02409 Multivariate Statistics

Lecture D, September 22 2025

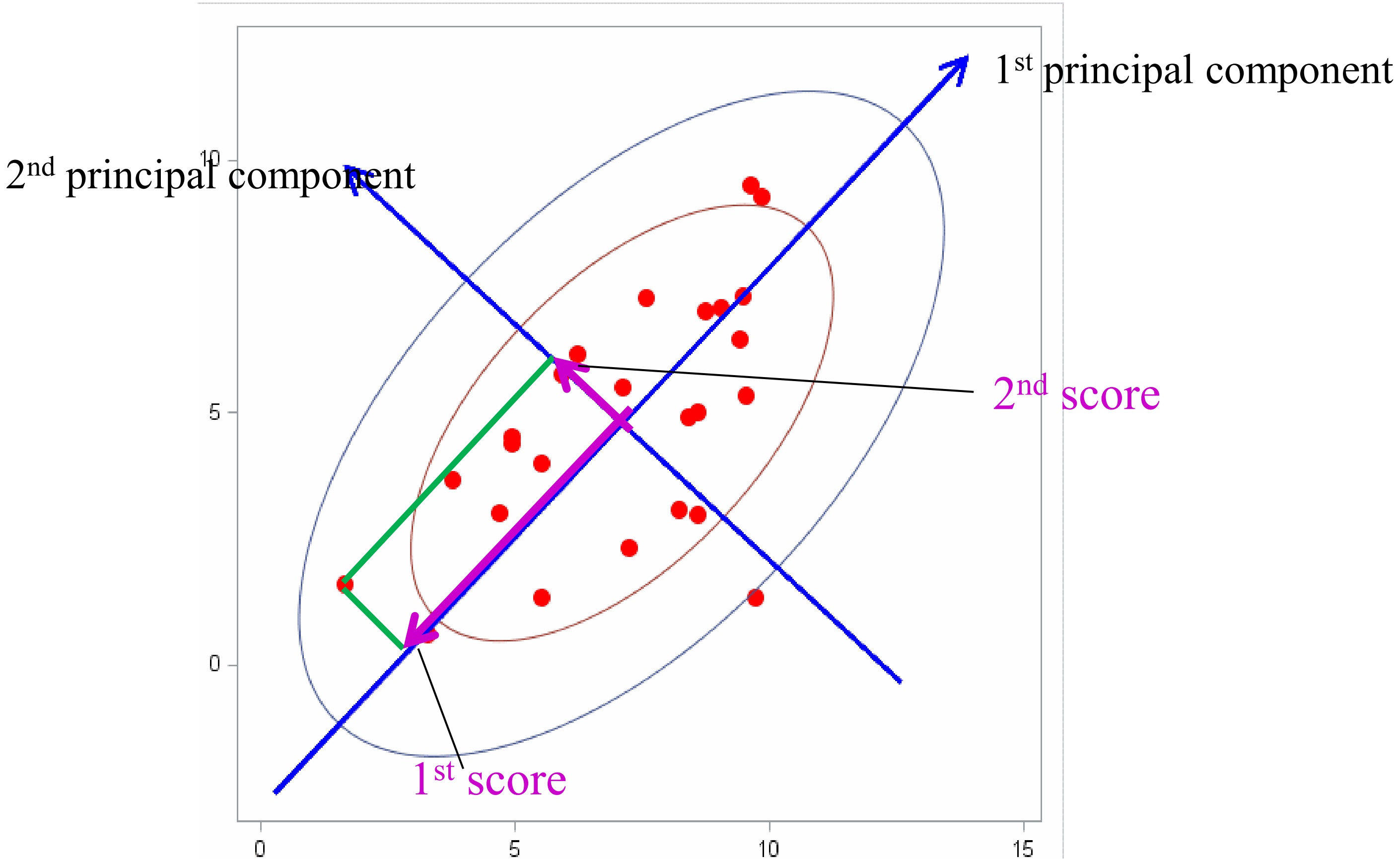
Anders Stockmarr
anst@dtu.dk

Course developers:
Anders Stockmarr
Anders Nymark Christensen

Agenda

- PCA review
  - The math
  - Some examples of use
- Factor analysis
  - Examples
  - The math

Principal Components I



The **scores** of the observation x_i of a multidimensional random variable X are the projections of x_i (adjusted for overall mean) on the principal components of the variance matrix of X

Principle Components II

- A good explanation of PCA in terms of Wine: https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

Picture: Winefolly.com

Some notation and background on eigenvalues and –vectors I

Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \text{ with } V(\boldsymbol{X}) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1k} \\ \vdots & & \vdots \\ \sigma_{k1} & \cdots & \sigma_k^2 \end{bmatrix}, \qquad \sigma_{ij} = \sigma_{ji}$$

Let the eigenvalues of $\boldsymbol{\Sigma}$ be $\lambda_1 \geq \cdots \geq \lambda_k$ and the corresponding eigenvectors $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$. Organize those in matrices

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}$$

$$\boldsymbol{P} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_k]$$

Then

$$\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}$$

$$\begin{bmatrix} 0.7662 & 0.4544 & 0.4544 \\ 0 & 0.7071 & -0.7071 \\ -0.6426 & 0.5418 & 0.5418 \end{bmatrix} \begin{bmatrix} 0.7662 & 0 & -0.6426 \\ 0.4544 & 0.7071 & 0.5418 \\ 0.4544 & -0.7071 & 0.5418 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_3 \end{bmatrix} \text{ with } V(\boldsymbol{X}) = \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & -0.25 \\ 0.5 & -0.25 & 1 \end{bmatrix},$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1.5931 & \cdots & 0 \\ \vdots & 1.25 & \vdots \\ 0 & \cdots & 0.1569 \end{bmatrix}$$

$$\boldsymbol{P} = \begin{bmatrix} 0.7662 & 0 & -0.6426 \\ 0.4544 & 0.7071 & 0.5418 \\ 0.4544 & -0.7071 & 0.5418 \end{bmatrix}$$

Roughly the average

Contrast between x2 and x3

Contrast between x1 against x2 and x3

Some notation and background on eigenvalues and –vectors II

lambda is the element from the random variable's variance-covariance matrix (Σ), obtained as an eigenvalue in the decomposition $\Lambda = P^T\Sigma P$

$$\boldsymbol{\Lambda} = \mathbf{P}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{P}$$

$$\begin{bmatrix} 1.5931 & \cdots & 0 \\ \vdots & 1.25 & \vdots \\ 0 & \cdots & 0.1569 \end{bmatrix} = \begin{bmatrix} 0.7662 & 0.4544 & 0.4544 \\ 0 & 0.7071 & -0.7071 \\ -0.6426 & 0.5418 & 0.5418 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & -0.25 \\ 0.5 & -0.25 & 1 \end{bmatrix} \begin{bmatrix} 0.7662 & 0 & -0.6426 \\ 0.4544 & 0.7071 & 0.5418 \\ 0.4544 & -0.7071 & 0.5418 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\boldsymbol{T}} = \lambda_1 \boldsymbol{p}_1 \boldsymbol{p}_1^T + \cdots + \lambda_k \boldsymbol{p}_k \boldsymbol{p}_k^T$$

The variance component 1 explains about x2

$1.5931 \cdot 0.4544 \cdot 0.4544$

$$\lambda_1: \quad 1.5931 \begin{bmatrix} 0.7662 \\ 0.4544 \\ 0.4544 \end{bmatrix} [0.7662\ 0.4544\ 0.4544] = \begin{bmatrix} 0.9352 & 0.5546 & 0.5546 \\ 0.5546 & 0.3289 & 0.3289 \\ 0.5546 & 0.3289 & 0.3289 \end{bmatrix}$$

$$\lambda_2: \quad 1.25 \begin{bmatrix} 0 \\ 0.7071 \\ -0.7071 \end{bmatrix} [0\ 0.7071\ -0.7071] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.6250 & -0.6250 \\ 0 & -0.6250 & 0.6250 \end{bmatrix}$$

$$\lambda_3: \quad 0.1569 \begin{bmatrix} -0.6426 \\ 0.5418 \\ -0.5418 \end{bmatrix} [-0.6426\ 0.5418\ 0.5418] = \begin{bmatrix} 0.0648 & -0.0546 & -0.0546 \\ -0.0546 & 0.0461 & 0.0461 \\ -0.0546 & 0.0461 & 0.0461 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\boldsymbol{T}} = \lambda_1 \boldsymbol{p}_1 \boldsymbol{p}_1^T + \cdots + \lambda_k \boldsymbol{p}_k \boldsymbol{p}_k^T =$$

$$\begin{bmatrix} 0.9352 & 0.5546 & 0.5546 \\ 0.5546 & 0.3289 & 0.3289 \\ 0.5546 & 0.3289 & 0.3289 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.6250 & -0.6250 \\ 0 & -0.6250 & 0.6250 \end{bmatrix} + \begin{bmatrix} 0.0648 & -0.0546 & -0.0546 \\ -0.0546 & 0.0461 & 0.0461 \\ -0.0546 & 0.0461 & 0.0461 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & -0.25 \\ 0.5 & -0.25 & 1 \end{bmatrix}$$

Question 4.2.

The fraction of the variation of slo910 that is explained by the first principal component may be written as $c\lambda_1$ where c is:

1 ☒ 0.1

2 ☐ 0.2

3 ☐ 0.3

4 ☐ 0.4

5 ☐ 0.5

6 ☐ Don't know.

-0.318683 * -0.318683 =0.1

Obs	_NAME_	slo500	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
1	slo500	1.00000	0.54123	0.50443	0.67020	0.62326	0.39537	0.12104	0.26611	-0.17035	-0.21488
2	slo580	0.54123	1.00000	0.96360	0.93989	0.92607	-0.22834	-0.63711	-0.50498	-0.75865	-0.81688
3	slo750	0.50443	0.96360	1.00000	0.96283	0.95643	-0.24279	-0.62413	-0.54306	-0.80164	-0.85952
4	slo830	0.67020	0.93989	0.96283	1.00000	0.99155	-0.00107	-0.42261	-0.30682	-0.65176	-0.71651
5	slo910	0.62326	0.92607	0.95643	0.99155	1.00000	-0.00749	-0.40729	-0.30009	-0.62054	-0.69500
6	int500	0.39537	-0.22834	-0.24279	-0.00107	-0.00749	1.00000	0.85029	0.90585	0.68150	0.65104
7	int580	0.12104	-0.63711	-0.62413	-0.42261	-0.40729	0.85029	1.00000	0.95477	0.91718	0.89811
8	int750	0.26611	-0.50498	-0.54306	-0.30682	-0.30009	0.90585	0.95477	1.00000	0.86818	0.85097
9	int830	-0.17035	-0.75865	-0.80164	-0.65176	-0.62054	0.68150	0.91718	0.86818	1.00000	0.98827
10	int910	-0.21488	-0.81688	-0.85952	-0.71651	-0.69500	0.65104	0.89811	0.85097	0.98827	1.00000

Eigenvectors										
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10
slo500	-.130694	0.487662	-.811231	0.109990	0.033337	0.153203	0.138534	-.148265	-.092048	0.026123
slo580	-.357271	0.189657	0.126734	0.522397	-.446694	0.167612	-.333231	0.414583	0.021404	0.190834
slo750	-.365486	0.179262	0.222800	-.055600	0.184634	0.290129	-.086913	-.470507	0.648618	0.116843
slo830	-.324513	0.319584	0.183316	-.052294	0.116497	-.093938	0.275549	0.309360	0.066098	-.746875
slo910	-.318683	0.313912	0.316282	0.088061	0.374134	-.351926	0.177417	-.110976	-.450071	0.426529
int500	0.200416	0.478488	0.283326	-.464216	-.543971	0.230356	0.085188	-.164085	-.217408	0.047501
int580	0.326552	0.304564	0.048824	-.197396	0.529566	0.346538	-.298041	0.493986	0.051325	0.152453
int750	0.297465	0.375375	-.049354	0.030567	-.099413	-.715831	-.333998	-.036931	0.363050	-.014743
int830	0.366532	0.142771	0.208174	0.558954	0.152600	0.208461	-.207220	-.422963	-.262308	-.363482
int910	0.377829	0.102013	0.118889	0.365509	-.046913	0.050611	0.710090	0.159796	0.336035	0.227297

The spectral decomposition of Σ *is* (*p.* **458**)

$$\Sigma = \lambda_1 \boldsymbol{p}_1 \boldsymbol{p}_1^T + \cdots + \lambda_m \boldsymbol{p}_m \boldsymbol{p}_m^T + \lambda_{m+1} \boldsymbol{p}_{m+1} \boldsymbol{p}_{m+1}^T + \cdots + \lambda_k \boldsymbol{p}_k \boldsymbol{p}_k^T$$

The fraction of the variation of slo910 that may be explained by the first principal component is on the form $\lambda_1 c$, where c is approximately:

A: 0.1 0%

B: 0.2 0%

C: 0.3 0%

D: 0.4 0%

E: 0.5 0%

F: Don't know 0%

Principal components II

Assume that the variance matrix of X is **known**. The vector $\boldsymbol{Y}$ of **scores** of $\boldsymbol{X}$ is

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} = \boldsymbol{P}^T \boldsymbol{X} = \begin{bmatrix} \boldsymbol{p}_1^T \\ \vdots \\ \boldsymbol{p}_k^T \end{bmatrix} \boldsymbol{X}$$

$$V(\boldsymbol{Y}) = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}$$

i.e. the **scores are uncorrelated and** $V(Y_i) = \lambda_i$.

Let $m \leq k$ and define

$$\boldsymbol{P}^{(m)} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m]$$

$$\boldsymbol{\Lambda}^{(m)} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$$

Then the vector of the **first m scores** is trivially

$$\boldsymbol{Y}^{(m)} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} = \boldsymbol{P}^{(m)T} \boldsymbol{X} = \begin{bmatrix} \boldsymbol{p}_1^T \\ \vdots \\ \boldsymbol{p}_m^T \end{bmatrix} \boldsymbol{X}$$

Theorem 6.5

The total variance i.e. the sum of variance of the original variables is equal to the sum of the variance of the principal components i.e.

$$\sum_i \mathrm{V}(X_i) = \sum_i \mathrm{V}(Y_i)$$

The correlations between the original variables and the first m scores:

$$Cov\left(\boldsymbol{X}, \boldsymbol{Y}^{(m)}\right) = \begin{bmatrix} Cov(X_1, Y_1) & \cdots & Cov(X_1, Y_m) \\ \vdots & & \vdots \\ Cov(X_k, Y_1) & \cdots & Cov(X_k, Y_m) \end{bmatrix} = Cov\left(\boldsymbol{X}, \boldsymbol{P}^{(m)^T}\boldsymbol{X}\right) = Cov(\boldsymbol{X}, \boldsymbol{X})\boldsymbol{P}^{(m)} = V(\boldsymbol{X})\boldsymbol{P}^{(m)} = \boldsymbol{\Sigma}\boldsymbol{P}^{(m)} = \boldsymbol{P}\Lambda\boldsymbol{P}^T\boldsymbol{P}^{(m)}$$

The inner ptoduckt

$$= [\boldsymbol{p}_1 \quad \cdots \quad \boldsymbol{p}_k] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_1^T \\ \vdots \\ \boldsymbol{p}_k^T \end{bmatrix} [\boldsymbol{p}_1 \quad \cdots \quad \boldsymbol{p}_m]$$

$$= [\lambda_1\boldsymbol{p}_1 \quad \cdots \quad \lambda_k\boldsymbol{p}_k] \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} = [\lambda_1\boldsymbol{p}_1 \quad \cdots \quad \lambda_m\boldsymbol{p}_m]$$

From page 8:

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

$$Corr\left(\boldsymbol{X}, \boldsymbol{Y}^{(m)}\right) = \begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\sigma_k} \end{bmatrix} [\lambda_1\boldsymbol{p}_1 \quad \cdots \quad \lambda_m\boldsymbol{p}_m] \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_m}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\sigma_k} \end{bmatrix} \left[\sqrt{\lambda_1}\boldsymbol{p}_1 \quad \cdots \quad \sqrt{\lambda_m}\boldsymbol{p}_m\right]$$

Reconstructing the original observations from the first m principal components, assuming a known covariance.

Remark 6.7: Since $\boldsymbol{X} = \boldsymbol{PY}$ an obvious way of **reconstructing $\boldsymbol{X}$ from the first m scores $\boldsymbol{Y}^{(m)}$** is

$$\boldsymbol{X}^* = \boldsymbol{P}^{(m)}\boldsymbol{Y}^{(m)} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m]\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} = Y_1\boldsymbol{p}_1 + \cdots + Y_m\boldsymbol{p}_m$$

The variance matrix of the reconstructed vector is

$$V(X^*) = PV(Y^*)P^T$$

The spectral decomposition of Σ is $(p.\,\mathbf{458})$

$$\Sigma = \lambda_1\boldsymbol{p}_1\boldsymbol{p}_1^T + \cdots + \lambda_m\boldsymbol{p}_m\boldsymbol{p}_m^T + \lambda_{m+1}\boldsymbol{p}_{m+1}\boldsymbol{p}_{m+1}^T + \cdots + \lambda_k\boldsymbol{p}_k\boldsymbol{p}_k^T$$

which means that

$$\Sigma - V(X^*) = \lambda_{m+1}\boldsymbol{p}_{m+1}\boldsymbol{p}_{m+1}^T + \cdots + \lambda_k\boldsymbol{p}_k\boldsymbol{p}_k^T$$

Reconstructing the original observations from the first m principal components, assuming a known covariance.

Remark 6.7: Since $\boldsymbol{X} = \boldsymbol{PY}$ an obvious way of **reconstructing $\boldsymbol{X}$ from the first m scores $\boldsymbol{Y}^{(m)}$** is

$$\boldsymbol{X}^* = \boldsymbol{P}^{(m)}\boldsymbol{Y}^{(m)} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m]\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} = Y_1\boldsymbol{p}_1 + \cdots + Y_m\boldsymbol{p}_m$$

The variance matrix of the reconstructed vector is

$$V(X^*) = PV(Y^*)P^T$$

The spectral decomposition of Σ *is* $(p.\,\mathbf{458})$

$$\Sigma = \lambda_1\boldsymbol{p}_1\boldsymbol{p}_1^T + \cdots + \lambda_m\boldsymbol{p}_m\boldsymbol{p}_m^T + \lambda_{m+1}\boldsymbol{p}_{m+1}\boldsymbol{p}_{m+1}^T + \cdots + \lambda_k\boldsymbol{p}_k\boldsymbol{p}_k^T$$

which means that

$$\Sigma - V(X^*) = \lambda_{m+1}\boldsymbol{p}_{m+1}\boldsymbol{p}_{m+1}^T + \cdots + \lambda_k\boldsymbol{p}_k\boldsymbol{p}_k^T$$

Example: Take (slide 5)

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & -0.25 \\ 0.5 & -0.25 & 1 \end{bmatrix}$$

Note the numerical values of the entries; much smaller than for the first two terms.

$$\boldsymbol{\Sigma} = \mathbf{P\Lambda P}^T = \lambda_1\boldsymbol{p}_1\boldsymbol{p}_1^T + \cdots + \lambda_k\boldsymbol{p}_k\boldsymbol{p}_k^T =$$

$$\begin{bmatrix} 0.9352 & 0.5546 & 0.5546 \\ 0.5546 & 0.3289 & 0.3289 \\ 0.5546 & 0.3289 & 0.3289 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.6250 & -0.6250 \\ 0 & -0.6250 & 0.6250 \end{bmatrix} + \begin{bmatrix} 0.0648 & -0.0546 & -0.0546 \\ -0.0546 & 0.0461 & 0.0461 \\ -0.0546 & 0.0461 & 0.0461 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & -0.25 \\ 0.5 & -0.25 & 1 \end{bmatrix}$$

Exam 2016 problem 4

The data here are based on sub-sampled digital images from the Thematic Mapper (TM) scanner onboard the Landsat series of satellites. We have data covering the greater Copenhagen area from two time points, 27 June 1986 and 18 August 2005. The data are geometrically co-registered so there is time-wise pixel correspondance, i.e., pixel (i, j) in 1986 represents the same area on the ground as pixel (i, j) in 2005. The variables are digital numbers representing reflected sunlight from the surface of the earth, measured in six wavelength bands, namely RGB - visible blue, green and red; and three infrared bands.

We label the six spectral bands from time point 1, 1985 X1-X6;

and the six spectral bands from time point 2, 2005 Y1-Y6.

Assume that the variance of (X, Y) is known.

Exam 2016 problem 4

The variance of the first score is the eigenvalue

Eigenvalues of the correlation matrix			
	Eigenvalue	Proportion	Cumulative
1	$\boxed{9.01626784}$	0.7514	0.7514
2	1.62996970	0.1358	0.8872
3	0.87459018	0.0729	0.9601
4	0.14911589	0.0124	0.9725
5	0.08653455	0.0072	0.9797
6	0.07875169	0.0066	0.9863
7	0.06162583	0.0051	0.9914
8	0.03445767	0.0029	0.9943
9	0.03319704	0.0028	0.9970
10	0.02235388	0.0019	0.9989
11	0.00796802	0.0007	0.9996
12	0.00516771	0.0004	1.0000

Exam 2016 problem 4

Principal components for the correlation matrix												
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
X_1	0.265382	-0.344758	0.341909	-0.542694	0.225310	0.417245	0.262426	0.201407	-0.024140	0.241094	0.072036	0.033266
X_2	0.304236	0.048394	0.337080	-0.044573	-0.761246	0.263464	-0.249231	-0.257100	0.042231	-0.090476	-0.003678	-0.048798
X_3	0.274342	0.434404	0.000887	-0.001131	-0.037508	-0.140833	-0.208649	0.298727	-0.101839	0.413721	0.220889	0.593032
X_4	0.293249	0.356342	0.052250	-0.070095	0.164756	-0.083087	-0.294205	0.272541	-0.059147	0.181582	-0.170100	-0.720482
X_5	0.305738	0.264857	0.089019	0.089019	0.413683	0.072865	-0.178829	-0.250859	-0.207321	-0.667211	0.058482	0.179070
X_6	0.253710	0.477851	-0.088451	-0.088451	-0.098089	-0.056597	0.767310	-0.179396	0.237796	0.002316	-0.059657	-0.065209
Y_1	0.265175	-0.157106	0.531964	0.690452	0.195213	-0.045306	0.112292	0.192174	0.207850	-0.099236	0.043122	0.026085
Y_2	0.291196	-0.328656	0.100210	-0.109384	-0.140987	-0.682585	0.168448	-0.096008	-0.498470	0.016830	-0.113453	-0.012796
Y_3	0.294705	-0.235797	-0.354534	-0.104648	-0.119393	-0.184028	-0.044614	0.229510	0.356028	-0.241625	0.639643	-0.154384
Y_4	0.310159	-0.197091	-0.247105	-0.107818	0.015822	-0.089014	-0.167410	0.114894	0.468531	-0.075725	-0-675043	0.248861
Y_5	0.312982	-0.161072	-0.213937	0.187527	0.272425	0.088748	-0.153441	-0.678689	0.070818	0.450246	0.135731	-0.052431
Y_6	0.285988	-0.107409	-0.474810	0.342581	-0.117167	0.451223	0.153737	0.249124	-0.491192	-0.086058	-0.110056	-0.007234

The variance of the first score is

A: 0.26538		0%
B: 9.01626784	Right answer	0%
C: 0.7514		0%
D: 1		0%
E: Don't know		0%

The variance of the first score is

Option B

Eigenvalues of the correlation matrix			
	Eigenvalue	Proportion	Cumulative
1	9.01626784	0.7514	0.7514
2	1.62996970	0.1358	0.8872
3	0.87459018	0.0729	0.9601
4	0.14911589	0.0124	0.9725
5	0.08653455	0.0072	0.9797
6	0.07875169	0.0066	0.9863
7	0.06162583	0.0051	0.9914
8	0.03445767	0.0029	0.9943
9	0.03319704	0.0028	0.9970
10	0.02235388	0.0019	0.9989
11	0.00796802	0.0007	0.9996
12	0.00516771	0.0004	1.0000

If we want to retain 95% of the variation of X, we should retain how many principal components?

A: 1 0%

B: 2 0%

C: 3 0%

D: 4 0%

E: 11 0%

F: Don't know 0%

If we want to retain at least 95% of the correlation, we should retain how many principal components?

Eigenvalues of the correlation matrix			
	Eigenvalue	Proportion	Cumulative
1	9.01626784	0.7514	0.7514
2	1.62996970	0.1358	0.8872
3	0.87459018	0.0729	0.9601
4	0.14911589	0.0124	0.9725
5	0.08653455	0.0072	0.9797
6	0.07875169	0.0066	0.9863
7	0.06162583	0.0051	0.9914
8	0.03445767	0.0029	0.9943
9	0.03319704	0.0028	0.9970
10	0.02235388	0.0019	0.9989
11	0.00796802	0.0007	0.9996
12	0.00516771	0.0004	1.0000

Option C

A good detector of changes over time will measure the contrasts between the 1986 data (X) and the 2005 data (Y). Which principal component is best change detector?

A: 1	0%
B: 2	0%
C: 3	0%
D: 4	0%
E: 11	0%
F: Don't know	0%

A good detector of changes over time will measure the contrasts between the 1986 data (X) and the 2005 data (Y). Which principal component is best change detector?

PC 2 resembles X-Y

Principal components for the correlation matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
X_1	0.265382	-0.344758	0.341909	-0.542694	0.225310	0.417245	0.262426	0.201407	-0.024140	0.241094	0.072036	0.033266
X_2	0.304236	0.048394	0.337080	-0.044573	-0.761246	0.263464	-0.249231	-0.257100	0.042231	-0.090476	-0.003678	-0.048798
X_3	0.274342	0.434404	0.000887	-0.001131	-0.037508	-0.140833	-0.208649	0.298727	-0.101839	0.413721	0.220889	0.593032
X_4	0.293249	0.356342	0.052250	-0.070095	0.164756	-0.083087	-0.294205	0.272541	-0.059147	0.181582	-0.170100	-0.720482
X_5	0.305738	0.264857	0.089019	0.089019	0.413683	0.072865	-0.178829	-0.250859	-0.207321	-0.667211	0.058482	0.179070
X_6	0.253710	0.477851	-0.088451	-0.088451	-0.098089	-0.056597	0.767310	-0.179396	0.237796	0.002316	-0.059657	-0.065209
Y_1	0.265175	-0.157106	0.531964	0.690452	0.195213	-0.045306	0.112292	0.192174	0.207850	-0.099236	0.043122	0.026085
Y_2	0.291196	-0.328656	0.100210	-0.109384	-0.140987	-0.682585	0.168448	-0.096008	-0.498470	0.016830	-0.113453	-0.012796
Y_3	0.294705	-0.235797	-0.354534	-0.104648	-0.119393	-0.184028	-0.044614	0.229510	0.356028	-0.241625	0.639643	-0.154384
Y_4	0.310159	-0.197091	-0.247105	-0.107818	0.015822	-0.089014	-0.167410	0.114894	0.468531	-0.075725	-0-675043	0.248861
Y_5	0.312982	-0.161072	-0.213937	0.187527	0.272425	0.088748	-0.153441	-0.678689	0.070818	0.450246	0.135731	-0.052431
Y_6	0.285988	-0.107409	-0.474810	0.342581	-0.117167	0.451223	0.153737	0.249124	-0.491192	-0.086058	-0.110056	-0.007234

Which principal component overall reflects the overall pixel intensity of the scores?

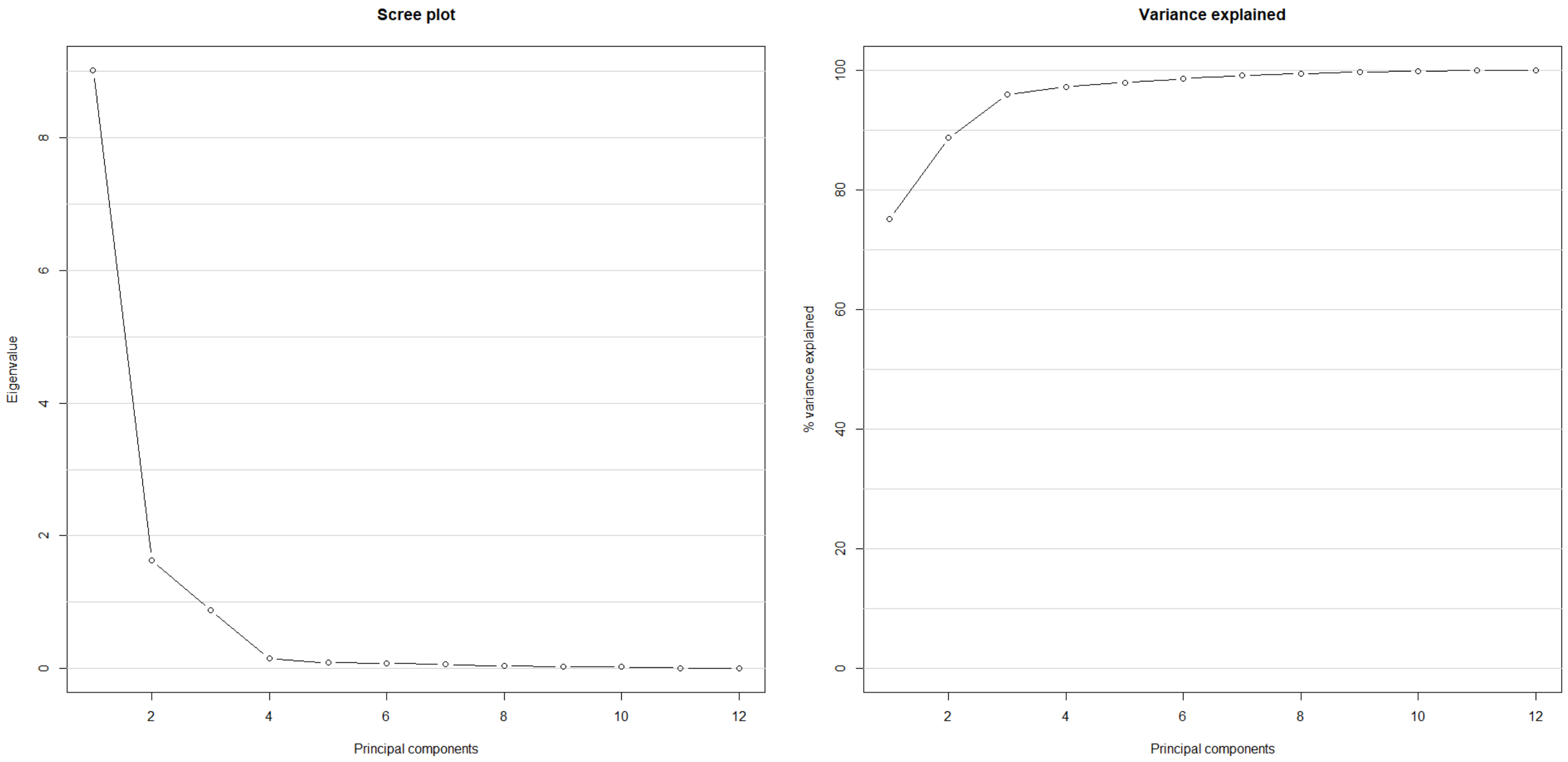
A: 1	0%
B: 2	0%
C: 3	0%
D: 4	0%
E: 11	0%
F: Don't know	0%

Which principal component overall reflects the overall pixel intensity of the scores?

PC 1 has more or less identical values

Principal components for the correlation matrix												
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**	**PC11**	**PC12**
X_1	0.265382	-0.344758	0.341909	-0.542694	0.225310	0.417245	0.262426	0.201407	-0.024140	0.241094	0.072036	0.033266
X_2	0.304236	0.048394	0.337080	-0.044573	-0.761246	0.263464	-0.249231	-0.257100	0.042231	-0.090476	-0.003678	-0.048798
X_3	0.274342	0.434404	0.000887	-0.001131	-0.037508	-0.140833	-0.208649	0.298727	-0.101839	0.413721	0.220889	0.593032
X_4	0.293249	0.356342	0.052250	-0.070095	0.164756	-0.083087	-0.294205	0.272541	-0.059147	0.181582	-0.170100	-0.720482
X_5	0.305738	0.264857	0.089019	0.089019	0.413683	0.072865	-0.178829	-0.250859	-0.207321	-0.667211	0.058482	0.179070
X_6	0.253710	0.477851	-0.088451	-0.088451	-0.098089	-0.056597	0.767310	-0.179396	0.237796	0.002316	-0.059657	-0.065209
Y_1	0.265175	-0.157106	0.531964	0.690452	0.195213	-0.045306	0.112292	0.192174	0.207850	-0.099236	0.043122	0.026085
Y_2	0.291196	-0.328656	0.100210	-0.109384	-0.140987	-0.682585	0.168448	-0.096008	-0.498470	0.016830	-0.113453	-0.012796
Y_3	0.294705	-0.235797	-0.354534	-0.104648	-0.119393	-0.184028	-0.044614	0.229510	0.356028	-0.241625	0.639643	-0.154384
Y_4	0.310159	-0.197091	-0.247105	-0.107818	0.015822	-0.089014	-0.167410	0.114894	0.468531	-0.075725	-0-675043	0.248861
Y_5	0.312982	-0.161072	-0.213937	0.187527	0.272425	0.088748	-0.153441	-0.678689	0.070818	0.450246	0.135731	-0.052431
Y_6	0.285988	-0.107409	-0.474810	0.342581	-0.117167	0.451223	0.153737	0.249124	-0.491192	-0.086058	-0.110056	-0.007234

The Scree and Variance Explained Plots



The **Scree and Variance explained plots** show the eigenvalues $\hat{\lambda}_i$ and the cumulative proportion of variance explained, $\frac{\hat{\lambda}_1+\cdots+\hat{\lambda}_m}{\hat{\lambda}_1+\cdots+\hat{\lambda}_m+\hat{\lambda}_{m+1}+\cdots+\hat{\lambda}_k}$

Selection of number of components

We consider the heptathlon data from exercise 7.1

1. **Variance explained**
   If we have some criteria, e.g. 80% of the variance explained we simply select that number of components (2)

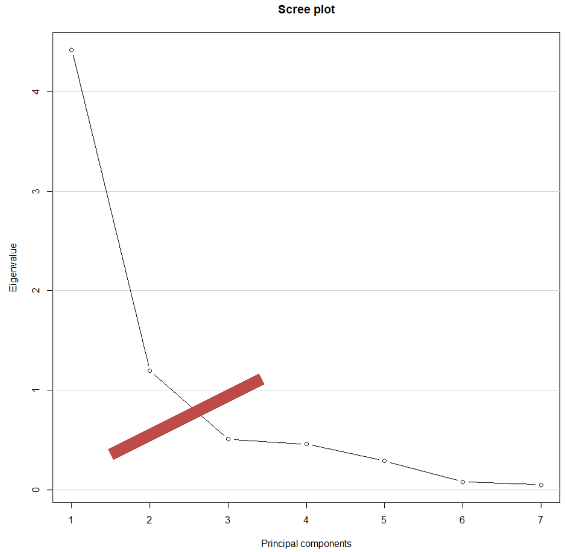
2. **All componenets with eigenvalue larger than 1**
   If we have used the correlation matrix, the componenets with eigenvalue larger than 1 explains more of the variance than the original variable (2)

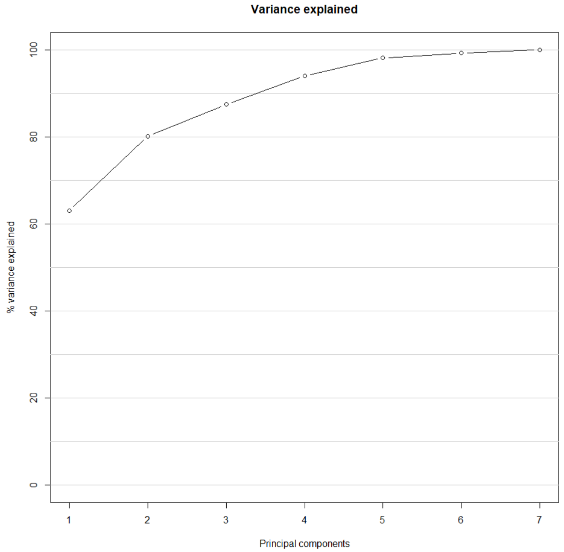
3. **Scree plot**
   Find the components above the scree (2)

4. **Permutation tests**
   Not covered in this course

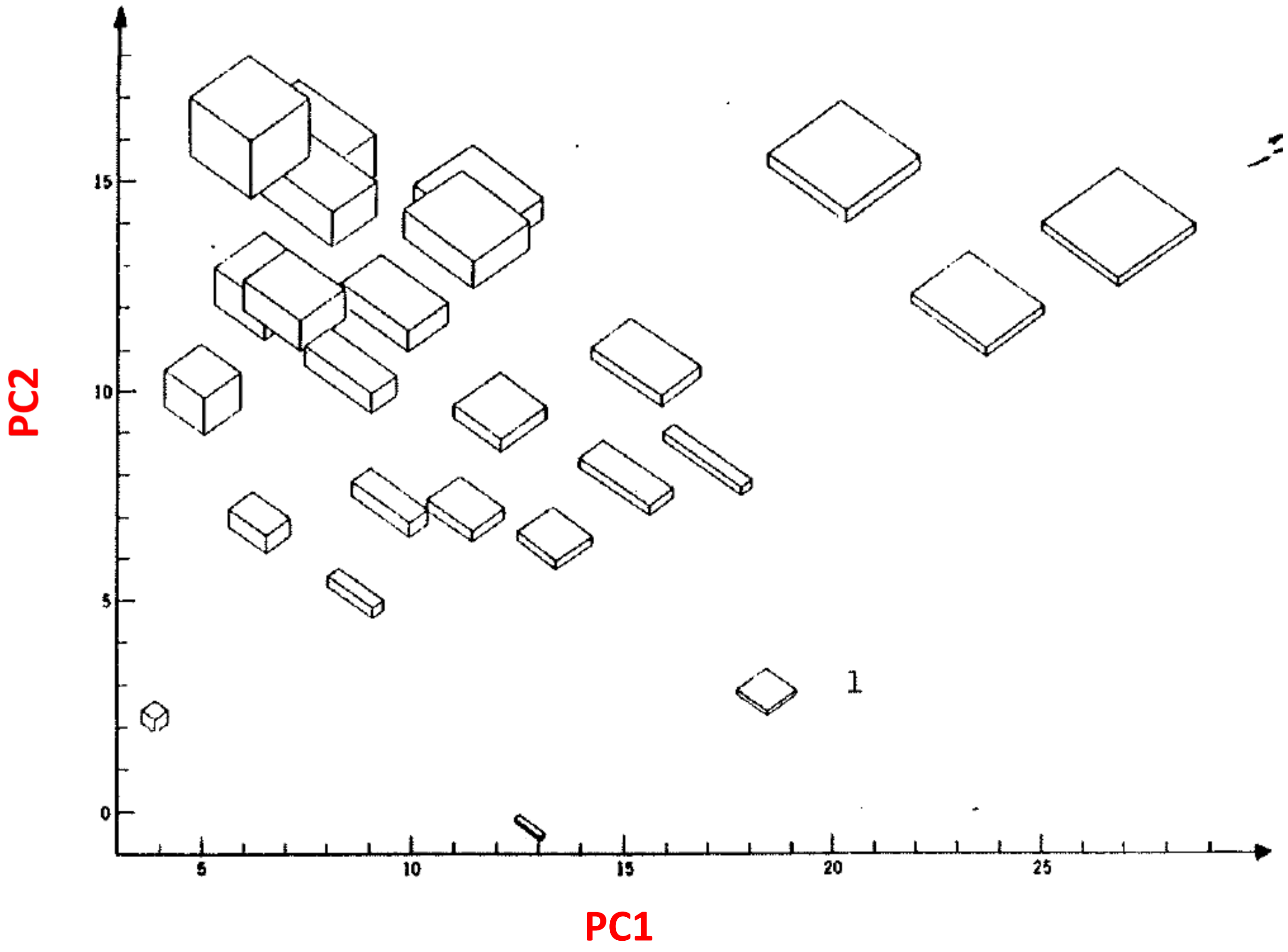
Eigenvalues of the Correlation Matrix

	Eigenvalue	Proportion	Cumulative
1	4.41699433	0.6310	0.6310
2	1.19602462	0.1709	0.8019
3	0.51002343	0.0729	0.8747
4	0.45861676	0.0655	0.9402
5	0.28957437	0.0414	0.9816
6	0.07902477	0.0113	0.9929
7	0.04974172	0.0071	1.0000



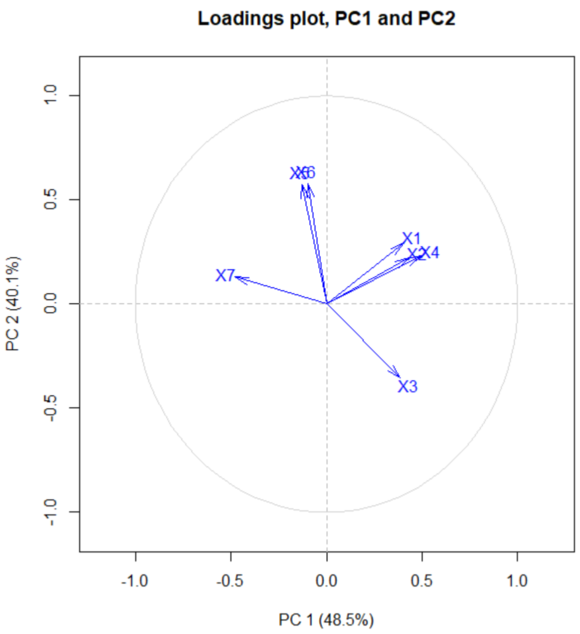


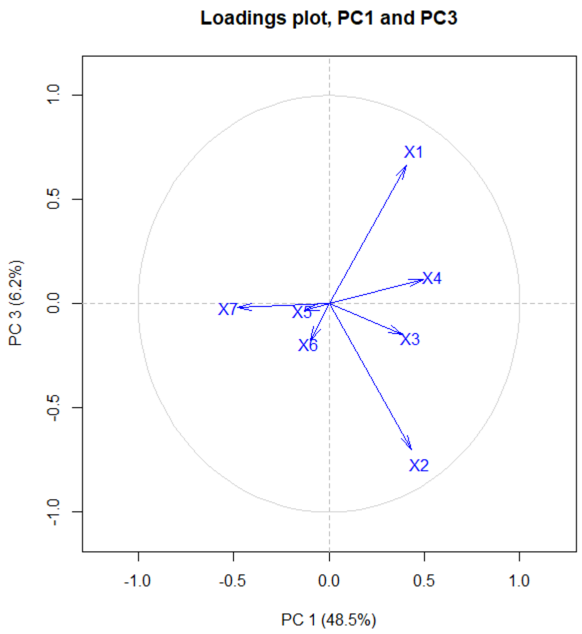
Revisiting the box data



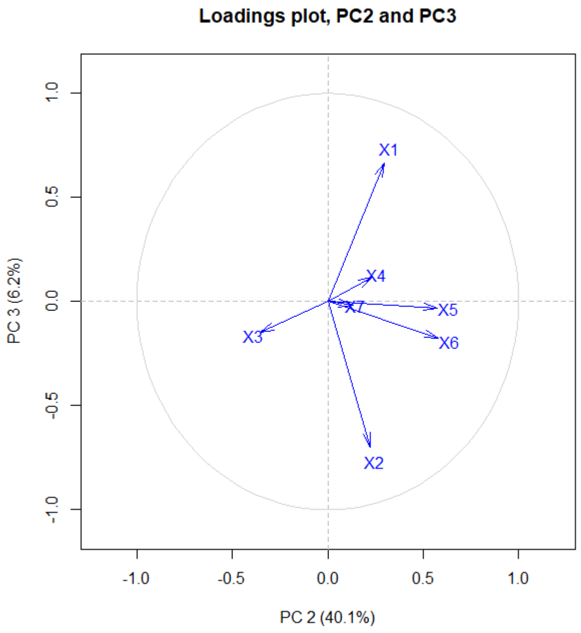
X_1: longest side
X_2: second longest side
X_3: smallest side
X_4: longest diagonal
X_5: radius in circumscribed sphere divided by radius in inscribed sphere
X_6: (longest side + second longest side)/ shortest side
X_7: surface area/volume

Let us now use the correlation matrix - Loadings plots





Label	Name	Loadings		
		PC1	PC2	PC3
X1	Longax	0.41	0.29	0.67
X2	intermax	0.43	0.22	-0.70
X3	shortax	0.39	-0.36	-0.15
X4	longdia	0.49	0.23	0.12
X5	ratspher	-0.12	0.58	-0.03
X6	ratax	-0.10	0.58	-0.17
X7	ratarvol	-0.48	0.13	-0.02



If the correlation matrix is known (it isn't here), the **Loadings plots** show the points

$$\begin{bmatrix} Cor(X_i, P_s) \\ Cor(X_i, P_t) \end{bmatrix}$$

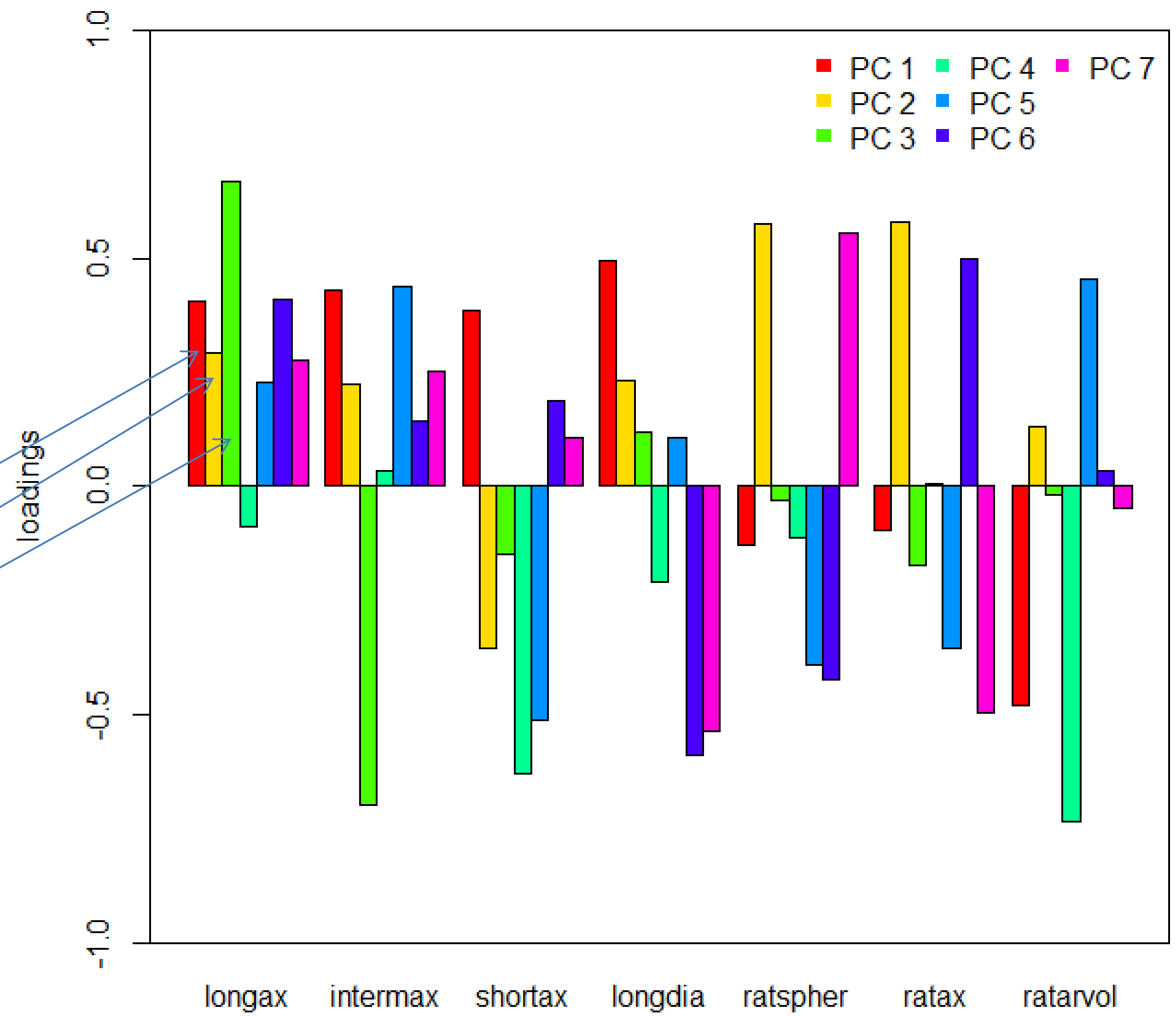
$$for\ i = 1, \cdots, 7\ and\ (s,t) = (1,2), \cdots (m-1, m).$$

Otherwise, it just shows how much the original variables impact on the PC's; which is often enough information.

Loadings as a barplot

The **bar plot shows** the impact of the original variables on the principal components.

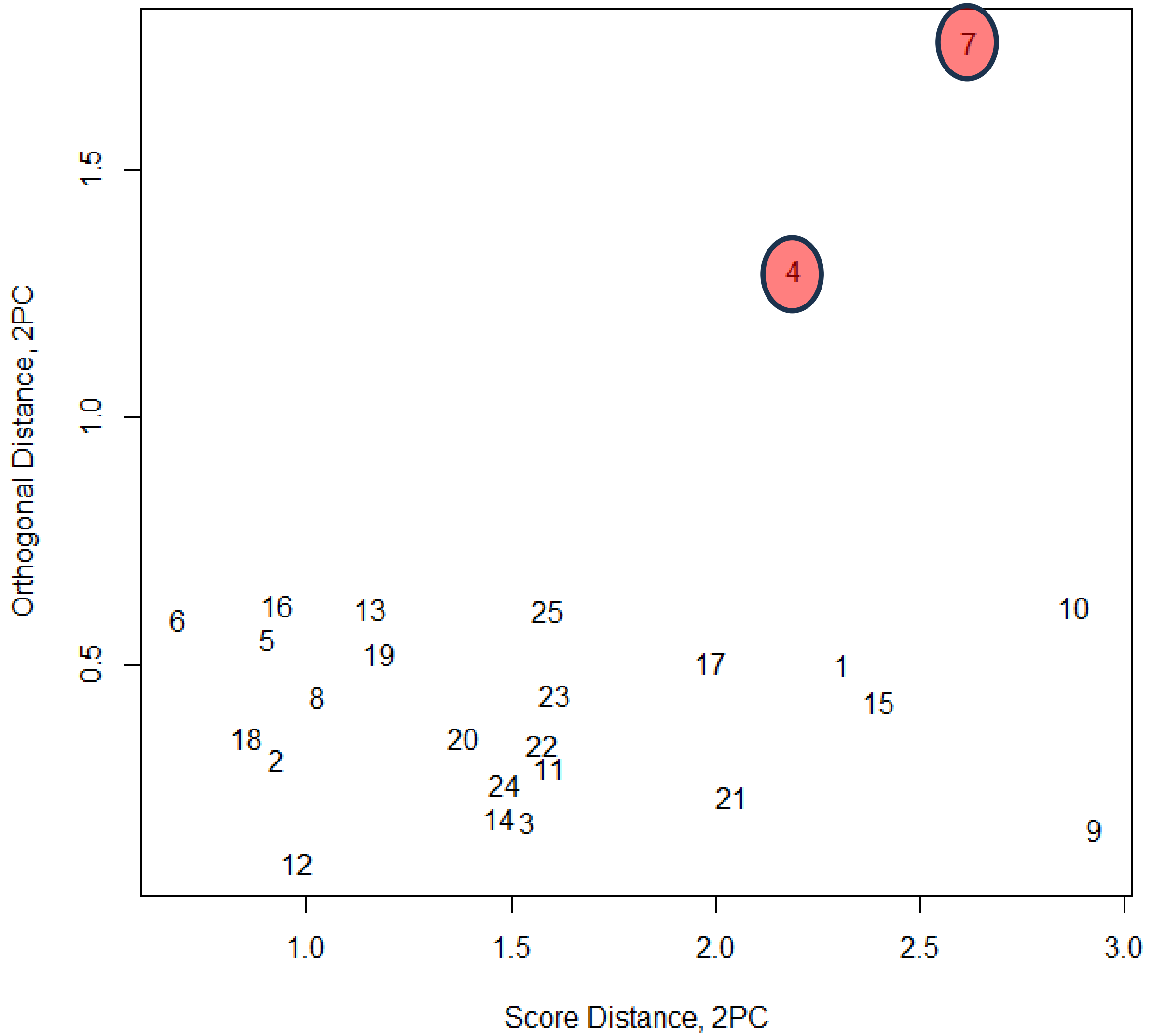
	Loadings		
	PC1	PC2	PC3
longax	0.41	0.29	0.67
intermax	0.43	0.22	-0.70
shortax	0.39	-0.36	-0.15
longdia	0.49	0.23	0.12
ratspher	-0.12	0.58	-0.03
ratax	-0.10	0.58	-0.17
ratarvol	-0.48	0.13	-0.02



Interpretation of the PCs

- **+PC1: Large boxes of cube form**
- **+PC2: flat,voluminous boxes**
- **-PC3: oblong boxes with a relatively large surface area**

Model Control



Investigating problematic boxes

Extreme!

Close to Extreme!

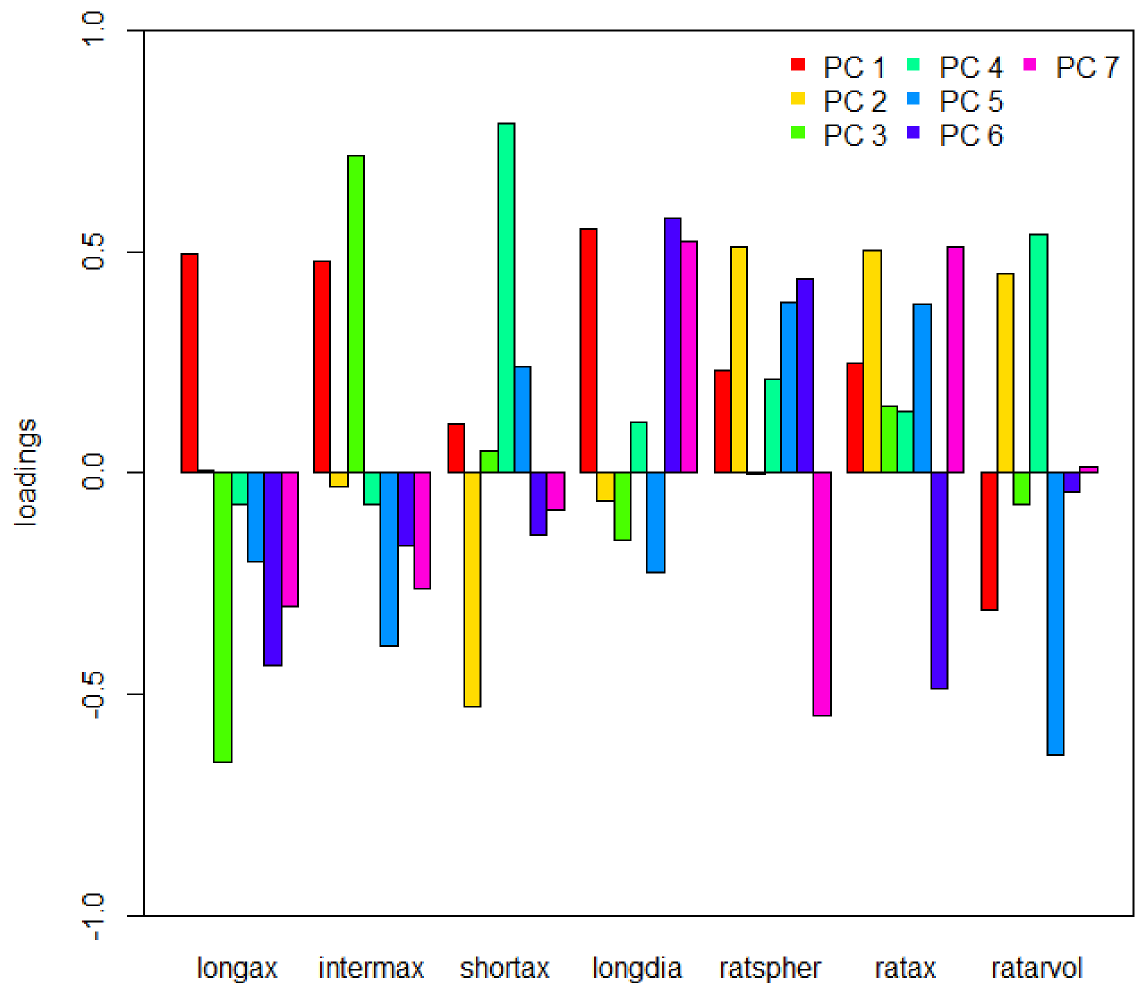
4: A large, nearly symmetric die

7: A tinderstick

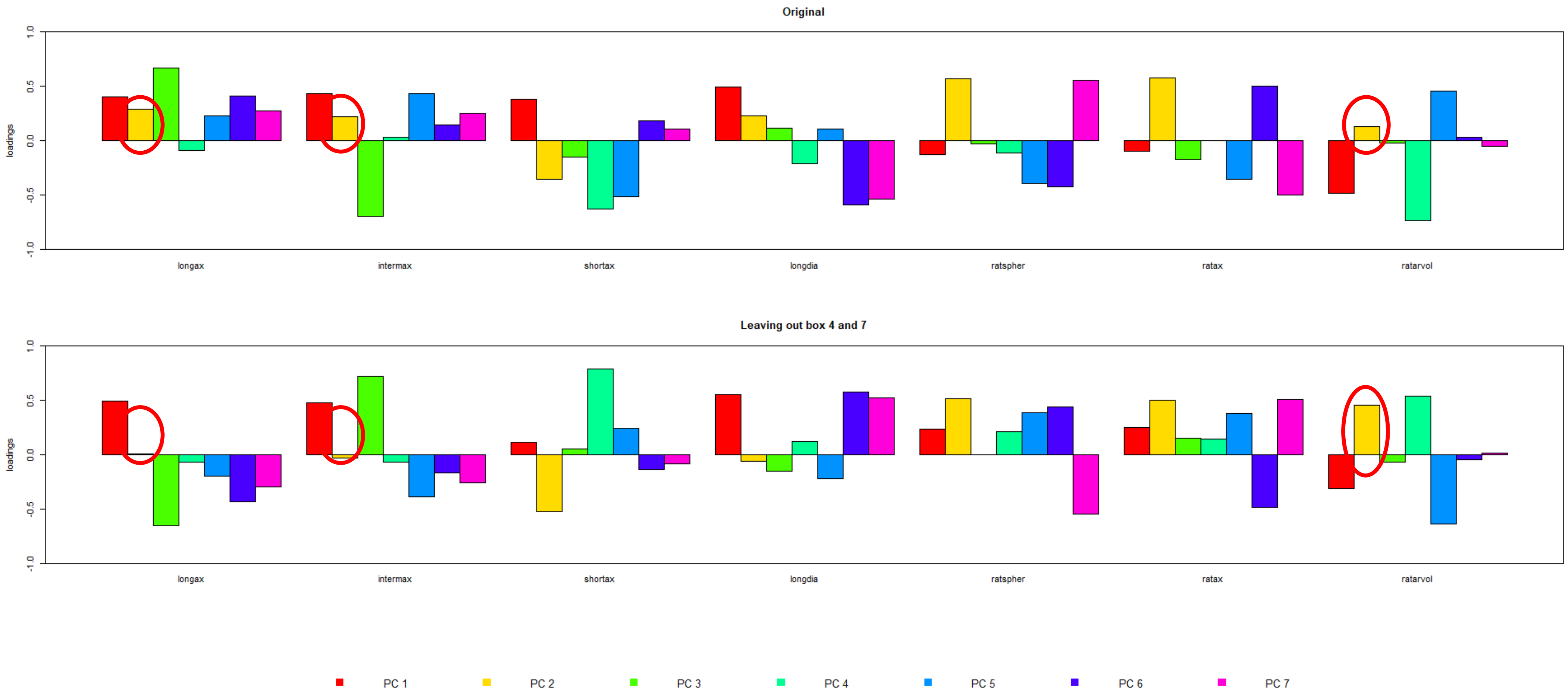
```
boxdata[c(4,7),]
  longax intermax shortax longdia ratspher ratax ratarvol
4   7.57     7.28    7.07  12.662    1.791 2.101    0.822
7   3.27     0.62    0.44   3.357    7.629 8.838    8.389
```

```
summary(boxdata)
     longax        intermax        shortax         longdia          ratspher          ratax           ratarvol
 Min.   :1.66   Min.   :0.620   Min.   :0.440   Min.   : 2.799   Min.   : 1.783   Min.   : 2.087   Min.   :0.822
 1st Qu.:5.51   1st Qu.:3.010   1st Qu.:1.170   1st Qu.: 6.924   1st Qu.: 2.760   1st Qu.: 3.509   1st Qu.:1.276
 Median :7.57   Median :4.920   Median :1.570   Median : 9.716   Median : 4.539   Median : 5.382   Median :2.013
 Mean   :7.10   Mean   :4.773   Mean   :2.349   Mean   : 9.134   Mean   : 5.458   Mean   : 7.167   Mean   :2.346
 3rd Qu.:9.03   3rd Qu.:6.440   3rd Qu.:3.310   3rd Qu.:11.742   3rd Qu.: 7.629   3rd Qu.: 9.909   3rd Qu.:2.616
 Max.   :9.84   Max.   :9.490   Max.   :7.070   Max.   :13.604   Max.   :13.133   Max.   :18.519   Max.   :8.389
```

Leaving out two extreme boxes



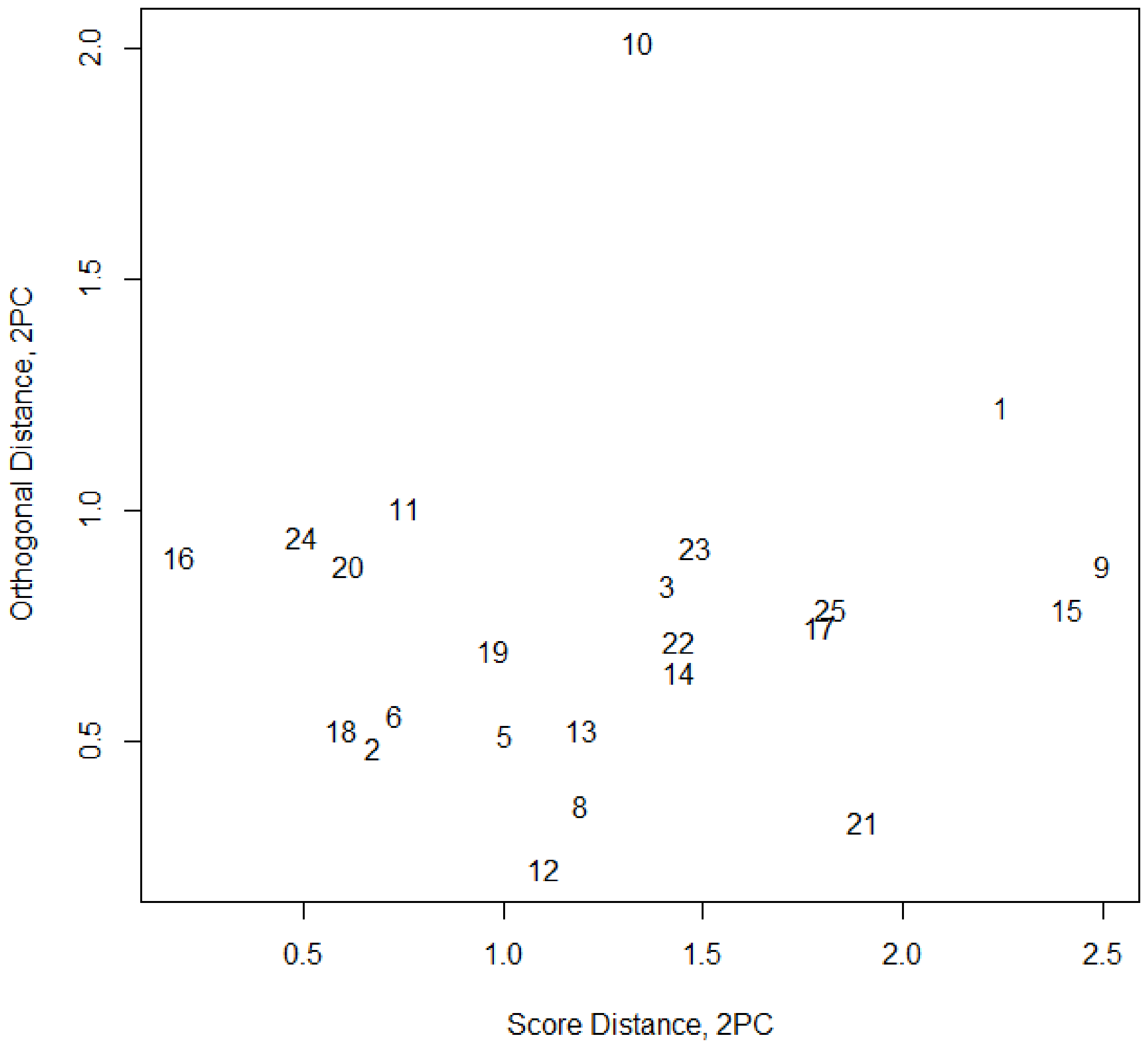
Leaving out two extreme boxes



Interpretation of the PCs

- +PC1: Unchanged
- +PC2: asymmetric voluminous boxes
- -PC3: Unchanged except for sign

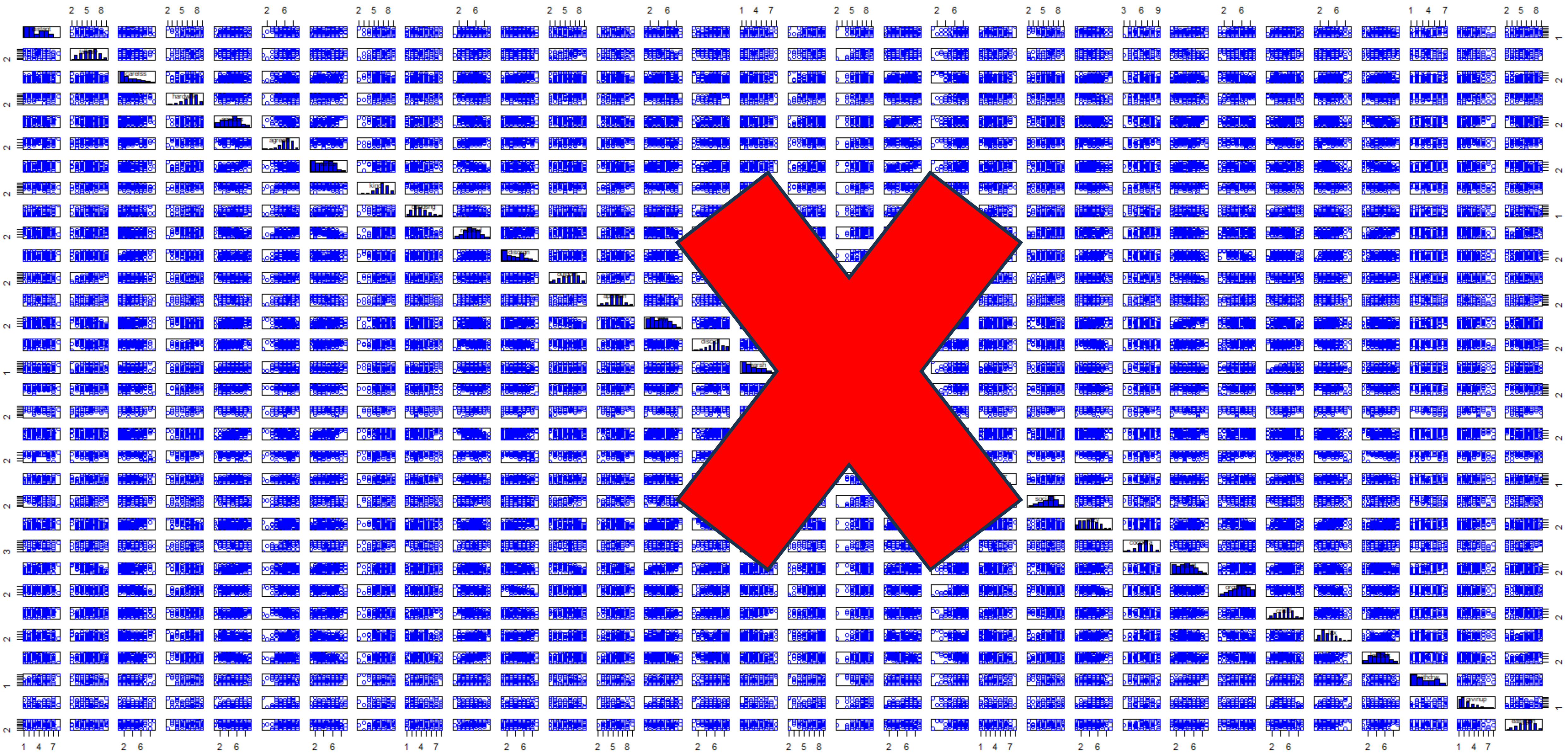
Model Control



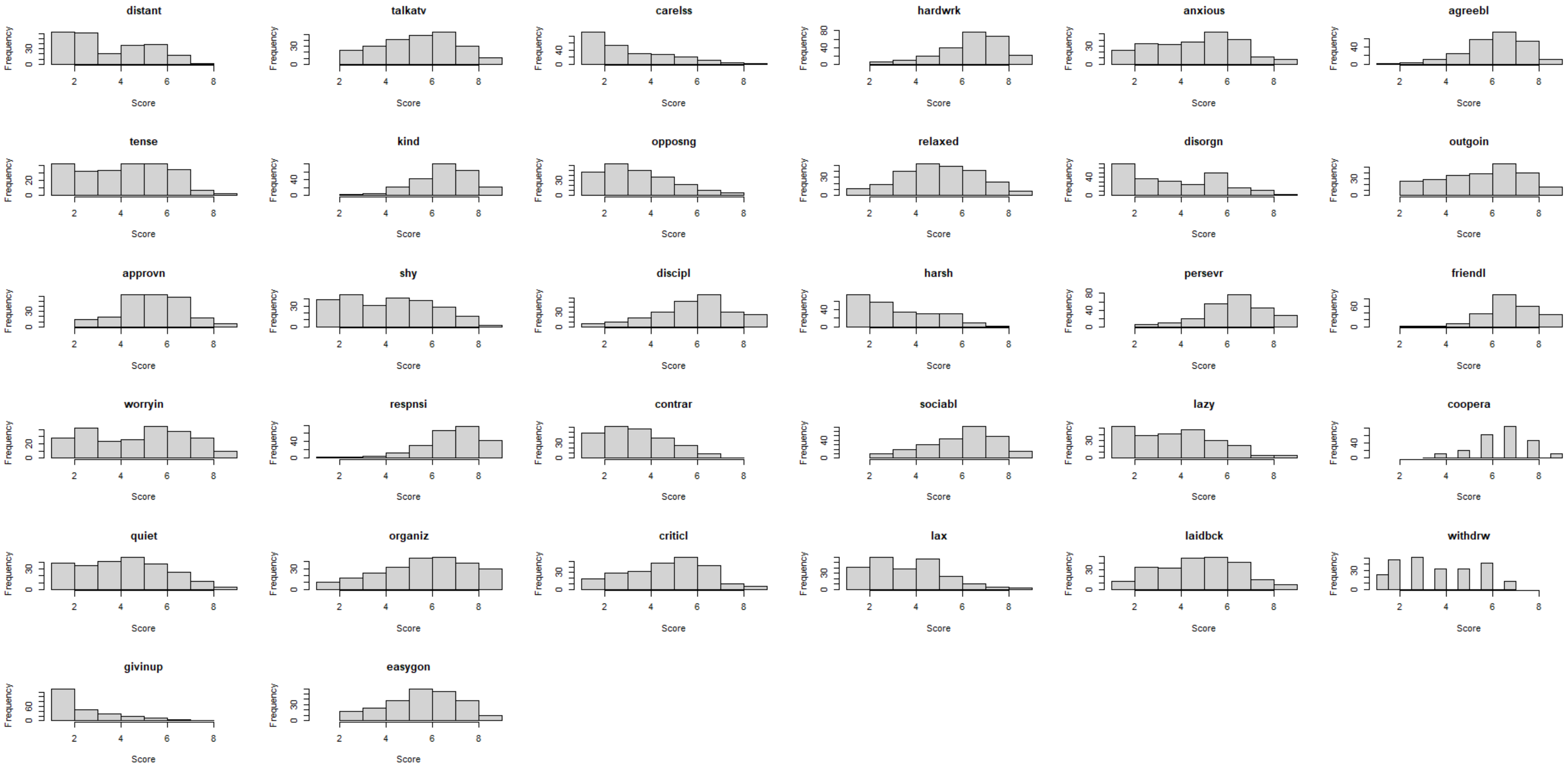
Personality Traits

- **self-ratings on 32 personality traits**, ‘distant’, ‘talkative’, ‘easygoing’,
- Scale 1-9
- 240 observations/persons

Personality Traits

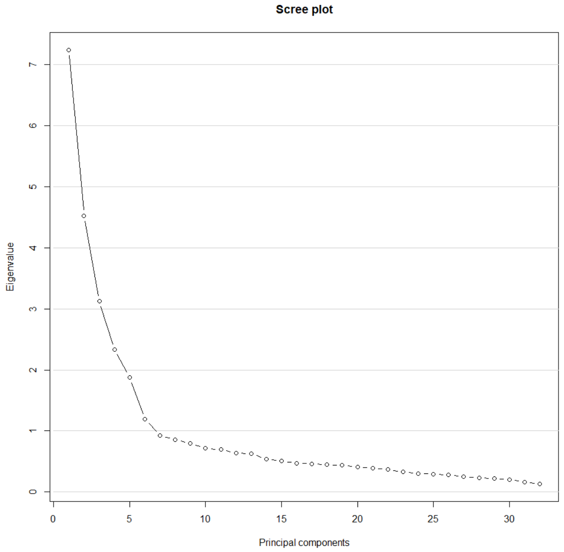


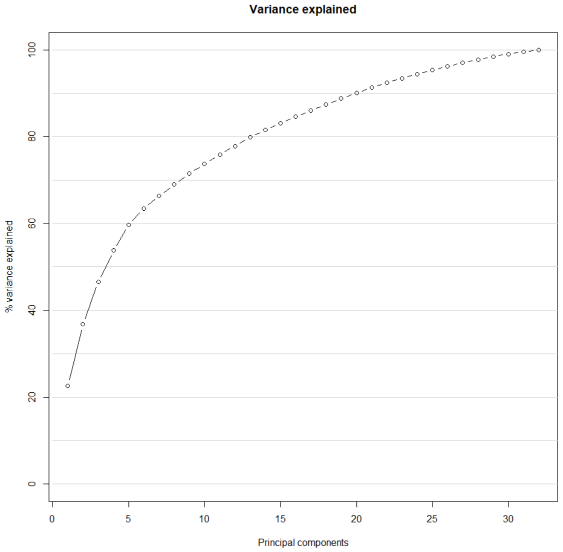
Personality Traits



Personality Traits: PCA

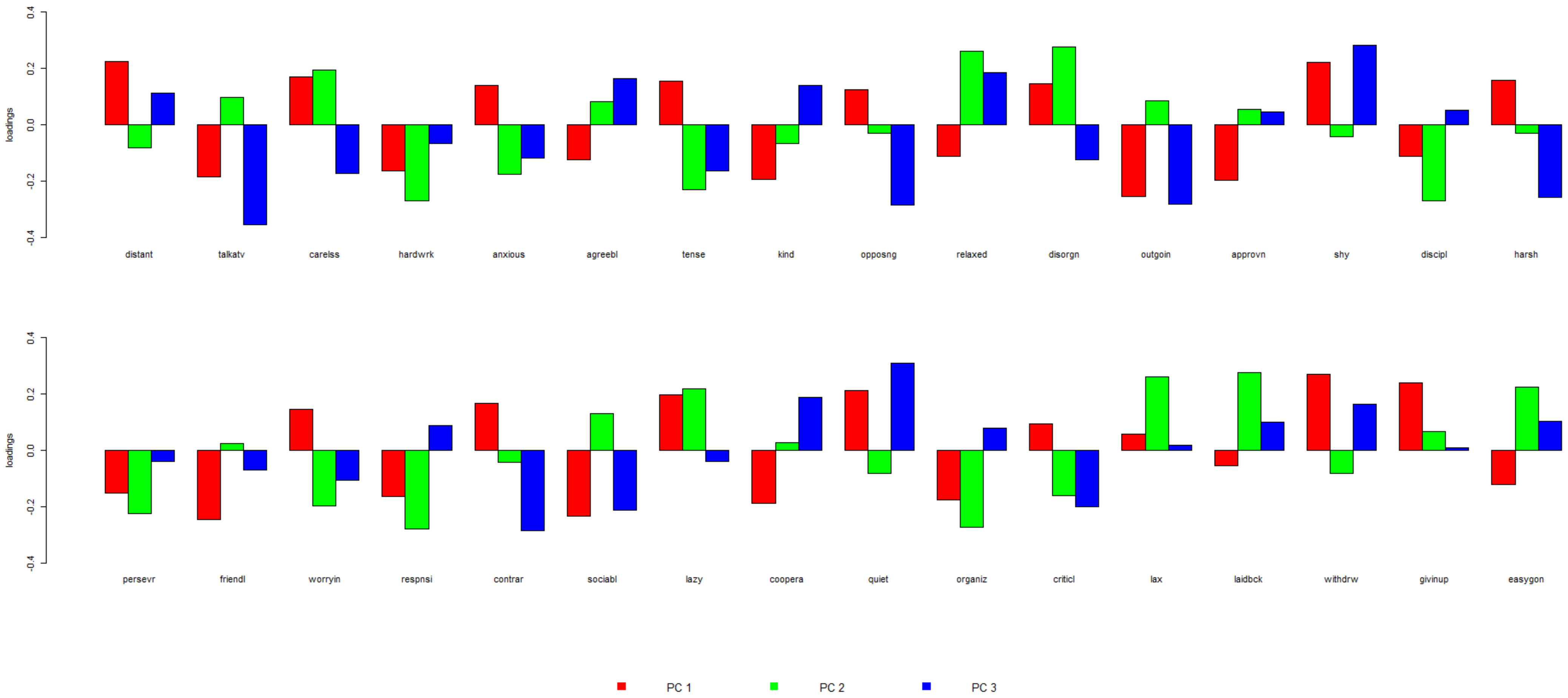
Eigenvalues of the Correlation Matrix			
	Eigenvalue	Proportion	Cumulative
1	7.24070683	0.2263	0.2263
2	4.52509007	0.1414	0.3677
3	3.12405726	0.0976	0.4653
4	2.33358902	0.0729	0.5382
5	1.87836113	0.0587	0.5969
6	1.19406360	0.0373	0.6342
7	0.92686362	0.0290	0.6632





Eigenvectors						
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
distant	0.226997	-0.081361	-0.114306	0.024592	0.223300	-0.162818
talkatv	-0.183983	0.099445	0.352939	-0.057073	0.012065	-0.037717
carelss	0.170559	0.194231	0.170853	-0.133912	0.044996	-0.379050
hardwrk	-0.161763	-0.267930	0.066105	0.001440	0.123786	-0.260102
anxious	0.139093	-0.174617	0.118623	-0.390410	-.028905	-0.023007
agreebl	-0.122470	0.083193	-0.166088	-0.355815	0.147211	-0.105861
tense	0.156159	-0.229244	0.163212	-0.313223	-.065918	-0.002087
kind	-0.192259	-0.065649	-0.141675	-0.300662	0.125512	0.150414
opposng	0.126339	-0.030217	0.282788	0.040404	0.309201	0.191352
relaxed	-0.112309	0.263051	-0.187558	0.090518	0.278131	0.021906
disorgn	0.147310	0.275758	0.122227	-0.152816	-.002830	-0.361896
outgoin	-0.253220	0.085468	0.282071	-0.073483	-.030810	0.019819
approvn	-0.196708	0.056857	-0.046837	-0.225867	0.139942	-0.030641
shy	0.222701	-0.041260	-0.283716	-0.104625	-.015987	0.058753
discipl	-0.111156	-0.268468	-0.053919	0.055134	0.193775	-0.337376
harsh	0.160029	-0.030730	0.257501	0.093247	0.294566	-0.032333
persevr	-0.151139	-0.224538	0.040638	-0.054423	0.161255	-0.303419
friendl	-0.245609	0.022557	0.069105	-0.292582	0.014932	0.063760
worryin	0.144471	-0.198020	0.108211	-0.343149	-.125297	0.119892
respnsi	-0.164575	-0.280249	-0.086345	-0.015934	0.177706	0.057653
contrar	0.165837	-0.042535	0.286971	0.016179	0.315792	0.104414
sociabl	-0.234683	0.129233	0.213917	-0.106759	-.063517	0.166108
lazy	0.197045	0.218976	0.039399	-0.169661	-.009965	0.236244
coopera	-0.189881	0.026983	-0.187554	-0.256469	0.085256	-0.037217
quiet	0.211313	-0.083010	-0.307913	-0.164834	0.113826	0.083750
organiz	-0.175597	-0.272531	-0.078816	0.113500	0.094405	0.319667
criticl	0.094313	-0.162252	0.201424	0.055516	0.337333	0.232152
lax	0.055978	0.260172	-0.016908	-0.127918	0.182034	0.031335
laidbck	-0.054764	0.275937	-0.097969	0.066258	0.328010	-0.029746
withdrw	0.268331	-0.081231	-0.163356	-0.021057	0.183392	-0.088726
givinup	0.238356	0.067285	-0.008421	-0.128839	-.042191	0.180871
easygon	-0.122908	0.223565	-0.101093	-0.068365	0.274823	0.153592

Personality Traits: PCA III



Personality traits: PC1-PC3

- PC1: Unseriousness
- - PC2: Good work skills
- - PC3: Socially dominating

Instead of just exploring data, let us see if we can set up a model for explaining them.

The Factor Model

$$[\boldsymbol{X}_1 \quad \cdots \quad \boldsymbol{X}_n] = \boldsymbol{A}[\boldsymbol{F}_1 \quad \cdots \quad \boldsymbol{F}_n] + [\boldsymbol{G}_1 \quad \cdots \quad \boldsymbol{G}_n]$$

$$\boldsymbol{X} = \boldsymbol{AF} + \boldsymbol{G}$$

$$\begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{km} \end{bmatrix} \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} G_1 \\ \vdots \\ G_k \end{bmatrix}$$

$$V(\boldsymbol{X}) = \boldsymbol{R} = \begin{bmatrix} 1 & \cdots & \rho_{1k} \\ \vdots & \ddots & \vdots \\ \rho_{k1} & \cdots & 1 \end{bmatrix}$$

$$V(\boldsymbol{F}) = \boldsymbol{I} = \boldsymbol{I}_m = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

$$V(\boldsymbol{G}) = \boldsymbol{\Delta} = \begin{bmatrix} \delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_k \end{bmatrix}$$

$$Cov(\boldsymbol{F}, \boldsymbol{G}) = \boldsymbol{0} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

Assumptions

The data has variance 1, i.e., the data is normalised / we use the correlation matrix

The factors are uncorrelated

The unique factors are uncorrelated

F and G are uncorrelated

$\boldsymbol{X}$ $(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)$ is (are) the observation(s),
$\boldsymbol{A}$ an paramtrized unknown matrix that we want to estimate,
$\boldsymbol{F}$ $(\boldsymbol{F}_1, \cdots, \boldsymbol{F}_n)$ the unobservable common factors (**factor scores**), (ar latent variables), and
$\boldsymbol{G}$ are unobservable unique factors, related to the individual measurements

Personality Traits: R code

We use principal components techniques to estimate the parameters in the model:

```
library(psych)

#### Factor analysis 5 Factors ####
fa5 <- principal(cor(stanford),nfactors = 5,rotate = "none")
fa5l <- fa5$loadings[,1:5]
fa5com <- fa5$communality
```

The estimated communalities $\hat{h}_i^2, i = 1, \dots k$, are the row sums of the squared estimates in A:

$$h_i^2 = a_{i1}^2 + \cdots + a_{ik}^2$$

$$\hat{h}_i^2 = \hat{a}_{i1}^2 + \cdots + \hat{a}_{ik}^2$$

Note that

$$V(X_i) = a_{i1}^2 + \cdots + a_{ik}^2 + \delta_i = h_i^2 + \delta_i = 1$$

Personality Traits: Factor Analysis I

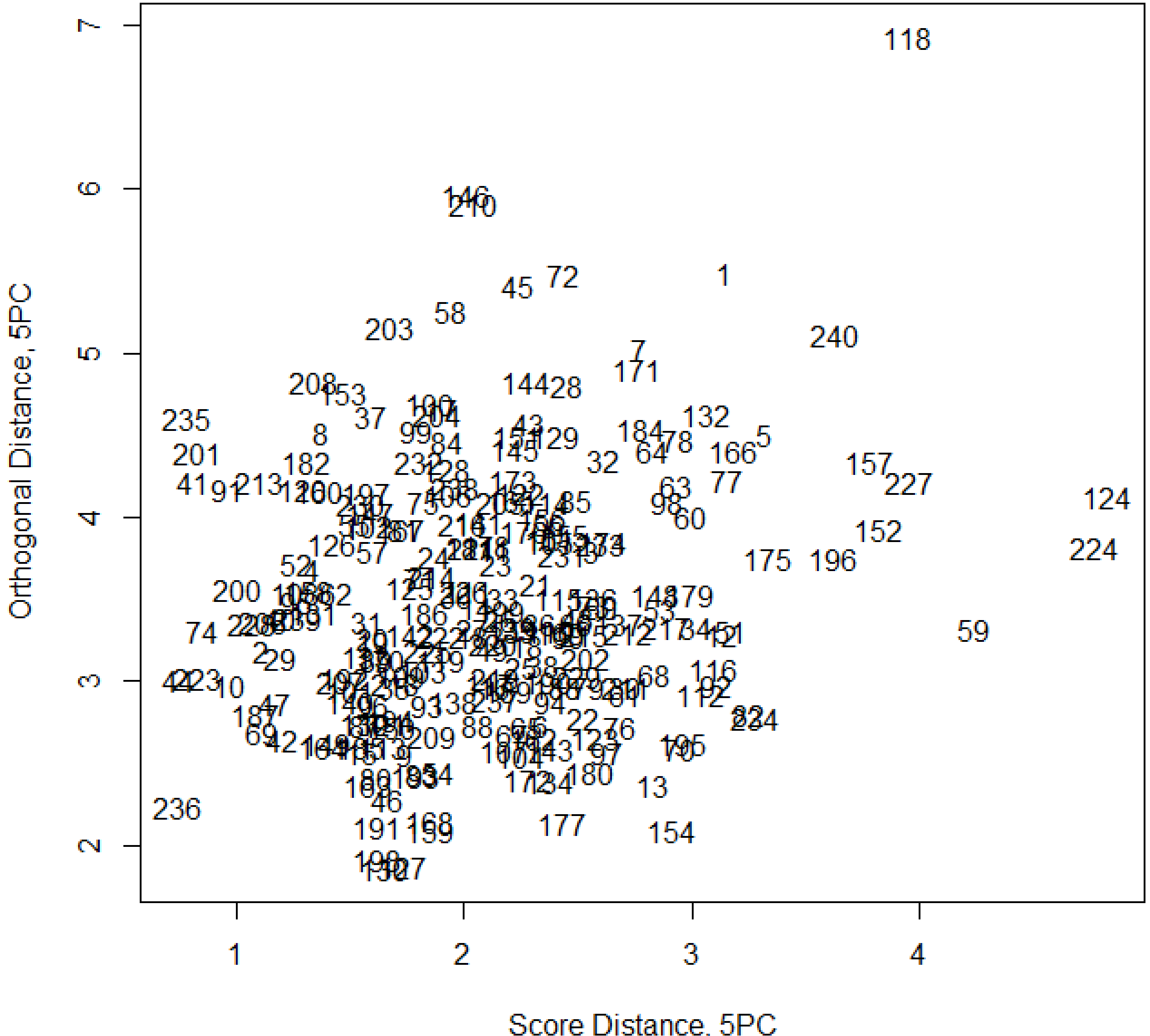
Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
Distant	0.61082	0.17307	-0.20204	-0.03757	0.30604
talkatv	-0.49507	-0.21154	0.62382	0.08719	0.01654
carelss	0.45895	-0.41317	0.30198	0.20456	0.06167
hardwrk	-0.43528	0.56995	0.11684	-0.00220	0.16965
anxious	0.37428	0.37145	0.20967	0.59639	-0.03961
agreebl	-0.32955	-0.17697	-0.29356	0.54355	0.20176
tense	0.42020	0.48765	0.28848	0.47848	-0.09034
kind	-0.51734	0.13965	-0.25041	0.45929	0.17202
opposng	0.33996	0.06428	0.49983	-0.06172	0.42377
relaxed	-0.30221	-0.55957	-0.33151	-0.13828	0.38119
disorgn	0.39639	-0.58660	0.21604	0.23344	-0.00388
outgoin	-0.68138	-0.18181	0.49856	0.11225	-0.04223
approvn	-0.52931	-0.12095	-0.08279	0.34504	0.19180
shy	0.59926	0.08777	-0.50147	0.15983	-0.02191
discipl	-0.29910	0.57109	-0.09530	-0.08422	0.26557
harsh	0.43061	0.06537	0.45513	-0.14244	0.40371
persevr	-0.40669	0.47764	0.07183	0.08314	0.22101
friendl	-0.66090	-0.04798	0.12214	0.44695	0.02046
worryin	0.38875	0.42123	0.19126	0.52420	-0.17172
respnsi	-0.44285	0.59615	-0.15262	0.02434	0.24355
contrar	0.44624	0.09048	0.50722	-0.02472	0.43280
sociabl	-0.63150	-0.27491	0.37810	0.16309	-0.08705
lazy	0.53022	-0.46581	0.06964	0.25918	-0.01366
coopera	-0.51094	-0.05740	-0.33150	0.39178	0.11685
quiet	0.56861	0.17658	-0.54424	0.25180	0.15600
organiz	-0.47251	0.57973	-0.13931	-0.17338	0.12939
criticl	0.25378	0.34515	0.35602	-0.08481	0.46233
lax	0.15063	-0.55344	-0.02989	0.19541	0.24948
laidbck	-0.14736	-0.58698	-0.17316	-0.10122	0.44955
withdrw	0.72204	0.17280	-0.28873	0.03217	0.25135
givinup	0.64138	-0.14313	-0.01488	0.19682	-0.05782
easygon	-0.33073	-0.47557	-0.17868	0.10443	0.37665

This is the ***A*** matrix in our model!

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
7.2407068	4.5250901	3.1240573	2.3335890	1.8783611

Final Communality Estimates: Total = 19.101804 out of 32, 60%.															
distant	talkatv	carelss	hardwrk	anxious	agreebl	tense	kind	opposng	relaxed	disorgn	outgoin	approvn	shy	discipl	harsh
0.538943	0.686870	0.518191	0.556748	0.679274	0.562248	0.734701	0.590391	0.552925	0.678768	0.602408	0.760278	0.457489	0.644307	0.502315	0.580122
persevr	friendl	worryin	respnsi	contrar	sociabl	lazy	coopera	quiet	organiz	criticl	lax	laidbck	withdrw	givinup	easygon
0.454454	0.654190	0.669419	0.634712	0.652524	0.651498	0.570319	0.541398	0.738434	0.625563	0.531219	0.430308	0.608584	0.698774	0.474158	0.520254

Personality Traits: Factor Analysis II



We should look into person 118, but we'll skip that for now.

Consequences of the model assumptions I

$$V(\boldsymbol{X}) = \boldsymbol{R} = V(\boldsymbol{AF} + \boldsymbol{G}) = \boldsymbol{AA}^T + \boldsymbol{\Delta}$$

$$V(X_i) = a_{i1}^2 + \cdots + a_{im}^2 + \delta_i = h_i^2 + \delta_i = 1$$

$$Cov(\boldsymbol{X}, \boldsymbol{F}) = Cov(\boldsymbol{AF} + \boldsymbol{G}, \boldsymbol{F}) = \boldsymbol{A}$$

$$Cov(X_i, F_j) = Corr(X_i, F_j) = a_{ij}$$

$$V\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{F} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{I} \end{bmatrix}$$

$$\boldsymbol{V}(\boldsymbol{X}|\boldsymbol{F}) = \boldsymbol{R} - \boldsymbol{AA}^T = \boldsymbol{\Delta}$$

$$V\begin{bmatrix} X_i \\ \boldsymbol{F} \end{bmatrix} = \begin{bmatrix} 1 & [a_{i1} \quad \cdots \quad a_{im}] \\ \begin{bmatrix} a_{i1} \\ \vdots \\ a_{im} \end{bmatrix} & \boldsymbol{I} \end{bmatrix}$$

$$V(X_i|\boldsymbol{F}) = 1 - [a_{i1} \quad \cdots \quad a_{im}]\boldsymbol{I}^{-1}\begin{bmatrix} a_{i1} \\ \vdots \\ a_{im} \end{bmatrix} =$$

$$1 - (a_{i1}^2 + \cdots + a_{im}^2) = \delta_i$$

$$V(X_i) - V(X_i|\boldsymbol{F}) = a_{i1}^2 + \cdots + a_{im}^2 = h_i^2$$

h_i^2 is the ith **communality**, the fraction of the variation in X_i that is explained by the common factors.

The **factor loading** a_{ij} is the correlation between the *i*th variable and the *j*th factor.

Remember that

$$E(X_1|X_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
$$V(X_1|X_2 = x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Reconstructing the factor scores

$$V(X_i|F_j) = 1 - a_{ij}^2$$

$$V(X_1|F_j) + \cdots + V(X_k|F_j) = k - (a_{1j}^2 + \cdots + a_{kj}^2)$$

$$\Sigma V(X_i) - \{V(X_1|F_j) + \cdots + V(X_k|F_j)\}$$

$$= a_{1j}^2 + \cdots + a_{kj}^2$$

The part of the total variation explained by the jth factor

$$V\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{R} \end{bmatrix}$$

Reconstruction:

$$E(\boldsymbol{F}|\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{\mu_F} + \boldsymbol{A}^T\boldsymbol{R}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_X}) = \boldsymbol{A}^T\boldsymbol{R}^{-1}\boldsymbol{x}$$

The reconstructed ***F***-values **(factor scores)** for known ***X***-values

Remember that

$$E(X_1|X_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
$$V(X_1|X_2 = x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Inversion of k by k $\boldsymbol{R}$ can be replaced by inversion of an m by m matrix, see text book p. 413

The Principal Factor Solution – version 1

Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \text{ with } \quad V(\boldsymbol{X}) = \boldsymbol{R} = \begin{bmatrix} 1 & \cdots & \rho_{1k} \\ \vdots & \ddots & \vdots \\ \rho_{k1} & \cdots & 1 \end{bmatrix}, \qquad \rho_{ij} = \rho_{ji}$$

Let the eigenvalues of $\boldsymbol{\Sigma} = \boldsymbol{R}$ be $\lambda_1 \geq \cdots \geq \lambda_k$ and the corresponding eigenvectors $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$. Let $m \leq k$ and define

$$\boldsymbol{P}^{(m)} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m]$$

$$\boldsymbol{\Lambda}^{(m)} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$$

Then version 1 of the **principal factor solution** is

$$\widehat{\boldsymbol{A}} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m] \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix} = P^{(m)} \Lambda^{(m)^{1/2}}$$

Remember from PCA: $\boldsymbol{\Sigma} = \boldsymbol{P\Lambda P}^{\mathrm{T}} = \boldsymbol{P\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P\Lambda}^{1/2}\left(\boldsymbol{P\Lambda}^{1/2}\right)^{\boldsymbol{T}} = \boldsymbol{AA}^{\boldsymbol{T}}$

The Principal Factor Solution – version 2

- Note that in version 1 og the principle factor solution we have for $m = k$ that $\hat{A}^T A = \widehat{\mathrm{R}}$.
- Since in general $\mathrm{R} = A^T A + \Delta$, we are assuming **NO CONTRIBUTION** from the unique factors; all variation is explained form the factors.
- In practice, that might not be realistic.
- Note that with $A^T A = R - \Delta$, the diagonal elements in this matrix are $1 - \delta_i = \mathrm{h}_i^2$. Here h_i^2 is the **degree of explainability on $\boldsymbol{X_i}$** by the factors F.
- Since G is assumed independent of X, it doesn't explain anything about X. So the total information about X_i in the system need to come from the factors F.
- So, a way to proceed will be to **REPLACE THE DIAGONAL ELEMENTS OF R** with the total explainability in the system, the multiple correlation copefficient of X_i given the other variables.
- Thus, we will look for eigenvectors of the matrix

$$\begin{pmatrix} r_{1|2,\ldots,k}^2 & \cdots & r_{1k} \end{pmatrix}$$

The Principal Factor Solution – version 2

- Thus, we will look for eigenvectors of the matrix

$$V = \begin{pmatrix} r^2_{1|2,\dots,k} & \cdots & r_{1k} \\ \vdots & \ddots & \vdots \\ r_{k1} & \cdots & r^2_{k|1,\dots,k-1} \end{pmatrix}$$

where the off diuagonal elements are the original elements of R, instead of R itself.

We will disregard that V might not be positive semidefinite, as we shall be looking for the larger eigenvalues anyway.

This leads to:

The Principal Factor Solution – version 1

Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \text{ with } \quad V(\boldsymbol{X}) = \boldsymbol{R} = \begin{bmatrix} 1 & \cdots & \rho_{1k} \\ \vdots & \ddots & \vdots \\ \rho_{k1} & \cdots & 1 \end{bmatrix}, \qquad \rho_{ij} = \rho_{ji}$$

Let $\boldsymbol{V}$ be constructed as on the previous slide, and let the eigenvalues of $\boldsymbol{V}$ be $\lambda_1 \geq \cdots \geq \lambda_k$ and the corresponding eigenvectors $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$. Choose $m \leq k$ and define

$$\boldsymbol{P}^{(m)} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m]$$

$$\boldsymbol{\Lambda}^{(m)} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$$

Then version 2 of the **principal factor solution** is

$$\widehat{\boldsymbol{A}} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m] \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix} = P^{(m)} \Lambda^{(m)^{1/2}}$$

The Principal Factor Solution

Theorem 6.19

We consider the factor model $\boldsymbol{X} = \mathbf{A}\boldsymbol{F} + \boldsymbol{G}$ where $\boldsymbol{X}$ is k-dimensional and $\boldsymbol{F}$ m-dimensional. The correlation matrix of $\boldsymbol{X}$ is denoted $\mathbf{R}$, and $\mathbf{V}$ is the matrix which we find by substituting the ones in the diagonal of $\mathbf{R}$ with estimates of the communalities. These should be chosen in the interval $[r^2, 1]$ where r^2 is the multiple correlation coefficient between the relevant variable and the rest of the variables. Usually one chooses either r^2 or 1. The ***principle factor solution*** to the estimation problem is then

$$\mathbf{P}\boldsymbol{\Lambda}_*^{\frac{1}{2}} = (\sqrt{\lambda_1}\boldsymbol{p}_1, \ldots, \sqrt{\lambda_m}\boldsymbol{p}_m),$$

where λ_i, $i = 1, \ldots, m$ are the m largest eigenvalues of $\mathbf{V}$ and where $\boldsymbol{p}_i$, $i = 1, \ldots, m$ are the corresponding normed eigenvectors.

Is Factor Analysis just a twist on PCA?

NO!

- In PCA we try to get insight into and/or compress our data.
- In Factor Analysis we have *assumptions* that there underlying factors that explain the data. We assume that the factors and unique factors are uncorrelated.
- PCA is just *a technique to estimate the factor loadings*!
- Many other options are available, *that can give different results*.

Factor Rotation

The Principle factor solution for A is not unique. For Q orthonormal, i.e. $QQ^T = I$ (Q is a rotation matrix), AQ is also a solution, since

$$(AQ)(AQ)^T = AQQ^TA^T = AA^T$$

The choice of rotation may reflect properties of the estimator

Factor Rotation

The VARIMAX criterion maximizes

$$\sum_j m \left\{ \sum_i \left(\frac{a_{ij}^2}{h_i^2} \right)^2 - \frac{1}{m} \left[\sum_i \left(\frac{a_{ij}^2}{h_i^2} \right) \right]^2 \right\}$$

Note that this is the empirical variance of the terms a_{ij}^2/h_j^2.
It turns out that maximizing this variance will cause many a_{ij}^2 and some a_{ij}^2 to be large, which gives a SIMPLE MODEL where dependencies are selected from a simplification criteria.

Personality Traits: Factor Analysis
R code I

```
library(psych)
#### Factor analysis 5 Factors ####
fa5 <- principal(cor(stanford),nfactors = 5,rotate = "none")
fa5l <- fa5$loadings[,1:5]
fa5com <- fa5$communality
#rotated: rfa5 <- principal(cor(stanford),nfactors = 5,rotate = "varimax")
rfa5l <- rfa5$loadings[,1:5]
rfa5com <- rfa5$communality
#Factors
Factors5 <- data.frame("FA" = fa4l,"Rot FA" = rfa4l)
```

Personality Traits: Factor Analysis
R code II

Plots for factor analysis with 5 factors, without invoking a package:

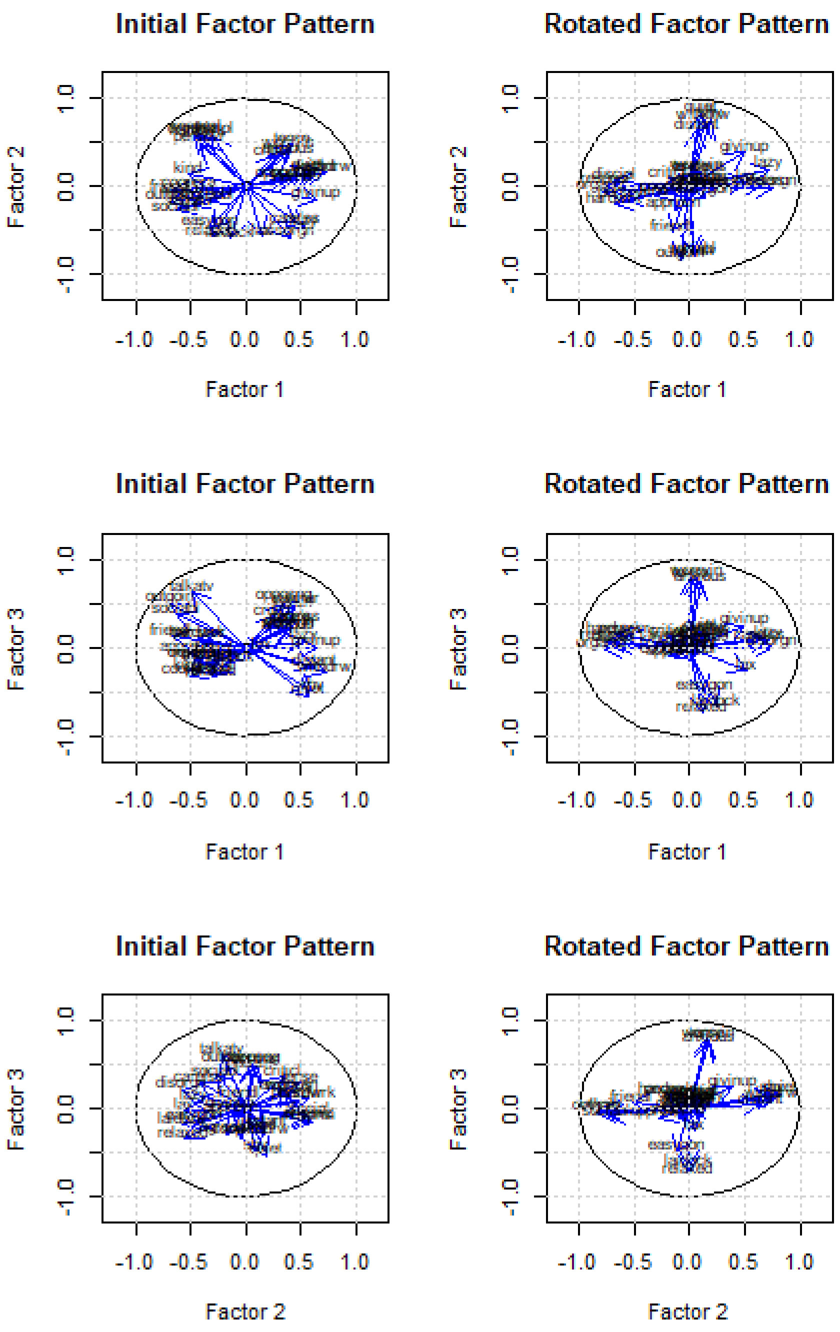
```
par(mfrow = c(3,2))
circle = seq(-3.2,3.2,by=0.1)
# Different combinations of plots
ij = matrix(c(1,1,2,2,3,3),ncol=2)
Names = names(stanford)
for (i in 1:3){
        l = ij[i,1]
        k = ij[i,2]
        #Plot for the Factors
        plot(0,0,xlim = c(-1.2,1.2),ylim = c(-1.2,1.2),xlab = paste0("Factor ",l),
             ylab = paste0("Factor ",k),main = "Initial Factor Pattern")
        points(1*cos(circle),1*sin(circle),type='l')
        arrows(c(rep(0,7)),c(rep(0,7)),fa5l[,l],fa5l[,k],length = 0.1)
        text(fa5l[,l],fa5l[,k]+0.1,Names,cex = 0.7)
        grid()

        #Plot for rotated Factors
        plot(0,0,xlim = c(-1.2,1.2),ylim = c(-1.2,1.2),xlab = paste0("Factor ",l),
             ylab = paste0("Factor ",k),main = "Rotated Factor Pattern")
        points(1*cos(circle),1*sin(circle),type='l')
        arrows(c(rep(0,7)),c(rep(0,7)),rfa5l[,l],rfa5l[,k],length = 0.1)
        text(rfa5l[,l],rfa5l[,k]+0.1,Names,cex = 0.7)
        grid()
}
```

Factor plots

- Very **uninformative** with so many variable

But the code can be used in exercises with less variables.

- In this situation, the barplot is preferred.



2012 Q4.4: What fraction of the total variance will be explained by the two varimax rotated factors?

Rotated Factor Pattern		
	Factor1	Factor2
slo500	0.80511	0.36531
slo580	0.88873	-0.39852
slo750	0.89224	-0.42571
slo830	0.97630	-0.18197
slo910	0.95884	-0.17862
int500	0.16135	0.94880
int580	-0.26389	0.94821
int750	-0.12800	0.98342
int830	-0.52466	0.82105
int910	-0.59265	0.79118

Variance Explained by Each Factor	
Factor1	Factor2
4.8452152	4.6050509

Final Communality Estimates: Total = 9.450266									
slo500	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
0.781651	0.948667	0.977319	0.986273	0.951277	0.926253	0.968747	0.983507	0.949383	0.977189

2012 Q4.4: What fraction of the total variance will be explained by the two varimax rotated factors?

A: 0.78165	0%
B: 0.940	0%
C: 0.78165^2	0%
D: $(4.8452152^2 + 4.6050509^2)/100$	0%
E: $0.80511^2 + 0.36531^2$	0%
F: Don't know	0%

2012 Q4.4 - What fraction of the total variance will be explained by the tw
VARIMAX rotated factors?

10 factors, i.e., total variance is 10 (data are normalized).

Rotated Factor Pattern		
	Factor1	Factor2
slo500	0.80511	0.36531
slo580	0.88873	-0.39852
slo750	0.89224	-0.42571
slo830	0.97630	-0.18197
slo910	0.95884	-0.17862
int500	0.16135	0.94880
int580	-0.26389	0.94821
int750	-0.12800	0.98342
int830	-0.52466	0.82105
int910	-0.59265	0.79118

Variance Explained by Each Factor	
Factor1	Factor2
4.8452152	4.6050509

Final Communality Estimates: Total = 9.450266									
)	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
51	0.948667	0.977319	0.986273	0.951277	0.926253	0.968747	0.983507	0.949383	0.977189

0.78165

$\boxed{0.9450}$

0.78165^2

$(4.8452152^2 + 4.6050509^2)/100$

$0.80511^2 + 0.36531^2$

Don't know

2012 Q4.5: How much of the variation of int500 is not explained by the two varimax rotated factors?

Rotated Factor Pattern		
	Factor1	Factor2
slo500	0.80511	0.36531
slo580	0.88873	-0.39852
slo750	0.89224	-0.42571
slo830	0.97630	-0.18197
slo910	0.95884	-0.17862
int500	0.16135	0.94880
int580	-0.26389	0.94821
int750	-0.12800	0.98342
int830	-0.52466	0.82105
int910	-0.59265	0.79118

Variance Explained by Each Factor	
Factor1	Factor2
4.8452152	4.6050509

Final Communality Estimates: Total = 9.450266									
slo500	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
0.781651	0.948667	0.977319	0.986273	0.951277	0.926253	0.968747	0.983507	0.949383	0.977189

2012 Q4.5: How much of the variation of int500 is not explained by the two varimax rotated factors?

A: $0.16135^2 + 0.94880^2$	0%
B: 0.926253^2	0%
C: $\frac{1}{2}(0.16135^2 + 0.94880^2)$	0%
D: $1 - (0.16135^2 + 0.94880^2)$	0%
E: $1 - \frac{1}{2}(0.16135^2 + 0.94880^2)$	0%
F: Don't know	0%

2012 Q4.5: How much of the variation of int500 is not explained by the two varimax rotated factors?

Rotated Factor Pattern		
	Factor1	Factor2
slo500	0.80511	0.36531
slo580	0.88873	-0.39852
slo750	0.89224	-0.42571
slo830	0.97630	-0.18197
slo910	0.95884	-0.17862
int500	0.16135	0.94880
int580	-0.26389	0.94821
int750	-0.12800	0.98342
int830	-0.52466	0.82105
int910	-0.59265	0.79118

Variance Explained by Each Factor	
Factor1	Factor2
4.8452152	4.6050509

Final Communality Estimates: Total = 9.450266									
	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
1	0.948667	0.977319	0.986273	0.951277	0.926253	0.968747	0.983507	0.949383	0.977189

$0.16135^2 + 0.94880^2$ A

0.926253^2 B

$\frac{1}{2}(0.16135^2 + 0.94880^2)$ C

$1 - (0.16135^2 + 0.94880^2)$ D

$1 - \frac{1}{2}(0.16135^2 + 0.94880^2)$ E

Don't know F

Exercises

- Exam 2011, problem 2, PCA and factor analysis
- Ex 7.5 ”Pen & Paper” on factor analysis
- Ex 7.7 R on factor analysis
- Exam 2013, problem 1, factor analysis