Statistics for the Sciences

Principal Component Analysis (PCA)

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- PCA Idea
- PCA Theory
- Example
- Lab
- Appendix: Derivation of PCs

Principal Component Analysis

- PCA (principal component analysis) is a dimension reduction method
- The central idea of PCA is to **reduce the dimensionality of a data set** consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set.
- This is achieved by transforming to a new set of variables, the **principal components (PCs)**, which are **uncorrelated**, and which are ordered so that the first few retain **most of the variation** present in all of the original variables.

PCA - Idea

Set-up:

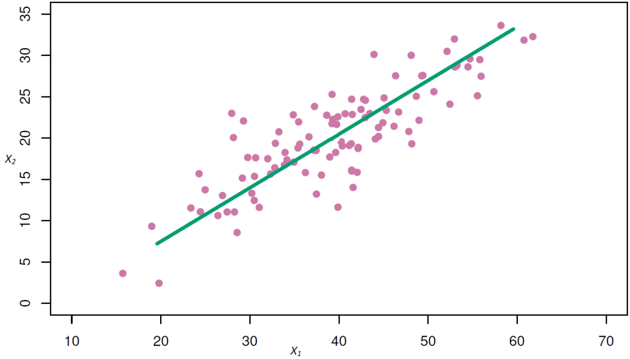
- Data matrix: $\mathbf{X}_{np}$, n observations and p features (variates).
  - There is no response variable so they are not called predictors any more.

Idea:

- Not all p variables are needed (much redundant info).
- Find **low-dimensional representations** that capture most of the variation in the data.

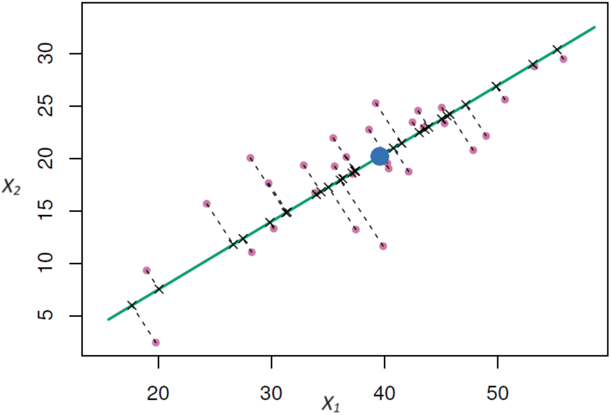
PCA - Idea

- Question: What is a good 1-dimensional representation (the first principal component) of the data?
  - Find a line which maximizes the variance of the data projected onto the line:



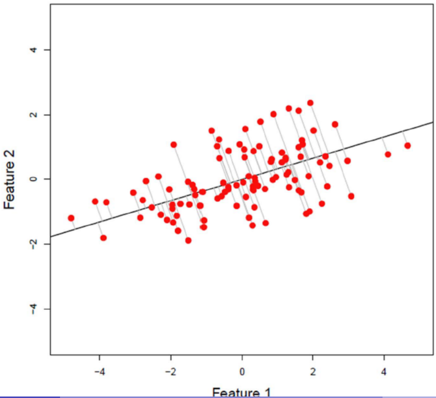
PCA - Idea

- Another interpretation is that the first principal component vector defines the line that is as close as possible to the data.



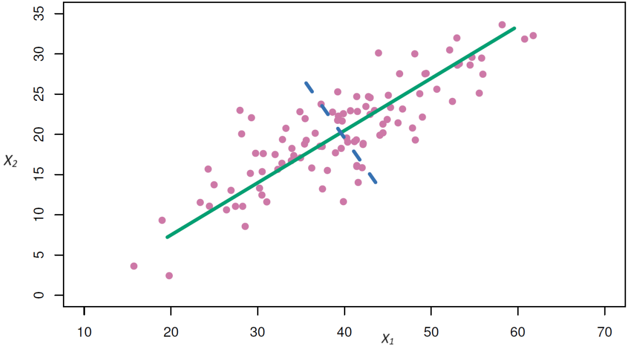
PCA - Idea

- PCA minimizes orthogonal projection onto line: $Z = v_1 x_1 + v_2 x_2$.
  - Note: Not same as LS line!
- If the features are centered, then
  - The direction of the line is the vector (v_1, v_2). The vector should be an unit vector $v_1^2 + v_2^2 = 1$, otherwise v_1 and v_2 could be increased arbitrarily in order to blow up the variance.
  - Equivalently, slope of line is $= v_2/v_1$.



PCA - Idea

- Subsequent components orthogonal (perpendicular).
  - No dimension reduction if there are two features only.



PCA - Theory

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_p)^{'}$ is a vector of p random variables, and that the variances of the p random variables and the structure of the covariances/correlations between the p variables are of interest.
- Instead of looking at the p variances and $\frac{1}{2}p(p-1)$ covariances, we look for a few $(< p)$ derived variables that preserve most of the information given by these variances and correlations/covariances.
  - PCA concentrates on variances.
- The first step is to look for a **linear combination** of the elements of $\mathbf{x}$ having **maximum variance**

$$\boldsymbol{v}_1^{'}\mathbf{x} = v_{11}x_1 + \cdots + v_{1p}x_p = \sum_{j=1}^{p} v_{1j}x_j.$$

PCA - Theory

- Let
$$\mathbf{\Sigma} = var(\mathbf{x})$$
be the variance-covariance matrix of $\mathbf{x}$.
- Using tools in *Linear Algebra* and *Calculus*, the 1st PC(principal component) of $\mathbf{x}$ is the eigenvector of the largest eigenvalue of $\mathbf{\Sigma}$.

PCA - Theory

- Next, look for a linear combination $\boldsymbol{v}_2^{'}\mathbf{x}$, **uncorrelated with** $\boldsymbol{v}_1^{'}\mathbf{x}$ having **maximum variance**, and so on. ... Up to p PCs could be found.
  - It is hoped, in general, that most of the variation in $\mathbf{x}$ will be accounted for by m PCs, where $m << p$.
- Again, it can be shown that the 2nd PC(principal component) of $\mathbf{x}$ is the eigenvector of second largest eigenvalue of $\boldsymbol{\Sigma}$.
- The 3rd PC(principal component) of $\mathbf{x}$ is the eigenvector of third largest eigenvalue of $\mathbf{x}$.
- And so on $\vdots$

PCA - Theory

- Denote all the PCs by: $\mathbf{Z} = \mathbf{X}\mathbf{V}$
  - $\mathbf{z}_k = \mathbf{X}\mathbf{v}_k$ is the k^{th} PC, $k = 1, \ldots, p$.
  - The coefficients of the kth PC are called the kth **PC loading** (feature weights), it defines the **direction of the PC**, $k = 1, \ldots, p$.

- If we project the ith data point $(x_{i1}, \ldots, x_{ip})$ onto this direction, the projected value $z_{i1} = v_{i1}x_{i1} + \cdots + v_{ip}x_{ip}$, a linear combination of the coordinates of the data point, is referred as the **principal component score**.

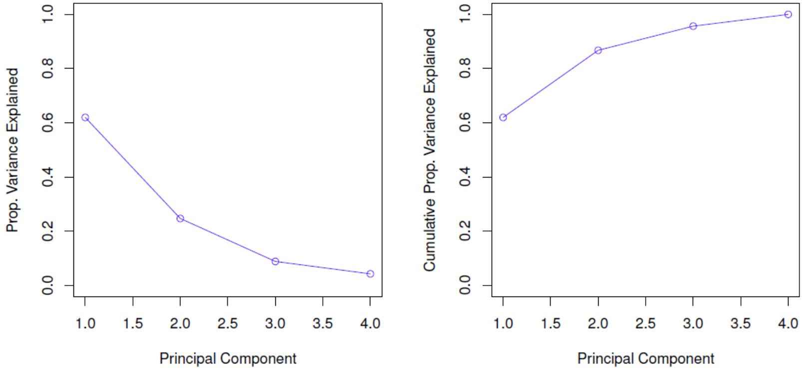
PCA - Theory

- 1. How much variance is explained by each PC? How much cumulative variance can be explained by the first k ($k << p$) PC's?
- It can be shown that the amount of variance explained by a PC component is exactly equal to the correpsonding eignvalue.
  - The variance of the first PC is equal to the largest eigenvalue of $\mathbf{\Sigma}$.
  - The variance of the second PC is equal to the second largest eigenvalue of $\mathbf{\Sigma}$.
  - And so on $\vdots$

PCA - Theory

- (2) How many PCs should we use? No simple answer to this question.
  - Take K PCs that explains at least 90% (95%, 99%, etc.) variance
  - the "scree plot" on the next slide can be used as a guide: we look for an "elbow".
    - This is done by eyeballing the scree plot, and looking for a point at which the proportion of variance explained by each subsequent principal component drops off. This is often referred to as an **elbow** in the scree plot.

PCA - Theory


Prop. Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Principal Component
Cumulative Prop. Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Principal Component

Example

- Example (`wu.csv`: https://mjkeough.github.io/examples/wu.nb.html):
  Wu et al. (2021) took soil samples at 300 sites near the Three Gorges Reservoir in China. Each site was classified into one of three categories of land use: orchard ($n = 75$), dry land ($n = 98$), and paddy field ($n = 127$). They also measured the concentrations of ten metals (Ca, Cd, Cr, Cu, Fe, Mg, Mn, Ni, Pb, and Zn; mg/kg) and five soil characteristics (pH, concentrations of N and P in mg/kg, and % of soil organic carbon [SOC] and K) for each site. Wu et al. (2021) did a PCA on the metals and used the components in further analyses. Using PCA, we will instead examine the pattern among sites and land use categories based on the metals and the other characteristics (15 variables in total).
- After removing the last two columns

```
## 'data.frame':    300 obs. of  15 variables:
##  $ ca : num  0.81 0.466 0.932 0.81 0.322 ...
##  $ cr : num  66.2 43.8 70 60.6 46.5 ...
##  $ cu : num  20.39 10.29 25.59 19.9 7.44 ...
##  $ fe : num  3.47 2.04 3.78 3.54 2.09 ...
##  $ mn : num  633 229 569 546 239 ...
##  $ pb : num  21.5 22.4 25.6 25.3 20.9 ...
##  $ zn : num  68.8 36.5 78.1 67.5 28.9 ...
##  $ cd : num  0.175 0.134 0.227 0.175 0.096 0.185 0.35 0.258 0.134 0.34 ...
##  $ mg : num  1.4 0.482 1.59 1.232 0.403 ...
##  $ ni : num  25.9 13.3 31.3 26.2 12.2 ...
##  $ soc: num  0.59 0.84 0.98 0.977 0.63 0.582 0.87 1.17 0.86 0.874 ...
##  $ ph : num  6.82 6.17 6.13 6.33 5.41 6.92 6.82 7.34 5.86 6.67 ...
##  $ n  : num  711 745 1123 1026 648 ...
```

Example

- Quick data summary

```
##   ca_Mean cr_Mean cu_Mean fe_Mean mn_Mean pb_Mean zn_Mean cd_Mean mg_Mean
## 1    0.75   64.39   20.82    3.23  527.03   25.53   68.22    0.22    1.12
##   ni_Mean soc_Mean ph_Mean n_Mean k_Mean p_Mean ca_SD cr_SD cu_SD fe_SD  mn_SD
## 1   25.79     0.79    6.26 888.34   2.06 627.64  0.29 10.82  5.88  0.67 198.19
##   pb_SD zn_SD cd_SD mg_SD ni_SD soc_SD ph_SD   n_SD k_SD   p_SD ca_Max cr_Max
## 1  2.31 18.21  0.07  0.42  7.63   0.21  0.83 243.65 0.42 253.08   2.08  95.68
##   cu_Max fe_Max mn_Max pb_Max zn_Max cd_Max mg_Max ni_Max soc_Max ph_Max  n_Max
## 1  36.63   4.56 1059.5   34.7 145.42   0.54   1.76   50.1     1.4   8.43 1660.2
##   k_Max   p_Max ca_Min cr_Min cu_Min fe_Min mn_Min pb_Min zn_Min cd_Min mg_Min
## 1  3.03 2474.91   0.16   36.4   7.44   1.53  129.7   17.7  21.94   0.07   0.29
##   ni_Min soc_Min ph_Min n_Min k_Min  p_Min
## 1   9.31    0.15   4.63   381     1 196.35
```

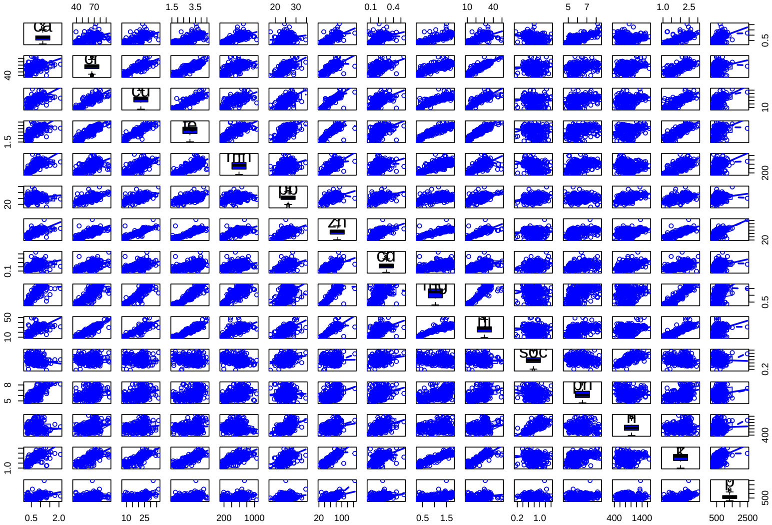
Example

- Correlation Matrix

```
##         ca    cr     cu    fe    mn    pb    zn   cd     mg    ni    soc     ph
## ca   1.000 0.503  0.670  0.69  0.72 0.200 0.737 0.41  0.809  0.68 -0.242  0.686
## cr   0.503 1.000  0.843  0.87  0.59 0.646 0.763 0.48  0.787  0.90  0.024  0.306
## cu   0.670 0.843  1.000  0.90  0.72 0.626 0.893 0.52  0.882  0.89 -0.057  0.346
## fe   0.691 0.869  0.904  1.00  0.78 0.631 0.878 0.46  0.918  0.93 -0.111  0.347
## mn   0.722 0.590  0.724  0.78  1.00 0.338 0.727 0.37  0.794  0.72 -0.340  0.383
## pb   0.200 0.646  0.626  0.63  0.34 1.000 0.589 0.49  0.470  0.61  0.311  0.024
## zn   0.737 0.763  0.893  0.88  0.73 0.589 1.000 0.62  0.887  0.85  0.013  0.393
## cd   0.414 0.483  0.515  0.46  0.37 0.487 0.624 1.00  0.489  0.51  0.246  0.324
## mg   0.809 0.787  0.882  0.92  0.79 0.470 0.887 0.49  1.000  0.92 -0.174  0.477
## ni   0.677 0.897  0.888  0.93  0.72 0.608 0.852 0.51  0.917  1.00 -0.103  0.434
## soc -0.242 0.024 -0.057 -0.11 -0.34 0.311 0.013 0.25 -0.174 -0.10  1.000 -0.098
## ph   0.686 0.306  0.346  0.35  0.38 0.024 0.393 0.32  0.477  0.43 -0.098  1.000
## n   -0.074 0.250  0.175  0.15 -0.18 0.500 0.242 0.43  0.075  0.17  0.722 -0.041
## k    0.712 0.735  0.813  0.84  0.73 0.494 0.841 0.48  0.919  0.87 -0.131  0.449
## p    0.473 0.252  0.466  0.36  0.40 0.202 0.559 0.30  0.425  0.32 -0.042  0.148
##          n     k     p
## ca  -0.074 0.712 0.473
## cr   0.250 0.735 0.252
## cu   0.175 0.813 0.466
## fe   0.151 0.844 0.363
## mn  -0.184 0.726 0.398
## pb   0.500 0.494 0.202
```

Example

- Scatter plot matrix



Example

- Test **significance of the correlations**. If this test is not statistically significant, you should not employ a PCA
- A popular test method is Bartlett test of the correlations

```
## $chisq
## [1] 5543
##
## $p.value
## [1] 0
##
## $df
## [1] 105
```

Example

- PCA will perform centering the variables.
- Centers of the variables prior to PCA

```
##     ca     cr     cu     fe     mn     pb     zn     cd     mg
##   0.75  64.39  20.82   3.23 527.03  25.53  68.22   0.22   1.12  2
##     ph      n      k      p
##   6.26 888.34   2.06 627.64
```

- Standard deviations of the variables

```
##      ca      cr      cu      fe      mn      pb      zn      cd
##   0.289  10.803   5.872   0.664 197.864   2.304  18.184   0.069
##     soc      ph       n       k       p
##   0.208   0.825 243.242   0.418 252.658
```

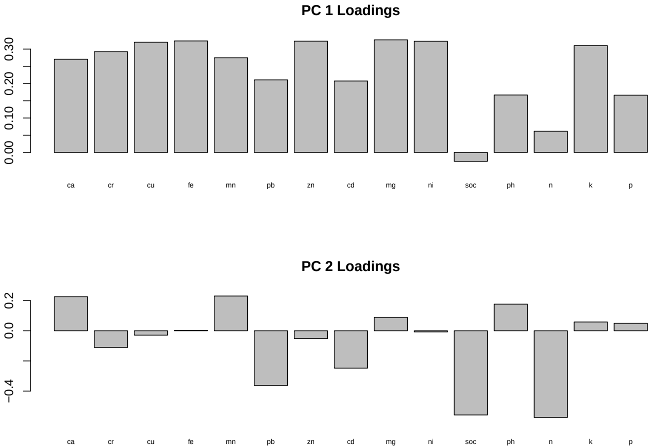
Example

- The loadings

```
##
## Loadings:
##     Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
## ca   0.27   0.23   0.35          0.12          0.17   0.25   0.20   0.53
## cr   0.29  -0.11  -0.25   0.18         -0.12  -0.35  -0.45   0.27  -0.15
## cu   0.32         -0.13                              -0.15   0.17   0.27
## fe   0.32         -0.21                                     0.16
## mn   0.27   0.23                -0.18   0.19   0.65          0.26  -0.51
## pb   0.21  -0.36  -0.29                -0.67   0.35   0.28  -0.19   0.12
## zn   0.32                -0.16          0.13                        0.22
## cd   0.21  -0.25   0.32         -0.85         -0.14  -0.14
## mg   0.33                               0.26         -0.14
## ni   0.32         -0.15   0.17                -0.20
## soc        -0.56   0.25          0.32   0.21   0.40  -0.51  -0.18   0.11
## ph   0.17   0.18   0.62   0.44   0.17  -0.45                       -0.25
## n          -0.57   0.16          0.16   0.22  -0.17   0.55   0.34  -0.31
## k    0.31                               0.22  -0.10   0.15  -0.75  -0.23
## p    0.17          0.25  -0.84   0.18  -0.25  -0.17                -0.21
##     Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
## ca   0.43    0.20    0.22    0.14    0.14
## cr   0.34            0.46           -0.17
## cu  -0.67    0.52                    0.13
## fe   0.11   -0.29   -0.27   -0.69    0.39
## mn                           0.13
## pb   0.11                           -0.14
## zn  -0.35   -0.73    0.25    0.25   -0.10
## cd
## mg           0.11   -0.35   -0.17   -0.78
## ni   0.14           -0.57    0.61    0.29
## soc                 -0.10
## ph  -0.21   -0.10           -0.11
## n
## k            0.15    0.33            0.25
## p    0.13           -0.11
##
##                 Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## SS loadings      0.995  0.999  1.000  1.005  1.006  1.005  0.999  0.997  1.004
```

Example

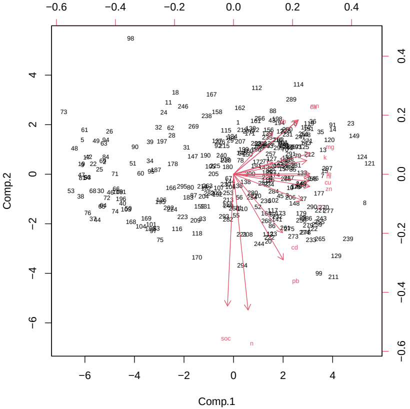
- Plot of the first two loadings lets us check how the variables contribute to the pattern



Example

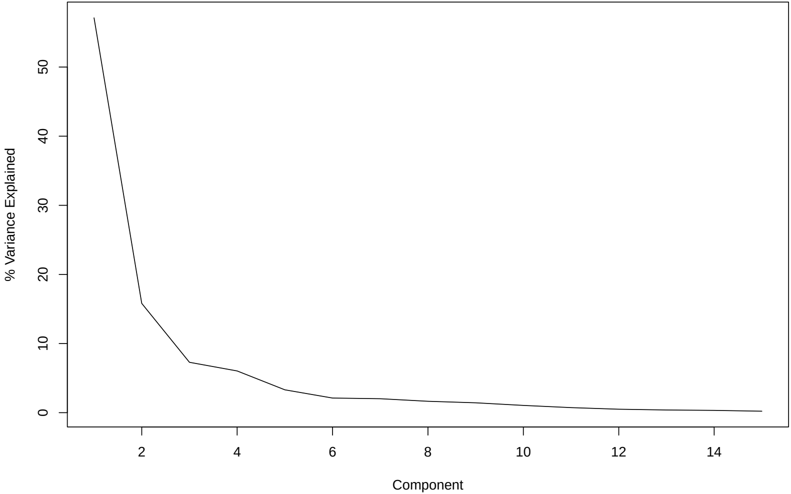
Visual representation of the data in the space of the first two principal components.

- Dimension of data is reduced from 15 to 2.
- Points in the plot represent the scores of the observations.
- Arrows represent the loadings of the original variables on the principal components.
- The direction and length of an arrow indicate the contribution of a variable to the principal components.
- Variables that are close to each other or point in the same direction are positively correlated, while those that are opposite each other are negatively correlated.



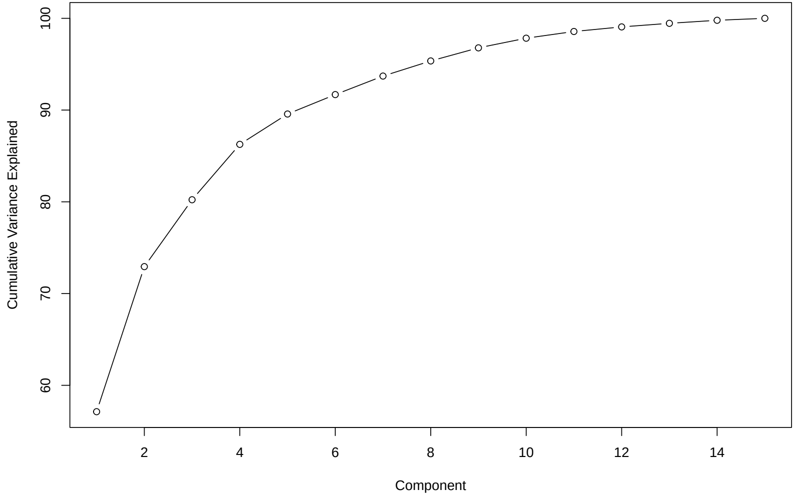
Example

- Plot of variance explained



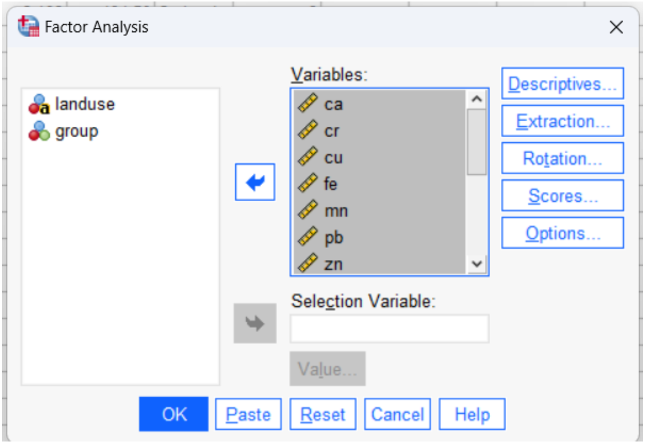
Example

- Plot of cumulative variance explained



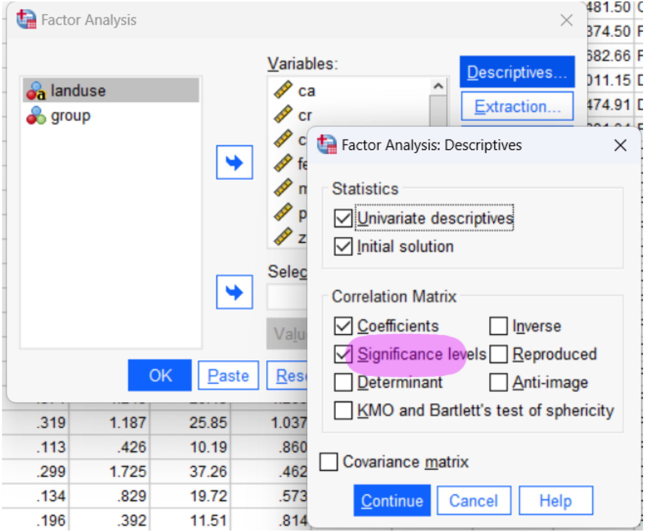
Lab

- After importing data, Click on `Analyze` → `Dimension Reduction` → `Factor`, move variables into the `Variables` box.



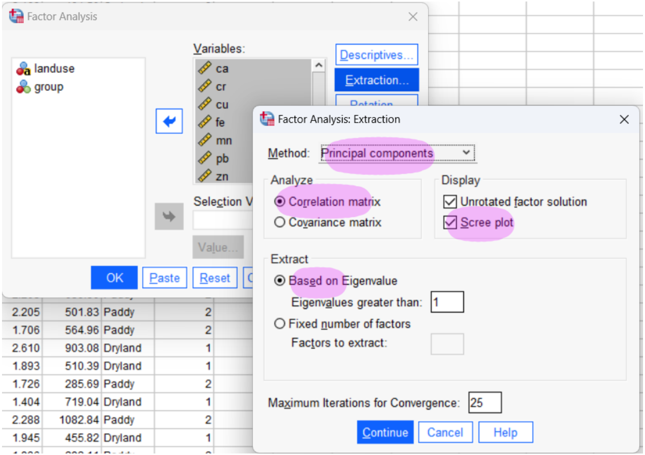
Lab

- Click on `Descriptives`,



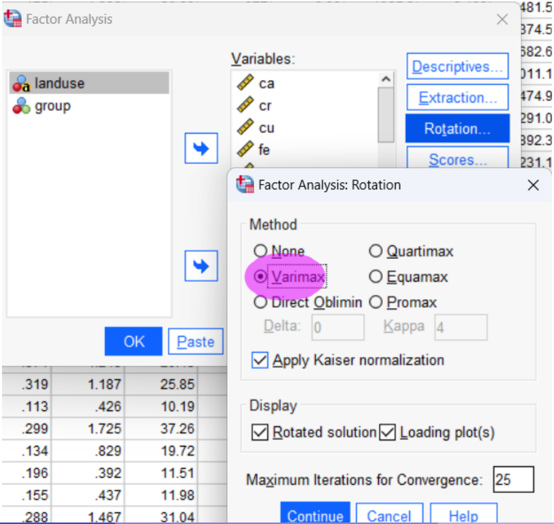
Lab

- Click `Extraction...` and ensure `Principal Components` is selected, with `Scree plot` checked, and `Eigenvalues greater than: 1`.



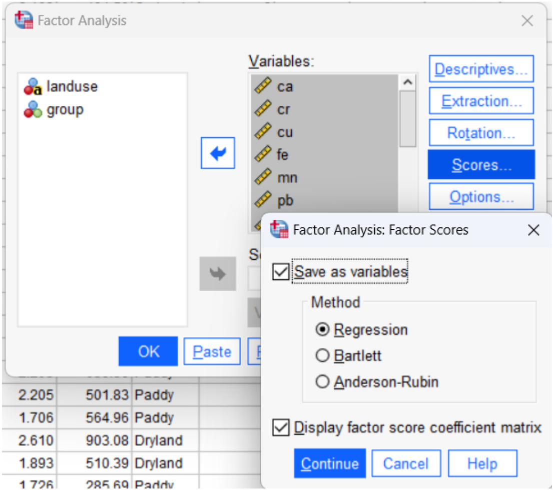
Lab

- Click `Rotation`, choose the `Varimax` rotation to make sure the PCs are orthogonal
  - and ensure `Loading plot(s)` checked,



Lab

- We may save the scores,



Lab

- Click `OK` run the analysis and we check some outputs

Descriptive Statistics

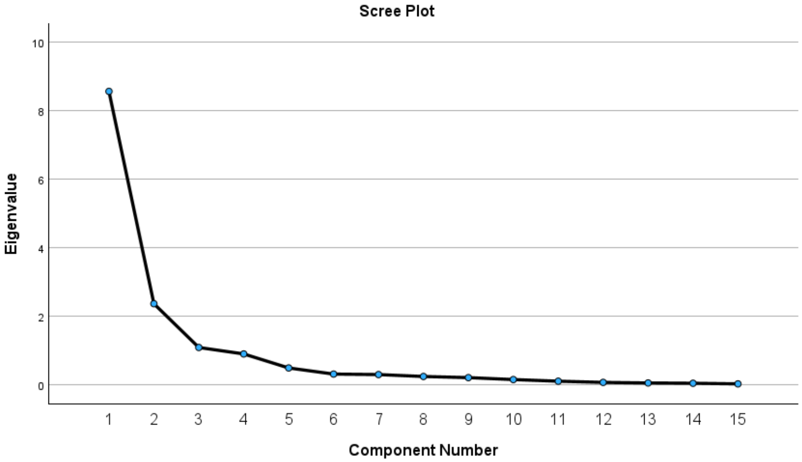
	Mean	Std. Deviation	Analysis N
ca	.74787	.289565	300
cr	64.3875	10.82058	300
cu	20.82442	5.882029	300
fe	3.23216	.665105	300
mn	527.030	198.1942	300
pb	25.5255	2.30803	300
zn	68.2228	18.21474	300
cd	.22119	.069238	300
mg	1.12374	.416714	300
ni	25.7934	7.62909	300
soc	.78680	.208733	300
ph	6.2557	.82683	300
n	888.335	243.6481	300
k	2.06397	.419128	300
p	627.6393	253.08023	300

Lab

Correlation Matrix

		ca	cr	cu	fe	mn	pb	zn	cd	mg	ni	soc	ph	n	k	p
Correlation	ca	1.000	.503	.670	.691	.722	.200	.737	.414	.809	.677	-.242	.686	-.074	.712	.473
	cr	.503	1.000	.843	.869	.590	.646	.763	.483	.787	.897	.024	.306	.250	.735	.252
	cu	.670	.843	1.000	.904	.724	.626	.893	.515	.882	.888	-.057	.346	.175	.813	.466
	fe	.691	.869	.904	1.000	.782	.631	.878	.464	.918	.928	-.111	.347	.151	.844	.363
	mn	.722	.590	.724	.782	1.000	.338	.727	.371	.794	.720	-.340	.383	-.184	.726	.398
	pb	.200	.646	.626	.631	.338	1.000	.589	.487	.470	.608	.311	.024	.500	.494	.202
	zn	.737	.763	.893	.878	.727	.589	1.000	.624	.887	.852	.013	.393	.242	.841	.559
	cd	.414	.483	.515	.464	.371	.487	.624	1.000	.489	.506	.246	.324	.426	.481	.297
	mg	.809	.787	.882	.918	.794	.470	.887	.489	1.000	.917	-.174	.477	.075	.919	.425
	ni	.677	.897	.888	.928	.720	.608	.852	.506	.917	1.000	-.103	.434	.168	.869	.315
	soc	-.242	.024	-.057	-.111	-.340	.311	.013	.246	-.174	-.103	1.000	-.098	.722	-.131	-.042
	ph	.686	.306	.346	.347	.383	.024	.393	.324	.477	.434	-.098	1.000	-.041	.449	.148
	n	-.074	.250	.175	.151	-.184	.500	.242	.426	.075	.168	.722	-.041	1.000	.088	.039
	k	.712	.735	.813	.844	.726	.494	.841	.481	.919	.869	-.131	.449	.088	1.000	.409
	p	.473	.252	.466	.363	.398	.202	.559	.297	.425	.315	-.042	.148	.039	.409	1.000

Lab


Scree Plot
Eigenvalue
10
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Component Number

Lab

- The loadings differ from results by R by a scale

Component Matrix[a]

	Component		
	1	2	3
ca	.792	-.348	.370
cr	.856	.171	-.257
cu	.936	.045	-.138
fe	.947	-.003	-.215
mn	.804	-.355	-.087
pb	.617	.559	-.307
zn	.945	.080	.052
cd	.607	.382	.340
mg	.957	-.137	-.021
ni	.945	.011	-.153
soc	-.075	.859	.264
ph	.488	-.272	.644
n	.180	.885	.169
k	.908	-.090	-.026
p	.487	-.076	.260

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Lab

- The loadings differ from results by R by a scale

Rotated Component Matrix[a]

	Component		
	1	2	3
ca	.598	-.159	.708
cr	.895	.163	
cu	.924		.191
fe	.961		.141
mn	.774	-.295	.309
pb	.698	.500	-.222
zn	.865	.172	.353
cd	.453	.513	.402
mg	.897		.356
ni	.937		.191
soc	-.145	.889	
ph	.218		.824
n	.128	.909	
k	.855		.320
p	.359		.424

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax without Kaiser Normalization.

Lab

Total Variance Explained

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	8.569	57.125	57.125
2	2.372	15.812	72.936
3	1.094	7.290	80.226

Extraction Method: Principal Component Analysis.

Appendix: Derivation of PCs

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_p)^{'}$ is a vector of p random variables, and that the variances of the p random variables and the structure of the covariances/correlations between the p variables are of interest.
- Instead of looking at the p variances and $\frac{1}{2}p(p-1)$ covariances, we look for a few $(< p)$ derived variables that preserve most of the information given by these variances and correlations/covariances.
  - PCA concentrates on variances.
- The first step is to look for a **linear combination** of the elements of $\mathbf{x}$ having **maximum variance**

$$\boldsymbol{v}_1^{'}\mathbf{x} = v_{11}x_1 + \cdots + v_{1p}x_p = \sum_{j=1}^{p} v_{1j}x_j.$$

- Next, look for a linear combination $\boldsymbol{v}_2^{'}\mathbf{x}$, **uncorrelated with** $\boldsymbol{v}_1^{'}\mathbf{x}$ having **maximum variance**, and so on. ... Up to p PCs could be found.
  - It is hoped, in general, that most of the variation in $\mathbf{x}$ will be accounted for by m PCs, where $m << p$.

Appendix: Derivation of PCs

We need **linear algebra** and **matrix calculus** to derive these uncorrelated PCs.

- Derivation of the form of the PCs: note that

$$var(\boldsymbol{v}_1^{'}\mathbf{x}) = \boldsymbol{v}_1^{'}\boldsymbol{\Sigma}\boldsymbol{v}_1,$$

where $\boldsymbol{\Sigma} = var(\mathbf{x})$. So the maximization must be subject to a normalization constraint

$$\boldsymbol{v}_1^{'}\boldsymbol{v}_1 = \sum_{j=1}^{p} v_{1j}^2 = 1.$$

Using the technique of Lagrange multipliers (Calculus III, λ_1 is called the Lagrange Multiplier), We maximize the function

$$\boldsymbol{v}_1^{'}\boldsymbol{\Sigma}\boldsymbol{v}_1 - \lambda_1(\boldsymbol{v}_1^{'}\boldsymbol{v}_1 - 1)$$

w.r.t. $\boldsymbol{v}_1$ by differentiating w.r.t. to $\boldsymbol{v}_1$.

Appendix: Derivation of PCs

This results in

$$\frac{\partial}{\partial \boldsymbol{v}_1}[\boldsymbol{v}_1^{'}\boldsymbol{\Sigma}\boldsymbol{v}_1 - \lambda_1(\boldsymbol{v}_1^{'}\boldsymbol{v}_1 - 1)] = 0$$
$$\boldsymbol{\Sigma}\boldsymbol{v}_1 - \lambda_1\boldsymbol{v}_1 = 0$$

and thus

$$\boldsymbol{\Sigma}\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_1.$$

- This should be recognizable as an eigenvector equation where $\boldsymbol{v}_1$ is an eigenvector of $\boldsymbol{\Sigma}$ and λ_1 is the associated eigenvalue.

Appendix: Derivation of PCs

- Which eigenvector should we choose?

$$\boldsymbol{v}_1^{'}\boldsymbol{\Sigma}\boldsymbol{v}_1 = \boldsymbol{v}_1^{'}\lambda_1\boldsymbol{v}_1 = \lambda_1$$

- Then we should choose λ_1 to be as big as possible. So λ_1 is the largest eigenvector of $\boldsymbol{\Sigma}$ and $\boldsymbol{v}_1$ is the corresponding eigenvector.
- Then the solution to

$$\boldsymbol{\Sigma}\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_1$$

is the 1st PC(principal component) of $\mathbf{x}$.

Appendix: Derivation of PCs

- The second PC, $\boldsymbol{v}_2^{'}\mathbf{x}$ maximizes $\boldsymbol{v}_2^{'}\boldsymbol{\Sigma}\boldsymbol{v}_2$ subject to

$$\boldsymbol{v}_2^{'}\boldsymbol{v}_2 = 1$$

  and being uncorrelated with $\boldsymbol{v}_1^{'}\mathbf{x}$:

$$cov(\boldsymbol{v}_1^{'}\mathbf{x}, \boldsymbol{v}_2^{'}\mathbf{x}) = \boldsymbol{v}_1^{'}\boldsymbol{\Sigma}\boldsymbol{v}_2 = \boldsymbol{v}_2^{'}\boldsymbol{\Sigma}\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_2^{'}\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_1^{'}\boldsymbol{v}_2 = 0.$$

- Using the technique of Lagrange multipliers with these two constraints, we maximize the function w.r.t $\boldsymbol{v}_2$

$$\boldsymbol{v}_2^{'}\boldsymbol{\Sigma}\boldsymbol{v}_2 - \lambda_2(\boldsymbol{v}_2^{'}\boldsymbol{v}_2 - 1) - \phi\boldsymbol{v}_2^{'}\boldsymbol{v}_1$$

Appendix: Derivation of PCs

- Differentiation of this quantity w.r.t. $\boldsymbol{v}_2$ (and setting the result equal to zero) yields

$$\frac{\partial}{\partial \boldsymbol{v}_2}[\boldsymbol{v}_2^{'}\boldsymbol{\Sigma}\boldsymbol{v}_2 - \lambda_2(\boldsymbol{v}_2^{'}\boldsymbol{v}_2 - 1) - \phi\boldsymbol{v}_2^{'}\boldsymbol{v}_1] = 0$$
$$\boldsymbol{\Sigma}\boldsymbol{v}_2 - \lambda_2\boldsymbol{v}_2 - \phi\boldsymbol{v}_1 = 0$$

- If we left multiply $\boldsymbol{v}_1$ into this expression

$$\boldsymbol{v}_1^{'}\boldsymbol{\Sigma}\boldsymbol{v}_2 - \lambda_2\boldsymbol{v}_1^{'}\boldsymbol{v}_2 - \phi\boldsymbol{v}_1^{'}\boldsymbol{v}_1 = 0$$
$$0 - 0 - \phi = 0$$

then we can see that ϕ must be zero.

Appendix: Derivation of PCs

- So we have
$$\boldsymbol{\Sigma} \boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2.$$
Furthermore,
$$var(\boldsymbol{v}_2^{'}\mathbf{x}) = \boldsymbol{v}_2^{'}\lambda_2\boldsymbol{v}_2 = \lambda_2\boldsymbol{v}_2^{'}\boldsymbol{v}_2 = \lambda_2.$$

Appendix: Derivation of PCs

- Thus, the second PC $\boldsymbol{v}_2^{'}\mathbf{x}$ is the solution to

$$\boldsymbol{\Sigma}\boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2,$$

  where λ_2 is the second largest eigenvalue of $\boldsymbol{\Sigma}$.
- This process can be repeated for $k = 1, 2, \ldots, p$ yielding up to p different eigenvectors of $\boldsymbol{\Sigma}$ along with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_p$.

Solution:

- Eigenvalue decomposition of $\mathbf{X}^{'}\mathbf{X}$. (`eigen()` in `R`).

Appendix: Derivation of PCs

In summary,

- The first PC is $\boldsymbol{v}_1^{'}\mathbf{x} = v_{11}\mathbf{x}_1 + \cdots + v_{1p}\mathbf{x}_p = \sum_{j=1}^{p} v_{1j}\mathbf{x}_j$, where $\boldsymbol{v}_1$ is the eigenvector correspoding to the largest eigenvalue of $\mathbf{X}^{'}\mathbf{X}$..
- The second PC is $\boldsymbol{v}_2^{'}\mathbf{x} = v_{21}\mathbf{x}_1 + \cdots + v_{2p}\mathbf{x}_p = \sum_{j=1}^{p} v_{2j}\mathbf{x}_j$, where $\boldsymbol{v}_2$ is the eigenvector correspoding to the second largest eigenvalue of $\mathbf{X}^{'}\mathbf{X}$.
- and so on.
    - The first PC is the linear combination of features that maximizes the variance.
    - Subsequent linear combinations are orthogonal to previous combinations which maximizes the variance.

License