

Penerapan Metode *Principal Component Analysis* (Studi Kasus: Tingkat Kemiskinan di Kepulauan Maluku)

Trifena Punana Lesnussa^{1*}, Meidy Kaseside², Everd Elseos Martin Utubira³
^{1,2,3} Fakultas Ilmu Alam dan Teknologi Rekayasa, Program Studi Matematika, Universitas Halmahera, Indonesia

Alamat Kampus: Jl. Raya Wari, Wari Ino, Kec. Tobelo, Kabupaten Halmahera Utara, Maluku Utara

*Korespondensi penulis: trifenapunanalesnussa@gmail.com

Abstract. *The Maluku Islands are a high-poverty region in Indonesia. The region consists of 2 provinces, namely Maluku province and North Maluku province. There are 21 districts/cities in the region, with 17 regencies and 4 municipalities. The poverty rate in this region is a challenge that always wants to be studied with a socio-population approach and a mathematical statistics approach. One method or approach in analyzing poverty is Principal Component Analysis (PCA). PCA has the advantage of simplifying information from various variables to several principal components without losing much information and can overcome the problem of multiple linearity by changing variables that correlate with freely related components. The purpose of this research is to identify poverty in districts/municipalities in Maluku Islands using the PCA approach. The results showed that the components formed by the PCA method were formed in 2 factors. Factor 1 consists of GRDP (X2), Life Expectancy Rate (X3), Unemployment Rate (X4) and Percentage of Population (X6). Meanwhile, factor 2 consists of 2 variables, namely the Poverty Level (X1) and TPAK (X5).*

Kata kunci: *Principal Component Analysis, factor analysis, poverty, Maluku islands.*

Abstrak. Kepulauan Maluku merupakan wilayah dengan kemiskinan tinggi di Indonesia. Wilayah ini terdiri dari 2 provinsi yaitu provinsi Maluku dan provinsi Maluku Utara. Kabupaten/Kota pada wilayah ini sebanyak 21 dengan 17 kabupaten dan 4 kota madya. Tingkat kemiskinan pada wilayah ini menjadi tantangan yang selalu ingin diteliti dengan pendekatan sosial kependudukan maupun pendekatan statistika matematika. Salah satu metode atau pendekatan dalam menganalisis kemiskinan adalah Principal Component Analysis (PCA). PCA memiliki keunggulan menyederhanakan informasi dari berbagai variabel ke beberapa komponen utama tanpa kehilangan banyak informasi dan dapat mengatasi masalah beberapa linearitas dengan mengubah variabel yang berkorelasi dengan komponen yang terkait secara bebas. Tujuan penelitian ini mengidentifikasi kemiskinan di kabupaten/kota di kepulauan Maluku dengan pendekatan PCA. Hasil penelitian menunjukkan bahwa komponen yang terbentuk dengan metode PCA terbentuk dalam 2 faktor. Faktor 1 terdiri dari PDRB (X2), Angka Harapan Hidup (X3), Tingkat Pengangguran (X4) dan Persentase penduduk (X6). Sedangkan untuk faktor 2 terdiri dari 2 variabel yakni Tingkat Kemiskinan(X1) dan TPAK (X5).

Kata kunci: Principal Component Analysis, analisis factor, kemiskinan, kepulauan Maluku.

1. LATAR BELAKANG

Salah satu tantangan utama dalam pembangunan ekonomi di dunia ini dengan karakteristik yang kompleks dan multidimensi adalah kemiskinan (Todaro & Smith, 2020). Untuk memahami dan mengatasi masalah kemiskinan ini membutuhkan analisis data yang efektif untuk mengidentifikasi pola, tren, dan faktor yang mempengaruhi tingkat kemiskinan untuk pengentasan kemiskinan (Bank Dunia, 2021). Data kemiskinan di suatu wilayah terdiri dari banyak variabel yang saling berkaitan, seperti tingkat pendidikan, Akses ke Layanan Kesehatan, Kondisi Kerja, dan Kualitas Perumahan.

Analisis Komponen Utama (PCA) adalah teknik yang digunakan untuk mengurangi ukuran data dan untuk mempertahankan sebagian besar informasi yang tersedia pada saat yang sama dalam statistik (Jolliffe & Cadima, 2016). PCA bertujuan memodifikasi pengumpulan variabel yang berkorelasi dengan beberapa komponen utama yang tidak berkorelasi, memfasilitasi interpretasi dan meningkatkan efektivitas visualisasi data. Dalam konteks penelitian kemiskinan, PCA dapat digunakan untuk mengidentifikasi indikator yang memberikan kontribusi terbesar bagi kemiskinan lokal, sehingga membantu memproduksi lebih banyak pedoman yang ditargetkan (Abdi & Williams, 2010).

Metode PCA memiliki beberapa keunggulan yang menjadikannya alat analitik yang efektif dalam penelitian kemiskinan. Pertama, PCA dapat menyederhanakan informasi dari berbagai variabel ke beberapa komponen utama tanpa kehilangan banyak informasi penting (Jolliffe, 2002). Kedua, PCA dapat mengatasi masalah beberapa linearitas dengan mengubah variabel yang berkorelasi dengan komponen yang terkait secara bebas. Ini meningkatkan keakuratan analisis regresi dan model statistik lainnya (Wold et al., 1987). Ketiga, PCA menyederhanakan strukturnya untuk meningkatkan pemahaman data dan dengan demikian memungkinkan identifikasi pola kemiskinan yang lebih jelas (Jolliffe & Cadima, 2016). Selain itu, PCA juga mendukung visualisasi data dengan mengurangi dimensi dan memungkinkan presentasi dalam bentuk dua atau tiga grafik dimensi yang lebih mudah dianalisis (Shrens, 2014). Akhirnya, metode ini membantu mengurangi data dengan menghilangkan variabel yang kurang relevan, pada akhirnya meningkatkan kualitas analisis dan mengurangi kemungkinan kesalahan estimasi (Jolliffe, 2002).

Banyak penelitian sebelumnya telah menerapkan PCA dalam penelitian sosial ekonomi, seperti mengukur sumur, mengklasifikasikan bidang berdasarkan tingkat kemiskinan, dan mencari tahu kemiskinan (Fieler & Pritchett, 2001). Namun, masih ada kemungkinan bahwa penggunaan PCA akan diperluas dalam analisis dataset yang lebih kompleks, termasuk yang termasuk faktor lingkungan, kemajuan teknologi dan kebijakan ekonomi. Oleh karena itu, dalam penelitian ini bertujuan mengimplementasikan PCA dalam analisis kemiskinan. Analisis faktor adalah teknik yang digunakan untuk menjelaskan hubungan atau korelasi antara berbagai variabel yang diamati. Prinsip dasar dari analisis faktor adalah menyederhanakan deskripsi data dengan mengurangi jumlah variabel atau dimensi yang ada. Salah satu metode dalam analisis faktor adalah Principal Component Analysis (PCA), di mana beberapa faktor yang terbentuk merupakan variabel laten yang belum ditentukan sebelumnya dan bersifat acak.

2. KAJIAN TEORITIS

Konsep dan Teknik Analisis PCA

Analisis Komponen Utama, atau PCA (Principal Component Analysis), merupakan metode statistik yang bertujuan mereduksi dimensi dari data yang besar. Metode ini menggabungkan variabel-variabel yang saling berkorelasi menjadi beberapa komponen utama yang simpel, tetapi tetap mampu mencakup sebagian besar variasi informasi dalam data. PCA sering digunakan untuk menemukan pola dalam data yang kompleks serta menganalisis hubungan antar variabel secara lebih mudah dipahami. Proses PCA dapat dijelaskan melalui beberapa langkah:

- 1) Standardisasi Data;
- 2) Matriks Kovarians atau Korelasi;
- 3) Eigenvalue dan Eigenvector;
- 4) Transformasi Data ke Ruang Komponen Utama.

PCA dalam Analisis Data Kemiskinan

Dalam konteks analisis kemiskinan, PCA sangat berguna untuk mereduksi dimensi data yang kompleks dan saling berkorelasi, seperti informasi mengenai pendapatan, pendidikan, akses layanan dasar, dan kondisi sosial-ekonomi lainnya. Ada beberapa alasan mengapa PCA bermanfaat dalam analisis data kemiskinan, antara lain:

- 1) Penggabungan Berbagai Variabel: Data kemiskinan sering kali mencakup banyak indikator. Dengan menggunakan PCA, indikator-indikator ini dapat digabungkan menjadi satu komponen utama yang lebih mudah dianalisis, tanpa kehilangan informasi yang penting.
- 2) Identifikasi Pola Kemiskinan: Dengan penerapan PCA, peneliti dapat mengidentifikasi pola atau karakteristik yang mendasari kemiskinan di suatu wilayah atau kelompok tertentu, yang mungkin tidak dapat dilihat melalui analisis univariat.
- 3) Reduksi Dimensi: Mengurangi dimensi data akan memudahkan dalam pemodelan statistik lebih lanjut, terutama ketika menghadapi dataset yang sangat besar dan kompleks.

Beberapa studi telah menggunakan PCA untuk menganalisis kemiskinan dengan berbagai tujuan. **Studi oleh Alonso et al. (2017):** Menggunakan PCA untuk menyederhanakan analisis multidimensi kemiskinan dan mengeksplorasi faktor-faktor penyebab kemiskinan di kawasan perkotaan dan pedesaan. **Studi oleh Ririhena et al (2020) :** pendekatan principal component analysis pada data demam berdarah dengue di provinsi

Maluku Utara. Studi-studi ini menunjukkan bahwa PCA sangat efektif dalam mengungkapkan faktor-faktor yang berkaitan dengan kasus social kependudukan, diantaranya kemiskinan yang tidak dapat dijelaskan hanya dengan satu indikator atau variabel.

Kelebihan dan Kekurangan PCA dalam Analisis Kemiskinan diantaranya (1). Efisiensi dalam Pengolahan Data: PCA memungkinkan analisis data yang kompleks menjadi lebih sederhana dengan mengurangi dimensi. (2). **Meningkatkan Pemahaman tentang Faktor Penyebab Kemiskinan:** Dengan mengidentifikasi komponen utama, PCA membantu peneliti untuk memahami variabel penting yang mempengaruhi kemiskinan secara lebih jelas. Namun PCA juga memiliki kekurangan diantaranya : (1). **kesulitan dalam Interpretasi Komponen** secara langsung, terutama jika data yang digunakan memiliki banyak variabel yang tidak selalu terkait secara intuitif. Kekurangan terakhir adalah (2). **asumsi normalitas**, PCA mengasumsikan bahwa data memiliki distribusi normal. Dalam beberapa kasus, data kemiskinan mungkin tidak memenuhi asumsi ini.

3. METODE PENELITIAN

Penelitian ini dilakukan dengan pendekatan kuantitatif berupa studi kasus untuk mengidentifikasi kemiskinan berdasarkan indikator ekonomi, sosial, dan demografis. Sumber Data diperoleh dari Badan Pusat Statistik Provinsi Maluku dan Maluku Utara. Variabel dan indikator dalam penelitian ini mencakup dimensi utama kemiskinan.

Tabel 1. Dimensi dan variabel penelitian

Dimensi	Variabel Penelitian
Ekonomi	Tingkat kemiskinan (x_1)
	Produk Domestik Regional Bruto (x_2)
	Tingkat pengangguran (x_4)
Pendidikan	Rata-rata Lama Sekolah (x_3)
Lapangan kerja	Tingkat partisipasi angkatan kerja (x_5)
	Persentase penduduk (x_6)

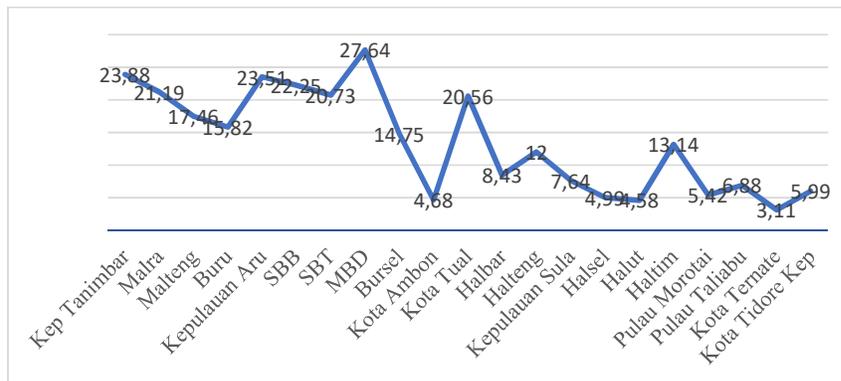
Metode analisis Principal Component Analysis dilakukan dengan mengikuti beberapa langkah penyelesaian sebagai berikut :

- 1) Standarisasi data dengan menghitung *mean* dan standar deviasi dari tiap-tiap fitur pada data
- 2) Pembentukan matriks kovarians atau korelasi
- 3) Menghitung eigenvalue dan eigenvector untuk menentukan komponen utama.
- 4) Transformasi original data dengan Eigen-vectors
- 5) Intrepretasi hasil dan kesimpulan

4. HASIL DAN PEMBAHASAN

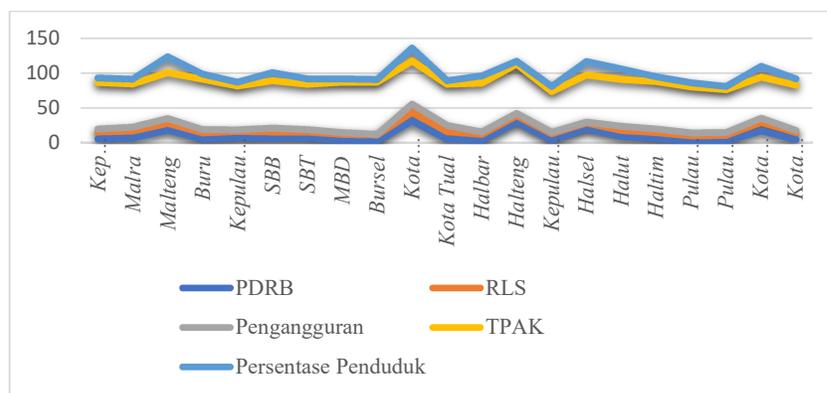
Statistika Deskriptif Kemiskinan di Kepulauan Maluku

Karakteristik kemiskinan di kepulauan Maluku secara nasional masih menempati wilayah yang dikategorikan sebagai wilayah dengan tingkat kemiskinan tinggi. Provinsi Maluku terdiri dari 11 kabupaten/kota sejak lama masuk dalam kategori provinsi termiskin di Indonesia dengan persentase 16,23% dan menempati posisi keempat pada tahun 2022, sedangkan untuk provinsi Maluku Utara sudah menunjukkan tren positif menuju pengentasan kemiskinan yang lebih.



Gambar 1. Tingkat Kemiskinan Tahun 2022

Berdasarkan Gambar 1, dapat dilihat bahwa untuk 11 kabupaten/kota yang berada di provinsi Maluku diperoleh rata-rata persentase kemiskinan sebesar 13,55%. Adapun kabupaten/kota yang menempati daerah dengan persentase kemiskinan tertinggi pada kabupaten Maluku Barat Daya (MBD), sedangkan untuk daerah dengan persentase kemiskinan terendah berada pada kota Ambon sebesar 4,68%. Untuk provinsi Maluku Utara, kabupaten Halmahera Timur (Haltim) menempati daerah dengan persentase kemiskinan tertinggi sebesar 13,14%, sedangkan persentase kemiskinan terendah sebesar 3,11% berada pada kota Ternate.



Gambar 2. Deskriptif variabel prediktor

Pada Gambar 2, dapat dilihat bahwa variabel PDRB mengalami fluktuatif untuk 21 kabupaten/kota di kepulauan Maluku. Daerah dengan PDRB tertinggi berada pada kota Ambon sebesar 31,53% dan terendah di kabupaten Pulau Morotai dengan persentase 2,47% serta memiliki nilai rata-rata sebesar 9,52%. Rata-rata Lama Sekolah (RLS) tertinggi berada pada kota Ambon dengan lamanya 12,21 tahun dan RLS terendah berada pada kabupaten Pulau Morotai dengan lamanya sekolah sebesar 7,42 tahun. Jika dirata-ratakan RLS secara wilayah kepulauan Maluku, diperoleh RLS sebesar 9,24 tahun, hal ini menyatakan bahwa RLS di wilayah kepulauan Maluku masih tergolong rendah.

Untuk tingkat pengangguran tertinggi berada pada kota Ambon dengan persentase 11,67% dan terendah pada kabupaten Buru Selatan. Adapun jika dirata-ratakan tingkat pengangguran di wilayah kepulauan Maluku maka diperoleh hasil sebesar 4,54%. Sedangkan untuk tingkat partisipasi angkatan kerja (TPAK) dapat dilihat bahwa persentase tertinggi berada pada kabupaten Buru Selatan sebesar 74,82% dan terendah pada kabupaten Kepulauan Sula. Rata-rata TPAK di wilayah kepulauan Maluku sebesar 65,96, ini menandakan bahwa TPAK di wilayah ini termasuk dalam kategori sedang. Variabel yang terakhir adalah persentase penduduk jika dirata-ratakan secara wilayah maka kepulauan Maluku memiliki persentase penduduk sebesar 9,52% yang terdiri dari 21 kabupaten/kota di wilayah ini. Kabupaten dengan persentase penduduk tertinggi berada di kabupaten Maluku Tengah dengan persentase 22,69% dan terendah sebesar 4,17 pada kabupaten Buru Selatan.

Penerapan PCA Pada Data Kemiskinan di Kepulauan Maluku

Dimensi dan variabel pada penelitian ini memiliki satuan yang berbeda maka langkah pertama adalah melakukan standarisasi data. Selanjutnya dilakukan pengujian KMO dan Bartlett's.

Tabel 2. Uji KMO dan Bartlett's

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,666
Bartlett's Test of Sphericity	Approx. Chi-Square	35,300
	Derajat bebas	15
	Signifikansi	0,002

Hasil yang diperoleh dari uji KMO sebesar $0,666 > 0,5$ dan Bartlett's sebesar $0,002 < 0,05$ maka semua variabel pada penelitian dapat dilakukan. Langkah setelah uji KMO dan Bartlett's adalah melakukan uji Anti Image Correlation.

Tabel 3. Uji Anti Image Correlation

Variabel	X1	X2	X3	X4	X5	X6
Tingkat Kemiskinan(X1)	0,508	0,227	-0,241	-0,056	-0,317	0,314
PDRB (X2)	0,227	0,663	-0,403	-0,019	-0,288	-0,360
Angka Harapan Hidup (X3)	-0,241	-0,403	0,641	-0,470	0,435	-0,078
Tingkat Pengangguran (X4)	-0,056	-0,019	-0,470	0,777	0,127	-0,167
TPAK (X5)	-0,317	-0,288	0,435	0,127	0,562	-0,049
Persentase penduduk (X6)	0,314	-0,360	-0,078	-0,167	-0,049	0,761

Tabel 3 diperoleh nilai korelasi antar variabel yang terdiri dari 6 variabel sangat kuat yakni $>0,5$ langkah selanjutnya yaitu uji *Communalities*.

Tabel 4. Uji *Communalities*

Variabel	Initial	Extraction
X1	1,000	0,575
X2	1,000	0,654
X3	1,000	0,819
X4	1,000	0,726
X5	1,000	0,513
X6	1,000	0,722

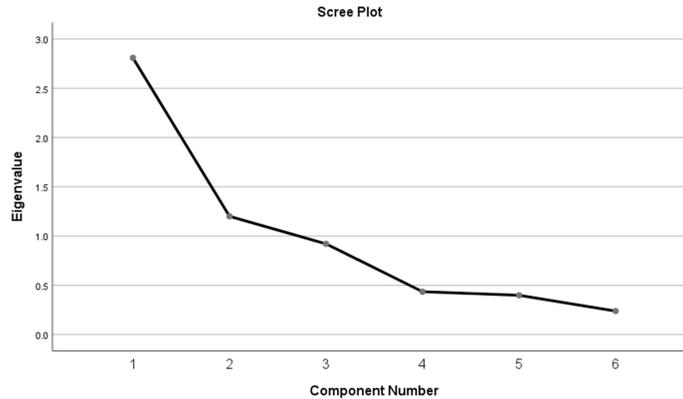
Tabel 4 di atas diperoleh initial untuk nilai *communalities* setiap variabel adalah 1,00 dan *extraction* menunjukkan seberapa besar faktor yang terbentuk dapat menerangkan varians suatu variabel. Nilai *communalities* tertinggi pada variabel Angka Harapan hidup (X3) yang artinya variabel Angka Harapan Hidup dapat menjelaskan 82% varians faktor yang terbentuk. Sedangkan nilai *communalities* terendah pada variabel TPAK (X5) sebesar 0,51% yang artinya variabel TPAK mampu menjelaskan 51% varians faktor yang terbentuk.

Tabel 5. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Variance	Cumulative	Total	Variance	Cumulative
1	2,808	46,806	46,806	2,808	46,806	46,806
2	1,200	20,003	66,809	1,200	20,003	66,809
3	0,920	15,339	82,148			
4	0,435	7,243	89,391			
5	0,398	6,639	96,030			
6	0,238	3,970	100,000			

Tabel 5 dapat dilihat bahwa dari keenam komponen masing-masing dengan *initial eigenvalues* total yakni 2,808; 1,200; 0,920; 0,435; 0,398; dan 0,238. Total varians dari semua komponen adalah 46,81%, 20%, 15,34%, 7,24%, 6,64% dan 3,97%. Pada tabel terlihat bahwa nilai ekstraksi terdiri dari dua yaitu 2,81 dan 1,20, maka semua variabel membentuk 2 faktor dengan total varians 46,81% untuk faktor pertama dan 20% untuk faktor kedua.

Sedangkan kumulatif total varians juga sebesar 46,81% untuk faktor 1 dan kumulatif sebesar 66,81 pada faktor kedua. Gambar 2 memperlihatkan *scree plot* yang membentuk 2 faktor.



Gambar 3. *Scree plot*

Berdasarkan hasil communalities, total varians, dan scree plot, dapat disimpulkan bahwa terdapat dua kelompok yang terbentuk dari faktor-faktor tersebut. Setelah faktor-faktor itu terbentuk, langkah selanjutnya adalah menganalisis nilai yang diperoleh dari komponen matriks.

Tabel 6. Component Matrix

Variabel	Component	
	1	2
Tingkat Kemiskinan(X1)	-0,423	0,629
PDRB (X2)	0,725	-0,359
Angka Harapan Hidup (X3)	0,823	0,376
Tingkat Pengangguran (X4)	0,767	0,371
TPAK (X5)	-0,555	0,453
Persentase penduduk (X6)	0,728	-0,438

Tabel 6 memperlihatkan faktor yang terbentuk berdasarkan komponen matriks dimana terdapat 2 nilai yang berbeda di bagian kolom komponen (1) dan kolom komponen (2). Untuk menyatakan bahwa masing-masing variabel termasuk pada faktor (1) atau (2) maka dapat dilihat dari nilai yang tertinggi diantara 2 faktor. Berikut hasil rotasi komponen matriks untuk mempermudah penjelasan.

Tabel 7. Rotasi Component Matrix

Variabel	Component	
	1	2
Tingkat Kemiskinan(X1)		0,629
PDRB (X2)	0,725	
Angka Harapan Hidup (X3)	0,823	
Tingkat Pengangguran (X4)	0,767	
TPAK (X5)		0,453
Persentase penduduk (X6)	0,728	

Tabel 7 memperlihatkan hasil rotasi komponen matriks dari variabel yang bersesuaian dengan faktor yang terbentuk. Faktor 1 terdiri dari PDRB (X2), Angka Harapan Hidup (X3), Tingkat Pengangguran (X4) dan Persentase penduduk (X6). Sedangkan untuk faktor 2 terdiri dari 2 variabel yakni Tingkat Kemiskinan(X1) dan TPAK (X5).

5. KESIMPULAN DAN SARAN

Faktor yang terbentuk pada penelitian ini sebanyak 2 faktor. Faktor 1 terdiri dari PDRB (X2), Angka Harapan Hidup (X3), Tingkat Pengangguran (X4) dan Persentase penduduk (X6). Sedangkan untuk faktor 2 terdiri dari 2 variabel yakni Tingkat Kemiskinan(X1) dan TPAK (X5). Adapun saran yang diberikan adalah perlu penerapan secara berkala untuk memperbarui data kemiskinan dan mengevaluasi efektivitas dan mengupayakan penambahan variabel dalam penelitian selanjutnya.

DAFTAR REFERENSI

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Bank Dunia. (2021). World Bank poverty and equity data. <https://www.worldbank.org/en/topic/poverty>
- Fieler, A. C., & Pritchett, L. (2001). The role of geography in explaining world poverty: Geography and poverty. *World Bank Economic Review*, 15(2), 229–242. <https://doi.org/10.1093/wber/15.2.229>
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.
- Jolliffe, I. T., & 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), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Shrens, J. (2014). *Visualization techniques for big data analysis*. Springer.
- Todaro, M. P., & Smith, S. C. (2020). *Economic development* (13th ed.). Pearson Education.
- Wold, S., Ruhe, A., Wold, H., & Dunn, I. (1987). The collinearity problem in linear regression: The partial least squares (PLS) approach to generalized inverse. *SIAM Journal on Scientific and Statistical Computing*, 6(3), 735–743. <https://doi.org/10.1137/0906049>