**Principal Component Analysis (PCA) dan Factor Analysis (FA) pada Data Diabetes**

**Nachla Fadilla (23031554072), Sofia Zahira R (23031554197)**

# BAB 1 Pendahuluan

Latar Belakang

Principal Component Analysis (PCA) dan Factor Analysis (FA) adalah dua teknik analisis multivariat yang digunakan untuk mereduksi dimensi data dengan tetap mempertahankan informasi penting. PCA digunakan untuk menemukan kombinasi linier terbaik dari variabel asli yang menjelaskan varians maksimum, sedangkan FA digunakan untuk mengidentifikasi faktor laten yang menjelaskan korelasi antar variabel. Pada penelitian ini, kami menerapkan PCA dan FA pada dataset diabetes untuk memahami pola yang ada di dalam data serta menginterpretasikan variabel-variabel utama yang mempengaruhi dataset tersebut

Rumusan Masalah

1. Bagaimana karakteristik dataset diabetes berdasarkan analisis deskriptif?
2. Apakah asumsi PCA dan FA terpenuhi dalam dataset ini?
3. Bagaimana interpretasi komponen utama dalam PCA?
4. Bagaimana hasil analisis faktor dibandingkan dengan PCA?

Tujuan

1. Melakukan eksplorasi awal terhadap dataset diabetes
2. Menguji asumsi KMO dan Bartlett untuk kelayakan analisis PCA dan FA
3. Mengidentifikasi komponen utama dalam PCA serta menginterpretasi hasilnya
4. Mengidentifikasi faktor laten yang terbentuk dalam analisis faktor

# BAB 2 Metodologi Penelitian

1. Dataset yang kami gunakan adalah dataset diabetes

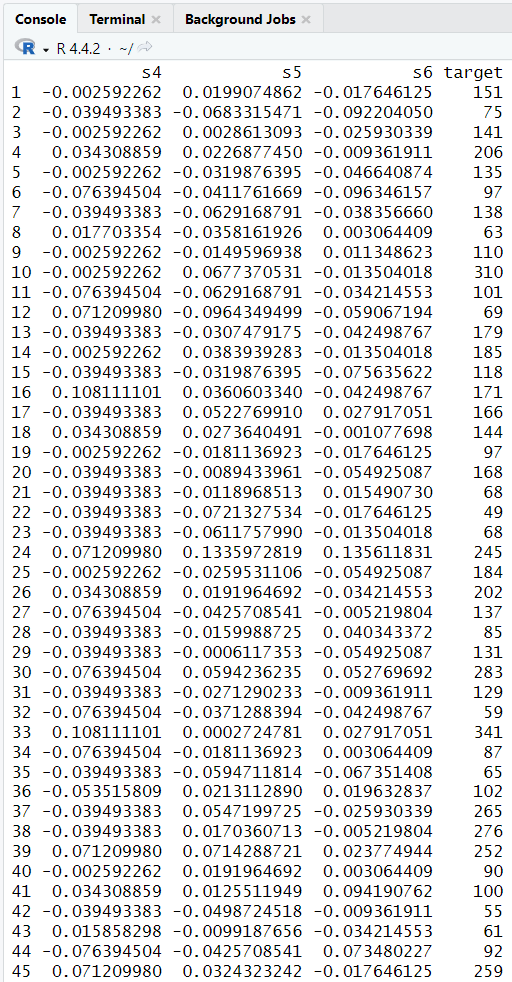
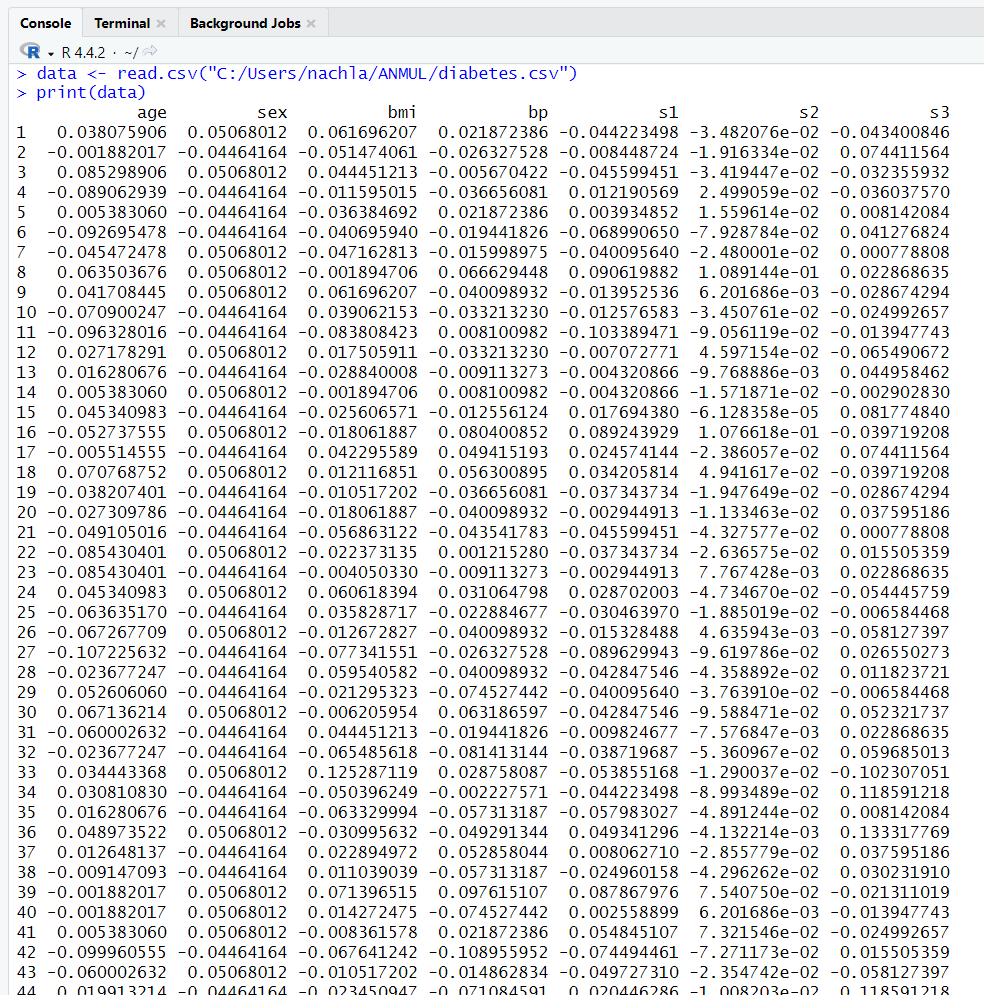
(<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html>)

1. Variabel dalam dataset ini terdiri dari beberapa fitur kuantitatif yang menggambarkan karakteristik pasien diabetes, seperti kadar glukosa, tekanan darah, BMI, dan lain lain

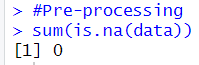
# BAB 3 Pembahasan

1. Summary
2. Cek Dataset

Dataset ini memiliki 10 kolom dan 1 kolom target



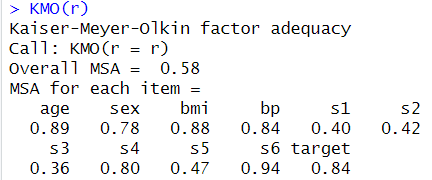
1. Cek Missing Values



Setelah di cek, dataset ini tidak memliki missing value. Jadi bisa langsung untuk di proses

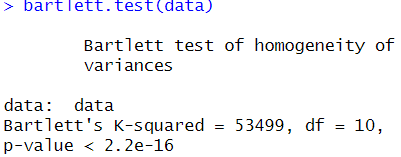
1. Asumsi
2. KMO Test

Mengukur kecukupan sampel dalam analisis factor. Dihasilkan nilai MSA sebesar 0,58 menunjukkan bahwa data memiliki kecukupan sedang untuk analisis factor. Tetapi, untuk nilai MSA per variable ada yang memiliki nilai rendah, misalnya s1=0,40; s2=0,42; dan s3=0,36 yang menunjukkan bahwa tidak cocok untuk analisis faktor



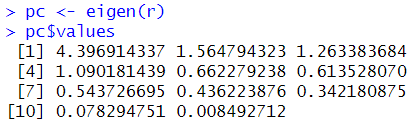
1. Bartlett’s Test

Menguji asumsi homogenitas varians dalam analisis factor. Dihasilkan K-squared= 5399, df=10, dan p-value<2,2e-16. Karena p-value yang dihasilkan sangat kecil, maka ini menunjukkan bahwa korelasi antar variable signifikan, sehingga dapat dilakukan analisis faktor



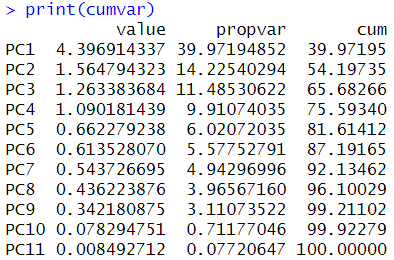
1. PCA
   1. Eigen values

Eigenvalues yang lebih dari 1 merupakan kandidat utama untuk dipertahankan dalam analisis komponen

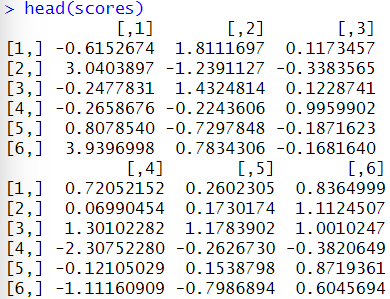


* 1. Cumulative Variance

Secara kumulatif, PC1 sampai PC5 menjelaskan sekitar 81,61% varians yang cukup untuk mereduksi dimensi data

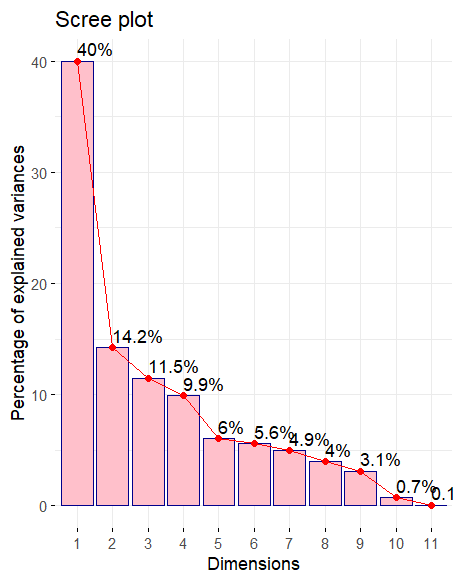


* 1. Hasil PCA



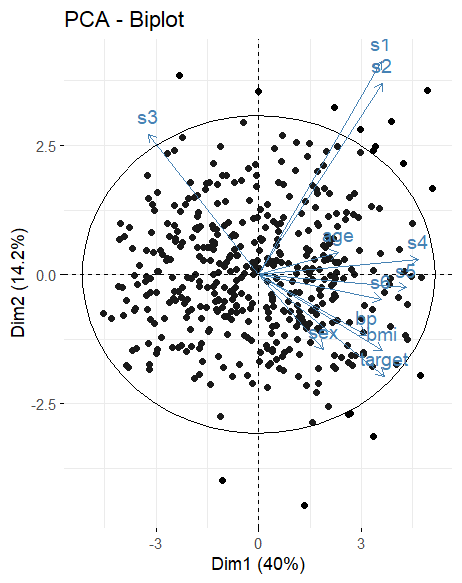
head(scores) diperoleh dari hasil perkalian antara scale\_data dengan pc$vectors dari matriks kovarians

* 1. Visualisasi Screeplot



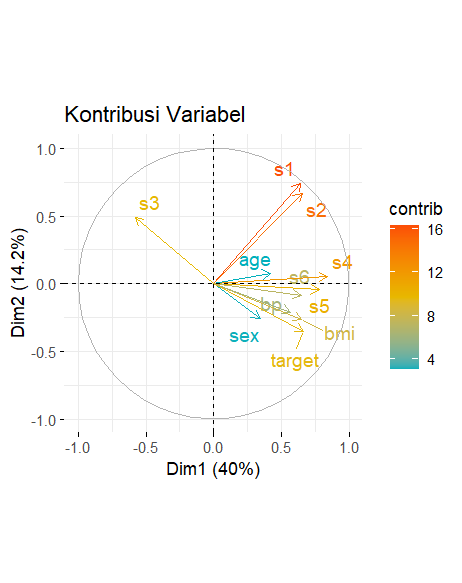
Dari gambar tersebut dapat disimpulkan bahwa PC1 tertinggi untuk variansnya. Dari PC1 sampai PC11 terlihat jelas bahwa ada penurunan varians yang sangat menonjol, yaitu dari PC1=40% ke PC11=0,1%. Sehingga komponen yang optimal berada di PC3 dan PC4, karena penurunya tidak terlalu drastis

* 1. Visualisasi Biplot



Dari gambar tersebut dapat disimpulkan bahwa, variabel bmi, bp, sex, dan target lebih banyak kontribusi pada Dimensi 1(sumbu X) sekitar 40%. Variabel s1, s2, dan s3 memiliki kontribusi yang cuup tinggi di Dimensi 2(sumbu Y) sekitar 14,2%.

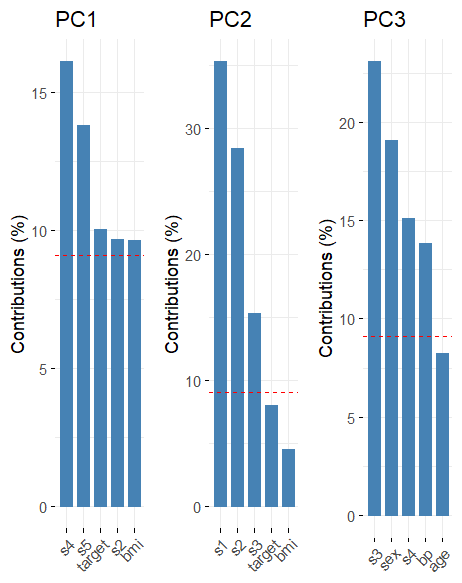
* 1. Correlation Circle



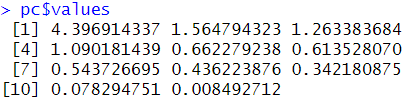
Dari gambar tersebut diketahui bahwa, variabel dengan warna orange(s1, s2, s4, s5) berkontribusi tinggi terhadap Dimensi 1 sekitar PC1=40%. Sedangkan variabel yang berwarna biru(sex, age, bp) berkontribusi lebih rendah dibandingkan variabel orange

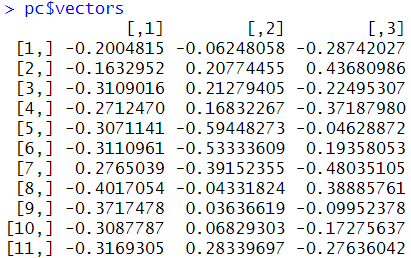
* 1. Contribution

Gambar di bawah menunjukkan bahwa, variabel yang berkontribusi dalam PC1 ada s4, s5, dan s2. Di PC2 yang berkontribusi ada s1, s2, s3. Di PC3 yang berkontribusi ada s3, s4, sex, bp , dan age

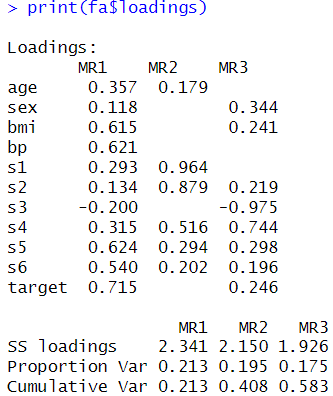


1. FA
   1. Menghitung Kovarians dan Eigen Values

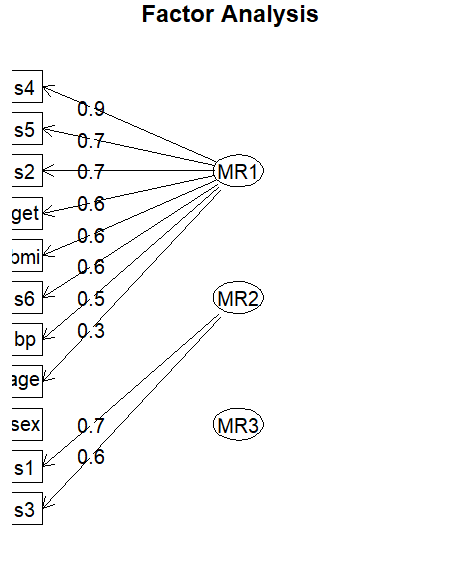




* 1. Menghitung Faktor Score



* 1. Visualisasi Diagram



Dari gambar tersebut diketahui bahwa, gambar kotak(variable observasi), oval(factor laten). MR1 menjadi factor utama karena memiliki banyak variabel, antara lain s4, s5, s2

# BAB 4 Kesimpulan

Berdasarkan hasil analisis yang kami lakukan, diketahui bahwa nilai Kaiser-Meyer-Olkin (KMO) sebesar 0.58 menunjukkan data cukup layak untuk analisis faktor, meskipun beberapa variabel seperti s1, s2, dan s3 memiliki nilai MSA rendah. Hasil uji Bartlett yang signifikan (p < 2.2e-16) mengindikasikan adanya korelasi antar variabel, sehingga analisis faktor dapat dilakukan. Dari hasil eigenvalues dan scree plot, PC1 menjelaskan 40% varians, diikuti oleh PC2 sebesar 14.2% dan PC3 sebesar 11.5%, sehingga tiga komponen utama pertama cukup untuk mewakili pola dalam data karena sudah menjelaskan lebih dari 65% total varians. Scree plot menunjukkan adanya "elbow" pada PC3, yang menegaskan bahwa dua hingga tiga komponen utama sudah cukup untuk analisis lebih lanjut.

# Daftar Pustaka

Revelle, W. (2023). Psych: Procedures for psychological, psychometric, and personality research. Northwestern University. <https://cran.r-project.org/package=psych>

Abdi, H., & Williams, L. J. (2020). Principal component analysis: Concepts, applications, and recent trends. Computational Statistics & Data Analysis, 150, 107012. <https://doi.org/10.1016/j.csda.2020.107012>

Wang, L., Zhang, H., & Xu, X. (2021). Factor analysis in modern data analysis: Challenges and solutions. Journal of Statistical Computation and Simulation, 91(13), 2512-2530. <https://doi.org/10.1080/00949655.2021.1916132>

Wickham, H., & Bryan, J. (2023). R for Data Science (2nd ed.). O’Reilly Media.

# Lampiran

Code diunggah ke github atau melampirkan link drive

Format: Bookman Old Style, 11, space 1.5

Maksimal 6 halaman

Link google docs

Share ke [ulfanuraini@unesa.ac.id](mailto:ulfanuraini@unesa.ac.id) dan [dindaguminta@unesa.ac.id](mailto:dindaguminta@unesa.ac.id)