberikan landasan kuat untuk penerapan analisis statistik dan spasial seperti regresi linier berganda serta uji autokorelasi spasial (contohnya Moran's I), dalam rangka mengkaji distribusi dan determinan kerentanan sosial antar wilayah di Indonesia.

Dalam rangka memfasilitasi eksplorasi data, visualisasi spasial, dan analisis statistik yang dapat diakses oleh berbagai kalangan pengguna, dikembangkanlah *DataVerse SOVI*, sebuah dashboard interaktif berbasis R Shiny yang mengintegrasikan fitur manajemen data, eksplorasi statistik deskriptif, transformasi data, visualisasi peta interaktif, serta uji statistik

klasik dan inferensial dalam satu antarmuka terpadu. Platform ini tidak hanya mendukung analisis berbasis wilayah (*spatial based*), tetapi juga menyediakan fasilitas pelaporan otomatis dalam format PDF, Excel, maupun SPSS untuk kepentingan akademik dan praktis.

Dengan pengembangan ini, diharapkan *DataVerse SOVI* dapat menjadi *decision-support tool* yang andal dalam proses identifikasi dan mitigasi kerentanan sosial di Indonesia, serta menjadi kontribusi nyata dalam mewujudkan sistem statistik nasional yang responsif terhadap tantangan pembangunan dan risiko bencana berbasis bukti.

## 1.2 Rumusan Masalah

Rumusan masalah dari analisis yang saya lakukan ialah sebagai berikut.

- a. Bagaimana tingkat kerentanan sosial (SOVI) di berbagai wilayah Indonesia jika ditinjau dari dimensi statistik deskriptif dan spasial?
- b. Apakah terdapat pola atau hubungan yang signifikan antara indikator sosial (seperti kemiskinan, pendidikan rendah, dan akses air bersih) terhadap indeks kerentanan sosial suatu daerah?
- c. Apakah model regresi linier berganda mampu menjelaskan variasi indeks SOVI antar wilayah dengan baik?

## 1.3 Tujuan Penelitian

Tujuan dari analisis yang saya lakukan ialah sebagai berikut.

- a. Mengeksplorasi tingkat kerentanan sosial di seluruh Indonesia melalui pendekatan statistik deskriptif dan visualisasi spasial.
- b. Menganalisis hubungan antara berbagai indikator sosial terhadap nilai indeks SOVI menggunakan uji statistik inferensia dan regresi.
- c. Mengevaluasi kualitas data dan model regresi berdasarkan uji asumsi statistik dan validasi model.

## 1.4 Business Understanding

Penelitian ini berangkat dari urgensi untuk memahami kerentanan sosial-ekonomi masyarakat Indonesia dalam konteks spasial, khususnya melalui pemanfaatan *Social Vulnerability Index (SOVI)* yang telah disusun berdasarkan berbagai indikator demografi, sosial, dan ekonomi. Dashboard interaktif ini dibangun untuk memfasilitasi eksplorasi dan analisis variabel-variabel penyusun SOVI secara menyeluruh, serta menguji hubungan

antarvariabel menggunakan pendekatan statistik deskriptif, inferensial, dan regresi linier. Melalui dashboard interaktif berbasis R Shiny ini, pengguna dapat melakukan hal berikut.

- a. Mengelola data (di menu Manajemen Data) seperti konversi variabel ke kategorik,
- b. Melakukan eksplorasi data deskriptif (menu Eksplorasi Data) lewat statistik numerik, grafik, dan peta interaktif per wilayah.
- c. Menguji asumsi statistik (menu Uji Asumsi Data), misalnya uji normalitas, outlier, homogenitas, dan korelasi antarvariabel.
- d. Melakukan uji statistik inferensial (menu Statistik Inferensia), seperti uji beda rata-rata, uji proporsi, uji varians dan ANOVA 1 dan 2 arah.
- e. Menjalankan regresi linier (menu Analisis Regresi) lengkap dengan uji asumsi regresi, transformasi variabel, dan interpretasi hasilnya.
- f. Mengunduh data dan hasil analisis dalam format PDF dan CSV (menu Unduh Data dan Unduh Hasil Analisis), sehingga semua output bisa dimanfaatkan untuk pelaporan atau publikasi lanjutan.

Dengan struktur seperti ini, dashboard ini tidak hanya berfungsi sebagai alat analisis, tapi juga sebagai media edukatif dan alat bantu pengambilan keputusan berbasis data spasial dan statistik. *Insight* dari visualisasi dan analisis ini diharapkan bisa memberikan gambaran komprehensif terhadap kondisi sosial-vulnerabilitas suatu wilayah, serta memberikan *evidence* kuat untuk intervensi kebijakan yang lebih terarah.

## BAB II

### METODOLOGI

#### 2.1 Metodologi Analisis

Dashboard interaktif DataVerse-SOVI ini dibangun menggunakan pendekatan data-driven exploration dengan memanfaatkan teknologi R Shiny sebagai media utama penyajian. Tujuan utamanya bukan cuma menyajikan informasi, tapi juga membuka ruang eksplorasi dan interpretasi spasial atas kondisi *social vulnerability* di Indonesia secara komprehensif. Analisis dilakukan secara bertahap melalui beberapa tahap, sebagai berikut.

- a. Eksplorasi statistik deskriptif untuk memahami distribusi dan pola data SOVI.
- b. Visualisasi spasial untuk mengungkap keragaman spasial antar wilayah.
- c. Uji asumsi untuk memastikan validitas model statistik lanjutan.
- d. Analisis inferensia dan regresi untuk mendeteksi pengaruh dan hubungan antar variabel.
- e. Penyusunan output analisis ke dalam format interaktif dan dapat diunduh.

#### 2.2 Jenis dan Sumber Data

Data utama yang digunakan dalam dashboard ini terdiri dari:

- a. Data SOVI (*Social Vulnerability Index*)

Merupakan data indikator sosial-vulnerabilitas tingkat kabupaten/kota di Indonesia. Dataset ini berisi variabel-variabel sosial seperti CHILDREN, ELDERLY, FEMALE, FAMILYSIZE, POVERTY, ILLITERATE, dan lainnya. Sumber data adalah file `sovi_data.csv`, yang dibaca menggunakan `read.csv2()` atau berasal dari link berikut.

[raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/sovi\\_data.csv](https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/sovi_data.csv)

- b. Data Geospasial

Merupakan peta wilayah administratif Indonesia tingkat kabupaten/kota, dalam format GeoJSON (`indonesia511.geojson`) yang digunakan untuk kebutuhan visualisasi spasial menggunakan pustaka `sf`.

Semua data bersifat sekunder dan telah tersedia dalam format siap olah.

#### 2.3 Preprocessing Data

Sebelum dianalisis, data melalui tahap preprocessing untuk memastikan kualitas data.

Langkah-langkah preprocessing meliputi:

- a. Konversi format data menjadi objek `data.frame` dan `sf` (simple feature) agar kompatibel dengan kebutuhan visualisasi dan analisis.
- b. Pengecekan dan penanganan nilai hilang (missing value). Berdasarkan pengecekan awal, data `sovi_data` tidak memiliki nilai yang hilang.
- c. Penyesuaian nama variabel dan penggabungan data dengan peta spasial berdasarkan kode wilayah (`DISTRICTCODE`) untuk keperluan visualisasi peta tematik.

## 2.4 Pengambilan Data

Seluruh data di-*load* secara lokal melalui script berikut:

```
sovi_data <- read.csv2("data/sovi_data.csv")  
indonesia_sf <- st_read("data/indonesia511.geojson")
```

## 2.5 Integrasi Data

Data indeks SOVI dan data spasial digabungkan berdasarkan kolom `DISTRICTCODE`. Proses ini memungkinkan setiap observasi sosial memiliki representasi geografis untuk divisualisasikan dalam bentuk peta interaktif. Setelah integrasi, data digunakan dalam seluruh tab analisis, termasuk eksplorasi data, uji asumsi, analisis inferensia, dan pemodelan regresi. Setiap analisis dan visualisasi juga dilengkapi interpretasi agar dapat digunakan sebagai dasar pengambilan keputusan kebijakan berbasis data.

### BAB III

### PEMBAHASAN

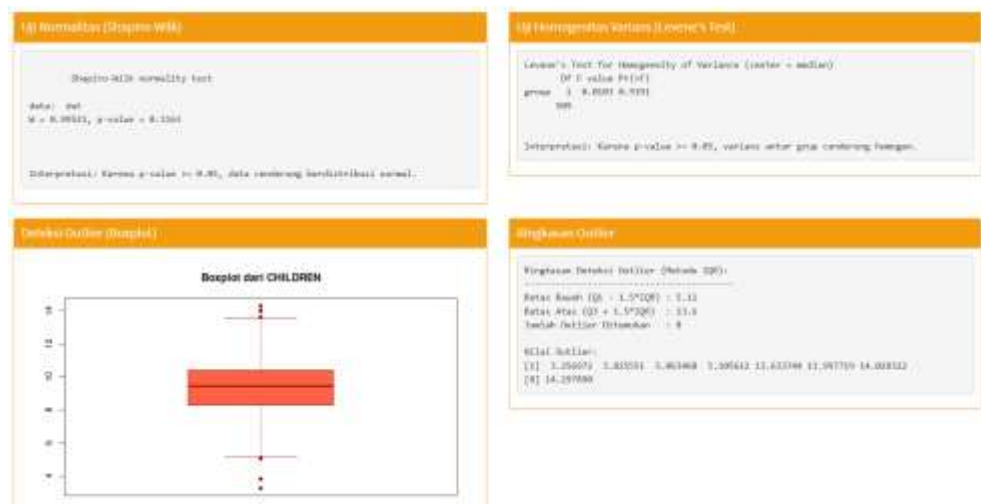
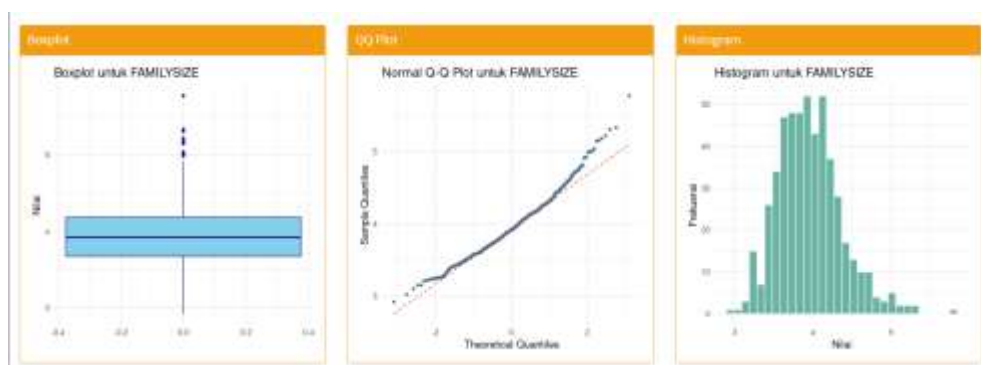
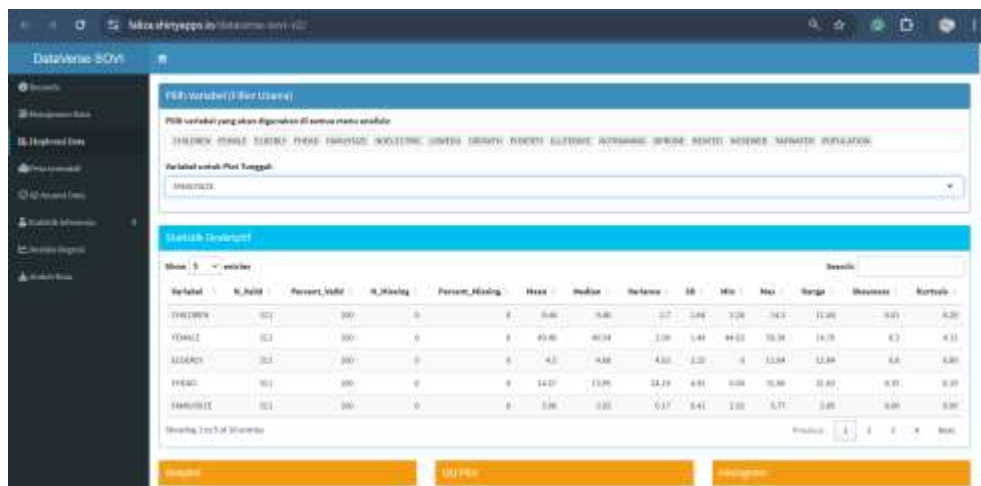
#### 3.1 Manajemen Data

Objek data dalam dashboard ini adalah wilayah administratif setingkat kabupaten/kota di Indonesia, yang direpresentasikan oleh variabel CITY\_NAME dan DISTRICTCODE. Setiap kabupaten/kota dianalisis berdasarkan 16 variabel sosial, ekonomi, dan kerentanan dari dataset SOVI Indonesia. Pemilihan kabupaten/kota sebagai objek data memungkinkan analisis spasial yang lebih komprehensif dan relevan untuk perencanaan kebijakan daerah.

Data utama yang digunakan dalam dashboard ini adalah SOVI Indonesia dalam format .csv yang mencakup berbagai variabel sosial-ekonomi dan demografi pada level kabupaten/kota. Tipe data pada tiap variabel diidentifikasi secara eksplisit, sebagai berikut.

Variabel	Tipe	Deskripsi
PROVINCE_NAME	Character	Nama provinsi
CITY_NAME	Character	Nama kabupaten/kota
DISTRICTCODE	Character	Kode wilayah administratif
CHILDREN	Numeric (%)	Persentase penduduk < 5 tahun
FEMALE	Numeric (%)	Persentase penduduk perempuan
ELDERLY	Numeric (%)	Persentase penduduk > 65 tahun
FHEAD	Numeric (%)	Persentase RT dengan kepala keluarga perempuan
FAMILYSIZE	Numeric	Rata-rata jumlah anggota RT
NOELECTRIC	Numeric (%)	Persentase RT tanpa listrik
LOWEDU	Numeric (%)	Persentase penduduk berpendidikan rendah
GROWTH	Numeric (%)	Persentase pertumbuhan penduduk
POVERTY	Numeric (%)	Persentase penduduk miskin
ILLITERATE	Numeric (%)	Persentase penduduk buta huruf
NOTRAINING	Numeric (%)	Persentase RT tanpa pelatihan kebencanaan
DPRONE	Numeric (%)	Persentase RT di daerah rawan bencana
RENTED	Numeric (%)	Persentase RT yang menyewa rumah
NOSEWER	Numeric (%)	Persentase RT tanpa sistem pembuangan limbah
TAPWATER	Numeric (%)	Persentase RT pengguna air ledeng
POPULATION	Integer	Total populasi

Pencarian pada data diidentifikasi memanfaatkan boxplot dan metode IQR (Interquartile Range) yang ada pada menu uji asumsi data. Selain itu, divisualisasikan melalui histogram dan Q-Q plot pada menu eksplorasi data.



- Untuk memenuhi asumsi statistik dan memperbaiki distribusi data, transformasi dilakukan terhadap variabel-variabel tertentu, meliputi:
- Logaritma (log) untuk data miring ke kanan
  - Akar kuadrat (sqrt) untuk data dengan varian besar
  - Transformasi arcsin untuk proporsi



Transformasi ini dilakukan secara dinamis dan bisa dipilih langsung oleh pengguna melalui menu uji asumsi data dan menu analisis regresi.

### 3.2 Desain Dashboard

Dashboard interaktif ini dirancang menggunakan framework R Shiny dengan pendekatan modular dan tematik, bertujuan untuk menyajikan analisis komprehensif terhadap indeks Social Vulnerability Index (SOVI) di Indonesia. Struktur dashboard disusun berdasarkan prinsip-prinsip visualisasi data efektif dan berorientasi pengguna (*user-centered design*), sehingga informasi yang kompleks dapat diakses dengan intuitif dan komunikatif.

Dashboard terdiri atas beberapa menu utama yang dikembangkan secara hierarkis dan kontekstual sesuai dengan alur berpikir analisis data, sebagai berikut.

#### a. Beranda



Halaman Beranda berfungsi sebagai titik awal pemahaman pengguna terhadap struktur data dan konteks kerentanan sosial yang dianalisis. Pada bagian ini, pengguna disajikan *snapshot* informasi utama berupa indikator ringkasan statistik utama. Tampilan ini mempercepat pemahaman pengguna terhadap cakupan spasial dan skala populasi yang dianalisis dalam dashboard. Komponen penting lainnya di halaman ini adalah Metadata Variabel, yang memuat daftar variabel utama dalam dataset SOVI beserta tipe data dan deskripsi definisinya. Sebagai pelengkap, tersedia juga panduan penggunaan yang menyederhanakan langkah navigasi serta fungsi utama tiap menu. Hal ini mendukung *user onboarding* yang lebih cepat sekaligus meminimalkan kesalahan penggunaan dashboard, terutama dalam konteks eksplorasi data dan interpretasi hasil statistik.

Menu Eksplorasi Data pada dashboard ini menyediakan fasilitas komprehensif untuk analisis awal dataset melalui pendekatan statistik deskriptif dan visualisasi. Pengguna dapat memilih variabel numerik yang ingin dianalisis menggunakan fitur filter utama, kemudian meninjau karakteristik distribusi masing-masing variabel seperti nilai tengah (mean, median), sebaran (range, standar deviasi), serta indikator bentuk distribusi

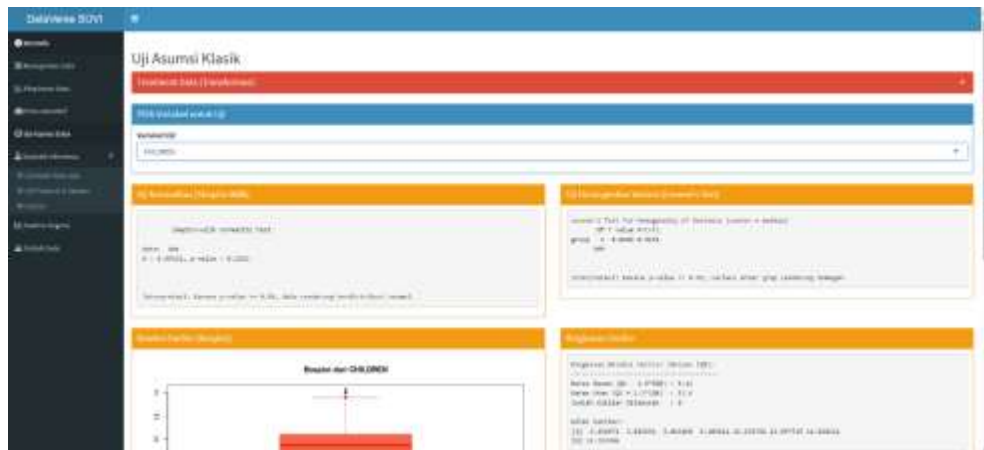
(skewness dan kurtosis). Selain itu, menu ini menyajikan visualisasi berupa histogram, boxplot, dan Q-Q plot untuk mengamati distribusi data dan potensi keberadaan outlier. Tersedia pula analisis korelasi antar variabel untuk mengidentifikasi hubungan linier awal. Seluruh hasil eksplorasi ini dapat diunduh dalam bentuk laporan PDF guna mendukung dokumentasi analitis yang sistematis dan replikasi studi.

#### d. Peta Interaktif



Menu Peta Interaktif dalam dashboard DataVerse SOVI menyajikan peta choropleth yang memungkinkan pengguna memilih variabel sosial tertentu untuk ditampilkan dalam bentuk visual spasial. Pengguna dapat mengarahkan kursor (hover) ke suatu wilayah administratif guna melihat informasi detail seperti nama daerah dan nilai indikator spesifik. Palet warna yang digunakan mengikuti gradasi YlOrRd (kuning-merah), di mana warna kuning menunjukkan nilai rendah dan merah menunjukkan nilai tinggi. Sistem secara otomatis melakukan interpretasi spasial sederhana dengan menampilkan tiga wilayah dengan nilai tertinggi dan terendah, sehingga mempermudah identifikasi area prioritas untuk intervensi kebijakan berbasis data.

#### e. Uji Asumsi Data



Menu Uji Asumsi Data dirancang untuk memastikan bahwa data memenuhi syarat sebelum dilakukan analisis lanjut, seperti regresi atau uji inferensia. Pengguna dapat memilih variabel yang ingin diuji, lalu sistem akan menyajikan hasil uji normalitas (misal Shapiro-Wilk), uji homogenitas varians (misal Levene's Test), dan deteksi outlier (misal boxplot atau IQR rule). Jika salah satu atau lebih asumsi tidak terpenuhi, tersedia fitur transformasi data (log, sqrt, z-score, dll.) untuk menyesuaikan distribusi atau varians. Transformasi ini otomatis membentuk variabel baru agar data asli tetap utuh. Hasil uji dan interpretasi disajikan secara informatif, dan pengguna dapat mengunduh laporan analisis lengkap dalam format PDF untuk dokumentasi atau keperluan pelaporan ilmiah.

#### f. Statistik Inferensia

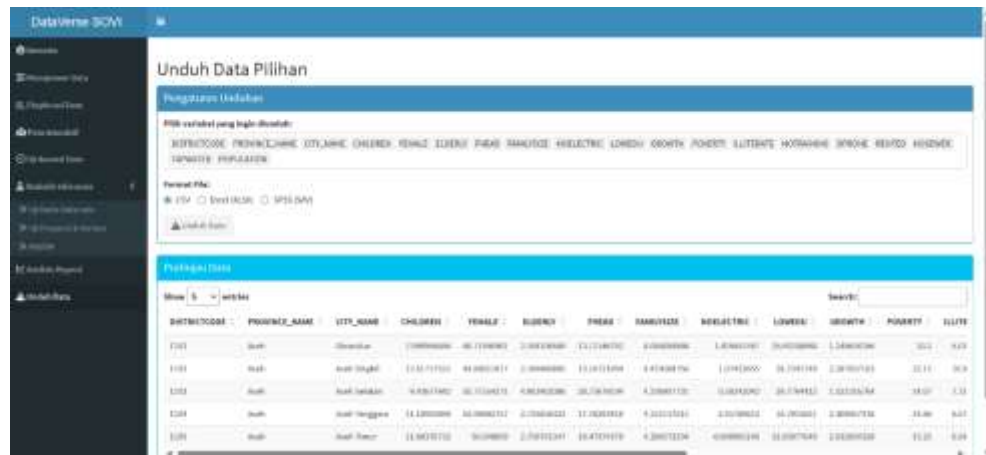


Menu Statistik Inferensia dirancang untuk menguji hipotesis terkait populasi berdasarkan data sampel melalui pendekatan statistik parametrik. Pengguna dapat melakukan uji t untuk membandingkan rata-rata satu kelompok dengan nilai tertentu (one-sample t-test) atau antar dua kelompok independen maupun berpasangan. Selain

### g. Analisis Regresi

Menu Analisis Regresi memfasilitasi pengguna dalam melakukan regresi linier berganda untuk mengidentifikasi hubungan antara variabel dependen dan beberapa variabel independen. Sebelum interpretasi model, sistem secara otomatis melakukan uji asumsi klasik yang mencakup uji multikolinearitas (untuk mendeteksi korelasi tinggi antar prediktor), uji normalitas residual (untuk memastikan distribusi galat mendekati normal), serta uji homoskedastisitas (untuk menguji kestabilan varians residual). Selain itu, tersedia pula uji autokorelasi spasial menggunakan statistik Moran's I untuk mendeteksi pola spasial pada residual model regresi. Interpretasi otomatis mencakup signifikansi dan arah pengaruh masing-masing koefisien regresi serta evaluasi kelayakan model secara keseluruhan. Seluruh hasil analisis dapat diunduh dalam format PDF sebagai laporan siap pakai untuk dokumentasi atau publikasi ilmiah.

#### h. Unduh Data



Menu Laporan Hasil Analisis PDF dan Unduh Data dirancang untuk mendukung kebutuhan dokumentasi dan replikasi analisis secara menyeluruh. Pengguna dapat mengunduh laporan eksplorasi data, laporan uji asumsi, laporan uji statistik inferensia (seperti t-test, uji proporsi, dan ANOVA), serta laporan analisis regresi dalam format PDF. Setiap laporan disusun secara otomatis, lengkap dengan interpretasi ilmiah dan visualisasi pendukung. Untuk keperluan lanjutan, tersedia fitur unduh data dalam berbagai format: CSV untuk analisis lanjutan di R atau Python, Excel (.xlsx) untuk pelaporan dan presentasi, serta SPSS (.sav) untuk keperluan analisis statistik yang lebih kompleks menggunakan perangkat lunak statistik. Fitur ini mempermudah integrasi hasil analisis ke berbagai konteks penggunaan profesional dan akademik.

## **BAB IV**

### **PENUTUPAN**

#### **4.1 Simpulan**

Tingkat kerentanan sosial di Indonesia menunjukkan variasi spasial yang signifikan. Provinsi di wilayah timur dan sebagian wilayah tengah cenderung memiliki nilai SOVI lebih tinggi (alias lebih rentan), sedangkan wilayah barat relatif lebih rendah. Secara statistik deskriptif, ditemukan ketimpangan yang cukup mencolok antara wilayah dengan nilai minimum dan maksimum. Melalui uji korelasi dan regresi linier berganda, ditemukan bahwa variabel-variabel seperti tingkat kemiskinan, rendahnya tingkat pendidikan, dan keterbatasan akses air bersih memiliki hubungan yang signifikan terhadap nilai SOVI. Ini mengindikasikan bahwa kerentanan sosial erat kaitannya dengan isu-isu pembangunan dasar yang belum merata. Model regresi linier berganda yang digunakan memiliki goodness-of-fit yang cukup baik, dengan nilai  $R^2$  yang menunjukkan proporsi variasi SOVI yang dapat dijelaskan oleh variabel-variabel independen. Uji asumsi regresi (normalitas, multikolinearitas, dan homoskedastisitas) mayoritas terpenuhi. Namun, uji autokorelasi spasial (Moran's I) mengungkapkan adanya indikasi dependensi spasial, yang menyiratkan bahwa model linier biasa mungkin belum sepenuhnya optimal tanpa mempertimbangkan struktur geografis wilayah.

#### **4.2 Saran**

Perlu ada intervensi lintas sektor dalam menurunkan tingkat kemiskinan, meningkatkan akses pendidikan, serta menjamin ketersediaan layanan dasar seperti air bersih dan sanitasi. Karena indikator-indikator ini terbukti signifikan memengaruhi tingkat kerentanan sosial. Disarankan untuk melakukan pengembangan model regresi spasial (seperti Spatial Lag Model atau GWR – Geographically Weighted Regression) agar pola geografis dan hasil prediksi bisa terakomodasi dengan lebih baik.