Applied matrix algebra for Multivariate Analysis

Jan Graffelman[1]

[1]Department of Statistics and Operations Research
Universitat Politècnica de Catalunya
Barcelona, Spain

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

jan.graffelman@upc.edu

February 8, 2020

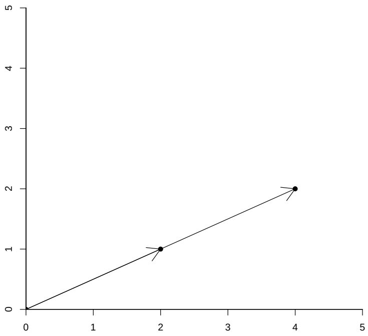
Contents

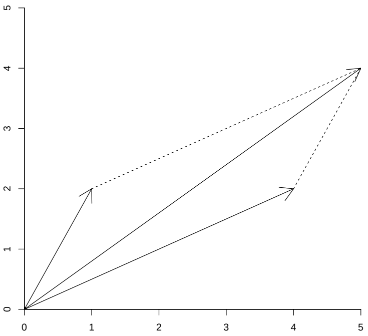
Vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix},$$

$$\mathbf{x}' = \left[\ x_1, x_2, \ldots, x_n\ \right] \qquad \mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

Geometric interpretation





Vectors

Norm or length

$$\| \mathbf{x} \| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$$

$$\| \alpha \mathbf{x} \| = \alpha \, \| \mathbf{x} \|$$

Scalar product:

$$\mathbf{x}'\mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

$$\mathbf{x}'\mathbf{y} = 0 \leftrightarrow \mathbf{x} \text{ and } \mathbf{y} \text{ perpendicular }.$$

Angle:

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\| \mathbf{x} \| \| \mathbf{y} \|}$$

Linear combination, linear (in)dependence

Linear combination of n vectors

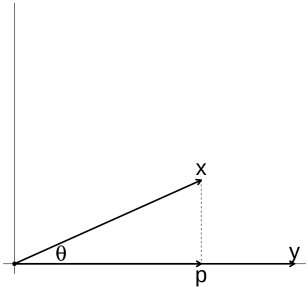
$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n$$

To investigate linear dependence:

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0} \tag{1}$$

- The set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n$ is linearly dependent iff Eq. (1) holds for some set $(a_1, a_2, \ldots, a_n)$ not all zero.
- The set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n$ is linearly independent iff Eq. (1) holds only for $(a_1, a_2, \ldots, a_n) = (0, 0, \ldots, 0)$.

Projection



$$\cos\theta = \frac{\| \mathbf{p} \|}{\| \mathbf{x} \|}, \quad \| \mathbf{p} \| = \frac{\mathbf{x}'\mathbf{y}}{\| \mathbf{y} \|}, \quad \mathbf{p} = \alpha \mathbf{y}, \quad \alpha = \frac{\| \mathbf{p} \|}{\| \mathbf{y} \|} \rightarrow \mathbf{p} = \left(\frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \right) \mathbf{y}$$

Matrix

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Basic matrix operations

- sum

$$\mathbf{C}_{n\times p} = \mathbf{A}_{n\times p} + \mathbf{B}_{n\times p} \qquad c_{ij} = a_{ij} + b_{ij}$$

- scalar multiplication

$$\mathbf{C} = \alpha \mathbf{A}_{n\times p} \qquad c_{ij} = \alpha a_{ij}$$

- product

$$\mathbf{C}_{n\times p} = \mathbf{A}_{n\times k}\mathbf{B}_{k\times p} \qquad c_{ij} = \sum_{l=1}^{k} a_{il} b_{lj}$$

- transposition

$$\mathbf{C} = \mathbf{A}' \qquad c_{ij} = a_{ji} \qquad \mathbf{E} = \mathbf{AB} \quad \mathbf{E}' = (\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

- inversion

$$\mathbf{A}_{k\times k} \qquad \mathbf{AB} = \mathbf{BA} = \mathbf{I} \qquad \mathbf{B}_{k\times k} = \mathbf{A}^{-1}$$

Some particular cases of matrix multiplication

- $\mathbf{A}_{p\times k}\mathbf{B}_{k\times l} = \mathbf{C}_{p\times l}$
- $\mathbf{A}_{p\times k}\mathbf{x}_{k\times 1} = \mathbf{y}_{p\times 1}$
- $\mathbf{x}'_{1\times p}\mathbf{A}_{p\times k} = \mathbf{y}'_{1\times k}$
- $\mathbf{X}_{n\times p}\mathbf{D}_{p\times p} = \left[\begin{array}{c|c|c|c} \mathbf{x}_1 d_1 & \mathbf{x}_2 d_2 & \cdots & \mathbf{x}_p d_p \end{array}\right]$
- $\mathbf{D}_{n\times n}\mathbf{X}_{n\times p} = \left[\begin{array}{c} d_1\mathbf{x}_1 \\ \hline d_2\mathbf{x}_2 \\ \hline \vdots \\ \hline d_n\mathbf{x}_n \end{array}\right]$

Some special matrices

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \mathbf{O} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

Symmetric and Orthogonal Matrices

- Symmetric matrix: $\mathbf{A} = \mathbf{A}'$

$$\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

- Orthogonal (orthonormal) matrix:

$$\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I} \qquad \mathbf{A}' = \mathbf{A}^{-1}$$

Determinant

Matrix property

$$|\mathbf{A}_{k\times k}| = \sum_{j=1}^{k} a_{ij} \, |\mathbf{A}_{ij}| \, (-1)^{i+j}$$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad |\mathbf{A}| = ad - bc$$

$k > 2$: by computer.

$|\mathbf{A}| = 0$ implies linear dependence, and $\mathbf{A}$ is singular.

Inverse

$$\mathbf{A}_{k\times k} \qquad \mathbf{AB} = \mathbf{BA} = \mathbf{I} \qquad \mathbf{B}_{k\times k} = \mathbf{A}^{-1}$$

Case 2×2

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If $k > 2$ then use a computer.

Inverse of a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \qquad \mathbf{D}^{-1} = \begin{bmatrix} 1/d_1 & 0 & \cdots & 0 \\ 0 & 1/d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/d_n \end{bmatrix}$$

Rank

- Row rank = maximum number of linearly independent rows.
- Column rank = maximum number of linearly independent columns.
- "The" rank = row rank = column rank.
- A rank k matrix can be represented exactly in a k dimensional space.

Trace

$$\mathbf{A}_{k \times k} \qquad \mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$$

- $\mathrm{tr}(\alpha \mathbf{A}) = \alpha \mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}) \qquad \mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB}) = \mathrm{tr}(\mathbf{BCA})$
- $\mathrm{tr}(\mathbf{AA}') = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}^2$

Eigenvalues & eigenvectors

$$\mathbf{Av} = \lambda \mathbf{v} \qquad \mathbf{A}_{k \times k}$$

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$$

The characteristic equation

$$|\,\mathbf{A} - \lambda \mathbf{I}\,| = 0$$

- There are k roots, not necessarily all distinct.
- If $\mathbf{A} = \mathbf{A}'$, all roots are real.
- Each root (eigenvalue) has an associated eigenvector.
- $\mathbf{v}$ usually scaled (normalized) to unit length such that $\mathbf{v}'\mathbf{v} = 1$.

Spectral decomposition

$$\mathbf{A}_{k \times k} \text{ and } \mathbf{A} = \mathbf{A}'$$

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{v}_i \mathbf{v}_i' = \lambda_1 \mathbf{v}_1 \mathbf{v}_1' + \lambda_2 \mathbf{v}_2 \mathbf{v}_2' + \cdots + \lambda_k \mathbf{v}_k \mathbf{v}_k'$$

$$\mathbf{A} = \mathbf{V} \mathbf{D}_\lambda \mathbf{V}'$$

$$\mathbf{V} = \left[\begin{array}{c|c|c|c} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{array} \right], \quad \mathbf{D}_\lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}, \quad \mathbf{V}'\mathbf{V} = \mathbf{I}.$$

- Eigenvalues usually ordered s.t. $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k$
- If $\mathbf{A}$ is not of full rank, there will be zero eigenvalues.
- $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{V}\mathbf{D}_\lambda\mathbf{V}') = \text{tr}(\mathbf{V}'\mathbf{V}\mathbf{D}_\lambda) = \text{tr}(\mathbf{I}\mathbf{D}_\lambda) = \text{tr}(\mathbf{D}_\lambda) = \sum_{i=1}^{k} \lambda_i$.
- $\lambda_1 \mathbf{v}_1 \mathbf{v}_1' + \lambda_2 \mathbf{v}_2 \mathbf{v}_2'$ provides a rank 2 least squares approximation to a $\mathbf{A}$.

Exercise

- We have a matrix

$$\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

- Find the eigenvalues and eigenvectors by hand and by using R.

```
> S <- matrix(c(2,1,1,1),ncol=2)
> S
     [,1] [,2]
[1,]    2    1
[2,]    1    1
> out <- eigen(S)
> V <- out$vectors
> V
           [,1]       [,2]
[1,] -0.8506508  0.5257311
[2,] -0.5257311 -0.8506508
> D <- diag(out$values)
> D
         [,1]     [,2]
[1,] 2.618034 0.000000
[2,] 0.000000 0.381966
> V%*%D%*%t(V)
     [,1] [,2]
[1,]    2    1
[2,]    1    1
>
```

A data matrix

	Husband age	Husband height	Wife age	Wife height
1	49	1809	43	1590
2	25	1841	28	1560
3	40	1659	30	1620
4	52	1779	57	1540
5	58	1616	52	1420
6	32	1695	27	1660
7	43	1730	52	1610
8	47	1740	43	1580
9	31	1685	23	1610
10	26	1735	25	1590

Age and height of husband and wife for 10 couples

The problem

- The data matrix $\mathbf{X}$ is 10×4, and of rank 4.
- Can we approximate $\mathbf{X}$ by a rank 2 matrix, say $\hat{\mathbf{X}}$
- Entries of $\hat{\mathbf{X}}$ must be as "close" as possible to $\mathbf{X}$
- Note: a rank 2 matrix can be represented in a two-dimensional graph.

The Solution

$$\mathbf{X} = \begin{bmatrix} 49 & 1809 & 43 & 1590 \\ 25 & 1841 & 28 & 1560 \\ 40 & 1659 & 30 & 1620 \\ 52 & 1779 & 57 & 1540 \\ 58 & 1616 & 52 & 1420 \\ 32 & 1695 & 27 & 1660 \\ 43 & 1730 & 52 & 1610 \\ 47 & 1740 & 43 & 1580 \\ 31 & 1685 & 23 & 1610 \\ 26 & 1735 & 25 & 1590 \end{bmatrix} \qquad \hat{\mathbf{X}} = \begin{bmatrix} 43.94 & 1809.21 & 43.72 & 1589.89 \\ 46.36 & 1838.22 & 48.00 & 1562.03 \\ 35.09 & 1659.32 & 29.27 & 1619.79 \\ 44.07 & 1780.44 & 44.75 & 1538.92 \\ 39.40 & 1618.03 & 39.32 & 1418.55 \\ 35.64 & 1694.60 & 29.49 & 1660.28 \\ 39.24 & 1731.53 & 36.00 & 1608.81 \\ 40.67 & 1740.69 & 38.73 & 1579.51 \\ 36.67 & 1683.96 & 31.92 & 1610.78 \\ 39.90 & 1733.25 & 37.32 & 1591.27 \end{bmatrix}$$

Least squares criterion

- In linear regression, we estimate the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ by minimizing $\sum e_i^2$, where $e_i = y_i - (b_0 + b_1 x_i)$.
- In this matrix approximation we minimize the errors in $\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}$.
- The least squares criterion amounts to $\sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 = \mathrm{tr}(\mathbf{E}'\mathbf{E})$.

Singular value decomposition (compact)

Any real matrix $n \times p$ matrix $\mathbf{X}$ can be decomposed as

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

- $\mathbf{U}$ $n \times r$ matrix of orthonormal left singular vectors. $\mathbf{U}'\mathbf{U} = \mathbf{I}_r$
- $\mathbf{D}$ $r \times r$ diagonal matrix of non-increasing positive singular values $(d_{11} \geq d_{22} \geq \cdots \geq d_{rr})$.
- $\mathbf{V}$ $p \times r$ matrix of orthonormal right singular vectors. $\mathbf{V}'\mathbf{V} = \mathbf{I}_r$

Alternatively

$$\mathbf{X} = \sum_{i=1}^{r} d_{ii}\mathbf{u}_i\mathbf{v}_i' = d_1\mathbf{u}_1\mathbf{v}_1' + d_2\mathbf{u}_2\mathbf{v}_2' + \cdots + d_r\mathbf{u}_r\mathbf{v}_r'$$

Singular value decompostion (theorem)

A rank k approximation $\hat{\mathbf{X}}$ to matrix $\mathbf{X}$, optimal in the least squares sense, is obtained as

$$\hat{\mathbf{X}} = \mathbf{U}_{[,1:k]}\mathbf{D}_{[1:k,1:k]}\mathbf{V}_{[,1:k]}{}'$$

E.g., a rank 2 approximation to matrix $\mathbf{X}$ is obtained by $\mathbf{U}_{n\times 2}\mathbf{D}_{(2\times 2)}\mathbf{V}_{p\times 2}{}'$

Singular vectors are eigenvectors

- $\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}'\mathbf{U}\mathbf{D}\mathbf{V}' = \mathbf{V}\mathbf{D}^2\mathbf{V}'$
- $\mathbf{X}\mathbf{X}' = \mathbf{U}\mathbf{D}\mathbf{V}'\mathbf{V}\mathbf{D}\mathbf{U}' = \mathbf{U}\mathbf{D}^2\mathbf{U}'$
- Eigenvalues of $\mathbf{X}\mathbf{X}'$ and $\mathbf{X}'\mathbf{X}$ are squared singular values.
- Singular vectors are eigenvectors, $\mathbf{U}$ of $\mathbf{X}\mathbf{X}'$ and $\mathbf{V}$ of $\mathbf{X}'\mathbf{X}$.

Singular value decomposition (extended)

Sometimes the svd is also written as

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

- $\mathbf{U}$ $n \times p$ matrix of orthonormal left singular vectors. $\mathbf{U}'\mathbf{U} = \mathbf{I}_p$
- $\mathbf{D}$ $p \times p$ diagonal matrix of non-increasing singular values.
- $\mathbf{V}$ $p \times p$ matrix of orthonormal right singular vectors. $\mathbf{V}'\mathbf{V} = \mathbf{I}_p$

where matrix $\mathbf{D}$ now has trailing zeros on the diagonal.

Singular value decomposition in R

```
X <- read.table("c:/data/HusbandsAndWives.dat")
X <- read.table("http://www-eio.upc.edu/~jan/data/MVA/HusbandsAndWives.dat")

X <- as.matrix(X)
X <- X[,1:4]

svd.results <- svd(X)
U <- svd.results$u
V <- svd.results$v
D <- diag(svd.results$d)
print(U)
print(V)
print(D)

U2 <- U[,1:2]
V2 <- V[,1:2]
D2 <- D[1:2,1:2]

Xhat <- U2%*%D2%*%t(V2)
print(Xhat)
```

Goodness-of-fit

- How good (or bad) is our approximation to $\mathbf{X}$?
- Some statistic expressing goodness of fit is needed (like R^2 in regression)
- The singular values are informative about the goodness-of-fit

Note that

$$\mathrm{tr}(\mathbf{X}'\mathbf{X}) = \mathrm{tr}(\mathbf{VDU}'\mathbf{UDV}') = \mathrm{tr}(\mathbf{VD}^2\mathbf{V}') = \mathrm{tr}(\mathbf{V}'\mathbf{VD}^2) = \mathrm{tr}(\mathbf{D}^2) = \sum_{j=1}^{p} d_{jj}^2 = \sum_{j=1}^{p} \lambda_j$$

And that for a rank 2 approximation

$$\begin{aligned}\mathrm{tr}(\hat{\mathbf{X}}'\hat{\mathbf{X}}) &= \mathrm{tr}(\mathbf{V}_{[,1:2]}\mathbf{D}_{[1:2,1:2]}\mathbf{U}'_{[,1:2]}\mathbf{U}_{[,1:2]}\mathbf{D}_{[1:2,1:2]}\mathbf{V}'_{[,1:2]}) = \mathrm{tr}(\mathbf{V}_{[,1:2]}\mathbf{D}^2_{[1:2,1:2]}\mathbf{V}'_{[,1:2]}) = \mathrm{tr}(\mathbf{V}'_{[,1:2]}\mathbf{V}_{[,1:2]}\mathbf{D}^2_{[1:2,1:2]}) \\ &= \mathrm{tr}(\mathbf{D}^2_{[1:2,1:2]}) = d_{11}^2 + d_{22}^2 = \lambda_1 + \lambda_2\end{aligned}$$

And that for the error matrix

$$\mathrm{tr}(\mathbf{E}'\mathbf{E}) = \mathrm{tr}((\mathbf{X} - \hat{\mathbf{X}})'(\mathbf{X} - \hat{\mathbf{X}})) = \mathrm{tr}(\mathbf{V}_{[,3:p]}\mathbf{D}^2_{[3:p,3:p]}\mathbf{V}'_{[,3:p]}) = \lambda_3 + \lambda_4 + \cdots \lambda_p$$

And a natural measure for goodness-of-fit is

$$\frac{\mathrm{tr}(\hat{\mathbf{X}}'\hat{\mathbf{X}})}{\mathrm{tr}(\mathbf{X}'\mathbf{X})} = \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^{p} \lambda_j}$$

Similar to the total, explained and residual sum-of-squares in regression.

Weighted singular value decomposition

- On occasions we may wish to use weights for cases (rows, r_i) and/or variables (columns, c_j)
- We normally minimize $\sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 = \text{tr}(\mathbf{E}'\mathbf{E})$
- Define $\mathbf{D}_r$ with weights for the rows $\mathbf{D}_c$ with weights for the columns.
- We now wish to minimize $\sum_{i=1}^{n} \sum_{j=1}^{p} r_i c_j e_{ij}^2 = \text{tr}(\mathbf{D}_c \mathbf{E}' \mathbf{D}_r \mathbf{E})$
- Note that $\sum_{i=1}^{n} \sum_{j=1}^{p} r_i c_j e_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\sqrt{r_i}\sqrt{c_j} e_{ij}\right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \tilde{e}_{ij}^2$
- $\sum_{i=1}^{n} \sum_{j=1}^{p} \tilde{e}_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\sqrt{r_i}\sqrt{c_j} x_{ij} - \sqrt{r_i}\sqrt{c_j} \hat{x}_{ij}\right)^2$
- Solution obtained by transforming the data prior to the svd, and backtransforming afterwards

$$\mathbf{X}_t = \mathbf{D}_r^{\frac{1}{2}} \mathbf{X} \mathbf{D}_c^{\frac{1}{2}} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

  Now compute $\tilde{\mathbf{U}} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U}$ and $\tilde{\mathbf{V}} = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V}$
- Note that $\tilde{\mathbf{U}}\mathbf{D}\tilde{\mathbf{V}}' = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U}\mathbf{D}\mathbf{V}' \mathbf{D}_c^{-\frac{1}{2}} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{D}_r^{\frac{1}{2}} \mathbf{X} \mathbf{D}_c^{\frac{1}{2}} \mathbf{D}_c^{-\frac{1}{2}} = \mathbf{X}$
- $\tilde{\mathbf{U}}_{[,1:k]} \mathbf{D}_{[1:k,1:k]} \tilde{\mathbf{V}}'_{[,1:k]}$ is a rank k approximation to $\mathbf{X}$ in the weighted least squares sense.

Quadratic forms

$$4x_1^2 + 5x_2^2 + 3x_3^2 + 2x_1x_2 + 4x_1x_3 + x_2x_3 = \mathbf{x}'\mathbf{A}\mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 5 & \frac{1}{2} \\ 2 & \frac{1}{2} & 3 \end{bmatrix}$$

In general

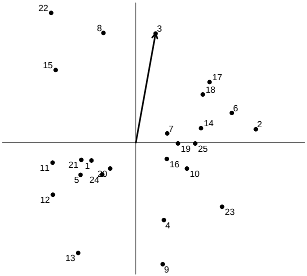
$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} x_i x_j = \mathbf{x}'\mathbf{A}\mathbf{x}$$

- $\mathbf{A}$ positive definite $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ positive semi-definite $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ negative definite $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ negative semi-definite $\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ indefinite

Quadratic forms and eigenvalues

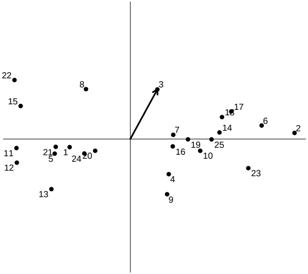
- **A** positive definite $\leftrightarrow$ all $\lambda_i > 0$
- **A** positive semi-definite $\leftrightarrow$ all $\lambda_i \geq 0$
- **A** negative definite $\leftrightarrow$ all $\lambda_i < 0$
- **A** negative semi-definite $\leftrightarrow$ all $\lambda_i \leq 0$
- **A** indefinite

Quadratic forms and distance



- $\| \mathbf{x} \|^2 = \mathbf{x}'\mathbf{x} = x_1^2 + x_2^2 = \mathbf{x}'\mathbf{A}\mathbf{x}$ with $\mathbf{A} = \mathbf{I}$
- $\mathbf{x}'\mathbf{A}\mathbf{x}$ is squared Euclidean distance from the origin.
- $(\mathbf{x} - \mathbf{y})'\mathbf{A}(\mathbf{x} - \mathbf{y})$ is squared Euclidean distance from $\mathbf{x}$ to $\mathbf{y}$.

Quadratic forms and distance



Account for difference in variability: $\tilde{\mathbf{x}} = \mathbf{D}^{-1}\mathbf{x}$ $\quad \mathbf{D} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}$

- $\| \tilde{\mathbf{x}} \|^2 = \tilde{\mathbf{x}}'\tilde{\mathbf{x}} = \left(\frac{x_1}{s_1}\right)^2 + \left(\frac{x_2}{s_2}\right)^2 = \mathbf{x}'\mathbf{A}\mathbf{x}$ with $\mathbf{A} = \mathbf{D}^{-2} = \begin{bmatrix} \frac{1}{s_1^2} & 0 \\ 0 & \frac{1}{s_2^2} \end{bmatrix}$
- $\mathbf{x}'\mathbf{A}\mathbf{x}$ is squared Weighted Euclidean distance from the origin.
- $(\mathbf{x} - \mathbf{y})'\mathbf{A}(\mathbf{x} - \mathbf{y})$ is squared Weighted Euclidean distance from $\mathbf{x}$ to $\mathbf{y}$.

Working with sample data matrices

- Data matrix $\mathbf{X}_{n \times p}$
- Sample mean vector $\mathbf{m}_{p \times 1} = (\frac{1}{n}\mathbf{1}'\mathbf{X})'$
- Centered data matrix
  $\mathbf{X}_c = \mathbf{X} - \mathbf{1}_{n \times 1}\mathbf{m}' = \mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X} = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{X}$
- Centring matrix $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$ $\quad \mathbf{X}_c = \mathbf{H}\mathbf{X}$
- Standardized data matrix
  $\mathbf{X}_s = \mathbf{X}_c\mathbf{D}_s^{-1}$ $\quad \mathbf{D}_s = diag(s_1, s_2, \ldots, s_p)$ $\quad \mathbf{X}_s = \mathbf{H}\mathbf{X}\mathbf{D}_s^{-1}$
- Sample covariance matrix $\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c{}'\mathbf{X}_c$
- Sample correlation matrix $\mathbf{R} = \mathbf{D}_s^{-1}\mathbf{S}\mathbf{D}_s^{-1} = \frac{1}{n-1}\mathbf{X}_s{}'\mathbf{X}_s$

Sample covariance matrix

$$\mathbf{S}_{n-1} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} = \frac{1}{n-1}\mathbf{X}_c{}'\mathbf{X}_c$$

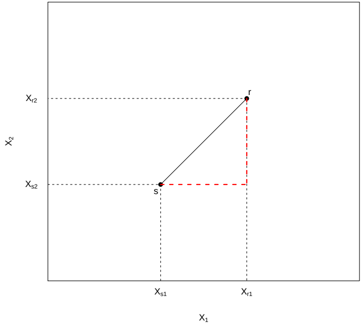
$$= \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})'$$

$$\mathbf{S}_n = \frac{n-1}{n}\mathbf{S}_{n-1}$$

Sample correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} = \frac{1}{n-1}\mathbf{X}_s{}'\mathbf{X}_s$$

Euclidean Distance



$$
\begin{aligned}
\delta_{rs}^2 &= (x_{r1} - x_{s1})^2 + (x_{r2} - x_{s2})^2 \\
&= (\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s)
\end{aligned}
$$

Generalizes to p variables.

Some dissimilarity measures (quantitative data)

- Euclidean distance:

$$\delta_{rs} = \sqrt{(\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s)} = \left\{ \sum_{i=1}^{p} (x_{ri} - x_{si})^2 \right\}^{\frac{1}{2}}$$

- Mahalanobis distance:

$$\delta_{rs} = \left\{ (\mathbf{x}_r - \mathbf{x}_s)'\mathbf{S}^{-1}(\mathbf{x}_r - \mathbf{x}_s) \right\}^{\frac{1}{2}}$$

- Minkowski distance

$$\delta_{rs} = \left\{ \sum_{i=1}^{p} |x_{ri} - x_{si}|^{\lambda} \right\}^{\frac{1}{\lambda}}$$

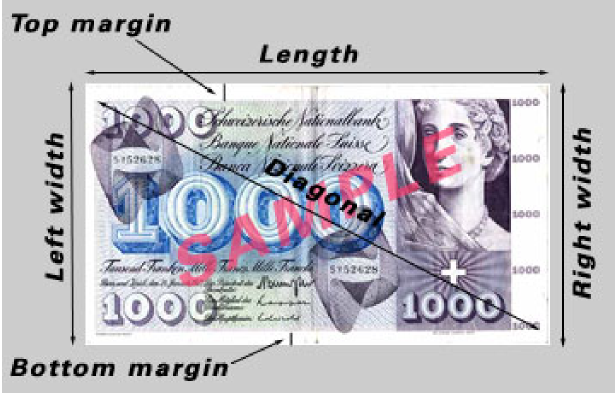
  - $\lambda = 1$ Manhattan distance
  - $\lambda = 2$ Euclidean distance

A duality

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \left[\begin{array}{c|c|c|c} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_p \end{array} \right]$$

- n points in a p-dimensional space
- p points in a n-dimensional space

Example data set: Swiss banknotes (1/8)



Download the Swiss banknote data

Example data set: Swiss banknotes (2/8)

Data matrix (all variables in mm):

```
> X <- read.table("http://www-eio.upc.es/~jan/Data/SwissBanknotes.dat",
                  header=TRUE)
> head(X,10)
   Length  Left Right Bottom  Top Diagonal
1   214.8 131.0 131.1    9.0  9.7    141.0
2   214.6 129.7 129.7    8.1  9.5    141.7
3   214.8 129.7 129.7    8.7  9.6    142.2
4   214.8 129.7 129.6    7.5 10.4    142.0
5   215.0 129.6 129.7   10.4  7.7    141.8
6   215.7 130.8 130.5    9.0 10.1    141.4
7   215.5 129.5 129.7    7.9  9.6    141.6
8   214.5 129.6 129.2    7.2 10.7    141.7
9   214.9 129.4 129.7    8.2 11.0    141.9
10  215.2 130.4 130.3    9.2 10.0    140.7
>
> nrow(X)
[1] 100
> ncol(X)
[1] 6
>
```

Example data set: Swiss banknotes (3/8)

```
> class(X)
[1] "data.frame"
>
> X <- as.matrix(X)
> m <- apply(X,2,mean)
> m
  Length     Left    Right   Bottom      Top Diagonal
 214.969  129.943  129.720    8.305   10.168  141.517
>
> colMeans(X)
  Length     Left    Right   Bottom      Top Diagonal
 214.969  129.943  129.720    8.305   10.168  141.517
>
> v <- apply(X,2,var)
> v
   Length      Left     Right    Bottom       Top  Diagonal
0.1502414 0.1325768 0.1262626 0.4132071 0.4211879 0.1998091
>
> s <- sqrt(v)
> s
   Length      Left     Right    Bottom       Top  Diagonal
0.3876099 0.3641109 0.3553345 0.6428118 0.6489899 0.4470001
>
> s <- apply(X,2,sd)
> s
   Length      Left     Right    Bottom       Top  Diagonal
0.3876099 0.3641109 0.3553345 0.6428118 0.6489899 0.4470001
>
```

Example data set: Swiss banknotes (4/8)

Centered data matrix:

```
> Xc <- scale(X,scale=FALSE)
> head(Xc,10)
      Length   Left Right Bottom    Top Diagonal
 [1,] -0.169  1.057  1.38  0.695 -0.468   -0.517
 [2,] -0.369 -0.243 -0.02 -0.205 -0.668    0.183
 [3,] -0.169 -0.243 -0.02  0.395 -0.568    0.683
 [4,] -0.169 -0.243 -0.12 -0.805  0.232    0.483
 [5,]  0.031 -0.343 -0.02  2.095 -2.468    0.283
 [6,]  0.731  0.857  0.78  0.695 -0.068   -0.117
 [7,]  0.531 -0.443 -0.02 -0.405 -0.568    0.083
 [8,] -0.469 -0.343 -0.52 -1.105  0.532    0.183
 [9,] -0.069 -0.543 -0.02 -0.105  0.832    0.383
[10,]  0.231  0.457  0.58  0.895 -0.168   -0.817
> apply(Xc,2,mean)
       Length          Left         Right        Bottom           Top
 5.400113e-15 -1.250554e-14  5.684084e-16  3.375219e-16  7.371637e-16
     Diagonal
 4.547419e-15
> apply(Xc,2,sd)
   Length      Left     Right    Bottom       Top  Diagonal
0.3876099 0.3641109 0.3553345 0.6428118 0.6489899 0.4470001
```

Example data set: Swiss banknotes (5/8)

Standardized data matrix:

```
> Xs <- scale(X,scale=TRUE)
> head(Xs,10)
          Length       Left       Right     Bottom        Top   Diagonal
 [1,] -0.43600541  2.9029615  3.88366425  1.0811873 -0.7211206 -1.1565993
 [2,] -0.95198813 -0.6673790 -0.05628499 -0.3189114 -1.0292918  0.4093959
 [3,] -0.43600541 -0.6673790 -0.05628499  0.6144877 -0.8752062  1.5279639
 [4,] -0.43600541 -0.6673790 -0.33770993 -1.2523105  0.3574786  1.0805367
 [5,]  0.07997732 -0.9420206 -0.05628499  3.2591185 -3.8028327  0.6331095
 [6,]  1.88591687  2.3536783  2.19511458  1.0811873 -0.1047782 -0.2617449
 [7,]  1.36993414 -1.2166622 -0.05628499 -0.6300444 -0.8752062  0.1856823
 [8,] -1.20997950 -0.9420206 -1.46340972 -1.7190100  0.8197354  0.4093959
 [9,] -0.17801404 -1.4913038 -0.05628499 -0.1633448  1.2819922  0.8568231
[10,]  0.59596005  1.2551120  1.63226468  1.3923203 -0.2588638 -1.8277401
> apply(Xs,2,mean)
      Length          Left         Right        Bottom           Top
 1.394436e-14 -3.432972e-14  1.600814e-15  5.194976e-16  1.129951e-15
     Diagonal
 1.016017e-14
> apply(Xs,2,sd)
  Length     Left    Right   Bottom      Top Diagonal
       1        1        1        1        1        1
```

Example data set: Swiss banknotes (6/8)

Covariance matrix:

```
> S <- cov(X)
> S
              Length       Left       Right        Bottom         Top      Diagonal
Length   0.150241414  0.05801313  0.05729293  0.0571262626  0.01445253  0.0054818182
Left     0.058013131  0.13257677  0.08589899  0.0566515152  0.04906667 -0.0430616162
Right    0.057292929  0.08589899  0.12626263  0.0581818182  0.03064646 -0.0237777778
Bottom   0.057126263  0.05665152  0.05818182  0.4132070707 -0.26347475 -0.0001868687
Top      0.014452525  0.04906667  0.03064646 -0.2634747475  0.42118788 -0.0753090909
Diagonal 0.005481818 -0.04306162 -0.02377778 -0.0001868687 -0.07530909  0.1998090909

> (1/(n-1))*t(Xc)%*%Xc
              Length       Left       Right        Bottom         Top      Diagonal
Length   0.150241414  0.05801313  0.05729293  0.0571262626  0.01445253  0.0054818182
Left     0.058013131  0.13257677  0.08589899  0.0566515152  0.04906667 -0.0430616162
Right    0.057292929  0.08589899  0.12626263  0.0581818182  0.03064646 -0.0237777778
Bottom   0.057126263  0.05665152  0.05818182  0.4132070707 -0.26347475 -0.0001868687
Top      0.014452525  0.04906667  0.03064646 -0.2634747475  0.42118788 -0.0753090909
Diagonal 0.005481818 -0.04306162 -0.02377778 -0.0001868687 -0.07530909  0.1998090909
```

Example data set: Swiss banknotes (7/8)

Correlation matrix:

```
> R <- cor(X)
> R
             Length       Left      Right       Bottom         Top      Diagonal
Length   1.00000000  0.4110529  0.4159765  0.2292752146  0.05745277  0.0316389581
Left     0.41105294  1.0000000  0.6639218  0.2420437898  0.20764186 -0.2645751130
Right    0.41597649  0.6639218  1.0000000  0.2547217369  0.13289390 -0.1497015279
Bottom   0.22927521  0.2420438  0.2547217  1.0000000000 -0.63156375 -0.0006503468
Top      0.05745277  0.2076419  0.1328939 -0.6315637468  1.00000000 -0.2595983041
Diagonal 0.03163896 -0.2645751 -0.1497015 -0.0006503468 -0.25959830  1.0000000000
> (1/(n-1))*t(Xs)%*%Xs
             Length       Left      Right       Bottom         Top      Diagonal
Length   1.00000000  0.4110529  0.4159765  0.2292752146  0.05745277  0.0316389581
Left     0.41105294  1.0000000  0.6639218  0.2420437898  0.20764186 -0.2645751130
Right    0.41597649  0.6639218  1.0000000  0.2547217369  0.13289390 -0.1497015279
Bottom   0.22927521  0.2420438  0.2547217  1.0000000000 -0.63156375 -0.0006503468
Top      0.05745277  0.2076419  0.1328939 -0.6315637468  1.00000000 -0.2595983041
Diagonal 0.03163896 -0.2645751 -0.1497015 -0.0006503468 -0.25959830  1.0000000000
>
```

Example data set: Swiss banknotes (8/8)

Euclidean distance matrix of standardized data:

```
> De <- as.matrix(dist(Xs,method="euclidean"))
> De[1:7,1:7]
         1        2        3        4        5        6        7
1 0.000000 5.748422 5.976534 6.494667 6.929410 3.118474 6.364723
2 5.748422 0.000000 1.553193 1.894791 4.656743 5.051210 2.421493
3 5.976534 1.553193 0.000000 2.298660 4.087481 4.858121 2.629393
4 6.494667 1.894791 2.298660 0.000000 6.187264 5.328595 2.519885
5 6.929410 4.656743 4.087481 6.187264 0.000000 6.197710 5.063220
6 3.118474 5.051210 4.858121 5.328595 6.197710 0.000000 4.669522
7 6.364723 2.421493 2.629393 2.519885 5.063220 4.669522 0.000000
>
```

Multivariate descriptive statistics: graphical summary

- For quantitative variables
  - Scatterplot matrix
  - Biplots
  - Chernoff faces
  - Star plots
  - ...
- For categorical variables
  - stratified bar chart
  - Biplots
  - ...

Numerical summary

	mec	vec	alg	ana	sta
1	77	82	67	67	81
2	63	78	80	70	81
3	75	73	71	66	81
4	55	72	63	70	68
5	63	63	65	70	63
6	53	61	72	64	73
7	51	67	65	65	68
8	59	70	68	62	56
9	62	60	58	62	70
10	64	72	60	62	45
11	52	64	60	63	54
12	55	67	59	62	44
13	50	50	64	55	63
14	65	63	58	56	37
15	31	55	60	57	73
16	60	64	56	54	40
17	44	69	53	53	53
18	42	69	61	55	45
19	62	46	61	57	45
20	31	49	62	63	62
⋮	⋮	⋮	⋮	⋮	⋮
84	15	38	39	28	17
85	5	30	44	36	18
86	12	30	32	35	21
87	5	26	15	20	20
88	0	40	21	9	14

Mean vector:

mec	vec	alg	ana	sta
38.95	50.59	50.60	46.68	42.31

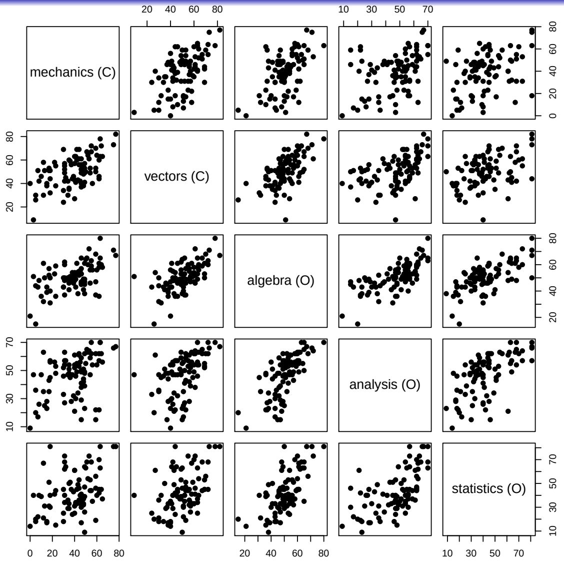
Covariance matrix

	mec	vec	alg	ana	sta
mec	305.77	127.22	101.58	106.27	117.40
vec	127.22	172.84	85.16	94.67	99.01
alg	101.58	85.16	112.89	112.11	121.87
ana	106.27	94.67	112.11	220.38	155.54
sta	117.40	99.01	121.87	155.54	297.76

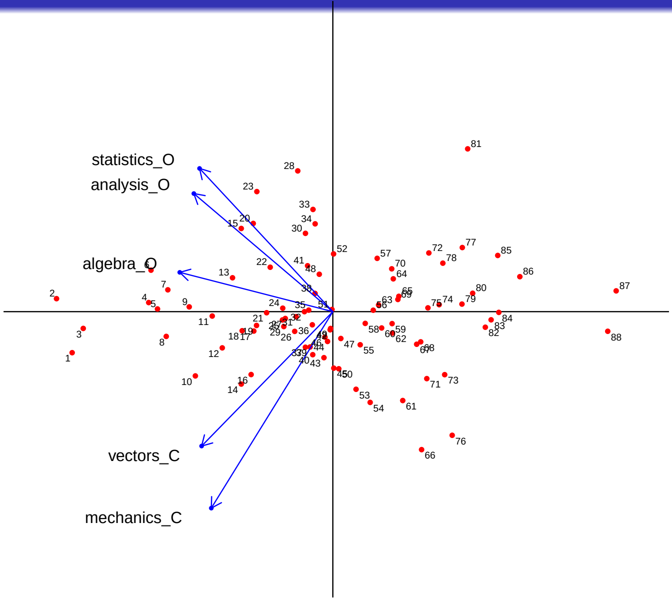
Correlation matrix

	mec	vec	alg	ana	sta
mec	1.00	0.55	0.55	0.41	0.39
vec	0.55	1.00	0.61	0.49	0.44
alg	0.55	0.61	1.00	0.71	0.66
ana	0.41	0.49	0.71	1.00	0.61
sta	0.39	0.44	0.66	0.61	1.00

Scatterplot matrix

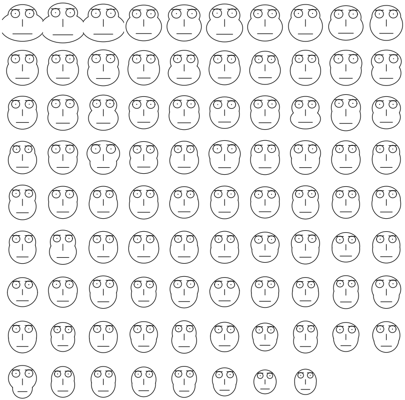


Biplot



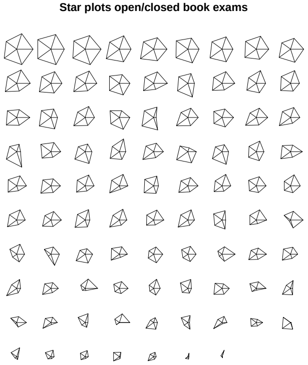
Chernoff faces

	mec	vec	alg	ana	sta
1	77	82	67	67	81
2	63	78	80	70	81
3	75	73	71	66	81
4	55	72	63	70	68
5	63	63	65	70	63
6	53	61	72	64	73
7	51	67	65	65	68
8	59	70	68	62	56
9	62	60	58	62	70
10	64	72	60	62	45
11	52	64	60	63	54
12	55	67	59	62	44
13	50	50	64	55	63
14	65	63	58	56	37
15	31	55	60	57	73
16	60	64	56	54	40
17	44	69	53	53	53
18	42	69	61	55	45
19	62	46	61	57	45
20	31	49	62	63	62
⋮	⋮	⋮	⋮	⋮	⋮
84	15	38	39	28	17
85	5	30	44	36	18
86	12	30	32	35	21
87	5	26	15	20	20
88	0	40	21	9	14



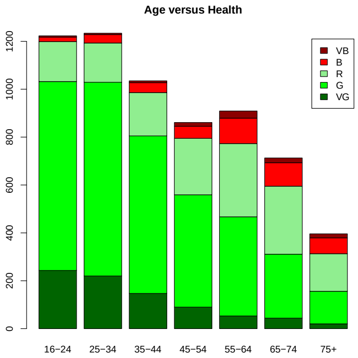
Star plots

	mec	vec	alg	ana	sta
1	77	82	67	67	81
2	63	78	80	70	81
3	75	73	71	66	81
4	55	72	63	70	68
5	63	63	65	70	63
6	53	61	72	64	73
7	51	67	65	65	68
8	59	70	68	62	56
9	62	60	58	62	70
10	64	72	60	62	45
11	52	64	60	63	54
12	55	67	59	62	44
13	50	50	64	55	63
14	65	63	58	56	37
15	31	55	60	57	73
16	60	64	56	54	40
17	44	69	53	53	53
18	42	69	61	55	45
19	62	46	61	57	45
20	31	49	62	63	62
⋮	⋮	⋮	⋮	⋮	⋮
84	15	38	39	28	17
85	5	30	44	36	18
86	12	30	32	35	21
87	5	26	15	20	20
88	0	40	21	9	14



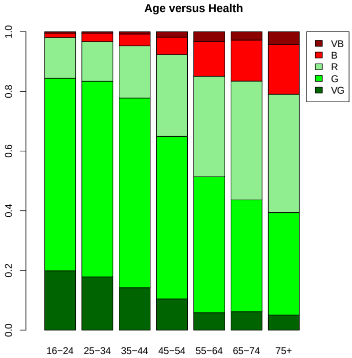
Categorical variables: stratified bar charts

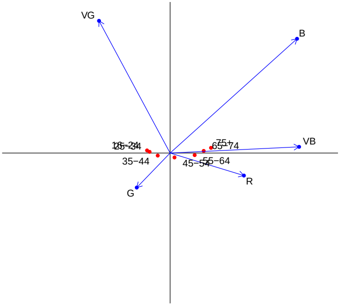
	VG	G	R	B	VB
16-24	243	789	167	18	6
25-34	220	809	164	35	6
35-44	147	658	181	41	8
45-54	90	469	236	50	16
55-64	53	414	306	106	30
65-74	44	267	284	98	20
75+	20	136	157	66	17



Health questionnaire of 6371 individuals

Categorical variables: barchart and biplot





Bibliography

- Manly, B.F.J. (1989) Multivariate statistical methods: a primer. 3rd edition. Chapman and Hall, London.
- Johnson & Wichern, (2002) Applied Multivariate Statistical Analysis, 5th edition, Prentice Hall.
- Peña, D. (2002) Análisis de datos multivariantes. McGraw-Hill, Madrid.