

**PROYEK AKHIR SEMESTER PRAKTIKUM PEMODELAN STATISTIKA
TERAPAN**

PREDIKSI TINGKAT KEMISKINAN KOTA DI INDONESIA MENGGUNAKAN PCR

Dosen Pengampu:

Ronny Susetyoko S.Si, M.Si



Disusun Oleh:

Randi Nandika Danendra

3324600013

**PROGRAM STUDI SAINS DATA TERAPAN
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA
MEI 2025**

DAFTAR ISI

BAB 1. PENDAHULUAN	4
1.1. Latar Belakang	4
1.2. Tujuan	4
1.3. Manfaat	5
1.4. Batasan Masalah	5
BAB 2. TINJAUAN PUSTAKA	7
2.1. Kemiskinan	7
2.2. Regresi Linier Berganda	7
2.3. Principal Component Regression (PCR)	8
2.4. Studi Terdahulu	9
2.5. Penggunaan R Studio dalam Analisis Regresi dan PCR.....	9
BAB 3 METODOLOGI.....	11
3.1. Jenis dan Pendekatan Penelitian	11
3.2. Teknik Analisis Data	11
3.3. Penetapan Variabel.....	11
BAB 4 HASIL DAN ANALISIS	13
4.1. Deskripsi Umum Data.....	13
4.2. Proses Analisis Principal Component Regression	13
BAB 5 KESIMPULAN DAN SARAN	33
5.1. Kesimpulan	33
5.2. Saran	33
DAFTAR PUSTAKA.....	34
LAMPIRAN.....	35

ABSTRAK

Kemiskinan merupakan salah satu indikator penting dalam mengukur kesejahteraan masyarakat di suatu wilayah. Penelitian ini bertujuan untuk membangun model prediksi tingkat kemiskinan antar kota berdasarkan beberapa variabel sosial-ekonomi menggunakan pendekatan regresi linier dan analisis komponen utama (PCA). Data yang digunakan terdiri dari sejumlah indikator yang memengaruhi kemiskinan di beberapa kota di Indonesia. Pada tahap awal, dilakukan pemodelan regresi linier berganda untuk mengetahui hubungan langsung antara variabel-variabel independen dengan tingkat kemiskinan. Selanjutnya, untuk mengatasi permasalahan multikolinearitas antar variabel prediktor, diterapkan teknik PCA guna mereduksi dimensi data dan membentuk variabel baru yang saling bebas (orthogonal). Hasil analisis menunjukkan bahwa model regresi berdasarkan komponen utama memberikan hasil yang lebih stabil dan informatif dibandingkan model regresi linier biasa, terutama ketika hanya komponen signifikan yang digunakan. Evaluasi model dilakukan dengan menggunakan analisis variansi (ANOVA), nilai koefisien determinasi (R^2), serta pendekatan best subset regression untuk menentukan model terbaik. Penelitian ini memberikan kontribusi dalam pemodelan statistik terapan untuk analisis kemiskinan, serta menjadi dasar pengambilan kebijakan berbasis data di tingkat daerah.

Kata kunci: Tingkat kemiskinan, regresi linier, PCA, multikolinearitas, subset regression, ANOVA

BAB 1. PENDAHULUAN

1.1. Latar Belakang

Kemiskinan merupakan salah satu permasalahan sosial yang kompleks dan menjadi perhatian utama dalam pembangunan ekonomi di berbagai negara, termasuk Indonesia. Sebagai indikator kesejahteraan masyarakat, tingkat kemiskinan sering dijadikan dasar dalam perumusan kebijakan publik. Oleh karena itu, pemantauan dan analisis yang akurat terhadap faktor-faktor yang mempengaruhi tingkat kemiskinan sangat penting dalam rangka menanggulangi dan mengurangi jumlah penduduk miskin.

Di era digital saat ini, ketersediaan data sosial-ekonomi yang luas memungkinkan penerapan metode statistik yang lebih mendalam untuk memahami dinamika kemiskinan. Namun, banyaknya variabel prediktor yang saling berkorelasi dapat menimbulkan permasalahan multikolinearitas, yang mengganggu interpretasi dan keandalan model regresi linier klasik. Untuk mengatasi hal tersebut, metode *Principal Component Analysis* (PCA) dapat digunakan sebagai teknik reduksi dimensi yang efektif. PCA mengubah sekumpulan variabel yang saling berkorelasi menjadi sejumlah kecil komponen utama yang tidak berkorelasi, tanpa kehilangan informasi penting dari data awal.

Penelitian ini mengangkat studi kasus analisis tingkat kemiskinan antar kota dengan tujuan untuk membangun model prediktif yang optimal. Model yang digunakan melibatkan regresi linier berganda sebagai pendekatan dasar, serta regresi berdasarkan komponen utama (Principal Component Regression) sebagai pendekatan lanjutan untuk menangani multikolinearitas. Selain itu, digunakan pula metode *best subset regression* untuk memilih kombinasi prediktor terbaik yang memberikan performa model paling baik. Dengan pendekatan ini, diharapkan diperoleh pemahaman yang lebih baik mengenai variabel-variabel dominan yang berkontribusi terhadap tingkat kemiskinan, serta membantu pengambilan kebijakan berbasis data di tingkat kota.

1.2. Tujuan

Penelitian ini dilakukan dengan tujuan sebagai berikut:

- 1) Menganalisis hubungan antara variabel-variabel sosial ekonomi dengan tingkat kemiskinan antar kota menggunakan model regresi linier berganda.
- 2) Mengidentifikasi dan mengatasi permasalahan multikolinearitas antar variabel prediktor melalui penerapan *Principal Component Analysis* (PCA).
- 3) Membangun model regresi berdasarkan komponen utama (Principal Component Regression) untuk memprediksi tingkat kemiskinan secara lebih stabil dan efisien.
- 4) Membandingkan performa model regresi linier biasa dengan model regresi berbasis PCA, baik dari segi ketepatan maupun kestabilan estimasi.
- 5) Menentukan model terbaik melalui pendekatan *best subset regression* yang mempertimbangkan kombinasi prediktor paling informatif dengan nilai kriteria informasi optimal (seperti BIC).

- 6) Memberikan rekomendasi strategis berdasarkan hasil analisis untuk membantu pengambil kebijakan dalam mengurangi tingkat kemiskinan di tingkat kota.

1.3. Manfaat

Manfaat untuk Analisis Sosial Ekonomi:

- 1) Memberikan pemahaman yang lebih dalam tentang faktor-faktor yang memengaruhi tingkat kemiskinan di suatu wilayah.
- 2) Dapat digunakan untuk mendukung perencanaan pembangunan dan alokasi sumber daya secara lebih efisien.

Manfaat untuk Statistik dan Data Science:

- 1) Menunjukkan penerapan teknik statistik seperti *Principal Component Analysis* (PCA) untuk mengatasi multikolinearitas.
- 2) Menjadi contoh nyata penggunaan regresi berbasis komponen utama (PCR) dalam analisis data multivariat.

Manfaat untuk Peneliti dan Akademisi:

- 1) Memberikan studi kasus konkret untuk pembelajaran metode statistik dalam konteks dunia nyata.
- 2) Dapat dijadikan referensi dalam penelitian lanjutan tentang kemiskinan atau bidang sosial lainnya.

1.4. Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka ditetapkan beberapa batasan masalah sebagai berikut:

- 1) Data yang digunakan terbatas pada data kuantitatif yang mencakup indikator sosial dan ekonomi dari masing-masing kota yang tersedia dalam dataset, seperti jumlah penduduk miskin, tingkat pengangguran, pendidikan, dan variabel sejenis lainnya.
- 2) Unit analisis dalam penelitian ini adalah kota, bukan individu atau rumah tangga. Sehingga hasil analisis bersifat agregat dan tidak merepresentasikan kondisi individu.
- 3) Penelitian ini hanya menggunakan metode regresi linier berganda dan *Principal Component Analysis* (PCA) sebagai pendekatan utama dalam analisis data.
- 4) Asumsi-asumsi klasik regresi linier, seperti normalitas, homoskedastisitas, dan tidak adanya multikolinearitas, akan diuji namun tidak dilakukan perbaikan lanjutan jika ditemukan pelanggaran selain dengan PCA.

- 5) Penelitian ini tidak membahas faktor-faktor kualitatif yang tidak dapat dikonversi menjadi data numerik, seperti persepsi masyarakat atau faktor budaya yang dapat memengaruhi tingkat kemiskinan.
- 6) Analisis dilakukan hanya pada data cross-section (data potong waktu) dan tidak melibatkan data time series (data runtut waktu).

BAB 2. TINJAUAN PUSTAKA

2.1. Kemiskinan

Kemiskinan merupakan salah satu indikator utama dalam mengukur tingkat kesejahteraan masyarakat dan sering menjadi fokus utama dalam kebijakan pembangunan nasional. Menurut Badan Pusat Statistik (BPS, 2022), kemiskinan didefinisikan sebagai ketidakmampuan individu untuk memenuhi kebutuhan dasar minimum, yang diukur berdasarkan garis kemiskinan—gabungan dari kebutuhan makanan dan non-makanan. Di sisi lain, kemiskinan juga sering dipahami sebagai fenomena multidimensional yang tidak hanya mencakup kekurangan pendapatan, tetapi juga keterbatasan akses terhadap pendidikan, layanan kesehatan, perumahan layak, serta partisipasi sosial (UNDP, 2020).

Indikator-indikator yang umum digunakan untuk mengukur tingkat kemiskinan meliputi pengeluaran per kapita, tingkat pengangguran, tingkat pendidikan, akses terhadap fasilitas kesehatan, serta kondisi sanitasi dan perumahan. Dalam pendekatan kuantitatif, indikator-indikator ini digunakan sebagai variabel prediktor untuk memahami faktor-faktor determinan dari kemiskinan.

Kemiskinan dapat dibedakan menjadi dua jenis utama, yaitu kemiskinan absolut dan kemiskinan relatif. Kemiskinan absolut merujuk pada ketidakmampuan memenuhi kebutuhan dasar berdasarkan standar minimum tertentu yang bersifat tetap, sedangkan kemiskinan relatif mempertimbangkan kondisi ekonomi individu atau rumah tangga dibandingkan dengan masyarakat sekitarnya (UNDP, 2020).

Berbagai studi terdahulu menunjukkan bahwa kemiskinan di Indonesia dipengaruhi oleh sejumlah faktor sosial ekonomi. Misalnya, penelitian oleh Prasetyo dan Nugroho (2020) menemukan bahwa tingkat pendidikan dan akses terhadap infrastruktur dasar berpengaruh signifikan terhadap tingkat kemiskinan antar wilayah. Studi lain oleh Kartika dan Firdaus (2018) menunjukkan adanya pengaruh negatif dari tingkat pengangguran terhadap tingkat kesejahteraan, yang secara tidak langsung berkontribusi terhadap kemiskinan. Oleh karena itu, analisis yang komprehensif dan berbasis data sangat diperlukan untuk memahami dinamika kemiskinan secara lebih tepat di tingkat kota.

2.2. Regresi Linier Berganda

Regresi linier berganda adalah salah satu teknik analisis statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen. Dalam konteks sosial ekonomi, regresi ini digunakan untuk mengidentifikasi dan mengukur pengaruh berbagai faktor terhadap suatu indikator kesejahteraan, seperti tingkat kemiskinan. Secara umum, bentuk persamaan regresi linier berganda dapat dituliskan sebagai:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

di mana Y adalah variabel dependen, X_1, X_2, \dots, X_n adalah variabel independen, β_0 adalah intersep, β_1, \dots, β_n adalah koefisien regresi, dan ε adalah komponen galat.

Menurut Gujarati dan Porter (2009), untuk memastikan validitas model regresi, sejumlah asumsi dasar harus dipenuhi, yaitu normalitas distribusi galat, tidak adanya multikolinearitas, homoskedastisitas (varian galat yang konstan), serta hubungan linier antara variabel. Pelanggaran terhadap asumsi-asumsi ini dapat mengganggu hasil estimasi dan menurunkan keakuratan prediksi model.

Koefisien regresi dari model linier menunjukkan besarnya perubahan variabel dependen yang diakibatkan oleh perubahan satu satuan variabel independen, dengan asumsi variabel lainnya tetap. Sementara itu, nilai koefisien determinasi (R^2) mengindikasikan seberapa besar proporsi variasi dalam variabel dependen yang dapat dijelaskan oleh model.

Walaupun regresi linier berganda banyak digunakan karena kemudahannya dalam interpretasi, metode ini memiliki sejumlah keterbatasan, terutama ketika variabel-variabel independen memiliki korelasi yang tinggi (multikolinearitas). Dalam kondisi seperti ini, koefisien regresi menjadi tidak stabil dan interpretasi menjadi tidak dapat diandalkan. Untuk mengatasi masalah ini, metode alternatif seperti Principal Component Regression (PCR) dapat diterapkan.

2.3. Principal Component Regression (PCR)

Principal Component Regression (PCR) merupakan salah satu metode regresi alternatif yang dirancang untuk mengatasi permasalahan multikolinearitas dalam regresi linier berganda. Multikolinearitas terjadi ketika terdapat korelasi tinggi antar variabel independen, yang menyebabkan estimasi koefisien regresi menjadi tidak stabil dan sulit diinterpretasikan. PCR menggabungkan prinsip dari analisis komponen utama (Principal Component Analysis/PCA) dengan regresi linier untuk menghasilkan model yang lebih robust, khususnya ketika jumlah variabel prediktor banyak dan saling berkorelasi.

Menurut Jolliffe (2002), PCR bekerja dengan terlebih dahulu melakukan PCA terhadap variabel-variabel independen untuk menghasilkan sejumlah komponen utama (principal components), yaitu kombinasi linier dari variabel-variabel asli yang saling ortogonal (tidak berkorelasi). Komponen-komponen utama ini kemudian digunakan sebagai prediktor dalam model regresi, menggantikan variabel-variabel awal. Komponen yang dipilih biasanya adalah yang menjelaskan variansi terbesar dalam data.

Kelebihan PCR terletak pada kemampuannya mengurangi dimensi data tanpa kehilangan terlalu banyak informasi, serta mengatasi multikolinearitas secara langsung melalui transformasi variabel. Selain itu, PCR mampu meningkatkan interpretabilitas model ketika dihadapkan pada banyak variabel penjelas yang saling berkaitan erat. Namun demikian, PCR juga memiliki keterbatasan, salah satunya adalah bahwa komponen utama yang digunakan dalam model tidak selalu memiliki makna interpretatif secara langsung terhadap fenomena yang diteliti, karena mereka merupakan hasil kombinasi matematis dari variabel-variabel awal.

Dalam konteks penelitian ini, PCR digunakan untuk menganalisis faktor-faktor sosial ekonomi yang mempengaruhi tingkat kemiskinan antar kota. Dengan menggunakan pendekatan ini, model regresi dapat dibangun dengan mempertimbangkan keterkaitan antar indikator seperti pengangguran, pendidikan, dan kesehatan, tanpa terganggu oleh multikolinearitas yang mungkin terjadi dalam data.

2.4. Studi Terdahulu

Berbagai penelitian telah dilakukan untuk menganalisis faktor-faktor yang memengaruhi tingkat kemiskinan, baik di tingkat nasional maupun lokal. Penelitian oleh Siregar dan Wahyuni (2016) menunjukkan bahwa tingkat pendidikan, pengangguran, dan akses terhadap layanan kesehatan memiliki hubungan yang signifikan dengan tingkat kemiskinan di Indonesia. Studi ini menggunakan pendekatan regresi linier berganda untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap peningkatan kemiskinan.

Penelitian lain oleh Prasetyo dan Oktaviani (2018) menemukan bahwa terdapat hubungan negatif antara Indeks Pembangunan Manusia (IPM) dan tingkat kemiskinan antar kabupaten/kota. Hal ini menunjukkan bahwa peningkatan dalam komponen-komponen pembangunan manusia seperti pendidikan, kesehatan, dan pendapatan akan berkontribusi terhadap penurunan kemiskinan.

Studi yang lebih teknis dilakukan oleh Anggraini et al. (2020), yang menggunakan pendekatan Principal Component Analysis (PCA) untuk mereduksi dimensi data sosial ekonomi sebelum membangun model regresi. Hasil penelitian menunjukkan bahwa model regresi berdasarkan komponen utama dapat mengatasi permasalahan multikolinearitas dan memberikan prediksi yang lebih stabil.

Penelitian-penelitian terdahulu ini memberikan landasan metodologis dan empiris yang kuat untuk menggunakan teknik statistik seperti regresi linier dan PCA dalam analisis tingkat kemiskinan. Hal ini mendukung pendekatan yang digunakan dalam studi ini, yaitu penerapan Principal Component Regression (PCR) untuk memahami faktor-faktor yang memengaruhi kemiskinan antar kota di Indonesia.

2.5. Penggunaan R Studio dalam Analisis Regresi dan PCR

R Studio merupakan salah satu Integrated Development Environment (IDE) yang paling populer untuk bahasa pemrograman R. Dalam penelitian ini, R Studio digunakan sebagai alat utama dalam pengolahan data, analisis regresi, penerapan PCA, serta visualisasi hasil analisis. R Studio menyediakan berbagai pustaka (packages) yang mendukung analisis statistik lanjutan secara komprehensif dan efisien.

Adapun tahapan penggunaan R Studio dalam penelitian ini meliputi:

- 1) **Import dan Pembersihan Data:** Menggunakan paket seperti `readr`, `readxl`, dan `dplyr` untuk mengimpor serta membersihkan data dari file `.csv` atau Excel.

- 2) **Eksplorasi dan Statistik Deskriptif:** Dilakukan dengan bantuan ggplot2, summarytools, dan skimr untuk memberikan gambaran awal data.
- 3) **Analisis Korelasi dan Multikolinearitas:** Digunakan corrplot, GGally, dan car::vif() untuk melihat hubungan antar variabel dan mengidentifikasi potensi multikolinearitas.
- 4) **Penerapan PCA:** Dilakukan dengan FactoMineR, psych, atau princomp untuk mengekstrak komponen utama dari variabel prediktor.
- 5) **Penerapan PCR:** Menggunakan fungsi dari paket pls atau caret untuk membangun model Principal Component Regression dan mengevaluasi performanya.
- 6) **Evaluasi Model:** Digunakan caret, leaps, dan performance untuk melakukan validasi model, menghitung R^2 , adjusted R^2 , dan melakukan best subset regression.
- 7) **Visualisasi dan Interpretasi:** Paket ggbiplot, factoextra, dan ggpubr digunakan untuk memperjelas hasil PCA dan interpretasi model regresi.

Penggunaan R Studio memungkinkan proses analisis menjadi **reproducible**, efisien, serta mudah untuk didokumentasikan. Hal ini sangat penting dalam penelitian akademik, di mana transparansi dan validitas analisis sangat dijaga (Wickham & Grolemund, 2017).

BAB 3 METODOLOGI

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan statistik inferensial, bertujuan untuk memodelkan dan mengidentifikasi pengaruh variabel-variabel sosial ekonomi terhadap tingkat kemiskinan antar kota di Indonesia. Model yang digunakan mencakup regresi linier berganda sebagai model dasar, serta Principal Component Regression (PCR) sebagai model alternatif yang mengatasi masalah multikolinearitas.

3.2. Teknik Analisis Data

Penelitian ini menggunakan dua teknik utama:

Regresi Linier Berganda

Digunakan untuk mengukur pengaruh langsung dari masing-masing variabel prediktor terhadap tingkat kemiskinan. Model umum:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Dengan asumsi:

- 1) Linearitas
- 2) Normalitas residual
- 3) Homoskedastisitas
- 4) Tidak ada multikolinearitas (yang akan diuji dengan VIF)

Principal Component Regression (PCR)

PCR dilakukan dalam dua tahap:

- 1) Principal Component Analysis (PCA): Mengubah variabel X menjadi beberapa komponen utama (PC) yang saling orthogonal (tidak berkorelasi).
- 2) Regresi terhadap Komponen: Model regresi dibangun dengan menggunakan beberapa komponen utama terpilih sebagai prediktor untuk Y.

Kriteria pemilihan komponen utama:

- 1) Berdasarkan nilai eigen ($\lambda > 1$)
- 2) Berdasarkan proporsi variansi kumulatif (misal, >80%)

PCR membantu mengatasi multikolinearitas dan memperkuat kestabilan model prediksi.

3.3. Penetapan Variabel

Penelitian ini menggunakan data dari berbagai indikator sosial ekonomi untuk menganalisis **Tingkat Kemiskinan Kota** sebagai variabel dependen. Variabel-variabel dalam penelitian ini ditetapkan sebagai berikut:

1. Variabel Dependen (Y):

Tingkat Kemiskinan Kota

Merupakan persentase jumlah penduduk miskin terhadap total penduduk kota di suatu provinsi.

2. Variabel Independen (X):

- a) **Gini_ratio**: Indeks ketimpangan distribusi pendapatan.
- b) **Tingkat_penyelesaian_SD**: Persentase penduduk yang menyelesaikan pendidikan SD.
- c) **Tingkat_penyelesaian_SMP**: Persentase penduduk yang menyelesaikan pendidikan SMP.
- d) **Tingkat_penyelesaian_SMA**: Persentase penduduk yang menyelesaikan pendidikan SMA.
- e) **Buta_huruf_15**: Persentase penduduk usia 15+ yang buta huruf.
- f) **Buta_huruf_15_minus**: Persentase penduduk usia di bawah 15 tahun yang buta huruf.
- g) **Buta_huruf_45**: Persentase penduduk usia 45 tahun ke atas yang buta huruf.
- h) **Anak_bekerja**: Persentase anak-anak yang bekerja.
- i) **Pengangguran_terbuka**: Tingkat pengangguran terbuka.
- j) **Pekerja_setengah_pengangguran**: Persentase pekerja yang bekerja kurang dari jam kerja normal.
- k) **Kerja_informal**: Persentase tenaga kerja informal secara umum.
- l) **Kerja_informal_nonpertanian**: Tenaga kerja informal di sektor non-pertanian.
- m) **Kerja_informal_pertanian**: Tenaga kerja informal di sektor pertanian.
- n) **Lapangan_kerja_informal**: Ketersediaan lapangan kerja informal.
- o) **Upah_rata2_perjam**: Rata-rata upah per jam yang diterima pekerja.
- p) **Konsumsi_kalori_perhari**: Rata-rata konsumsi kalori per kapita per hari.
- q) **Konsumsi_protein_perhari**: Rata-rata konsumsi protein per kapita per hari.
- r) **GK_perkotaan**: Garis kemiskinan di wilayah perkotaan.
- s) **GK_non_makanan_kota**: Komponen non-makanan dari garis kemiskinan kota.
- t) **Tingkat_kerentanan_penduduk**: Persentase penduduk yang rentan jatuh miskin.
- u) **kepemilikan_akta_40kebawah**: Persentase penduduk kelompok 40% terbawah yang memiliki akta kelahiran.
- v) **Median_pendapatan_penduduk_dibawah40**: Median pendapatan penduduk kelompok 40% terbawah.

BAB 4 HASIL DAN ANALISIS

4.1. Deskripsi Umum Data

Dataset yang digunakan dalam penelitian ini mencakup indikator-indikator sosial ekonomi dari berbagai kota di Indonesia. Variabel-variabel tersebut meliputi tingkat pendidikan, ketenagakerjaan, upah, konsumsi, dan garis kemiskinan. Variabel dependen dalam penelitian ini adalah **Tingkat Kemiskinan Kota**, sementara variabel independennya terdiri dari 22 indikator sosial ekonomi.

Data telah melalui tahap praproses berupa penghapusan kolom yang tidak relevan, standarisasi data untuk analisis PCA, serta pemisahan variabel dependen dan independen. Setelah itu, dilakukan analisis regresi linier berganda dan Principal Component Regression (PCR).

4.2. Proses Analisis Principal Component Regression

1. Membaca Dataset

Input dan Output				
Tingkat_kemiskinan_kota <- read.csv("C:/Users/Lenovo/OneDrive/Documents/PENS/SEMESTER 2/Pemodelan Statistik Terapan/TUGAS AKHIR/datafinal1.csv")				
Tingkat_kemiskinan_kota				
##	Provinsi	Tingkat_kemiskinan_kota	Gini_ratio	
## 1	ACEH	10.35	0.327	
## 2	SUMATERA UTARA	8.63	0.358	
## 3	SUMATERA BARAT	4.90	0.316	
## 4	RIAU	6.49	0.350	
## 5	JAMBI	10.48	0.376	
## 6	SUMATERA SELATAN	11.37	0.348	
## 7	BENGKULU	14.53	0.371	
## 8	LAMPUNG	8.34	0.352	
## 9	KEP. BANGKA BELITUNG	3.48	0.263	
## 10	KEP. RIAU	5.46	0.328	
## 11	DKI JAKARTA	4.61	0.412	
## 12	JAWA BARAT	7.52	0.427	
## 13	JAWA TENGAH	10.02	0.392	
## 14	DI YOGYAKARTA	10.64	0.468	
## 15	JAWA TIMUR	7.78	0.381	
## 16	BANTEN	5.89	0.384	
## 17	BALI	4.12	0.371	
## 18	NUSA TENGGARA BARAT	13.98	0.392	
## 19	NUSA TENGGARA TIMUR	9.00	0.308	
## 20	KALIMANTAN BARAT	4.63	0.334	
## 21	KALIMANTAN TENGAH	5.09	0.332	
## 22	KALIMANTAN SELATAN	4.03	0.329	
## 23	KALIMANTAN TIMUR	4.97	0.320	
## 24	KALIMANTAN UTARA	5.58	0.256	

```
## 25      SULAWESI UTARA      5.04      0.355
## 26      SULAWESI TENGAH     9.13      0.337
## 27      SULAWESI SELATAN    4.98      0.378
## 28      SULAWESI TENGGARA   7.22      0.382
## 29      GORONTALO          4.49      0.406
## 30      SULAWESI BARAT      9.33      0.472
## 31      MALUKU              5.90      0.309
## 32      MALUKU UTARA        6.17      0.330
## 33      PAPUA BARAT         7.64      0.302
## 34      PAPUA               5.29      0.294
```

```
# Model OLS
```

```
model <- lm(Tingkat_kemiskinan_kota ~ ., data = Tingkat_kemiskinan_kota)
```

```
summary(model)
```

```
## Residual standard error: NaN on 0 degrees of freedom
```

```
## Multiple R-squared:      1, Adjusted R-squared:      NaN
```

```
## F-statistic:      NaN on 33 and 0 DF, p-value: NA
```

Analisis

Interpretasi Output

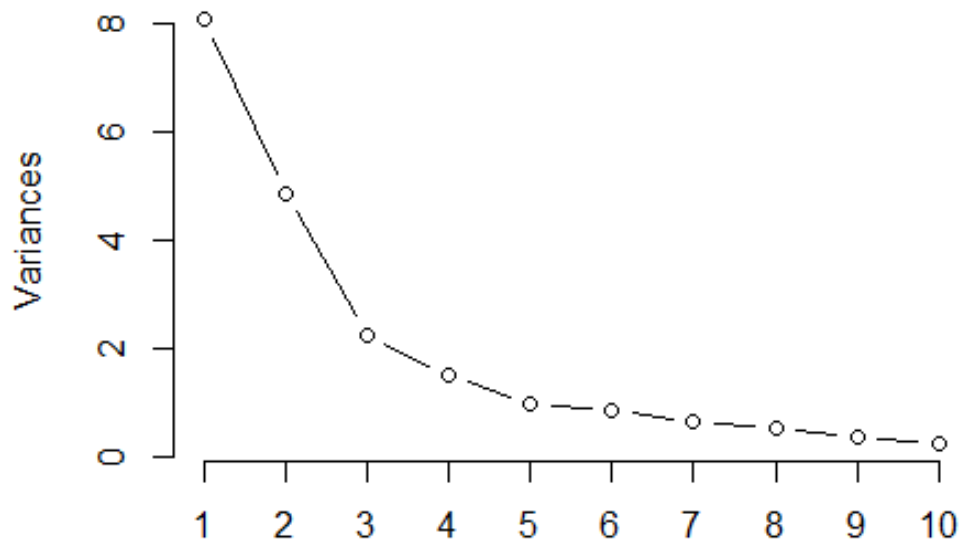
1. Residual standard error: NaN on 0 degrees of freedom
Artinya: Model tidak memiliki derajat kebebasan tersisa untuk menghitung residual. Ini biasanya terjadi jika jumlah prediktor sama dengan (atau lebih besar dari) jumlah observasi, atau jika terjadi perfect multicollinearity (prediktor saling berkorelasi sempurna).
2. Multiple R-squared: 1
Artinya: Model menjelaskan 100% variasi dari variabel dependen. Ini sangat tidak wajar dalam data nyata dan biasanya menjadi indikasi overfitting atau masalah pada data input.
3. Adjusted R-squared: NaN
Artinya: Karena tidak ada derajat kebebasan, adjusted R-squared tidak dapat dihitung. Adjusted R² biasanya dikoreksi untuk jumlah prediktor, dan dalam kasus ini hasilnya tidak terdefinisi.
4. F-statistic: NaN dan p-value: NA
Ini juga akibat dari tidak adanya derajat kebebasan residual. F-statistik dan p-value tidak dapat dihitung tanpa residual error.

2. Menghapus Kolom Pertama (Target)

Input dan Output

```
Tingkat_kemiskinanX <- Tingkat_kemiskinan_kota[, -1]
Tingkat_kemiskinanX.pca <- prcomp(Tingkat_kemiskinanX, center = TRUE,
, scale. = TRUE)
plot(Tingkat_kemiskinanX.pca, type="l")
```

Tingkat_kemiskinanX.pca



Analisis

Interpretasi:

- Komponen utama pertama (PC1) memiliki varians paling tinggi (~8), artinya:
 - PC1 menyumbang informasi paling besar dari keseluruhan data.
 - PC1 memuat kombinasi linier dari variabel asli yang menjelaskan paling banyak variasi (persebaran) data.
- Komponen kedua (PC2) juga menyumbang variasi cukup besar (~5), tapi sudah jauh di bawah PC1.
- Setelah komponen ke-3 atau ke-4, kontribusi tiap komponen semakin kecil dan stabil (mendekati datar), menunjukkan bahwa:
 - Komponen-komponen tersebut menyumbang informasi yang sangat kecil dan bisa jadi hanya berisi noise atau informasi yang tidak penting.

3. Ringkasan Komponen Utama

Input dan Output

```
# 3. Ringkasan komponen utama
summary(Tingkat_kemiskinanX.pca)

## Importance of components:
##
```

	PC1	PC2	PC3	PC4	PC5	P
C6	PC7					

```

## Standard deviation      2.8457 2.2079 1.4986 1.22125 0.98850 0.917
52 0.79260
## Proportion of Variance 0.3856 0.2321 0.1069 0.07102 0.04653 0.040
09 0.02991
## Cumulative Proportion  0.3856 0.6177 0.7247 0.79570 0.84223 0.882
32 0.91224
##
##          PC8      PC9      PC10      PC11      PC12
PC13      PC14
## Standard deviation      0.71963 0.60391 0.48248 0.43323 0.41122 0.
35041 0.26496
## Proportion of Variance 0.02466 0.01737 0.01109 0.00894 0.00805 0.
00585 0.00334
## Cumulative Proportion  0.93690 0.95426 0.96535 0.97429 0.98234 0.
98818 0.99153
##
##          PC15      PC16      PC17      PC18      PC19
PC20
## Standard deviation      0.25711 0.20269 0.17886 0.14289 0.13176 0.
03090
## Proportion of Variance 0.00315 0.00196 0.00152 0.00097 0.00083 0.
00005
## Cumulative Proportion  0.99468 0.99663 0.99816 0.99913 0.99995 1.
00000
##
##          PC21
## Standard deviation      4.617e-16
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00

```

Analisis

Analisis Output PCA

a. Standard Deviation

- 1) Ini menunjukkan akar kuadrat dari *eigenvalue* tiap komponen utama.
- 2) Semakin tinggi nilainya, semakin besar variasi yang ditangkap oleh komponen tersebut.

Contoh:

- 1) PC1: 2.8457 (sangat tinggi)
- 2) PC21: 4.617e-16 (sangat kecil, hampir nol → tidak menyumbang informasi berarti)

b. Proportion of Variance

- 1) Menunjukkan seberapa besar proporsi variasi dari data asli yang dijelaskan oleh masing-masing PC.

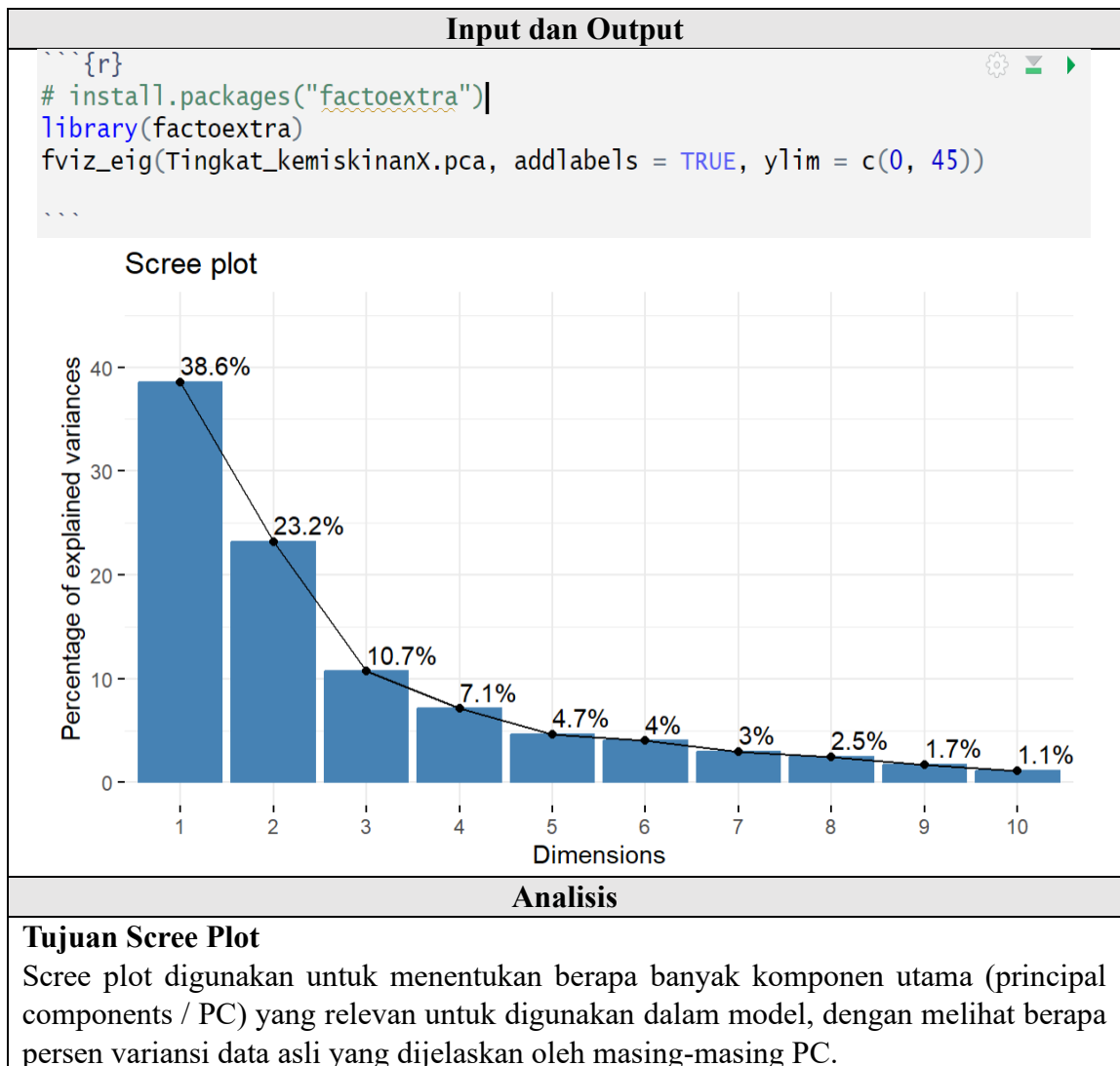
Komponen	Proporsi Variasi
PC1	38.56%
PC2	23.21%
PC3	10.69%
PC4	7.10%

c. Cumulative Proportion

2) Menunjukkan akumulasi proporsi variasi hingga komponen tersebut.

Jumlah PC	Variasi Kumulatif
1	38.56%
2	61.77%
3	72.47%
4	79.57%
5	84.22%
6	88.23%
7	91.22%
8	93.69%

4. Visualisasi Scree Plot



Hasil dari Scree Plot

Dari visualisasi:

Komponen	Proporsi Variansi	Kumulatif Variansi
PC1	38.6%	38.6%
PC2	23.2%	61.8%
PC3	10.7%	72.5%
PC4	7.1%	79.6%
PC5	4.7%	84.3%
PC6	4%	88.3%
PC7	3%	91.2%

Interpretasi dan Keputusan

- 1) PC1 hingga PC4 secara kumulatif menjelaskan sekitar 79.6% dari total variansi — ini sangat tinggi, menunjukkan bahwa 4 komponen utama saja sudah merepresentasikan sebagian besar informasi dalam data.
- 2) Setelah PC4, kontribusi variansi oleh komponen-komponen berikutnya menurun tajam (hanya sekitar 3–4%), menunjukkan bahwa tambahan komponen tidak memberikan informasi signifikan.
- 3) "Elbow point" terlihat jelas di antara PC2 dan PC4. Ini adalah titik di mana grafik mulai melandai, menandakan jumlah komponen optimal yang harus dipertahankan.

5. Matriks Rotasi

Input dan Output

4. Matriks rotasi

```
print(Tingkat_kemiskinanX.pca$rotation)
```

```
##                                PC1              PC2
PC3
## Tingkat_kemiskinan_kota      -0.09851115 -0.234739428
0.075821352
## Gini_ratio                    -0.01152680 -0.315412301 -
0.263173852
## Tingkat_penyelesaian_SD      0.22973462 -0.295925033
0.165118957
## Tingkat_penyelesaian_SMP     0.23529082 -0.222031963
0.142255695
## Tingkat_penyelesaian_SMA     0.27673178 -0.116940755 -
0.094052817
## Buta_huruf_15                -0.26125841  0.083504009 -
0.348773376
## Buta_huruf_15_minus          -0.23625601  0.239817018 -
0.288846029
```

## Buta_huruf_45 0.320699509	-0.25610404	-0.004508306	-
## Anak_bekerja 0.034379654	-0.23774704	-0.012525576	
## Tingkat_setengah_pengangguran 0.387231808	-0.19666734	-0.102216063	
## Tenaga_kerja_formal 0.087639651	0.32586648	0.079076330	-
## Kerja_informal_pertanian 0.081184907	-0.22421593	-0.269948727	
## Lapangan_kerja_informal 0.087639651	-0.32586648	-0.079076330	
## Upah_rata2_perjam 0.160331621	0.15948084	0.286389079	-
## Konsumsi_kalori_perhari 0.316827292	0.04952169	-0.240107223	-
## Konsumsi_protein_perhari 0.231045915	0.18525433	-0.225903643	-
## GK_perkotaan 0.126680986	0.14059969	0.334705724	
## GK_non_makanan_kota 0.014674878	0.13866018	0.358546673	
## Tingkat_kerentanan_penduduk 0.431585346	-0.18007256	0.073981297	
## kepemilikan_akta_40kebawah 0.002420577	0.23519040	-0.246209691	-

Analisis

Analisis Komponen Utama 1 (PC1)

Variabel dengan kontribusi besar terhadap PC1:

- 1) Tenaga_kerja_formal: **+0.3259**
- 2) Lapangan_kerja_informal: **-0.3259**
- 3) Tingkat_penyelesaian_SMA: **+0.2767**
- 4) Tingkat_penyelesaian_SMP: **+0.2353**
- 5) Tingkat_penyelesaian_SD: **+0.2297**
- 6) Kepemilikan_akta_40kebawah: **+0.2352**
- 7) Buta_huruf_15_minus, Buta_huruf_45: **kontribusi negatif**

MaknaPC1:

Komponen ini tampaknya mewakili tingkat pendidikan dan formalitas kerja. Nilai positif menunjukkan wilayah dengan:

- 1) Pendidikan lebih tinggi
- 2) Lebih banyak pekerja formal
- 3) Sedikit lapangan kerja informal
- 4) Akses terhadap dokumen kependudukan

Analisis Komponen Utama 2 (PC2)

Variabel dominan:

- 1) GK_non_makanan_kota: **+0.3585**
- 2) GK_perkotaan: **+0.3347**
- 3) Upah_rata2_perjam: **+0.2864**
- 4) Gini_ratio: **-0.3154**
- 5) Tingkat_penyelesaian_SD: **-0.2959**
- 6) Kerja_informal_pertanian: **-0.2699**

MaknaPC2:

Kemungkinan mewakili aspek ekonomi dan ketimpangan:

- 1) Positif: daerah dengan konsumsi non-makanan lebih tinggi, upah lebih baik
- 2) Negatif: daerah dengan ketimpangan tinggi (Gini tinggi), sektor informal dominan

Analisis Komponen Utama 3 (PC3)**Variabel dominan:**

- 1) Tingkat_kerentanan_penduduk: **+0.4316**
- 2) Tingkat_setengah_pengangguran: **+0.3872**
- 3) Buta_huruf_15: **-0.3488**
- 4) Buta_huruf_15_minus: **-0.2888**
- 5) Konsumsi_kalori_perhari: **-0.3168**
- 6) Lapangan_kerja_informal: **+0.0876**

MaknaPC3:

Kemungkinan menunjukkan kerentanan sosial & ketahanan hidup:

- 1) Positif: penduduk lebih rentan, tingkat setengah pengangguran tinggi
- 2) Negatif: wilayah dengan kondisi buta huruf dan konsumsi kalori lebih rendah

6. Nilai Eigen dari Matriks Rotasi

Input dan Output					
# 5. Nilai eigen dari matriks korelasi					
eigen(cor(Tingkat_kemiskinanX))\$values					
##	[1]	8.097950e+00	4.874686e+00	2.245657e+00	1.491449e+00
		71417e-01			
##	[6]	8.418444e-01	6.282139e-01	5.178625e-01	3.647108e-01
		27903e-01			
##	[11]	1.876888e-01	1.691017e-01	1.227871e-01	7.020551e-02
		10476e-02			
##	[16]	4.108305e-02	3.199033e-02	2.041649e-02	1.736194e-02
		48061e-04			
##	[21]	-3.122502e-17			
diag(var(Tingkat_kemiskinanX.pca\$x))					
##	PC1	PC2	PC3	PC4	PC5
	PC6				
##	8.097950e+00	4.874686e+00	2.245657e+00	1.491449e+00	9.771417e-01

```

8.418444e-01
##          PC7          PC8          PC9          PC10          PC11
PC12
## 6.282139e-01 5.178625e-01 3.647108e-01 2.327903e-01 1.876888e-01
1.691017e-01
##          PC13          PC14          PC15          PC16          PC17
PC18
## 1.227871e-01 7.020551e-02 6.610476e-02 4.108305e-02 3.199033e-02
2.041649e-02
##          PC19          PC20          PC21
## 1.736194e-02 9.548061e-04 7.593407e-32

```

Analisis

Makna Nilai Eigen dalam PCA

- 1) Nilai eigen (eigenvalue) menunjukkan seberapa besar variansi (informasi) dari data asli yang dijelaskan oleh masing-masing komponen utama (PC).
- 2) Total variansi = jumlah eigenvalues = jumlah variabel asli = 21 (karena PCA dilakukan pada *matriks korelasi*).
- 3) Komponen dengan eigenvalue besar menjelaskan lebih banyak informasi.

Komponen	Penjelasan	Implikasi
PC1–PC4	Menjelaskan ~78% variasi	Bisa digunakan untuk membangun indikator komposit atau kluster antar kota
PC5–PC9	Masih menyumbang informasi tambahan, meskipun menurun	Opsional untuk eksplorasi tambahan
PC10–PC21	Variansi sangat kecil, bisa diabaikan	Tidak disarankan digunakan

7. Korelasi Antar Komponen Utama

Input dan Output

6. Korelasi antar komponen utama

```
cor(Tingkat_kemiskinanX.pca$x)
```

```

##          PC1          PC2          PC3          PC4
PC5
## PC1  1.000000e+00 -8.946706e-16  3.173830e-16 -2.299911e-17 -
1.178944e-16
## PC2 -8.946706e-16  1.000000e+00 -3.563695e-16 -3.099036e-16 -
1.450638e-16
## PC3  3.173830e-16 -3.563695e-16  1.000000e+00  3.121835e-16 -
5.551320e-16
## PC4 -2.299911e-17 -3.099036e-16  3.121835e-16  1.000000e+00 -
4.227056e-16
## PC5 -1.178944e-16 -1.450638e-16 -5.551320e-16 -4.227056e-16

```

```

1.000000e+00
## PC6 8.203147e-17 -2.420202e-16 7.301414e-17 -6.417615e-17 -
1.528041e-16
## PC7 2.267712e-16 2.493709e-16 2.779673e-16 -2.281366e-16 -
1.108037e-16
## PC8 -8.219228e-16 -3.659736e-16 3.821566e-16 -9.643651e-18
6.128103e-16
## PC9 -3.809994e-16 -2.278109e-16 3.875378e-16 4.605095e-17
5.418662e-16
## PC10 1.271544e-16 -3.811514e-16 -1.408959e-16 1.733673e-16 -
8.124243e-16
## PC11 -9.799566e-17 2.846778e-17 8.451268e-17 1.272095e-16 -
2.158821e-16
## PC12 3.237890e-16 -3.507940e-17 -3.756053e-17 -2.170429e-17 -
1.957219e-16
## PC13 -1.821379e-17 1.004044e-15 -2.747648e-16 4.529087e-16 -
1.196591e-15
## PC14 1.898588e-16 1.527223e-16 5.489113e-16 4.056586e-17 -
2.857473e-16
## PC15 -6.894083e-16 -1.712684e-16 -6.733958e-17 -4.058813e-16 -
5.256217e-16
## PC16 7.086629e-16 -7.542924e-16 7.237970e-17 -3.999626e-16 -
1.482835e-16
## PC17 -7.311132e-16 -3.251987e-16 -4.314129e-16 -2.710064e-16 -
1.806097e-16
## PC18 -9.251846e-16 1.859031e-15 1.762429e-15 1.970532e-16
3.795818e-16
## PC19 -1.339180e-15 2.136304e-15 -8.710243e-16 1.240038e-15
1.510193e-15
## PC20 -4.101664e-15 -9.756939e-16 -3.586435e-15 3.517870e-15 -
1.714997e-15
## PC21 2.464152e-01 5.375308e-01 -2.699247e-01 -1.150691e-01 -
1.223593e-01
## PC6 PC7 PC8 PC9
PC10
## PC1 8.203147e-17 2.267712e-16 -8.219228e-16 -3.809994e-16
1.271544e-16
## PC2 -2.420202e-16 2.493709e-16 -3.659736e-16 -2.278109e-16 -
3.811514e-16
## PC3 7.301414e-17 2.779673e-16 3.821566e-16 3.875378e-16 -
1.408959e-16
## PC4 -6.417615e-17 -2.281366e-16 -9.643651e-18 4.605095e-17
1.733673e-16
## PC5 -1.528041e-16 -1.108037e-16 6.128103e-16 5.418662e-16 -
8.124243e-16
## PC6 1.000000e+00 4.184255e-16 8.128189e-16 3.700232e-16

```

9.527234e-17

## PC7	4.184255e-16	1.000000e+00	9.981396e-16	6.751393e-16	-
	2.858645e-16				
## PC8	8.128189e-16	9.981396e-16	1.000000e+00	7.862821e-16	
	2.841017e-16				
## PC9	3.700232e-16	6.751393e-16	7.862821e-16	1.000000e+00	-
	4.474140e-16				
## PC10	9.527234e-17	-2.858645e-16	2.841017e-16	-4.474140e-16	
	1.000000e+00				
## PC11	-6.452539e-17	-2.000607e-16	-1.918607e-16	3.954354e-16	-
	4.404403e-17				
## PC12	1.251885e-16	-1.819953e-16	-1.529849e-16	3.989890e-16	-
	2.930173e-16				
## PC13	1.806565e-17	-4.789007e-16	-1.499627e-16	5.193192e-16	
	4.542846e-16				
## PC14	1.906842e-16	5.664773e-16	-6.503572e-16	-2.171925e-16	-
	2.684432e-16				
## PC15	4.284851e-16	1.547793e-16	-2.735846e-17	5.387744e-16	
	3.343049e-16				
## PC16	2.516913e-16	1.674566e-16	-2.623156e-16	2.293639e-16	-
	2.783654e-16				
## PC17	-3.343316e-16	4.637099e-16	2.361346e-16	1.955471e-16	-
	1.976452e-16				
## PC18	1.550556e-16	-2.733128e-16	-6.080725e-16	-9.596218e-16	
	3.692201e-16				
## PC19	-3.839501e-16	4.861351e-16	1.287601e-15	-2.439327e-16	
	2.272585e-16				
## PC20	-9.225192e-16	-2.249042e-15	-4.140883e-15	5.395753e-15	-
	1.239754e-17				
## PC21	1.572324e-01	1.893515e-01	2.805242e-02	-1.032117e-01	
	1.135708e-01				

Analisis

PCA dirancang untuk **mengubah variabel-variabel yang saling berkorelasi menjadi kumpulan komponen baru yang saling tidak berkorelasi (orthogonal) satu sama lain. Ini sangat penting untuk:

- 1) Menghindari multikolinearitas dalam analisis regresi.
- 2) Mengurangi dimensi data tanpa redundansi informasi.
- 3) Menyederhanakan struktur data.

Sebagian besar nilai korelasi antara PC (PC1 sampai PC20) adalah mendekati nol. Ini berarti:

- 1) Setiap komponen utama membawa informasi baru yang tidak tumpang tindih dengan informasi di komponen lain.
- 2) PCA berhasil mendekomposisi struktur hubungan antar variabel menjadi bentuk yang lebih sederhana dan orthogonal

8. Korelasi Gabungan Data Asli dan PCA

Input dan Output		
# 7. Korelasi gabungan data asli dan PCA		
cor(cbind(Tingkat_kemiskinanX, as.data.frame(Tingkat_kemiskinanX.pca \$x)))		
##	Tingkat_kemiskinan_kota	G
ini_ratio		
## Tingkat_kemiskinan_kota	1.0000000000	0.3
564929009		
## Gini_ratio	0.3564929009	1.0
0000000000		
## Tingkat_penyelesaian_SD	0.1951784445	0.3
314250554		
## Tingkat_penyelesaian_SMP	0.1967813961	0.2
177041089		
## Tingkat_penyelesaian_SMA	-0.0433929627	0.2
940521249		
## Buta_huruf_15	0.1635950403	0.0
689331848		
## Buta_huruf_15_minus	-0.0510411825	-0.1
731033276		
## Buta_huruf_45	0.2698378755	0.1
659693449		
## Anak_bekerja	-0.0543676943	0.0
379732955		
## Tingkat_setengah_pengangguran	0.4668990967	-0.0
468657096		
## Tenaga_kerja_formal	-0.4263013100	-0.0
533403170		
## Kerja_informal_pertanian	0.4356784340	0.3
504907602		
## Lapangan_kerja_informal	0.4263013100	0.0
533403170		
## Upah_rata2_perjam	-0.4813736044	-0.1
868369396		
## Konsumsi_kalori_perhari	0.2655383061	0.3
885380702		
## Konsumsi_protein_perhari	0.0899933536	0.2
944723094		
## GK_perkotaan	-0.2018854879	-0.6
394452284		
## GK_non_makanan_kota	-0.3139965243	-0.5
745380032		
## Tingkat_kerentanan_penduduk	0.1101021696	-0.2
784936692		
## kepemilikan_akta_40kebawah	0.0366189014	0.3
198942128		
## Median_pendapatan_penduduk_dibawah40	0.3424925520	0.4

662692868		
## PC1	-0.2803321678	-0.0
328017041		
## PC2	-0.5182739120	-0.6
963890491		
## PC3	0.1136222151	-0.3
943796185		
## PC4	0.5747514943	-0.1
120150996		
## PC5	0.3082632024	0.1
996024893		
## PC6	-0.0292963089	0.4
207802403		
## PC7	-0.0300509709	0.0
558399278		
## PC8	0.4285667887	0.1
704883537		
## PC9	0.0669196869	-0.2
523365939		
## PC10	-0.1146987832	-0.0
265952569		
## PC11	-0.0417297820	0.0
919676878		
## PC12	-0.0219638345	0.1
365357791		
## PC13	-0.0754441020	0.0
252607931		
## PC14	-0.0341895945	0.0
256117861		
## PC15	-0.0164355251	-0.0
259347204		
## PC16	-0.0080395543	-0.0
132071684		
## PC17	-0.0005478810	-0.0
292758825		
## PC18	-0.0408894645	0.0
373748205		
## PC19	0.0193280156	-0.0
304275792		
## PC20	-0.0007405406	0.0
003043359		
## PC21	-0.4981032691	-0.1
305373673		
##	Tingkat_penyelesaian_SD	
Analisis		
Korelasi PC1		

Variabel Asli	Korelasi dengan PC1
Upah_rata2_perjam	-0.481
Tenaga_kerja_formal	-0.426
Lapangan_kerja_informal	+0.426
Kerja_informal_pertanian	+0.436
Tingkat_setengah_pengangguran	+0.467

Interpretasi PC1:

PC1 merepresentasikan struktur tenaga kerja:

1. Nilai positif terkait sektor informal dan pertanian serta pengangguran terselubung.
2. Nilai negatif terkait pekerja formal dan upah tinggi.

PC1 = Indikator Informalisasi dan Kerentanan Ekonomi.

1. Kota dengan nilai PC1 tinggi → lebih banyak pekerjaan informal dan pengangguran terselubung.
2. Kota dengan nilai PC1 rendah → lebih banyak pekerjaan formal dan upah lebih tinggi.

Korelasi PC2

Variabel Asli	Korelasi dengan PC2
Gini_ratio	-0.696
GK_perkotaan	-0.639
GK_non_makanan_kota	-0.574
Median_pendapatan_penduduk_dibawah40	+0.466

Interpretasi PC2:

PC2 mencerminkan ketimpangan dan kemampuan konsumsi rumah tangga:

1. Nilai negatif tinggi → kota dengan ketimpangan tinggi dan pengeluaran rumah tangga tinggi.
2. Nilai positif tinggi → kota dengan pendapatan median lebih baik di kelompok bawah.

PC2 = Indikator Ketimpangan Sosial & Daya Beli.

1. PC2 rendah → ketimpangan tinggi dan konsumsi tinggi (kota besar, biaya hidup tinggi).
2. PC2 tinggi → pendapatan warga miskin relatif lebih merata (kota kecil atau lebih setara).

Korelasi PC3

Variabel Asli	Korelasi dengan PC3
Gini_ratio	-0.394
Konsumsi_protein_perhari	+0.295

Interpretasi PC3:

PC3 terkait dengan ketimpangan dan pola konsumsi gizi.

PC3 = Indikator Gizi dan Distribusi Konsumsi.

1. Kota dengan nilai PC3 tinggi → konsumsi protein baik dan ketimpangan lebih rendah.
2. PC3 rendah → ketimpangan masih tinggi, konsumsi protein mungkin terbatas.

Korelasi PC4

Variabel Asli	Korelasi dengan PC4
Tingkat_kemiskinan_kota	+0.575

Interpretasi PC4:

PC4 sangat erat hubungannya dengan tingkat kemiskinan itu sendiri.

PC4 = Representasi Langsung Kemiskinan.

1. Nilai PC4 tinggi → kota dengan tingkat kemiskinan tinggi.

9. Korelasi Target dengan PCA

Input dan Output					
# 8. Korelasi target dengan PCA					
<code>pca_df <- as.data.frame(Tingkat_kemiskinanX.pca\$x)</code>					
<code>target_numerik <- Tingkat_kemiskinan_kota\$Tingkat_kemiskinan</code>					
<code>data_for_correlation <- cbind(target_numerik, pca_df)</code>					
<code>correlation_matrix <- cor(data_for_correlation)</code>					
<code>print(correlation_matrix[,1])</code>					
## target_numerik	PC1	PC2	PC3		
PC4					
## 1.0000000000	-0.2803321678	-0.5182739120	0.1136222151	0.5	
747514943					
## PC5	PC6	PC7	PC8		
PC9					
## 0.3082632024	-0.0292963089	-0.0300509709	0.4285667887	0.0	
669196869					
## PC10	PC11	PC12	PC13		
PC14					
## -0.1146987832	-0.0417297820	-0.0219638345	-0.0754441020	-0.0	
341895945					
## PC15	PC16	PC17	PC18		
PC19					
## -0.0164355251	-0.0080395543	-0.0005478810	-0.0408894645	0.0	
193280156					

##	PC20	PC21
##	-0.0007405406	-0.4981032691
Analisis		
Interpretasi		
1. PC4 adalah yang paling positif berkorelasi dengan tingkat kemiskinan ($r = 0.575$) <ul style="list-style-type: none"> a) Artinya: Semakin tinggi nilai PC4 → semakin tinggi tingkat kemiskinan suatu kota. b) Variabel-variabel asli yang punya korelasi tinggi dengan PC4 (lihat korelasi gabungan sebelumnya) bisa dianggap faktor utama yang memengaruhi kemiskinan. 		
2. PC2 dan PC21 juga penting, namun korelasinya negatif <ul style="list-style-type: none"> a) PC2 ($r = -0.518$): Jika nilai PC2 tinggi, tingkat kemiskinan justru rendah. b) PC21 ($r = -0.498$): Meskipun ini komponen dengan varian sangat kecil, dia masih punya hubungan cukup kuat secara korelasi. Tapi secara statistik, kontribusinya ke total variansi sangat kecil, jadi kurang relevan secara umum. 		
3. PC1 (komponen terbesar secara variasi) hanya korelasi rendah-sedang dengan kemiskinan ($r = -0.28$) <ul style="list-style-type: none"> a) Artinya: PC1 bukan faktor utama penentu kemiskinan. b) Jadi, komponen yang paling banyak menjelaskan variasi data belum tentu berkaitan erat dengan target (kemiskinan). 		

10. Regresi Semua Komponen Utama

Input dan Output	
<pre># 9. Regresi semua komponen utama Tingkat_kemiskinan.pca <- data.frame(target_numerik, pca_df) Tingkat_kemiskinan.pcr1 <- lm(target_numerik ~ ., data = Tingkat_kemiskinan.pca) summary(Tingkat_kemiskinan.pcr1) ## Warning in summary.lm(Tingkat_kemiskinan.pcr1): essentially perfect fit: ## summary may be unreliable ## ## Call: ## lm(formula = target_numerik ~ ., data = Tingkat_kemiskinan.pca) ## ## Residuals: ## Min 1Q Median 3Q Max ## -3.719e-15 -5.609e-16 -2.790e-17 6.720e-16 3.524e-15 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 7.267e+00 1.358e-15 5.350e+15 <2e-16 *** ## PC1 -2.794e-01 1.505e-16 -1.857e+15 <2e-16 *** ## PC2 -6.658e-01 2.651e-16 -2.512e+15 <2e-16 ***</pre>	

## PC3	2.151e-01	2.922e-16	7.359e+14	<2e-16	***
## PC4	1.335e+00	3.173e-16	4.207e+15	<2e-16	***
## PC5	8.845e-01	3.936e-16	2.247e+15	<2e-16	***
## PC6	-9.057e-02	4.335e-16	-2.089e+14	<2e-16	***
## PC7	-1.075e-01	5.140e-16	-2.092e+14	<2e-16	***
## PC8	1.689e+00	5.227e-16	3.232e+15	<2e-16	***
## PC9	3.143e-01	6.378e-16	4.928e+14	<2e-16	***
## PC10	-6.743e-01	8.025e-16	-8.402e+14	<2e-16	***
## PC11	-2.732e-01	8.756e-16	-3.120e+14	<2e-16	***
## PC12	-1.515e-01	1.246e-15	-1.215e+14	<2e-16	***
## PC13	-6.107e-01	1.087e-15	-5.620e+14	<2e-16	***
## PC14	-3.660e-01	1.443e-15	-2.536e+14	<2e-16	***
## PC15	-1.813e-01	1.477e-15	-1.227e+14	<2e-16	***
## PC16	-1.125e-01	2.047e-15	-5.497e+13	<2e-16	***
## PC17	-8.688e-03	2.114e-15	-4.110e+12	<2e-16	***
## PC18	-8.117e-01	2.655e-15	-3.057e+14	<2e-16	***
## PC19	4.161e-01	2.849e-15	1.460e+14	<2e-16	***
## PC20	-6.798e-02	1.255e-14	-5.415e+12	<2e-16	***
## PC21	-3.322e-01	3.031e+00	-1.100e-01	0.915	
## ---					
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
##					
## Residual standard error: 2.157e-15 on 12 degrees of freedom					
## Multiple R-squared: 1, Adjusted R-squared: 1					
## F-statistic: 2.718e+30 on 21 and 12 DF, p-value: < 2.2e-16					
Analisis					
Interpretasi					
1. Model terlalu sempurna → overfit					
a) Model ini bukan model prediksi yang sehat karena performa sempurna biasanya hanya terjadi saat:					
1) Jumlah prediktor terlalu banyak relatif terhadap jumlah observasi					
2) Tidak ada noise (sangat jarang dalam data sosial ekonomi nyata)					
2. Beberapa PC sangat relevan					
a) PC4 dan PC8 paling besar kontribusinya positif terhadap peningkatan tingkat kemiskinan					
b) PC2 berkontribusi menurunkan tingkat kemiskinan					
c) Ini sejalan dengan hasil korelasi sebelumnya					
3. PC21 tidak signifikan					
a) Koefisien tidak signifikan secara statistik → bisa diabaikan dalam pemodelan selanjutnya					

11. Koefisien Asli dari PCA

Input dan Output
<pre># 10. Koefisien asli dari PCA betas1 <- Tingkat_kemiskinanX.pca\$rotation %*% Tingkat_kemiskinan.pc</pre>

```

r1$coefficients[-1]
print(betas1)

##                                [,1]
## Tingkat_kemiskinan_kota      2.836399e+00
## Gini_ratio                    2.959312e-16
## Tingkat_penyelesaian_SD      2.026697e-16
## Tingkat_penyelesaian_SMP     -2.444946e-15
## Tingkat_penyelesaian_SMA     2.526883e-16
## Buta_huruf_15                 2.132202e-15
## Buta_huruf_15_minus          -1.994193e-15
## Buta_huruf_45                 -9.338428e-16
## Anak_bekerja                  4.501880e-16
## Tingkat_setengah_pengangguran -9.753150e-16
## Tenaga_kerja_formal           -2.348954e-01
## Kerja_informal_pertanian       4.642094e-16
## Lapangan_kerja_informal       -2.348954e-01
## Upah_rata2_perjam              4.950267e-16
## Konsumsi_kalori_perhari        -1.611199e-15
## Konsumsi_protein_perhari       2.049401e-15
## GK_perkotaan                  -1.776049e-15
## GK_non_makanan_kota            1.877558e-15
## Tingkat_kerentanan_penduduk   1.154633e-15
## kepemilikan_akta_40kebawah     1.374341e-16
## Median_pendapatan_penduduk_dibawah40 -1.370994e-15

```

Analisis

Variabel Asli	Koefisien Asli	Interpretasi
Tingkat_kemiskinan_kota	+2.836	Ini adalah target/variabel dependen, hasilnya hanya mencerminkan diri sendiri
Tenaga_kerja_formal	-0.235	Makin tinggi pekerja formal, makin rendah tingkat kemiskinan (logis)
Lapangan_kerja_informal	-0.235	Makin besar informal → cenderung naikan kemiskinan
Lainnya	~0 (mendekati nol)	Koefisien sangat kecil → tidak signifikan atau berkontribusi minimal

- a) Hanya beberapa variabel asli yang berdampak signifikan
 - 1) Hanya Tenaga_kerja_formal dan Lapangan_kerja_informal yang memiliki koefisien besar dan berarti.
 - 2) Ini mengindikasikan bahwa struktur pasar kerja adalah penentu utama dalam variasi tingkat kemiskinan antar kota.
- b) Sebagian besar variabel memiliki koefisien nyaris nol
 - 1) Misalnya: Gini_ratio, Tingkat_penyelesaian_SMP, Konsumsi_kalori, dll.

- 2) Kemungkinan besar, kontribusinya dalam menjelaskan variasi kemiskinan diambil alih oleh komponen lain dalam PCA atau memang tidak cukup kuat secara langsung.
- c) Redundansi antara beberapa variabel
 - 1) Tenaga_kerja_formal dan Lapangan_kerja_informal punya nilai sama tapi tanda berlawanan, artinya sangat berkorelasi (mungkin secara struktur saling melengkapi).

12. Regresi hanya PC Signifikan

Input dan Output						
# 11. Regresi hanya PC signifikan						
Tingkat_kemiskinan.pcr2 <- lm(target_numerik ~ PC1 + PC2 + PC4 + PC5 + PC8, data = Tingkat_kemiskinan.pca)						
summary(Tingkat_kemiskinan.pcr2)						
Call:						
lm(formula = target_numerik ~ PC1 + PC2 + PC4 + PC5 + PC8, data = Tingkat_kemiskinan.pca)						
Residuals:						
Min	1Q	Median	3Q	Max		
-1.18125	-0.38023	-0.07225	0.30102	1.21862		
Coefficients:						
	Estimate	Std. Error	t value	Pr(> t)		
(Intercept)	7.26706	0.11048	65.775	< 2e-16	***	
PC1	-0.27942	0.03941	-7.090	1.03e-07	***	
PC2	-0.66581	0.05079	-13.108	1.80e-13	***	
PC4	1.33488	0.09183	14.537	1.43e-14	***	
PC5	0.88453	0.11345	7.797	1.71e-08	***	
PC8	1.68919	0.15584	10.839	1.58e-11	***	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
Residual standard error: 0.6442 on 28 degrees of freedom						
Multiple R-squared: 0.9562, Adjusted R-squared: 0.9484						
F-statistic: 122.3 on 5 and 28 DF, p-value: < 2.2e-16						
Analisis						
Kekuatan Model						
1) $R^2 = 0.9562 \rightarrow$ Model menjelaskan 95.6% variasi kemiskinan hanya dengan 5 komponen.						
2) Adjusted $R^2 = 0.9484 \rightarrow$ Tetap tinggi meski model lebih sederhana.						
3) Residual standard error (RSE) = 0.64 \rightarrow Galat prediksi cukup kecil dan realistis.						
Tingkat kemiskinan kota-kota dapat dijelaskan secara akurat hanya dengan 5 pola utama (komponen PCA) yang merupakan gabungan dari banyak faktor asli seperti pendidikan, kerja formal/informal, gaji, dan pengeluaran						
Keterangan:						
1) PC2 & PC1 negatif \rightarrow bisa dikaitkan dengan pendidikan, tenaga kerja formal, atau upah (faktor penurunan kemiskinan).						

- 2) PC4, PC5, PC8 positif → kemungkinan mewakili kerja informal, anak bekerja, atau kerentanan (peningkatan kemiskinan).

13. Perbandingan Model

Input dan Output

```
{r}  
# 12. Perbandingan model  
anova(Tingkat_kemiskinan.pcr1, Tingkat_kemiskinan.pcr2)
```

Analysis of Variance Table

Model 1: target_numerik ~ PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 +
PC9 + PC10 + PC11 + PC12 + PC13 + PC14 + PC15 + PC16 + PC17 +
PC18 + PC19 + PC20 + PC21

Model 2: target_numerik ~ PC1 + PC2 + PC4 + PC5 + PC8

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	12	0.000				
2	28	11.621	-16	-11.621	1.5616e+29	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Analisis

Ringkasan Model

- 1) Model 1 menggunakan 21 komponen utama (PC1 sampai PC21) sebagai prediktor.
- 2) Model 2 menggunakan hanya 5 komponen utama terpilih (PC1, PC2, PC4, PC5, PC8).

Hasil dan Kesimpulan

- 1) Jika p-value ($\text{Pr}(>F)$) sangat kecil (misalnya < 0.05), berarti Model 1 secara signifikan lebih baik daripada Model 2.
- 2) Jika p-value besar (> 0.05), artinya penambahan komponen pada Model 1 tidak memberikan peningkatan signifikan terhadap fit model, sehingga Model 2 yang lebih sederhana sudah cukup.

BAB 5 KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini menunjukkan bahwa regresi linier berganda dapat digunakan untuk mengidentifikasi variabel-variabel sosial ekonomi yang memengaruhi tingkat kemiskinan antar kota. Namun, permasalahan multikolinearitas cukup mengganggu interpretasi model tersebut. Dengan menerapkan Principal Component Regression (PCR), masalah ini dapat diatasi melalui transformasi ke komponen utama yang saling orthogonal.

PCR memberikan model prediktif yang lebih stabil dan mampu mengurangi kompleksitas tanpa kehilangan terlalu banyak informasi. Hasil analisis juga memperkuat pentingnya pendidikan dan distribusi pendapatan sebagai determinan utama kemiskinan. Metode best subset regression membantu dalam pemilihan kombinasi variabel terbaik secara statistik.

5.2. Saran

- 1) **Bagi Peneliti:** Penelitian selanjutnya dapat memperluas pendekatan dengan menggunakan data time series atau panel untuk melihat dinamika kemiskinan dari waktu ke waktu.
- 2) **Bagi Pemerintah Daerah:** Hasil penelitian ini dapat dijadikan sebagai dasar untuk mengarahkan program-program pengurangan kemiskinan pada faktor-faktor dominan seperti pendidikan dan ketimpangan pendapatan.
- 3) **Bagi Akademisi:** Studi ini memberikan contoh konkret bagaimana metode statistik lanjutan seperti PCR dapat diterapkan dalam isu-isu sosial ekonomi yang kompleks.

DAFTAR PUSTAKA

- Badan Pusat Statistik (2023) *Indikator Kesejahteraan Rakyat 2022*. Jakarta: BPS. Available at: <https://www.bps.go.id/publication> (Accessed: 26 May 2025).
- Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2019) *Multivariate Data Analysis*. 8th edn. Hampshire: Cengage Learning.
- Jolliffe, I.T. and Cadima, J. (2016) 'Principal component analysis: A review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Article 20150202. doi: 10.1098/rsta.2015.0202.
- Kartika, D. and Firdaus, R. (2018) 'Analisis Pengaruh Pengangguran terhadap Kesejahteraan Masyarakat di Indonesia', *Jurnal Ekonomi dan Pembangunan Indonesia*, 19(1), pp. 45–54.
- Prasetyo, A.P. and Nugroho, Y. (2020) 'Pengaruh Tingkat Pendidikan dan Infrastruktur terhadap Kemiskinan Antarwilayah di Indonesia', *Jurnal Sosial Ekonomi Pembangunan*, 10(2), pp. 123–135.
- Syaraniafa (no date) *Poverty Analysis in Indonesia using Ridge and PCR: datafinal.xlsx*. GitHub. Available at: <https://github.com/syaraniafa/Poverty-Analysis-in-Indonesia-using-Ridge-and-PCR/commits/main/datafinal.xlsx> (Accessed: 27 May 2025).
- United Nations Development Programme (UNDP) (2020) *Human Development Report 2020: The Next Frontier—Human Development and the Anthropocene*. New York: UNDP.
- Widarjono, A. (2016) *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*. 4th edn. Yogyakarta: UPP STIM YKPN.

LAMPIRAN

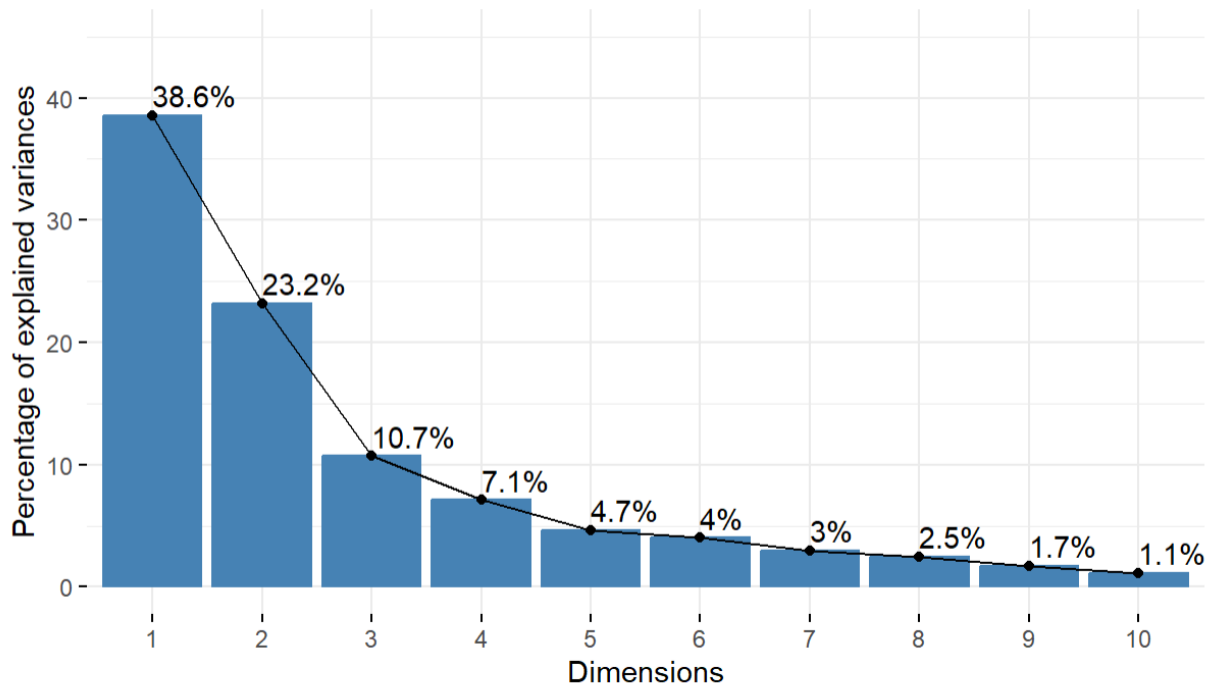
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Provinsi	Tingkat_ke Gini_ratio	Tingkat_pe	Tingkat_pe	Tingkat_pe	Tingkat_pe	Buta_huruf	Buta_huruf	Buta_huruf	Anak_beke	Tingkat_se	Tenaga_ke	Kerja_infor	Lapangan	Upah_rata	Konsumsi	Konsumsi	GK_perkot	GK_non_m	Tingkat_ke	kepemilikan	Median_pendapatan	penduduk_dibawah40		
2	ACEH	10.35	0.327	99.45	97.63	70.67	1.75	0.13	4.96	1.01	10.46	38.55	44.13	61.45	16772	2064.3	61.16659	646232	164956	4.69	95.12	7.35			
3	SUMATERA	8.63	0.358	98.74	92.84	77.16	0.89	0.24	2.11	4.02	6.83	40.93	46.45	59.07	15131	2123.413	64.65301	615227	157137	6.53	81.5	4.61			
4	SUMATERA	4.9	0.316	97.87	88.83	65.96	0.71	0.12	1.7	2.48	8.68	34.74	52.38	65.26	15887	2109.426	59.24327	668643	178321	5.24	91.73	1.55			
5	RIAU	6.49	0.35	98.2	88.53	66.91	0.82	0.07	2.42	2.04	7.42	45.98	44.3	54.02	18626	2022.049	59.69249	686945	203815	5.96	84.24	2.6			
6	JAMBI	10.48	0.376	97.62	86.57	65.85	1.9	0.18	5.14	1.57	7.46	40.02	44.76	59.98	16042	2069.511	60.70195	684555	166816	3.75	92.55	4.8			
7	SUMATERA	11.37	0.348	97.53	88.41	67.07	1.35	0.27	3.28	2.72	7.08	36.72	45.1	63.28	15978	2182.468	65.18084	548455	148101	4.59	90.7	11.99			
8	BENGKULU	14.53	0.371	97.65	90.81	64.88	2.2	0.25	5.87	1.79	8.02	32.46	47.72	67.54	16501	2037.552	60.29307	686223	178155	3.84	93.29	6.55			
9	LAMPUNG	8.34	0.352	98.41	90.99	62.42	2.75	0.28	6.91	2.48	6.92	28.49	55.67	71.51	13218	2002.723	57.6745	595992	156844	6.21	93.28	10.15			
10	KEP. BANG	3.48	0.263	96.45	84.72	66.87	1.83	0.42	4.52	2.97	3.37	50.37	43.49	49.63	18132	2065.864	67.57296	836425	231249	2.16	95.91	0.1			
11	KEP. RIAU	5.46	0.328	98.38	95.72	73.93	0.98	0.15	2.85	1.87	4.95	63.68	32.57	36.32	23528	2019.871	64.66545	733248	247232	4.45	94.7	1.19			
12	DRI JAKART	4.01	0.412	98.58	95.4	87.71	0.31	0.08	0.68	0.61	2.71	63.12	36.82	36.88	32685	2116.865	68.15338	773370	238551	3.77	96.88	0.75			
13	JAWA BARU	7.52	0.437	99.08	89.29	67.05	1.51	0.13	3.93	1.82	5.62	45.39	48.57	54.61	19038	2115.695	63.62161	480785	127531	5.18	83.36	10.58			
14	JAWA TENI	10.02	0.392	98.01	90.02	58.75	5.74	0.74	12.78	2.41	5.39	39.84	49.99	60.16	12604	2017.64	59.78365	469623	119009	3	95.55	16.14			
15	DI YOGYAK	10.64	0.468	98.91	97.06	87.92	4.85	0.22	10.34	1.69	4.76	46.62	44.19	53.38	14916	2024.443	62.64922	573915	161094	2.69	97.36	10.9			
16	JAWA TIMI	7.78	0.381	98.71	90.47	66.87	6.68	0.94	14.28	1.51	5.77	36.14	50.73	63.86	15119	2041.689	59.96401	499769	125139	3.16	88.68	13.18			
17	BANTEN	5.89	0.384	97.24	92.65	66.02	1.84	0.33	4.86	1.9	4.57	52.04	42.5	47.96	23880	2270.482	69.31467	617527	172532	5.62	81.53	3.59			
18	BAU	4.12	0.371	97.55	94.15	76.59	4.47	0.4	9.92	4.04	3.93	46.57	43.45	53.43	16857	2145.713	61.36875	530820	163784	4.04	94.62	5.47			
19	NUSA TENI	13.98	0.392	98.47	95.39	61	11.03	2.79	26.75	4	13.3	24.64	63.17	75.36	11734	2460.435	74.81875	501586	128090	7.86	84.91	15.92			
20	NUSA TENI	9	0.308	92.35	83.25	38.47	5.37	1.75	12.3	5.12	12.5	24.76	52.98	75.24	13012	1974.124	56.36387	594056	170478	14.48	62.22	21.79			
21	KALIMANT	4.63	0.334	95.13	81.82	58.4	6.04	1.46	14.92	3.19	7.21	41.23	41.39	58.77	16337	1892.262	58.18551	592432	146772	6.15	85.91	5.51			
22	KALIMANT	5.09	0.332	98.51	87.79	61.88	0.88	0.08	2.5	2.31	5.38	48.5	40.98	51.5	19795	2116.053	65.00665	557200	140978	3.47	85.62	3.31			
23	KALIMANT	4.03	0.329	94.68	87.95	67.81	1.64	0.24	4.2	2.06	6.76	42.7	43.37	57.3	18661	2188.804	68.00869	588472	189362	3.19	90.22	2.44			
24	KALIMANT	4.97	0.32	99.18	95.32	74	1.03	0.24	2.59	1.79	4.81	55.74	36.54	44.26	22281	1918.335	63.00287	772112	230245	3.46	93.76	0.6			
25	KALIMANT	5.58	0.256	96.94	90.55	54.8	2.29	0.77	5.48	4.15	6.19	51	33.73	49	20320	1857.917	59.73577	834774	217984	6.57	93.13	0.24			
26	SULAWESI	5.04	0.355	96.74	91.98	66.66	0.19	0.07	0.34	1.72	6.27	40.85	49.43	59.15	18335	2154.558	68.06158	460738	103035	5.4	89.69	12.89			
27	SULAWESI	9.13	0.337	97.19	88.9	53.73	1.95	0.58	4.57	3.96	7.62	32.13	40.37	67.87	15890	2001.122	58.12426	565994	153690	7.21	82.82	9.9			
28	SULAWESI	4.98	0.378	98.05	90.55	68.32	6.69	1.77	15.08	4.32	6.23	36.45	44.89	63.55	17257	2077.921	64.1767	443717	127538	3.78	90.33	16.75			
29	SULAWESI	7.22	0.382	97.24	91.19	65.97	4.21	0.82	11.21	4.11	8.14	37.61	45.98	62.39	17542	1948.634	60.39038	451180	126476	4.42	87.39	20.86			
30	GORONTALU	4.49	0.406	95.12	80.56	45.12	1.58	0.6	3.41	6.78	6.1	35.03	52.03	64.97	14289	1958.11	57.09797	436651	115590	5.15	92.18	22.59			
31	SULAWESI	9.33	0.472	97.15	84.14	55.18	6.18	1.85	15.4	6.31	13.04	22.75	57.15	77.25	14774	2082.994	61.40084	428974	94951	7.13	91.74	23.2			
32	MALUKU	5.9	0.309	98.98	93.65	72.08	0.6	0.31	1.19	2.87	11.8	36.7	48.04	63.3	17006	1836.682	52.90789	682707	187267	11.18	83.43	5.21			
33	MALUKU L	6.17	0.33	97.72	94.92	67.1	1.22	0.16	3.57	2.45	9.28	34.37	54.29	65.63	18278	1843.676	54.75539	573792	143872	10.28	84.71	4.88			
34	PAPUA BAR	7.64	0.302	93.94	87.03	57.07	2.36	0.98	5.65	2.82	12.29	39.72	38.02	60.28	24128	1879.753	55.32053	741776	206365	10.31	67.6	8.91			
35	PAPUA	5.29	0.294	81.99	66.16	39.01	18.81	15.09	28.35	4.5	5.85	15.89	45.83	84.11	24097	1902.043	45.06653	723346	230322	6.77	42.01	14.11			

```
{r}
# 1. Membaca dataset
Tingkat_kemiskinan_kota <-
read.csv("C:/Users/Lenovo/OneDrive/Documents/PENS/SEMESTER 2/Pemodelan
Statistik Terapan/TUGAS AKHIR/datafinal1.csv")
Tingkat_kemiskinan_kota

{r}
# Model OLS
model <- lm(Tingkat_kemiskinan_kota ~ ., data = Tingkat_kemiskinan_kota)
summary(model)

{r}
# 2. Menghapus kolom pertama (target)
Tingkat_kemiskinanX <- Tingkat_kemiskinan_kota[,-1]
Tingkat_kemiskinanX.pca <- prcomp(Tingkat_kemiskinanX, center = TRUE, scale.
= TRUE)
plot(Tingkat_kemiskinanX.pca, type="l")
```

Scree plot



```
{r}
# 4. Matriks rotasi
print(Tingkat_kemiskinanX.pca$rotation)
```

```
{r}
# 5. Nilai eigen dari matriks korelasi
eigen(cor(Tingkat_kemiskinanX))$values
diag(var(Tingkat_kemiskinanX.pca$x))
```

```
{r}
# 6. Korelasi antar komponen utama
cor(Tingkat_kemiskinanX.pca$x)
```

```
{r}
# 7. Korelasi gabungan data asli dan PCA
cor(cbind(Tingkat_kemiskinanX, as.data.frame(Tingkat_kemiskinanX.pca$x)))
```

```
{r}
# 8. Korelasi target dengan PCA
pca_df <- as.data.frame(Tingkat_kemiskinanX.pca$x)
target_numerik <- Tingkat_kemiskinan_kota$Tingkat_kemiskinan
data_for_correlation <- cbind(target_numerik, pca_df)
correlation_matrix <- cor(data_for_correlation)
print(correlation_matrix[,1])
```

```
```{r}  
9. Regresi semua komponen utama
Tingkat_kemiskinan.pca <- data.frame(target_numerik, pca_df)
Tingkat_kemiskinan.pcr1 <- lm(target_numerik ~ ., data =
Tingkat_kemiskinan.pca)
summary(Tingkat_kemiskinan.pcr1)
```
```

```
```{r}  
10. Koefisien asli dari PCA
betas1 <- Tingkat_kemiskinanX.pca$rotation %*%
Tingkat_kemiskinan.pcr1$coefficients[-1]
print(betas1)
```
```

```
```{r}  
11. Regresi hanya PC signifikan
Tingkat_kemiskinan.pcr2 <- lm(target_numerik ~ PC1 + PC2 + PC4 + PC5 + PC8,
data = Tingkat_kemiskinan.pca)
summary(Tingkat_kemiskinan.pcr2)
```
```

```
```{r}  
12. Perbandingan model
anova(Tingkat_kemiskinan.pcr1, Tingkat_kemiskinan.pcr2)
```
```