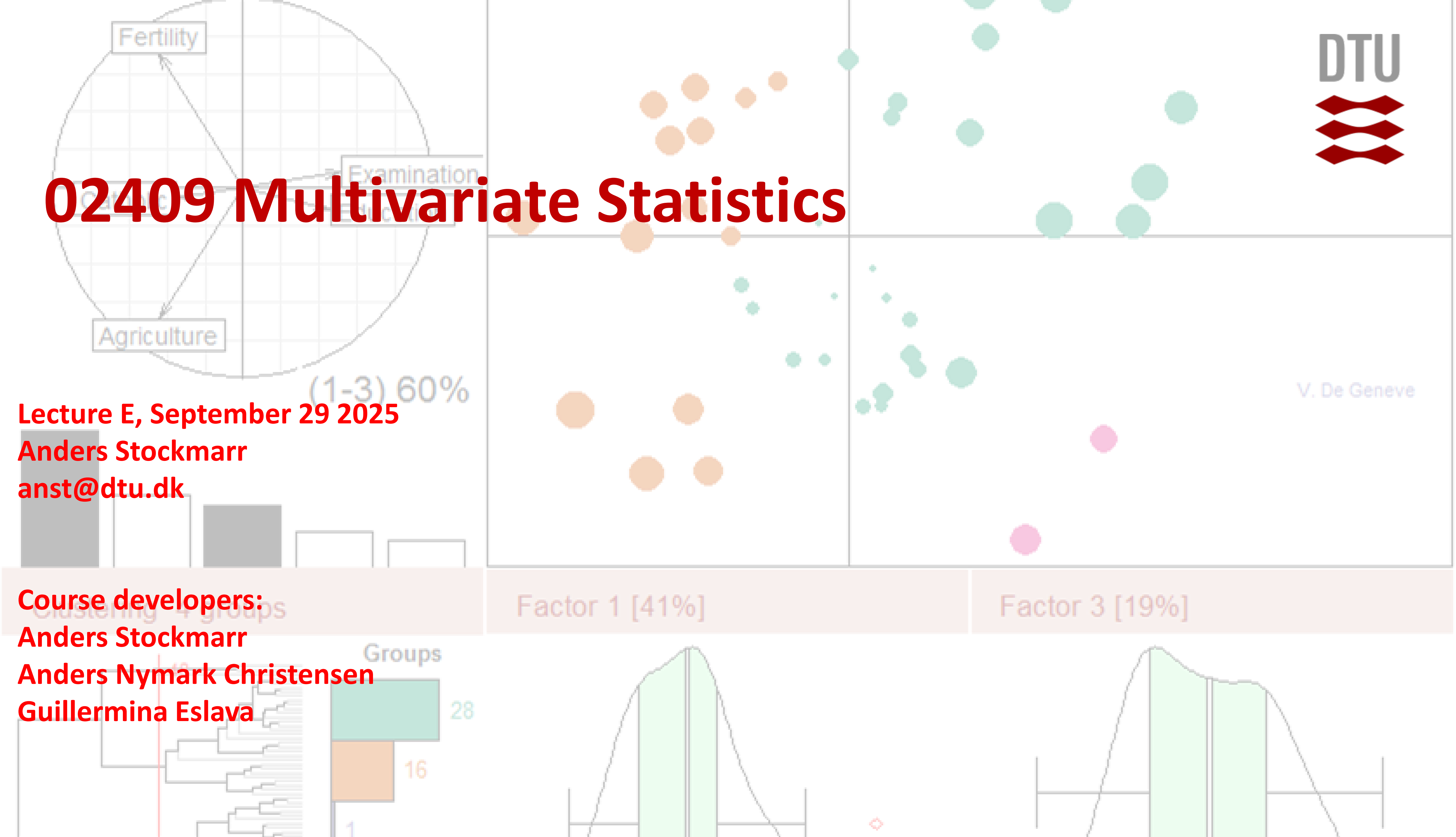


02409 Multivariate Statistics

Lecture E, September 29 2025
Anders Stockmarr
anst@dtu.dk

Course developers:
Anders Stockmarr
Anders Nymark Christensen
Guillermina Eslava

Agenda

- Factor analysis
- Canonical Correlation Analysis (CCA)
  - what can it be used for

The Factor Model

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

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

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

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

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

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

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

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

Assumptions

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 unknown matrix we want to determine,
$\boldsymbol{F}$ $(\boldsymbol{F}_1, \cdots, \boldsymbol{F}_n)$ the unobservable common factors (**factor scores**), that we (perhaps) want to determine (aka latent variables), and
$\boldsymbol{G}$ the unobservable unique factors

The Factor Model

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

1) Determine A

principal factor solution version 1 is

$$\boldsymbol{A} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m] \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix}$$

with $\lambda_1, \ldots, \lambda_m$ the m largest eigenvalues for $\boldsymbol{R}$, and $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ the corresponding eigenvectors

2) Rotate A

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

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

3) Estimate F

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

$\boldsymbol{X}$ $(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)$ is (are) the standardized observation(s),
R is the variance matrix (correlation matrix) of $\boldsymbol{X}$,
$\boldsymbol{A}$ a parametrized unknown matrix that we want to estimate,
$\boldsymbol{F}$ $(\boldsymbol{F}_1, \cdots, \boldsymbol{F}_n)$ the unobservable common factors (or latent variables),
$\boldsymbol{G}$ the unobservable unique factors

PCA is generally used to uncover structure, and reduce the number of variables

Factor Analysis estimates a **latent** model
PCA is just one way to estimate the factor loadings **A**; another option is e.g. maximum likelihood

In **Factor Analysis** we have a direct interpretations of **A**, e.g. communalities

Consequences of the model assumptions I

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

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

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

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

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

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

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

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

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

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

h_i^2 is the i$^{\text{th}}$ **communality**, the fraction of the variation in X_i that is explained by the common factors F.

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

Remember that

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

Consequences of the model assumptions II

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

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

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

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

The part of the total variation explained by the jth factor

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

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

The estimated ***F***-value **(factor score)** for known ***X***-value

Again, 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 the **principal factor solution version 1** is

$$\boldsymbol{A} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m] \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix}$$

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 the **principal factor solution version 1** is

$$\boldsymbol{A} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m] \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix}$$

Remember from PCA:

$$\mathbf{R} = \boldsymbol{P\Lambda P}^{\mathrm{T}} = \boldsymbol{P\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P\Lambda}^{1/2}\left(\boldsymbol{P\Lambda}^{1/2}\right)^{\boldsymbol{T}} = \boldsymbol{AA^T}$$

However, in general in the factor model,

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

Version 1 of the principle factor model **disregards contribution from the independent factor $\boldsymbol{G}$.**

Consequences of the model assumptions I

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

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

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

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

$$V\begin{bmatrix}\boldsymbol{X} \\ \boldsymbol{G}\end{bmatrix} = \begin{bmatrix}\boldsymbol{R} & \boldsymbol{\Delta} \\ \boldsymbol{\Delta} & \boldsymbol{\Delta}\end{bmatrix}$$

$$\boldsymbol{V}(\boldsymbol{X}|\boldsymbol{G}) = \boldsymbol{R} - \boldsymbol{\Delta}\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta} = \boldsymbol{R} - \boldsymbol{\Delta} = \boldsymbol{AA^T}$$

Model assumption:
For given $\boldsymbol{G}$, h_i^2 is the level of variation of X_i.

We do not know h_i^2. But when we condition on $\boldsymbol{G}$, all variation originates from F, and the diagonal elements of the variance matrix are less than 1.

To take effects of $\boldsymbol{G}$ into account when estimating, it is thus custom to replace the diagonal elements of $\boldsymbol{R}$ with lower values. The choice we make is the total level of explainability in the system, ie. the (squared) multiple correlation coefficient.

Remember that

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

The Principal Factor Solution version 2

Let

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

Define the matrix $\boldsymbol{V}$ by

$$V = \begin{bmatrix} \rho^2_{1|2,\ldots,k} & \cdots & \rho_{1k} \\ \vdots & \ddots & \vdots \\ \rho_{k1} & \cdots & \rho^2_{k|1,\ldots,k-1} \end{bmatrix}$$

Let the eigenvalues of $\boldsymbol{V}$ be $\lambda_1 \geq \cdots \geq \lambda_k$ and the corresponding eigenvectors $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$. 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 the **principal factor solution version 2** is

$$\boldsymbol{A} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m] \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix}$$

Example: Heptathlon data

	Name	Points	Hurdles	High.Jump	Shot	Run200	Longjump	Javelin	Run800
1	Joyner_Kersee,(USA)	7291	12.69	1.86	15.80	22.56	7.27	45.66	128.51
2	John,(GDR)	6897	12.85	1.80	16.23	23.65	6.71	42.56	126.14
3	Behmer,(GDR)	6858	13.20	1.83	14.20	23.10	6.68	44.54	124.20
4	Choubenkova,(URS)	6540	13.51	1.74	14.76	23.93	6.32	47.46	127.90
5	Sablovskaite,(URS)	6456	13.61	1.80	15.23	23.92	6.25	42.78	132.24
6	Schulz,(GDR)	6411	13.75	1.83	13.50	24.65	6.33	42.82	125.79
7	Fleming,(AUS)	6351	13.38	1.80	12.88	23.59	6.37	40.28	132.54
8	Greiner,(USA)	6297	13.55	1.80	14.13	24.48	6.47	38.00	133.65
9	Lajbnerova,(CZE)	6252	13.63	1.83	14.28	24.86	6.11	43.30	136.05
10	Bouraga,(URS)	6232	13.25	1.77	12.62	23.59	6.28	39.06	134.74
11	Wijnsma,(HOL)	6205	13.75	1.86	13.01	25.03	6.34	37.86	131.49
12	Dimitrova,(BUL)	6171	13.24	1.80	12.02	23.59	6.19	37.62	135.73
13	Scheider,(SWI)	6157	13.85	1.86	11.58	24.87	6.05	47.50	134.93
14	Braun,(FRG)	6109	13.71	1.83	13.16	24.78	6.12	44.58	142.82
15	Ruotsalainen,(FIN)	6101	13.79	1.80	12.32	24.61	6.08	45.44	137.06
16	Yuping,(CHN)	6087	13.93	1.86	14.21	25.00	6.40	38.60	146.67
17	Hagger,(GB)	5975	13.47	1.80	12.75	25.47	6.34	35.78	138.48
18	Brown,(USA)	5972	14.07	1.83	12.69	24.83	6.13	44.34	146.43
19	Mulliner,(GB)	5746	14.39	1.71	12.68	24.92	6.10	37.76	138.02
20	Hautenauve,(BEL)	5734	14.04	1.77	11.81	25.61	5.99	35.68	133.90
21	Kytola,(FIN)	5686	14.31	1.77	11.66	25.69	5.75	39.48	133.35
22	Geremias,(BRA)	5508	14.23	1.71	12.95	25.50	5.50	39.64	144.02
23	Hui-Ing,(TAI)	5290	14.85	1.68	10.00	25.23	5.47	39.14	137.30
24	Jeong-Mi,(KOR)	5289	14.53	1.71	10.83	26.61	5.50	39.26	139.17
25	Launa,(PNG)	4566	16.42	1.50	11.78	26.16	4.88	46.38	163.43

The `principal` Function

principal {psych} R Documentation

Principal components analysis (PCA)

Description

Does an eigen value decomposition and returns eigen values, loadings, and degree of fit for a specified number of components. Basically it is just doing a principal components analysis (PCA) for n principal components of either a correlation or covariance matrix. Can show the residual correlations as well. The quality of reduction in the squared correlations is reported by comparing residual correlations to original correlations. Unlike princomp, this returns a subset of just the best nfactors. The eigen vectors are rescaled by the sqrt of the eigen values to produce the component loadings more typical in factor analysis.

Factor analysis coding of the Hepthathlon data

```
library(psych)

# Correlation matrix
R<-cor(heptathlon2)

# squared multiple correlation coefficients

smc<-numeric(7)
for(i in 1:7){
  smc[i]<- R[i,-i]%*%solve(R[-i,-i])%*%R[-i,i]
  }
smc
[1] 0.9108174 0.8140824 0.6881388 0.8192189 0.9151310 0.4128242 0.6152957

# Construction of V matrix
V<-R
diag(V)<-smc
```

Factor analysis coding of the Hepthathlon data

```
principal(cor(heptathlon2),nfactors =3,
          rotate = "none")$loadings[,]

Loadings:
             PC1    PC2    PC3
Hurdles    -0.953  0.184
High.Jump   0.800 -0.260 -0.297
Shot        0.755  0.361  0.457
Run200     -0.852 -0.244
Longjump    0.963         0.103
Javelin     0.150  0.920 -0.354
Run800     -0.790  0.237  0.260
```

```
principal(V,nfactors = 3,
          rotate = "none")$loadings[,]

Loadings:
                  PC1         PC2          PC3
Hurdles    -0.9959647  0.24495436  0.037990372
High.Jump   0.8579436 -0.31007568  0.423588642
Shot        0.8734063  0.40661231 -0.003678345
Run200     -0.9309107 -0.28694991  0.241219073
Longjump    1.0075876 -0.07838114  0.006429187
Javelin     0.2358304  0.95223816  0.158935316
Run800     -0.9339401  0.27609909  0.151799610

Warning:
I cor.smooth(r) : Matrix was not positive
definite, smoothing was done
```

Factor analysis coding of the Hepthathlon data

```
eigen(V)$vectors[,1:3]%*%diag(sqrt(eigen(V)$va
lues[1:3]))
```

```
           [,1]        [,2]       [,3]
[1,]  0.9545446 -0.18912611 0.05947206
[2,] -0.7848444  0.31255696 0.32566158
[3,] -0.7164747 -0.36680153 0.07108174
[4,]  0.8339174  0.29750117 0.20663545
[5,] -0.9651381  0.02517499 0.02657049
[6,] -0.1302156 -0.59494431 0.16000578
[7,]  0.7384005 -0.18720366 0.16781646
```

The acutua eigenvectors

```
principal(V,nfactors = 3,
          rotate = "none")$loadings[,]
```

```
Loadings:
                 PC1         PC2          PC3
Hurdles   -0.9959647  0.24495436  0.037990372
High.Jump  0.8579436 -0.31007568  0.423588642
Shot       0.8734063  0.40661231 -0.003678345
Run200    -0.9309107 -0.28694991  0.241219073
Longjump   1.0075876 -0.07838114  0.006429187
Javelin    0.2358304  0.95223816  0.158935316
Run800    -0.9339401  0.27609909  0.151799610
```

What we get from the eigenvectors

- Smoothing effects undesirable, we can work out the loadings directly as in the left table.

Factor analysis coding of the Hepthathlon data

The original

```
eigen(R)$vectors[,1:3]%*%diag(sqrt(eigen(R)$va
lues[1:3]))
```

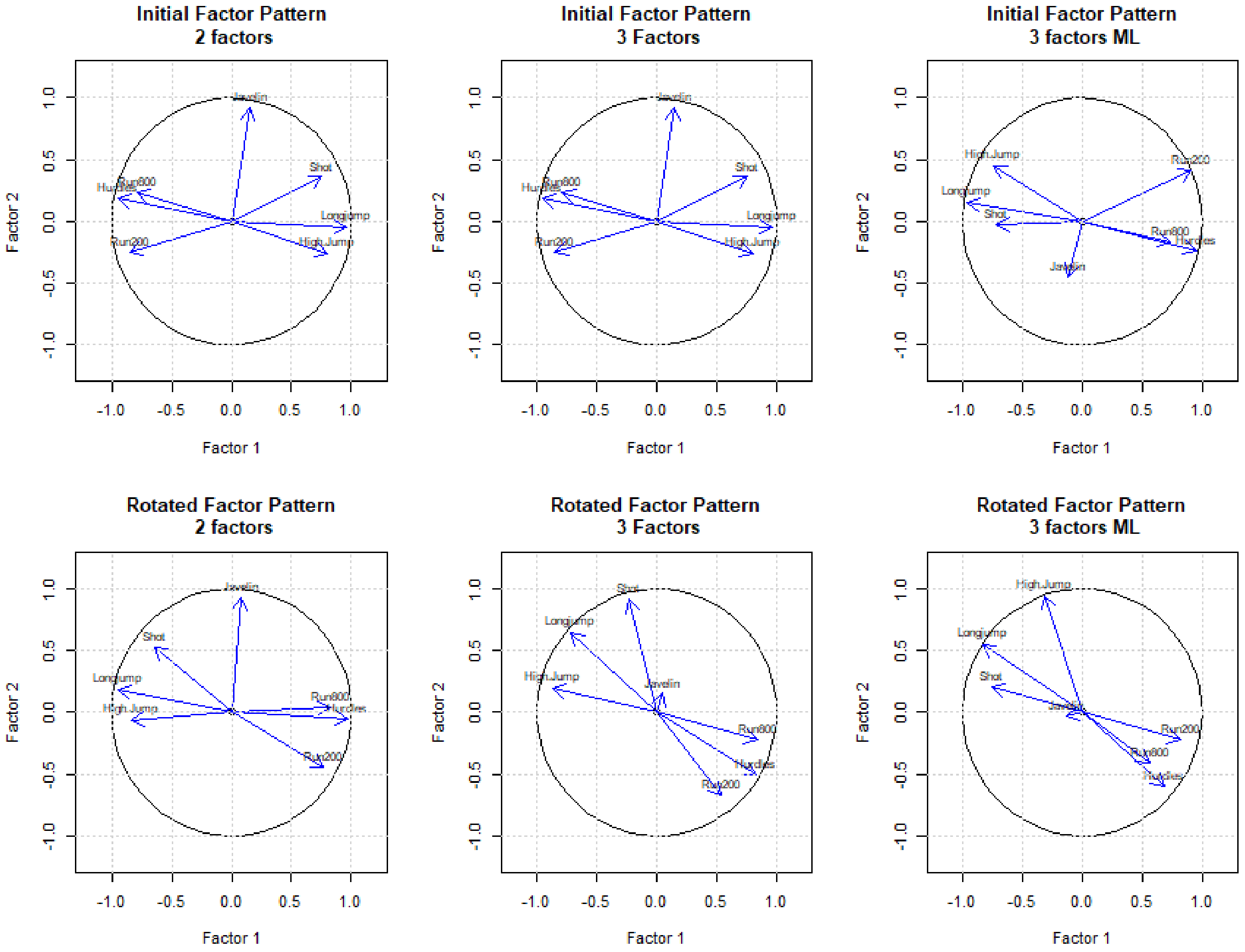
```
            [,1]        [,2]        [,3]
[1,]  0.9527330 -0.18394262  0.02701241
[2,] -0.7998971  0.25979027 -0.29669430
[3,] -0.7550672 -0.36064144  0.45730099
[4,]  0.8516293  0.24433641 -0.09159858
[5,] -0.9626588  0.05003306  0.10251349
[6,] -0.1500479 -0.91996733 -0.35422467
[7,]  0.7904642 -0.23683966  0.26032215
```

```
principal(R,nfactors = 3,
          rotate = "none")$loadings[,]
```

```
Loadings:
                PC1         PC2         PC3
Hurdles   -0.9527330  0.18394262  0.02701241
High.Jump  0.7998971 -0.25979027 -0.29669430
Shot       0.7550672  0.36064144  0.45730099
Run200    -0.8516293 -0.24433641 -0.09159858
Longjump   0.9626588 -0.05003306  0.10251349
Javelin    0.1500479  0.91996733 -0.35422467
Run800    -0.7904642  0.23683966  0.26032215
```

- Disregarding sign changes, these vectors are identical. No smoothing.

Factor loadings



maximum likelihood

On the ML estimation method

Pro:

- Asymtotically Unbiased;
- Model founded – distributions of related statistics may be derived from that;
- Minimized variation subject to model adherence;
- Solution is scale-independent (see page 417-18).
- Makes the factor structure testable.

Con:

- No closed form expression of the estimators $\hat{A}, \widehat{\Delta}$ exists; solutions to the likelihood equation

$$\hat{A} = \left(\hat{A}\hat{A}^T + \widehat{\Delta}\right)R^{-1}\hat{A}$$

  where the solution does not have an analytic expression.
- Subject to model assumptions on distributions (normality) that may fail.

Testing the factor structure

Test statistic:

$$Z = \left(n - 1 - \frac{1}{6}(2k+5) - \frac{2}{3}m\right) log\left(\frac{\left|\hat{A}\hat{A}^T + \widehat{\Delta}\right|}{|R|}\right)$$

Which asymptotically follows a $\chi^2\left(\overbrace{\frac{1}{2}\left((k-m)^2 - k - m\right)}^{\text{The degree}}\right)$ distribution.

n: # observations

k: # variables

m: # factors

$R = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$

k(k-1) / 2

m(m-1) / 2

$= AA^T + \Delta$

k(m+1)
- m(m-1) / 2

$A = k \overset{m}{\begin{bmatrix} \quad \end{bmatrix}}$

Testing the factor structure – 3 factors

$$Z = \left(n - 1 - \frac{1}{6}(2k + 5) - \frac{2}{3}m\right) log\left(\frac{|\hat{A}\hat{A}^T + \hat{\Delta}|}{|R|}\right)$$

R:

```
temp<- fa(R,nfactors = 3,rotate = "varimax",smc=T,n.obs=25,scores="regression",fm="ml")

A<-temp$loadings[,]
n<-25
k<-7
m<-3

my.Z<-(n-1-(2*k+5)/6-2*m/3)*log(det(A%*%t(A)+diag(temp$uniquenesses))/
                                det(R))
my.Z
[1] 7.900366
```

P-value:

```
1-pchisq(my.Z,df=((k-m)^2-k-m)/2)
[1] 0.04811636
```

Borderline significant

Note: `temp$STATISTIC=7.81` should essentially be the same as the above, except for number representation. Gives `p=0.0501`, borderline insignificant.

Testing the factor structure – 3 factors

What to do?

Increase the number of factors?

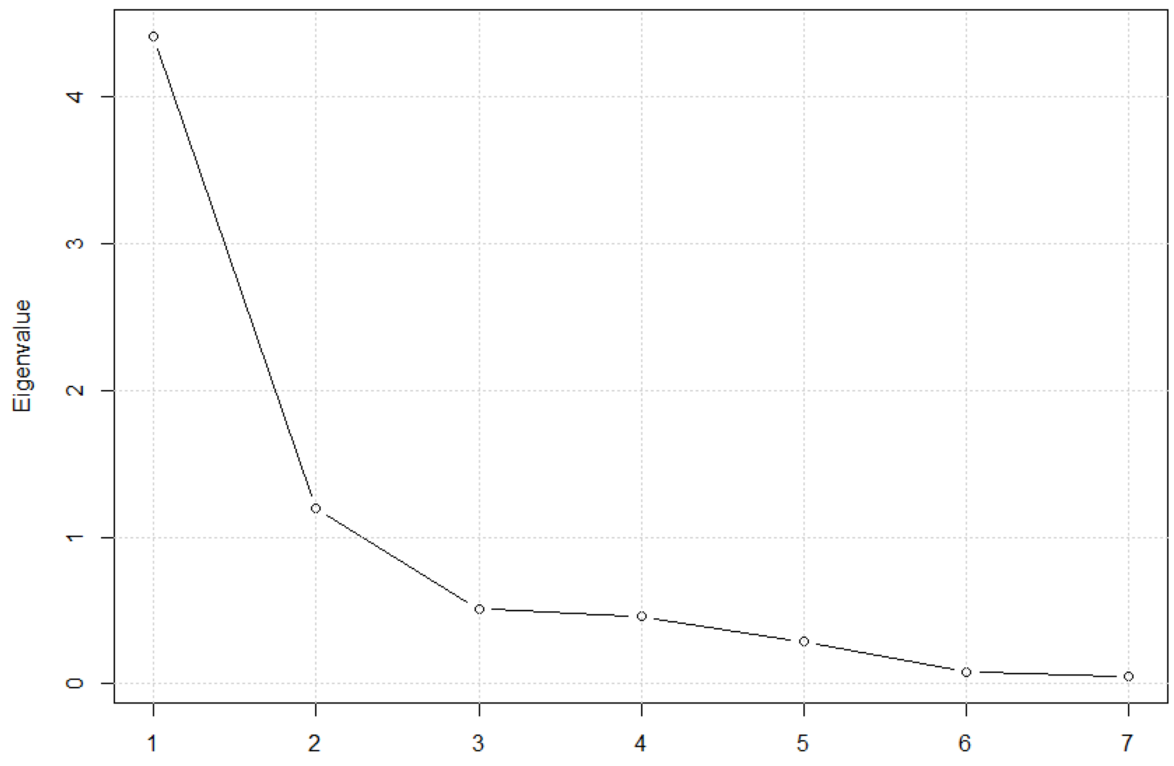
- There is a catch – the asymptotic distribution is $\chi^2\left(\frac{1}{2}\left((k-m)^2-k-m\right)\right)$.
- $\frac{1}{2}\left((7-3)^2-7-3\right)=3,$
- $\frac{1}{2}\left((7-4)^2-7-4\right)=-1$

- **Describing 7 variables from 4 factors (and independent factors) is not testable**; the model is overparametrized.

- Parameters under the model with free ranging correlation: $k(k-1)/2$.
- Parameters under the factor model: $mk-m(m-1)/2$.
- Difference: $\frac{1}{2}\left((k-m)^2-k-m\right)$.
- **More parameters under the factor model with 4 factors.**

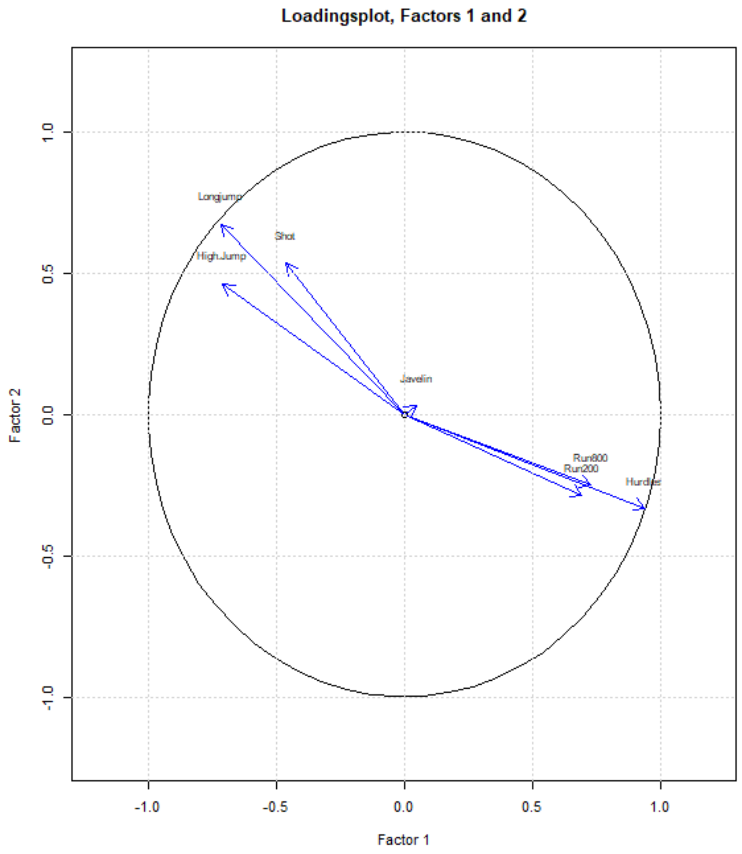
Testing the factor structure – 3 factors

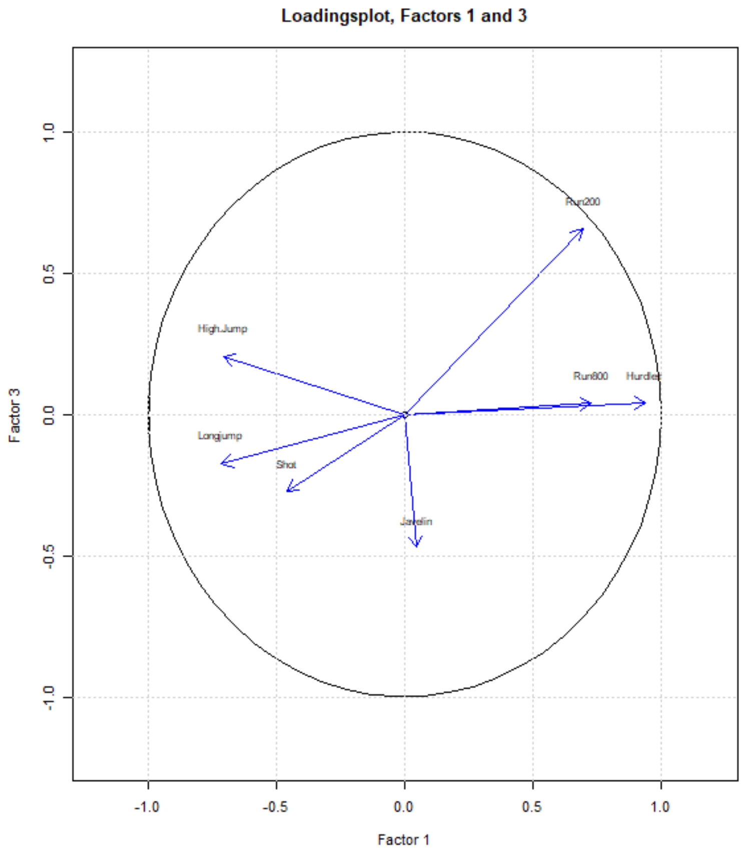
- Bordeline case – and:

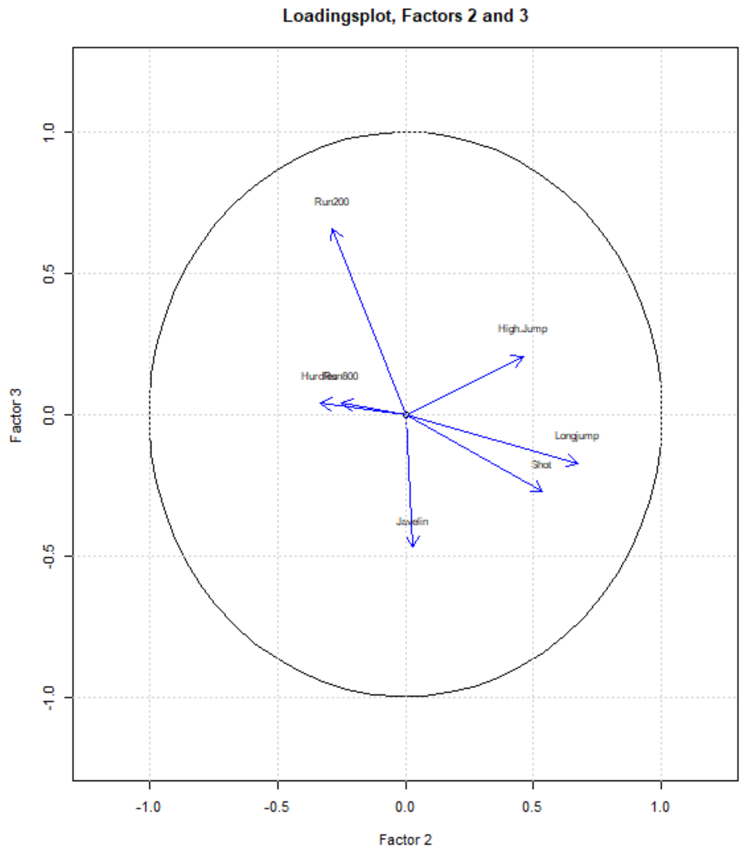


- Suggests 2-3 eigenvectors describing $\boldsymbol{R}$.

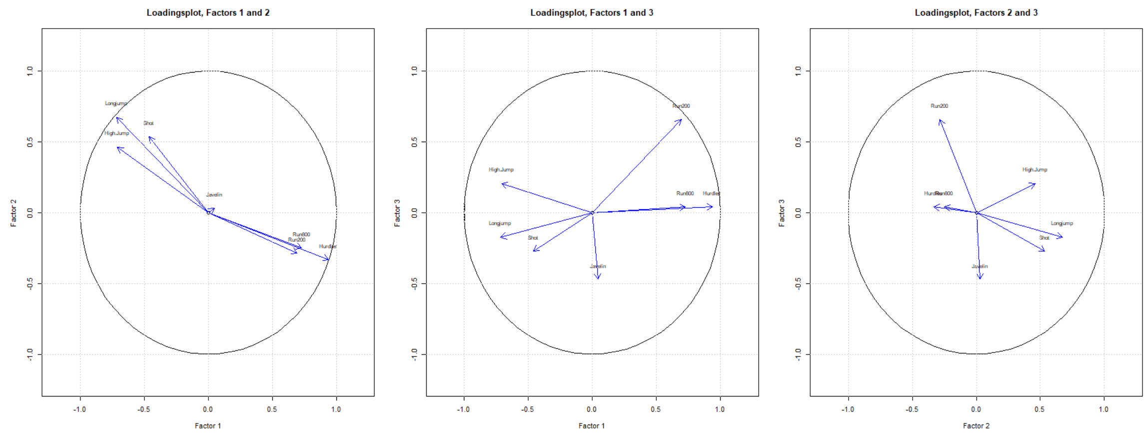
Factor loadings
rotated ml estimation with smc







Factor loadings
rotated ml estimation with smc



- Factor 1: Technical running vs. technical jump/shot;
- Factor 2: Powerful jump/shot vs. running;
- Factor 3: Sprint vs. Power/jarvelin.

Testing the factor structure – 3 factors

- Estimating the unknown factor values

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

R code (column representation):

```
data.frame(Name=heptathlon[,1],scale(heptathlon2)[,]%*%solve(R)%*%A)
```

Testing the factor structure – 3 factors

R code:

```
data.frame(Name=heptathlon[,1],  scale(heptathlon2)[,]%*%solve(R)%*%A)
```

	Name	ML1	ML3	ML2
1	Joyner_Kersee,(USA)	-0.8051687	2.2694725	-1.44028787
2	John,(GDR)	-1.2491403	0.4329102	-0.08238012
3	Behmer,(GDR)	-0.6699509	0.5876323	-1.44924673
4	Choubenkova,(URS)	-0.4789514	-0.1335285	-0.69458387
5	Sablovskaite,(URS)	-0.3933279	-0.2766225	-0.83698712
6	Schulz,(GDR)	0.0998068	0.7339487	0.20996216
7	Fleming,(AUS)	-0.7421209	-0.3391833	-1.00248387
8	Greiner,(USA)	-0.0899822	0.9712517	0.24420982
9	Lajbnerova,(CZE)	-0.4603487	-0.4069532	0.62882274
10	Bouraga,(URS)	-1.1577746	-1.0205545	-0.86337954
11	Wijnsma,(HOL)	0.1622804	0.9931730	0.85477131
12	Dimitrova,(BUL)	-1.3386984	-1.4823043	-0.86010112
13	Scheider,(SWI)	-0.1252746	-0.3341698	0.33535314
14	Braun,(FRG)	-0.2840782	-0.2646298	0.38499499
15	Ruotsalainen,(FIN)	-0.2195077	-0.4356765	-0.01936244
16	Yuping,(CHN)	0.6716668	1.6437404	0.55416741
17	Hagger,(GB)	-0.3431494	0.7216600	1.92827417
18	Brown,(USA)	0.4902519	0.4651493	-0.02554368
19	Mulliner,(GB)	1.1114100	0.9139482	-0.34752665
20	Hautenauve,(BEL)	0.2677671	0.1073350	1.25848564
21	Kytola,(FIN)	0.4450698	-0.5231686	0.93028026
22	Geremias,(BRA)	-0.1227801	-1.9577561	0.60779667
23	Hui-Ing,(TAI)	1.0451993	-1.1919612	-0.68240623
24	Jeong-Mi,(KOR)	0.6259630	-0.8437596	2.01574359
25	Launa,(PNG)	3.5608389	-0.6299534	-1.64857267

- Joyner-Kersee excels in both technical and powerful jumps/shot vs. running; she excels in power over sprint;

a very strong and skilled athlete.

- Launa excels in running over jumps/shot, but not sprint;

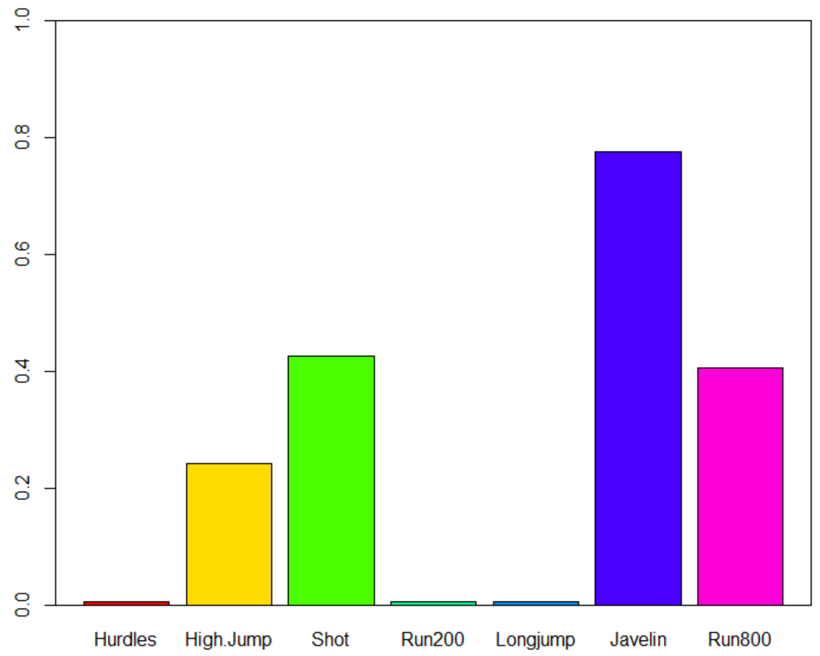
  a strong and technically skilled runner;

- Kytola excels in running and power/jarvelin;

  Kytola is Finnish, their national sport is the Jarvelin.

Heptathlon data –uniqueness factors

```
barplot(temp$uniquenesses,col=rainbow(7),ylim=c(0,1))
box()
```



- Hurdles, 200 Meter Run and Long Jump are essentially completely explained from the factors.
- High Jump, Shot Put, Jarvelin Throw and 800 Meter Run have a larger degree of unique variation, in particular Jarvelin Throw.
- High Jump, Shot Put and Jarvelin Throw are all disciplines with a technical element. 800 Meter Run has a tactical element where conservation of power is a main asset.
- Hurdles, 200 Meter Run and Long Jump, while also for Hurdles and Long Jump containing a technical element, are discipline where direct exposure of strength is a key element.

Canonical Correlation

$\mathbf{y}$, $\mathbf{x}$: Correlation. Data are divided into two variables $\mathbf{x}$ and $\mathbf{y}$.

$\mathbf{y}$, $\mathbf{x}_1 + \ldots + \mathbf{x}_p$: Multiple correlation.
Maximal correlation between y and a linear combination of x. Data are divided into one variable ***y****, and a set of variables* $\mathbf{x}_1, \ldots, \mathbf{x}_p$.

$\mathbf{y}_1 + \mathbf{y}_2 + \ldots + \mathbf{y}_p$, $\mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_q$: Canonical correlation.
Maximal correlation between a linear combination of y and a linear combination of x; data are divided into sets of variables ***x*** *and* ***y****.*

The most ”multivariate” of the correlation forms that we are going to study.

Canonical correlation identifies and quantifies the relations between *two sets of variables*.

Canonical correlation analysis identifies the *linear combination* of one set that is *maximally correlated* to a linear combination of the other set.

Canonical Correlation

- Suppose that we want to investigate the relationship between two parts $Y = (Y_1, \dots, Y_p)$ and $X = (X_1, \dots, X_q)$ of a dataset;

- We assume that a 2nd order moment representation of the simultaneous distribution of $Z = (Y^T, X^T)^T$ is

$$Z = A = \begin{bmatrix} Y \\ X \end{bmatrix}, \mu = E[Z] = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, V[Z] = \Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix},$$

Where we assume that Σ_{yy}, Σ_{xx} both are regular so that the inverse exists;

  - If not, one of the variables may be expressed as a linear relation from the others (essentially Exercise E1); reduce the number of variables studied, or work with generalized inverse matrices in the formulas to follow.

Canonical Correlation

Definition 6.11

Consider Z as above. Then the first *pair of canonical variables* is the pair of linear combinations

$$V_1 = a_1^T Y \text{ and } W_1 = b_1^T X$$

each having variance 1 that maximize the correlation $\rho\left(a^T Y,\, b^T X\right)$ for all (a, b). The maximum correlation ϱ_1 is the *first canonical correlation*. For $r \leq p$ we define the *r'th pair of canonical variables* as the pair of linear combinations

$$V_r = a_r^T Y \text{ and } W_r = b_r^T X \;,$$

which each has the variance 1, which are uncorrelated with the previous $r-1$ pairs of canonical variables, and which maximizes the correlation $\rho\left(a^T Y,\, b^T X\right)$ under those constraints. The maximum correlation ϱ_r is the *r'th canonical correlation*.

Canonical Correlation

Lets us look at the covariance between a linear combination $\boldsymbol{a}$ of Y and a linear combination $\boldsymbol{b}$ of X:

$$Cov\left(\boldsymbol{a}^T Y, \boldsymbol{b}^{\boldsymbol{T}} X\right) = \boldsymbol{a}^T cov(Y, X)\boldsymbol{b} = \boldsymbol{a}^T \Sigma_{\mathrm{yx}} \boldsymbol{b}$$

Cor(2_1, 2_2) = Corv(2_1, 2_2) / sqrt(V(2_1)V(2_2)

V(A\mathbb{Z}))= AV(Z)A^T

The correlation is

V(a^TY) = a^T V(Y) a = a^T \Sigma_{yy}a

$$Cor\left(\boldsymbol{a}^T Y, \boldsymbol{b}^{\boldsymbol{T}} X\right) = \frac{\boldsymbol{a}^T \Sigma_{\mathrm{yx}} \boldsymbol{b}}{\sqrt{\boldsymbol{a}^T \Sigma_{\mathrm{yy}} \boldsymbol{a} \boldsymbol{b}^T \Sigma_{\mathrm{xx}} \boldsymbol{b}}}$$

\tilde{a} = 1 / sqrt(a^T \Sigma_{yy}a)*a
\tilde{b} = 1 / sqrt(b^T \Sigma_{xx}b)*b

\tilde{a}^T \Sigma_{xy}\tilde{b}

We want to find linear combinations so that the correlation between the linear combinations is maximal; ie. Solve the maximization problem

$$\max_{\boldsymbol{a},\boldsymbol{b}} \boldsymbol{a}^T \Sigma_{\mathrm{yx}} \boldsymbol{b}$$

subject to $\boldsymbol{a}^T \Sigma_{\mathrm{yy}} \boldsymbol{a} = 1, \quad \boldsymbol{b}^T \Sigma_{\mathrm{xx}} \boldsymbol{b} = 1$

Canonical Correlation

Technique: Lagrange multipliers: The problem transforms into the maximization problem

$$\max_{\boldsymbol{a},\boldsymbol{b}} \left\{ \boldsymbol{a}^T \Sigma_{\mathrm{yx}} \boldsymbol{b} - \frac{\lambda_1}{2} \left(\boldsymbol{a}^T \Sigma_{\mathrm{yy}} \boldsymbol{a} - 1 \right) - \frac{\lambda_2}{2} \left(\boldsymbol{b}^T \Sigma_{\mathrm{xx}} \boldsymbol{b} - 1 \right) \right\} \qquad (1)$$

Differentiating (1) with respect to $\boldsymbol{a}, \boldsymbol{b}$, and putting equal to 0 gives

$$\begin{aligned} \Sigma_{yx} \boldsymbol{b} - \lambda_1 \Sigma_{yy} \boldsymbol{a} &= 0 \\ \boldsymbol{a}^T \Sigma_{yx} - \lambda_2 \Sigma_{xx} \boldsymbol{b} &= 0 \end{aligned} \qquad (2)$$

ie.

$$\begin{aligned} \boldsymbol{a} &= \frac{1}{\lambda_1} \Sigma_{yy}^{-1} \Sigma_{yx} \boldsymbol{b} \\ \boldsymbol{b} &= \frac{1}{\lambda_2} \Sigma_{xx}^{-1} \Sigma_{xy} \boldsymbol{a} \end{aligned} \qquad (3)$$

Inserting bottom expression of (3) into top expression of (2) and vice versa gives

$$\begin{aligned} \left(\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} - \lambda_1 \lambda_2 \Sigma_{yy} \right) \boldsymbol{a} &= 0 \\ \left(\Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} - \lambda_1 \lambda_2 \Sigma_{xx} \right) \boldsymbol{b} &= 0 \end{aligned} \qquad (4)$$

Or

$$\begin{aligned} \left(\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} - \lambda_1 \lambda_2 \right) \boldsymbol{a} &= 0 \\ \left(\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} - \lambda_1 \lambda_2 \right) \boldsymbol{b} &= 0 \end{aligned} \qquad (5)$$

Canonical Correlation

$$\begin{aligned}\left(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} - \lambda_1\lambda_2\right)\boldsymbol{a} &= 0 \\ \left(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \lambda_1\lambda_2\right)\boldsymbol{b} &= 0\end{aligned} \tag{5}$$

Take

$$\begin{aligned}E_1 &= \Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\ E_2 &= \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\end{aligned}$$

Note that E_1 is a $p \times p$ matrix, while E_2 is a $q \times q$ matrix.

Equation (5) then reads that **a,b** are **eigenvectors** for respectively $\mathrm{E1}$, $\mathrm{E2}$, with the **same** eigenvalue $\boldsymbol{\varrho}^2 = \lambda_1\lambda_2$. This is optimized if $\boldsymbol{\varrho}^2$ is taken as the largest eigenvalue for E_1 (and thus for E_2).

Obviously, this procedure may be iterated to find the maximal correlations perpendicular to **a,b**, respectively, as the remaining eigenvectors of $\mathrm{E1}$, $\mathrm{E2}$.

Canonical Correlation

Theorem 6.12

Let the situation be given in the above mentioned definition and let $\mathrm{D}(Z) = \Sigma$ be partitioned analogously

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}$$

Then the r'th canonical correlation is equal to the r'th largest root ϱ_r of

$$\det \begin{bmatrix} -\rho\Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & -\rho\Sigma_{xx} \end{bmatrix} = 0$$

and the coefficients in the r'th pair of canonical variables satisfies

$$\begin{array}{ll} \text{(i)} & \begin{bmatrix} -\varrho_r\Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & -\varrho_r\Sigma_{xx} \end{bmatrix} \begin{bmatrix} a_r \\ b_r \end{bmatrix} = 0 \\ \text{(ii)} & a_r^T\Sigma_{yy}a_r = 1 \\ \text{(iii)} & b_r^T\Sigma_{xx}b_r = 1 \end{array}$$

Theorem 6.13

Let the situation be as in the previous theorem. Then we have

$$(\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} - \varrho_r^2\Sigma_{yy})a_r = 0$$
$$\det(\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} - \varrho_r^2\Sigma_{yy}) = 0$$

respectively

$$(\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \varrho_r^2\Sigma_{xx})b_r = 0$$
$$\det(\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \varrho_r^2\Sigma_{xx}) = 0$$

- Note that Theorem 6.12 concerns $\boldsymbol{\varrho}$, and Theorem 6.13 concerns $\boldsymbol{\varrho^2}$. Finding the solutions through E_1, E_2 thus means that the canonical correlation $\boldsymbol{\varrho_r}$ is obtained as the square root of the eigenvalue $\boldsymbol{\varrho_r^2}$.

Canonical Correlation

- Assume wlog that $p < q$, so that we may represent the correlation between Y and X though p pairs of eigenvectors **a**,**b**. Then there are at most p posotive eigenvalues for E_1, E_2. The eigenvectors $a_1, \dots, a_p$ and $b_1, \dots, b_p$ for respectively E_1, E_2 are called *the canonical vectors, the linear combinations* $A^T Y, B^T X$ *the canonical variables* for Y, X, respectively, while the square root of the eigenvalues $\varrho_1, \dots, \varrho_p$ are called the *canonical correlations* between Y and X.

x and y is symmetric

Canonical Correlation

- Take

$$\Delta_y = \begin{bmatrix} \sigma_{y_1}^{-1} & & \\ & \ddots & \\ & & \sigma_{y_p}^{-1} \end{bmatrix}, \Delta_x = \begin{bmatrix} \sigma_{x_1}^{-1} & & \\ & \ddots & \\ & & \sigma_{x_q}^{-1} \end{bmatrix} \Delta = \begin{bmatrix} \Delta_y & \\ & \Delta_x \end{bmatrix}$$

The the correlation matrix for Z is

$$R = \Delta\Sigma\Delta$$

Note that

$$V\left(\Delta_y Y\right) = \Delta_y \Sigma_{\mathrm{yy}} \Delta_y = R_y, V(\Delta_x X) = \Delta_x \Sigma_{xx} \Delta_x$$

Canonical Correlation

- Take $A = (a_1; \cdots; a_p)$, $B = (b_1; \cdots; b_p)$ to be the canonical vectors $V = A^T Y$, $W = B^T X$ to be the canonical variables. Then

$$V\begin{bmatrix} Y \\ X \\ V \\ W \end{bmatrix} = \begin{bmatrix} R_{yy} & R_{yx} & R_{yv} & R_{yw} \\ R_{xy} & R_{xx} & R_{xv} & R_{xw} \\ R_{vy} & R_{vx} & I_p & R_{vw} \\ R_{wy} & R_{wx} & R_{wv} & I_p \end{bmatrix}$$

With

$$R_{vw} = R_{wv} = \begin{pmatrix} \varrho_1 & & \\ & \ddots & \\ & & \varrho_p \end{pmatrix}$$

The remaining 4 elements may be found from the defintion of V and W:

$$Cor(Y, V) = Cov(\Delta_y Y, A^T Y) = \Delta_y \Sigma_{yy} A = R_{yy} \Delta_y^{-1} A$$

$$Cor(Y, W) = Cov(\Delta_y Y, B^T X) = \Delta_y \Sigma_{yx} B = R_{yx} \Delta_x^{-1} B$$

$$Cor(X, V) = Cov(\Delta_x X, A^T Y) = \Delta_x \Sigma_{xy} A = R_{xy} \Delta_y^{-1} A$$

$$Cor(X, W) = Cov(\Delta_x X, B^T X) = \Delta_y \Sigma_{yy} B = R_{xx} \Delta_x^{-1} B$$

Canonical Correlation

$$Cor(Y,V) = Cov(\Delta_y Y, A^T Y) = \Delta_y \Sigma_{yy} A = R_{yy} \Delta_y^{-1} A$$

$$Cor(Y,W) = Cov(\Delta_y Y, B^T X) = \Delta_y \Sigma_{yx} B = R_{yx} \Delta_x^{-1} B$$

$$Cor(X,V) = Cov(\Delta_x X, A^T Y) = \Delta_x \Sigma_{xy} A = R_{xy} \Delta_y^{-1} A$$

$$Cor(X,W) = Cov(\Delta_x X, B^T X) = \Delta_y \Sigma_{yy} A = R_{xx} \Delta_x^{-1} B$$

Note that the canonical vectors above enter through the standization

$$\Delta_y^{-1} A = \begin{pmatrix} \sigma_{y_1} & & \\ & \ddots & \\ & & \sigma_{y_p} \end{pmatrix} A,\ \Delta_x^{-1} B = \begin{pmatrix} \sigma_{x_1} & & \\ & \ddots & \\ & & \sigma_{x_q} \end{pmatrix} B$$

This leads to that one often represent the canonical variables in the so-called *standardized form*:

$$V^* = (\Delta_y^{-1} A)^T Y = A^T \Delta_y^{-1} Y, \qquad W^* = (\Delta_x^{-1} B)^T X = B^T \Delta_x^{-1} X$$

Testing that a canonical correlation is 0

- So far, we have not invoked distributional properties of the data apart from the 2nd order moment representation.

- We now make the further assumption that data follows a normal distribution. The distribution is completely specified by mean and variance.
- Simultaneous distribution of normaliuzed $\left(a_p^T Y, b_p^T X\right)$, corresponding to the smallest canonical correlation:

$$N\left(\begin{pmatrix} a_p^T \mu_y \\ b_p^T \mu_x \end{pmatrix}, \begin{pmatrix} 1 & \varrho_p \\ \varrho_p & 1 \end{pmatrix}\right)$$

Testing that a canonical correlation is 0

$$N\left(\begin{pmatrix} a_p^T\mu_y \\ b_p^T\mu_x \end{pmatrix}, \begin{pmatrix} 1 & \varrho_p \\ \varrho_p & 1 \end{pmatrix}\right)$$

Testing

$$H_0\colon \varrho_p = 0 \qquad vs. \qquad H\colon \varrho_p > 0$$

Thus conforms with testing that the variance matrix is diagonal.

Test statistic (Bartlett's test):

$$Q = \left(n - \frac{5}{2}\right)\log\left(1 - \varrho_p^2\right) \sim \chi_1^2$$

Because the canonical variables are independent, the test may be carried out sequentially. The cc function derives simultaneous tests based on the Wilks distribution; see the following example.

The fewer the non-zero canonical correlations, the less and the simpler the dependencies between Y and X.

- We will get back to tests for covariance structure with the multivariate analysis of variance

Canonical correlation

Canonical correlation analysis has a reputation of being the most difficult multivariate technique to interpret. In many respects it is a well earned reputation! Certainly one has to know the variables involved very well to have any hope of extracting a convincing explanation. But in some circumstances (…), CCA does provide a useful description of the association between two sets of variables.

- Brian Everitt, 2005.

Example: Heads measurements data

- The data consist of measurements of the length and breadth of the heads of pairs of adult brothers in 25 randomly sampled families. All measurements are expressed in millimeters.

```
frets<-read.csv2("Data/frets.csv")
head(frets)
   l1  b1  l2  b2
1 191 155 179 145
2 195 149 201 152
3 181 148 185 149
4 183 153 188 149
5 176 144 171 142
6 208 157 192 152
```

L1: The head length of the eldest son.
B1: The head breadth of the eldest son.
L2: The head length of the second son.
B2: The head breadth of the second son.

Example: Heads measurements data

Correlation structure:

```
frets2<-scale(frets,center=T,scale=T)
round(cov(frets2),digits=2)
     l1   b1   l2   b2
l1 1.00 0.73 0.71 0.70
b1 0.73 1.00 0.69 0.71
l2 0.71 0.69 1.00 0.84
b2 0.70 0.71 0.84 1.00
```

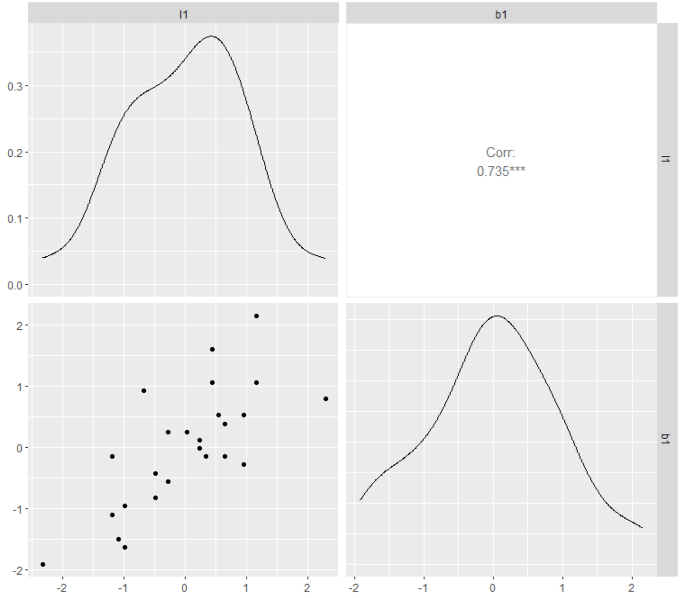
We want to study the correlation between the eldest and the second son:

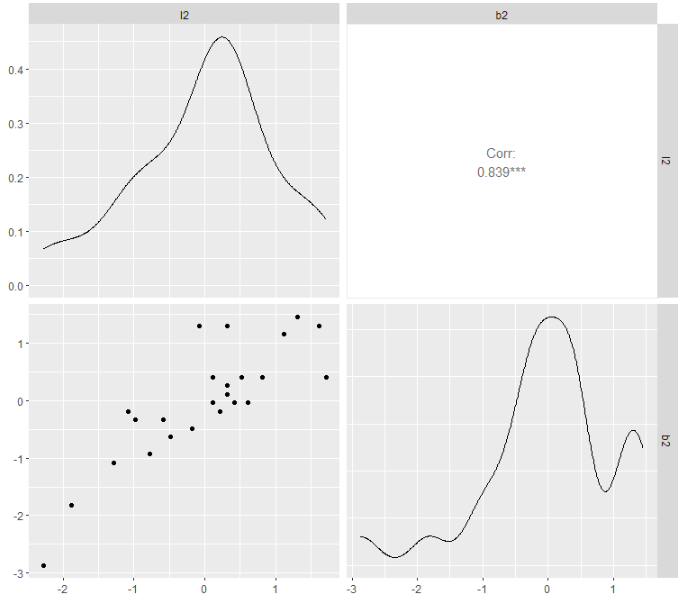
$$Y = (l1, b1)^T, X = (l2, b2)^T$$

Example: Heads measurements data

Correlation structure:

```
brother1<-frets2[,1:2]
brother2<-frets2[,3:4]

ggpairs(brother1)
ggpairs(brother2)

round(cov(brother1,brother2),digits=2)

      l2   b2
l1 0.71 0.70
b1 0.69 0.71
```





Finding the canonical vectors and correlations the cc function

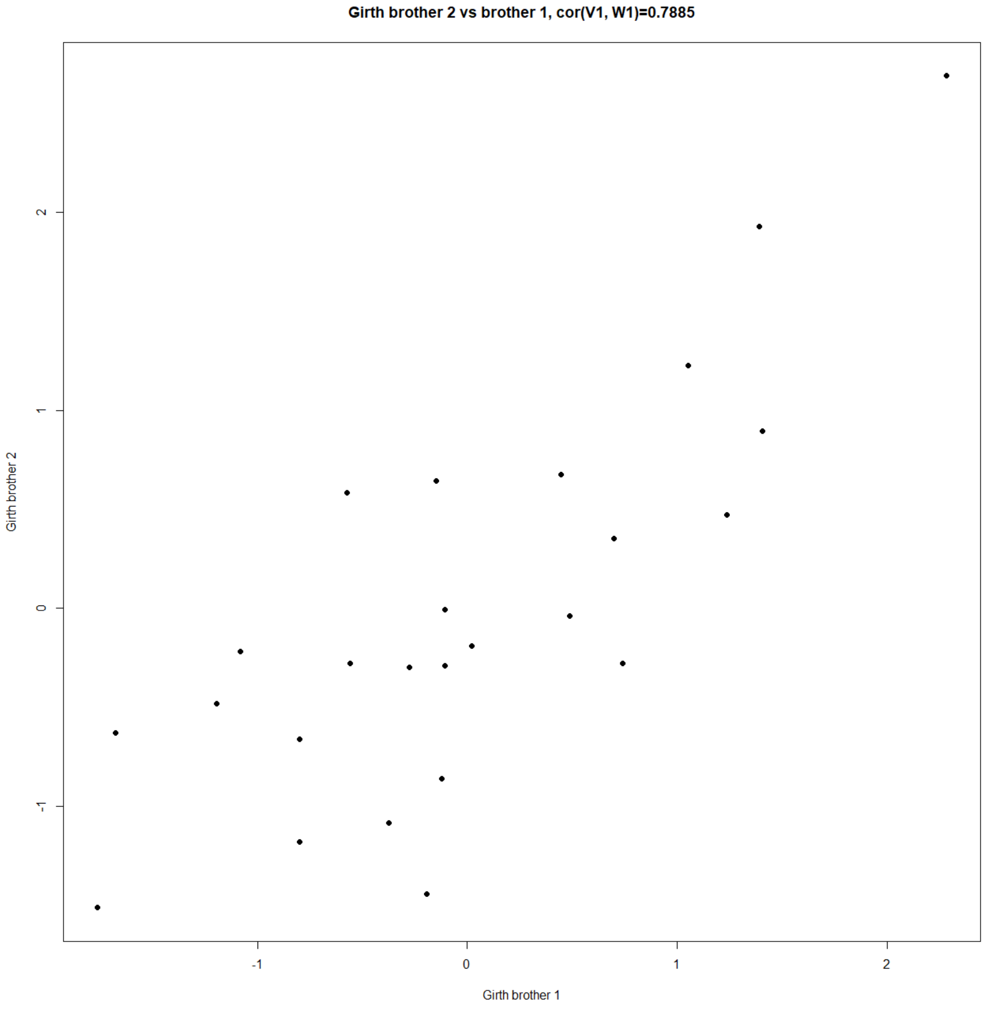
```
my.cca<-cc(brother1,brother2)

# The A matrix:
(A<-my.cca$xcoef)
           [,1]      [,2]
l1 -0.5521896 -1.366374
b1 -0.5215372  1.378365

# The B matrix:
(B<-my.cca$ycoef)
           [,1]      [,2]
l2 -0.5044484 -1.768570
b2 -0.5382877  1.758566

# canonincal correlations:

my.cca$cor
[1] 0.7885079 0.0537397

# note:
round(t(A)%*%cor(brother1)%*%A,digits=2)
     [,1] [,2]
[1,]    1    0
[2,]    0    1

round(t(B)%*%cor(brother2)%*%B,digits=2)
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

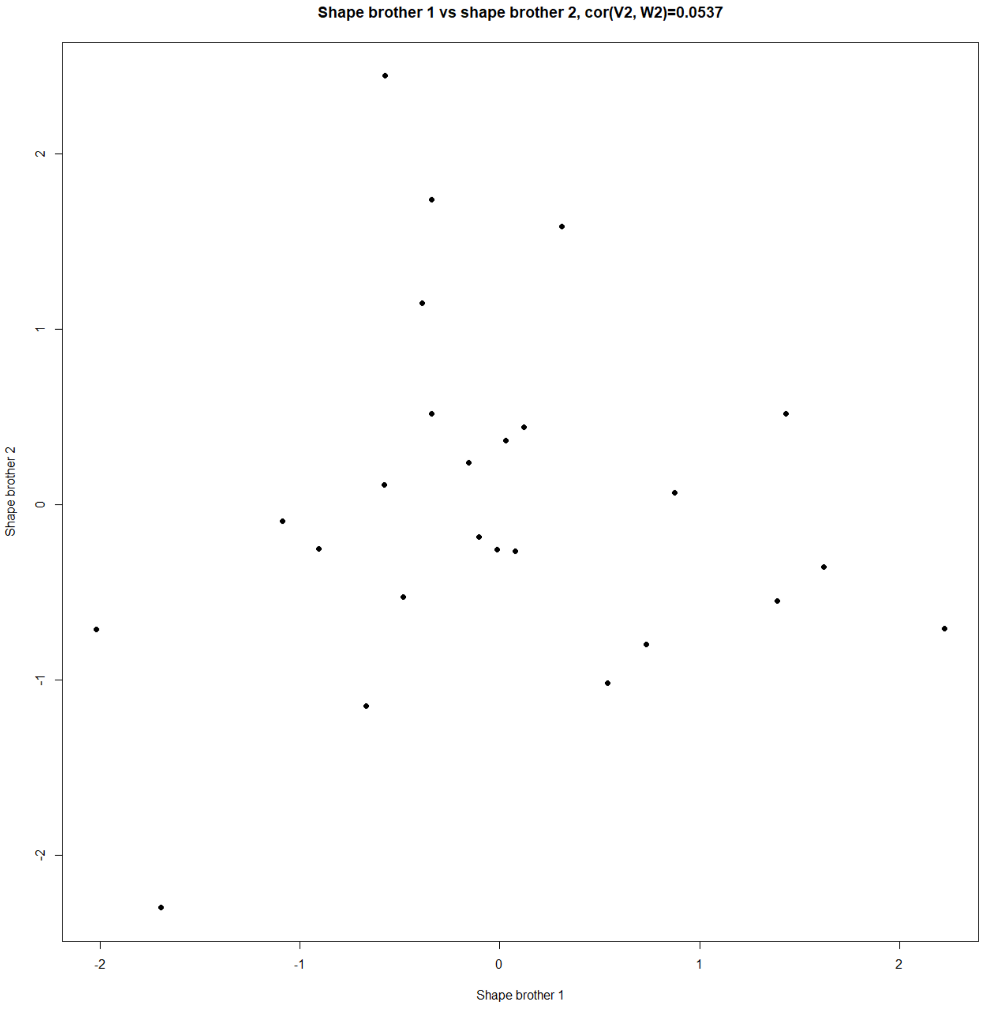
More or less (minus) the sum of the two variables; we label it ‘girth’

The contrast between length and breath; we label it ‘shape’

first ccc=0.79, second ccc=0.05

Example: Heads measurements data





Heads measurements data correlations between variables

```
my.cca$scores$corr.X.xscores  #(l1, b1) (V_1, V_2)
          [,1]        [,2]
l1 -0.9352877 -0.3538884
b1 -0.9271512  0.3746875
my.cca$scores$corr.X.yscores  #(l1, b1) (W_1, W_2)
          [,1]         [,2]
l1 -0.7374817 -0.01901786
b1 -0.7310660  0.02013559

my.cca$scores$corr.Y.xscores  #(l2, b2) (V_1, V_2)
     [,1]     [,2]
l2 -0.7539771 -0.01572908
b2 -0.7582663  0.01474027
my.cca$scores$corr.Y.yscores  #(l2, b2) (W_1, W_2)
     [,1]    [,2]
l2 -0.9562074 -0.2926900
b2 -0.9616470  0.2742901

# note:
cor(brother2)%*%B
     [,1]    [,2]
l2 -0.9562074 -0.2926900
b2 -0.9616470  0.2742901
```

Heads measurements data: Summing up

$$T = Y^T Y = (n-1)\widehat{\Sigma}_{yy}$$
$$H = Y^T X (X^T X)^{-1} X^T Y = (n-1)\widehat{\Sigma}_{yx}\widehat{\Sigma}_{xx}^{-1}\widehat{\Sigma}_{xy}$$
$$E = T - H = (n-1)(\widehat{\Sigma}_{yy} - \widehat{\Sigma}_{yx}\widehat{\Sigma}_{xx}^{-1}\widehat{\Sigma}_{xy})$$

We see that T corresponds to the total variation and E to the residual variation after having predicted Y by means of X.

- Summing up from the cc fu nation rather than Chisquare:

Theorem 6.16

Testing whether the canonical correlations are zero is equivalent to test whether the eigenvalues of $E^{-1}H$ are zero.

```
     CanCor Squared.CanCor eigenvaluesInvEH  proportion cumulative    id
1 0.7885079    0.621744734       1.64371733 0.998241044   0.998241 Wilks
2 0.0537397    0.002887956       0.00289632 0.001758956   1.000000 Wilks
       stat     approx df1 df2      p.value
1 0.3771629 6.59719349   4  42 0.0003256458
2 0.9971120 0.06371905   1  22 0.8030550074
```

- One canonical coefficient is sufficient; girth correlates between brothers but there is no evidence that shape does.

Exercises

- Exam 2011, problem 4
  - CCA, pen and paper
- Exam 2011, problem 5
  - CCA, SAS-outputs
- EX 7.6: Canonical Correlation Analysis on Beef Characterization.
  - SAS / R
  - Physical and chemical measurements vs. tasting panel