Topik : Analisis Komponen Utama

1. Buka dataset (https://ipb.link/data-lkp8)

2. Berdasarkan data yang tersedia, analisis apa yang dapat dilakukan?
   Analisis komponen utama (PCA) dapat dilakukan pada data yang tersedia. Teknik ini menyederhanakan data dengan cara mereduksi dimensi, tanpa mengurangi karakteristik data secara signifikan. Analisis Komponen Utama (Principal Component Analysis) adalah analisis *multivariate* yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi dengan mereduksi sejumlah variabel tersebut sehingga mempunyai dimensi yang lebih kecil namun dapat menerangkan sebagian besar keragaman variabel aslinya.

3. Apakah terdapat multikolinearitas pada variabel independen pada data yang tersedia? Jelaskan!
   Ya, dalam dataset tersebut terdapat masalah multikolinearitas. Multikolinearitas merujuk pada adanya korelasi linear antara dua atau lebih variabel independen dalam model regresi. Dalam dataset ini, terlihat bahwa variabel STBM memiliki hubungan yang erat dengan variabel Tsampah dan variabel Limbah. Tsampah dan Limbah merupakan indikator dari konsep STBM (Sanitasi Total Berbasis Masyarakat) dalam pengelolaan sampah dan limbah yang benar. Konsep STBM sendiri merupakan bagian dari upaya untuk meningkatkan Kesehatan Lingkungan (Kesling), sehingga variabel-variabel tersebut saling terkait.

   Multikolinearitas dapat dicek dengan cara berikut
   install.packages("PerformanceAnalytics")
   # check multicollinearity
   library(PerformanceAnalytics)
   chart.Correlation(dataA)

   dan nantinya akan muncul grafik berikut
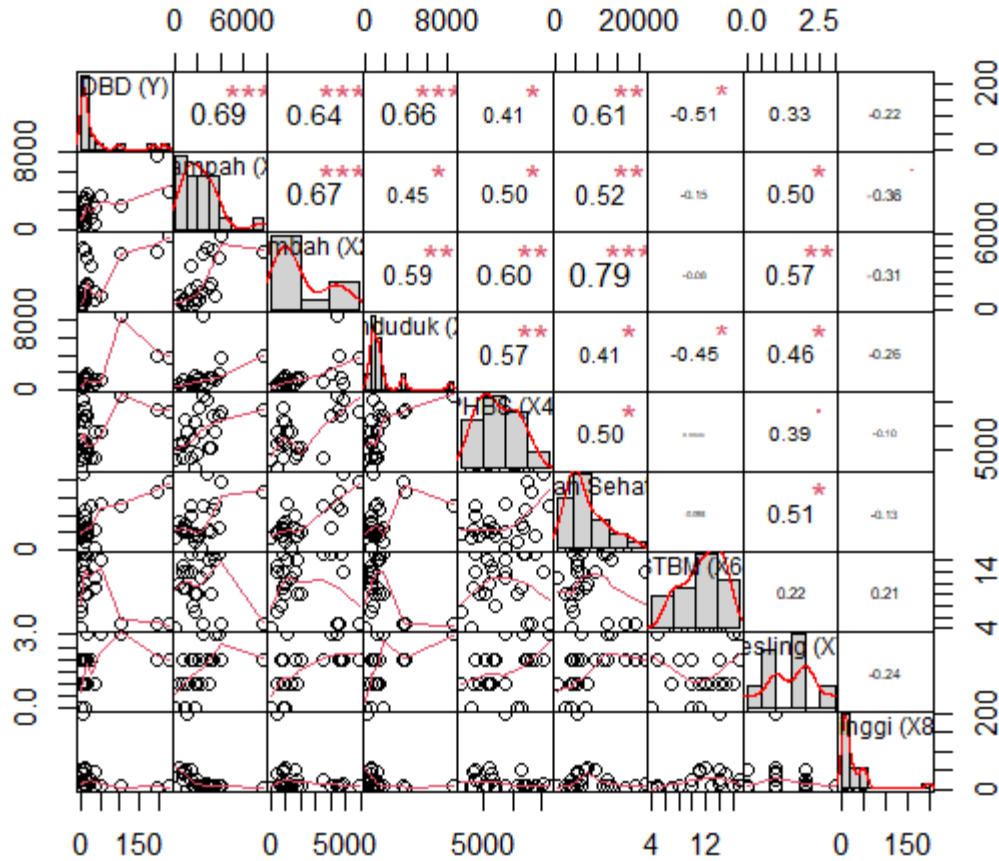
4. Jika terdapat multikolinearitas, maka lakukan langkah PCA (Principal Component
   Analysis) untuk menghilangkan multikolinearitasnya!
   #variabel kecamatan tidak digunakan, simpan data baru sebagai dataA
   dataA <- Tugas[,c(2:10)]

#Cov dataA

```
> cov(dataA)
                   DBD (Y)   Tsampah (X1)  Limbah (X2) Penduduk (X3)       PHBS (X4) Rumah Sehat (X5)      STBM (X6) Kesling (X7)
DBD (Y)         3520.21558     66729.0091   66818.4312    68781.1812     107998.3152       194202.455  -128.5543478   17.3677536
Tsampah (X1)    66729.00906  2662693.9982 1933483.2790  1308514.0072    3631046.1196      4576468.205 -1025.4891304  713.9873188
Limbah (X2)     66818.43116  1933483.2790 3091074.6884  1817695.1014    4694041.5435      7384995.518  -604.7173913  878.3442029
Penduduk (X3)   68781.18116  1308514.0072 1817695.1014  3111456.1449    4411663.5217      3902001.703 -3381.2608696  724.1376812
PHBS (X4)      107998.31522  3631046.1196 4694041.5435  4411663.5217   19593882.3696     11811387.141   182.7173913 1520.1847826
Rumah Sehat (X5) 194202.45471 4576468.2047 7384995.5181 3902001.7029   11811387.1413    28542015.563 -2183.5543478 2390.9547101
STBM (X6)        -128.55435    -1025.4891    -604.7174    -3381.2609        182.7174        -2183.554    18.3695652    0.8369565
Kesling (X7)       17.36775      713.9873     878.3442      724.1377       1520.1848         2390.955     0.8369565    0.7807971
Tinggi (X8)      -520.96377   -23630.6594  -21855.0580   -18706.7536     -18534.7391       -28779.833    35.5652174   -8.3985507
                  Tinggi (X8)
DBD (Y)           -520.963768
Tsampah (X1)    -23630.659420
Limbah (X2)     -21855.057971
Penduduk (X3)   -18706.753623
PHBS (X4)       -18534.739130
Rumah Sehat (X5) -28779.833333
STBM (X6)           35.565217
Kesling (X7)        -8.398551
Tinggi (X8)       1625.014493
```

#Nilai eigen, vektor eigen

```
> eigen(cov(dataA))
eigen() decomposition
$values
[1] 4.084121e+07 1.166569e+07 2.353457e+06 1.518984e+06 6.245623e+05 1.382425e+03 9.875116e+02 9.471304e+00 3.012286e-01

$vectors
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
 [1,] -6.032479e-03  1.521747e-03 -0.0187491554  1.265915e-03  1.360134e-02  0.4422350471  8.951464e-01 -5.058772e-02 -2.255669e-03
 [2,] -1.617448e-01 -3.694143e-02 -0.5200118849  7.503886e-01  3.723522e-01 -0.0008146573 -1.820231e-02  5.794438e-04  1.736262e-04
 [3,] -2.348306e-01  3.088303e-02 -0.3558815395  1.468907e-01 -8.919607e-01  0.0088025698 -4.198788e-05  8.991293e-04 -4.084206e-05
 [4,] -1.603990e-01 -1.420882e-01 -0.7072841826 -6.384531e-01  2.143245e-01 -0.0040771118 -1.604116e-02 -8.646528e-04  3.328049e-04
 [5,] -5.428310e-01 -7.959833e-01  0.2665965661  2.235408e-02  1.266097e-02 -0.0005350701  3.329133e-03  1.763096e-04 -4.836315e-05
 [6,] -7.734742e-01  5.864306e-01  0.1765228612 -8.481949e-02  1.395076e-01 -0.0034732130 -2.794995e-03  1.542691e-05  6.246637e-05
 [7,]  5.976262e-05 -7.942032e-05  0.0011923037  9.805360e-04 -1.394722e-03 -0.0059927699 -5.278339e-02 -9.922118e-01  1.126473e-01
 [8,] -7.621115e-05  7.712519e-06 -0.0001568387  2.218899e-05 -1.503066e-05 -0.0034570201 -7.152064e-03 -1.124072e-01 -9.936305e-01
 [9,]  1.084236e-03  6.268670e-05  0.0099008957 -4.595457e-03  3.898947e-03  0.8968127885 -4.418915e-01  1.787012e-02 -1.962892e-03
```

#PCA

```
> A.pca <- prcomp(dataA, center=TRUE, scale.=TRUE)
> summary(A.pca)
Importance of components:
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7     PC8    PC9
Standard deviation     2.1092 1.2062 0.9776 0.81196 0.73841 0.67904 0.45837 0.40850 0.3132
Proportion of Variance 0.4943 0.1616 0.1062 0.07325 0.06058 0.05123 0.02334 0.01854 0.0109
Cumulative Proportion  0.4943 0.6560 0.7621 0.83540 0.89598 0.94722 0.97056 0.98910 1.0000
```

#Menghitung nilai PC1, PC2,..., PC9

```
> Pcnya <- predict(A.pca, newdata=dataA)
> head(Pcnya)
            PC1        PC2         PC3          PC4         PC5         PC6          PC7        PC8         PC9
[1,] -2.32083032 -0.9846328  3.68601590 -0.001661839  0.37938631 -1.00516951  0.085604190  0.3926581 -0.07281131
[2,] -1.18853518 -0.5817649 -0.35587436 -0.191428602  0.61612881  0.21894353  0.322632155 -0.4037565 -0.28065642
[3,] -0.07706518 -1.2478369 -0.38219148  0.999502653  0.79474861  0.79778971 -0.807075996 -0.1680514  0.06751038
[4,] -1.78156228 -0.2897543  0.13252755 -0.126592366 -0.03402684  0.09226083 -0.175197802 -0.5540405  0.11774687
[5,] -0.89973591 -0.6956211  0.08477488  0.675165860  1.02825318  0.14318114 -0.008417346  0.0234148  0.12344791
[6,] -1.44521181  0.4630959  1.01583010 -0.320170541 -1.35981271  0.52940093  0.230474281 -0.1458881  0.15858052
```

Setelah melakukan pca, kita lihat kembali grafik multikolinearitas tadi dengan cara memasukkan data yang telah di PCA (Pcnya).

chart.Correlation(Pcnya)

Chart di bawah tidak menunjukkan adanya bintang merah dan angka yang mengindikasikan kolinear. Maka dapat kita simpulkan multikolinearitas sudah berhasil dihilangkan dengan menggunakan PCA.
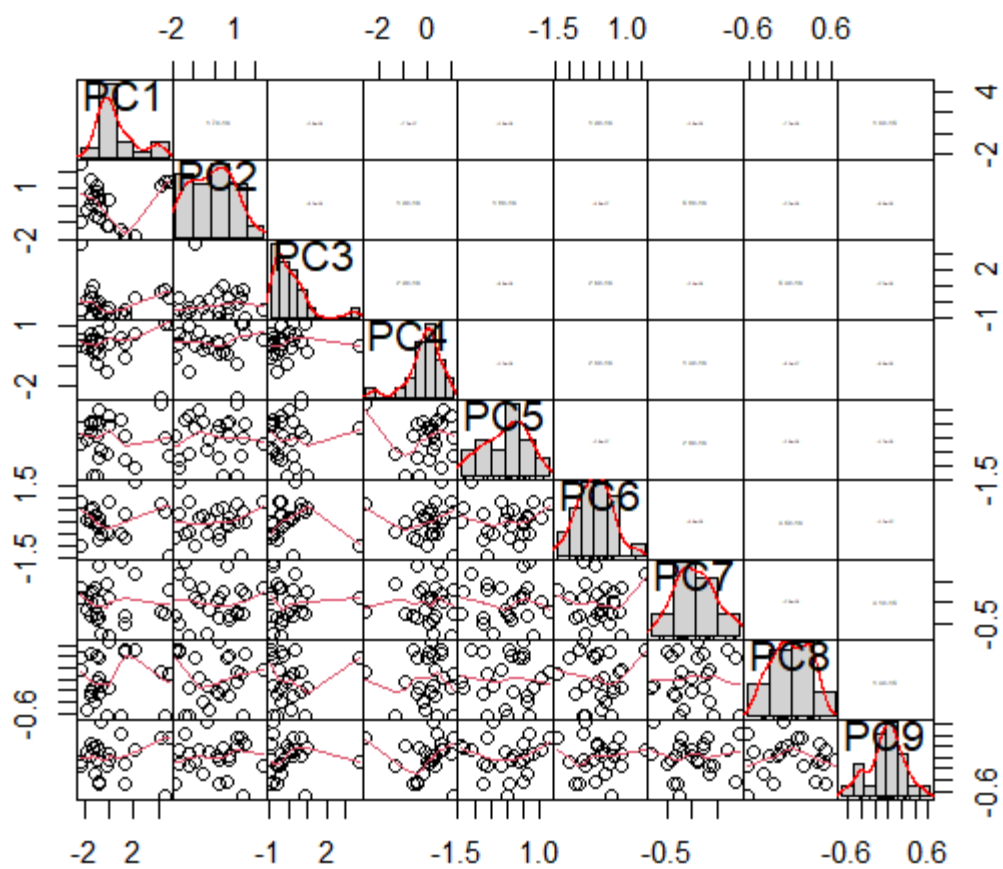
5. Tampilkan data akhir yang diperoleh dari reduksi variabel menggunakan PCA!

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
1	-2.32083032	-0.9846328	3.68601590	-0.001661839	0.37938631	-1.005169513	0.085604190	0.39265812	-0.07281131
2	-1.18853518	-0.5817649	-0.35587436	-0.191428602	0.61612881	0.218943528	0.322632155	-0.40375645	-0.28065642
3	-0.07706518	-1.2478369	-0.38219148	0.999502653	0.79474861	0.797789707	-0.807075996	-0.16805139	0.06751038
4	-1.78156228	-0.2897543	0.13252755	-0.126592366	-0.03402684	0.092260830	-0.175197802	-0.55404051	0.11774687
5	-0.89973591	-0.6956211	0.08477488	0.675165860	1.02825318	0.143181138	-0.008417346	0.02341480	0.12344791
6	-1.44521181	0.4630959	1.01583010	-0.320170541	-1.35981271	0.529400935	0.230474281	-0.14588807	0.15858052
7	-1.01633993	0.7515536	-0.28531454	0.272697461	0.10944328	-0.186233977	-0.423065716	0.41993294	-0.02483204
8	-1.52152178	0.7023126	0.27955473	0.290125808	0.62310073	0.112505495	-0.331563419	-0.30363538	0.03830431
9	-1.59537609	1.5021391	-0.89161537	0.558587279	0.49966579	-0.042076548	0.220215756	0.42998326	0.23002368
10	2.06575189	-1.7822708	0.34851701	0.606022897	-1.07813883	1.614874149	0.684911176	0.50530035	0.14578156
11	-0.84910612	0.5904146	-0.75213798	-0.396688880	0.49846581	-0.136519468	0.847578263	0.36753009	-0.42956106
12	-1.46544945	0.1004177	0.69144754	0.177967780	0.21200861	0.502380593	-0.009455063	-0.20510751	0.01332748
13	-2.35860616	2.4655294	-0.46983484	0.294015065	-0.20841897	0.787650253	0.329252228	0.17100969	-0.15183496
14	-0.93964743	0.2909680	-0.60457032	-0.436850706	-0.73149634	-0.783267241	-0.359543087	-0.26590527	-0.45640576
15	0.99394317	-1.4242794	-0.73830432	0.417335523	0.66474393	-0.505321897	0.929436454	-0.65535084	0.44154148
16	-0.68346035	-0.3238909	-0.78336392	0.217423192	-0.42069809	-0.493697625	-0.749114252	-0.11759891	0.62646852
17	-1.06359268	1.1707243	-0.08713299	-0.901093620	-1.40635782	-0.006823708	-0.261080035	-0.06075739	0.20049164
18	-0.02600837	0.2684653	-0.88721195	0.110901411	0.46212644	-0.522126853	0.086257323	0.64336799	-0.02625272
19	1.35107752	-1.8854552	-0.85692982	-0.628372738	0.04939750	-0.674340384	-0.303937514	0.50580944	-0.48548688
20	-0.28947067	-1.2369324	-0.54400387	-1.318095740	-0.60709733	-0.462025095	0.411426975	-0.63443278	-0.13408258
21	1.29577665	-1.7459791	-0.39278909	0.058358808	-0.76420191	0.355929848	-0.504954594	0.46720270	-0.10359966
22	4.29415630	1.1327531	0.64599614	-2.572012977	1.28185454	0.715106988	-0.225598683	0.14058944	0.35495033
23	4.88275321	1.3684026	0.30210098	1.110445073	-0.62056270	-1.476882723	0.283802165	0.09009317	0.34319416
24	4.63806099	1.3916417	0.84451001	1.104419199	0.01148800	0.424461569	-0.272587461	-0.64236750	-0.69584545