Module 18 Multivariate Analysis for Genetic data
Session 01: Matrix Algebra

Jan Graffelman

jan.graffelman@upc.edu

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

[2]Department of Biostatistics
University of Washington
Seattle, WA, USA

26th Summer Institute in Statistical Genetics (SISG 2021)

Multivariate Analysis (MVA) and Matrix algebra

- Multivariate analysis is, to a considerable extent, applied matrix algebra.
- A good knowledge of matrix algebra facilitates your understanding of MVA.

Outline

- Vectors
- Matrices
- Quadratic forms
- Matrices commonly used in MVA
- Example data set

Vectors

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

Scalar multiplication

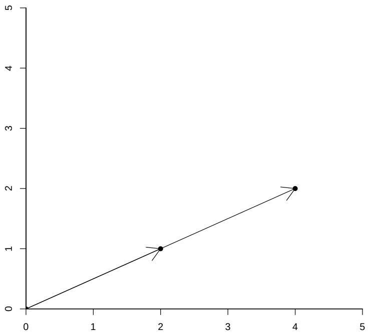
$$\alpha \mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix},$$

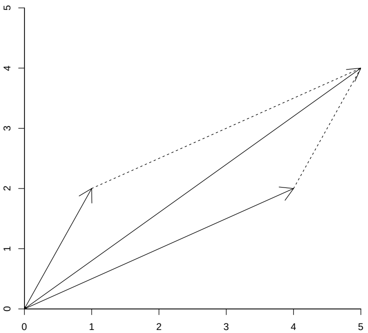
Vectors

Addition

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

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 or inner product:

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

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

Angle:

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

Notes

- The norm is the Euclidean distance from the origin in an n dimensional space
- Let $\mathbf{x}$ and $\mathbf{y}$ be two centred quantitative variables

$$
\begin{aligned}
\cos\theta &= \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \\
&= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \\
&= \frac{s_{xy}}{s_x \cdot s_y} \\
&= r_{xy}.
\end{aligned}
$$

- This property is widely used in biplot interpretation (see later).

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 $(a_1, a_2, \ldots, a_n) = (0, 0, \ldots, 0)$.

Investigating linear (in)dependence

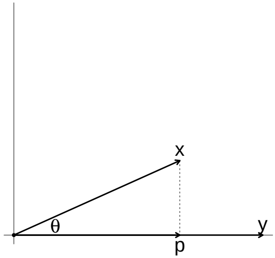
Is the following set of vectors linearly independent or not?

$$\begin{bmatrix} 1 & 5 & -2 \\ 3 & 7 & 2 \\ 2 & 9 & -3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 5 & -2 \\ 0 & -8 & 8 \\ 0 & -1 & 1 \end{bmatrix}$$

Later:

- determine the rank of the matrix
- compute eigenvalues
- do principal component analysis

Projection



$$\cos\theta = \frac{\|\mathbf{p}\|}{\|\mathbf{x}\|}, \quad \|\mathbf{p}\| = \cos\theta\,\|\mathbf{x}\| = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\,\|\mathbf{x}\| = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|}.$$

- Note that for a given vector **y**, the length of the projection is proportional to the scalar product between the vectors.
- This property is also widely used in biplot interpretation.

Matrix

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

About n and p:

- Classical $n >> p$
- Limiting $n = p$
- Today often $n << p$

A duality

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

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

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

- $\mathbf{A}_{p \times k} \mathbf{B}_{k \times l} = \mathbf{C}_{p \times l}$
- $\mathbf{A}_{p \times k} \mathbf{x}_{k \times 1} = \mathbf{y}_{p \times 1}$
- $\mathbf{x}'_{1 \times p} \mathbf{A}_{p \times k} = \mathbf{y}'_{1 \times k}$
- $\mathbf{X}_{n \times p} \mathbf{D}_{p \times p} = \left[\begin{array}{c|c|c|c} \mathbf{x}_1 d_1 & \mathbf{x}_2 d_2 & \cdots & \mathbf{x}_p d_p \end{array} \right]$
- $\mathbf{D}_{n \times n} \mathbf{X}_{n \times p} = \left[\begin{array}{c} d_1 \mathbf{x}_1 \\ \hline d_2 \mathbf{x}_2 \\ \hline \vdots \\ \hline d_n \mathbf{x}_n \end{array} \right]$
- Pre or post-multiplication of $\mathbf{X}$ by a diagonal matrix amounts to weighting observations or variables.

Some special matrices

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

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

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.

$|\mathbf{A}| \neq 0$ implies linear independence, and $\mathbf{A}$ is regular.

Example

Calculate the determinant of

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 6 \\ 4 & 1 & 0 \\ 5 & 4 & 1 \end{bmatrix}$$

$$|\,\mathbf{A}\,| = 4 \cdot \begin{vmatrix} 2 & 6 \\ 4 & 1 \end{vmatrix} \cdot (-1)^3 + 1 \cdot \begin{vmatrix} 3 & 6 \\ 5 & 1 \end{vmatrix} \cdot (-1)^4 = 88 - 27 = 61$$

in R:

```
> X <- matrix(c(3,2,6,4,1,0,5,4,1),ncol=3,byrow=TRUE)
> det(X)
[1] 61
```

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.
- Example:

$$\begin{bmatrix} 1 & 3 & 5 \\ 0 & 2 & 2 \\ 1 & 1 & 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 1 \\ 0 & 2 & 3 \end{bmatrix}$$

have rank 2 and 3 respectively.

Trace

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

- $\mathrm{tr}(\alpha \mathbf{A}) = \alpha \mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}) \qquad \mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB}) = \mathrm{tr}(\mathbf{BCA})$
- $\mathrm{tr}(\mathbf{AA}') = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}^2$
- Note the latter trace is a sum-of-squares which is a measure of variability.

Quadratic forms

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

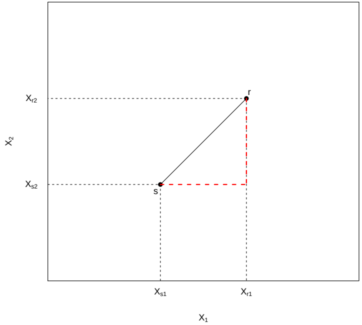
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 5 & \frac{1}{2} \\ 2 & \frac{1}{2} & 3 \end{bmatrix}$$

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

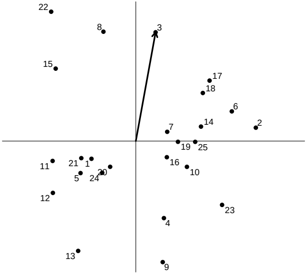
Euclidean Distance



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

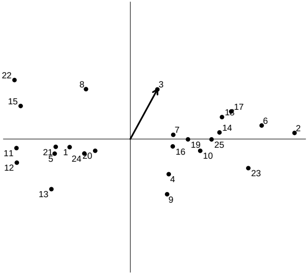
Generalizes to p variables.

Quadratic forms and distance



- Let $\mathbf{x}$ be a single observation written as column vector.
- $\| \mathbf{x} \|^2 = \mathbf{x}'\mathbf{x} = x_1^2 + x_2^2 = \mathbf{x}'\mathbf{A}\mathbf{x}$ 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 $\mathbf{y}$.

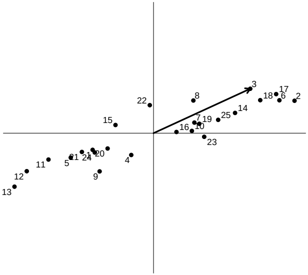
Quadratic forms and distance



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 $\mathbf{y}$.

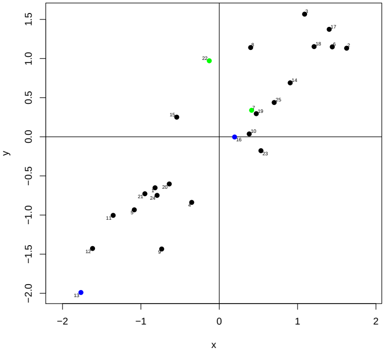
Quadratic forms and distance



Account for difference in variability and for correlation: $\tilde{\mathbf{x}} = \mathbf{S}^{-\frac{1}{2}}\mathbf{x}$ Mahalanobis' transformation

- $\| \tilde{\mathbf{x}} \|^2 = \tilde{\mathbf{x}}'\tilde{\mathbf{x}} = \mathbf{x}'\mathbf{S}^{-1}\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x}$ with $\mathbf{A} = \mathbf{S}^{-1} = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix}^{-1}$
- $\mathbf{x}'\mathbf{A}\mathbf{x}$ is the squared Mahalanobis distance from the origin.
- $(\mathbf{x} - \mathbf{y})'\mathbf{A}(\mathbf{x} - \mathbf{y})$ is squared Mahalanobis distance from $\mathbf{x}$ to $\mathbf{y}$.

Mahalanobis distance and Euclidean distance



	D_E	D_M
1	1.05	0.84
2	1.98	1.79
3	1.91	1.73
4	0.91	1.25
5	1.43	1.09
6	1.84	1.48
7	0.53	0.42
8	1.21	1.83
9	1.61	1.92
10	0.38	0.80
11	1.69	1.44
12	2.16	1.62
13	2.66	1.99
14	1.14	0.95
15	0.60	1.78
16	0.20	0.46
17	1.96	1.43
18	1.67	1.22
19	0.56	0.56
20	0.88	0.64
21	1.20	0.99
22	0.98	2.49
23	0.56	1.60
24	1.09	0.80
25	0.83	0.83

Multivariate statistics: population and sample

Population

(conceptual)

Random vector $\mathbf{X}_{p\times 1}$

Population mean vector

$$\mu_{p\times 1} = E(\mathbf{X})$$

Population covariance matrix

$$\boldsymbol{\Sigma}_{p\times p} = E\left((\mathbf{X}-\mu)(\mathbf{X}-\mu)'\right)$$

Sample

(observed)

Data matrix $\mathbf{X}_{n\times p}$

Sample mean vector

$$\mathbf{m}_{p\times 1} = \frac{1}{n}\left(\mathbf{1}'\mathbf{X}\right)'$$

Sample covariance matrix

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

Frequently used matrices in MVA

- 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

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

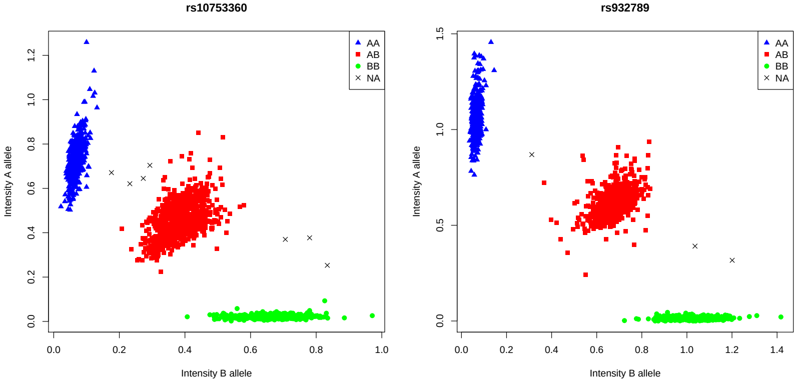
$$= \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})'$$

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

Sample correlation matrix

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

Example data set: genotype intensity data



Data taken from the GENEVA project at UW Biostatistics.

Example data set: genotype intensity data

Raw data matrix:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB	SNP1	SNP2
1	0.080	0.812	0.070	1.019	0	0
2	0.746	0.022	0.996	0.010	2	2
3	0.402	0.440	0.725	0.569	1	1
4	0.071	0.782	0.057	0.950	0	0
5	0.618	0.028	1.072	0.016	2	2
6	0.362	0.473	0.739	0.564	1	1
7	0.372	0.469	0.752	0.645	1	1
8	0.433	0.455	0.728	0.615	1	1
9	0.364	0.521	0.653	0.646	1	1
10	0.624	0.017	1.056	0.018	2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Mean vector:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
m	0.372	0.426	0.597	0.584

Standard deviations:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
s	0.220	0.263	0.352	0.377

Example data set: genotype intensity data

Centered data matrix:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
1	-0.292	0.386	-0.527	0.435
2	0.374	-0.404	0.399	-0.574
3	0.030	0.014	0.128	-0.015
4	-0.301	0.356	-0.540	0.366
5	0.246	-0.398	0.475	-0.568
6	-0.010	0.047	0.142	-0.020
7	-0.000	0.043	0.155	0.061
8	0.061	0.029	0.131	0.031
9	-0.008	0.095	0.056	0.062
10	0.252	-0.409	0.459	-0.566
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Mean vector:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
m	0.000	0.000	0.000	0.000

Standard deviations:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
s	0.220	0.263	0.352	0.377

Example data set: genotype intensity data

Standardized data matrix:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
1	-1.328	1.467	-1.494	1.154
2	1.700	-1.532	1.134	-1.524
3	0.136	0.055	0.365	-0.040
4	-1.369	1.353	-1.531	0.971
5	1.118	-1.509	1.349	-1.508
6	-0.046	0.180	0.404	-0.054
7	-0.001	0.165	0.441	0.161
8	0.277	0.112	0.373	0.082
9	-0.037	0.362	0.160	0.164
10	1.145	-1.551	1.304	-1.503
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Mean vector:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
m	0.000	0.000	0.000	0.000

Standard deviations:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
s	1.000	1.000	1.000	1.000

Example data set: genotype intensity data

Covariance matrix:

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
SNP1_IA	0.048	-0.052	0.075	-0.078
SNP1_IB	-0.052	0.069	-0.082	0.095
SNP2_IA	0.075	-0.082	0.124	-0.122
SNP2_IB	-0.078	0.095	-0.122	0.142

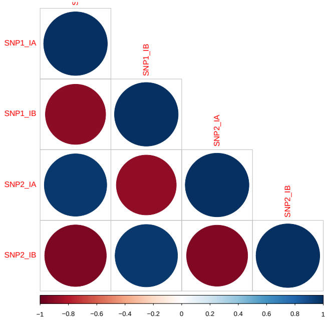
Correlation matrix:

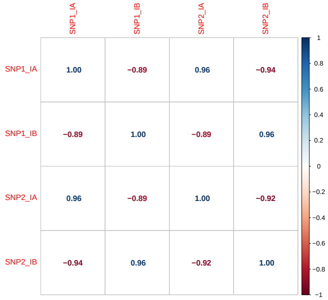
	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
SNP1_IA	1.000	-0.895	0.962	-0.937
SNP1_IB	-0.895	1.000	-0.886	0.962
SNP2_IA	0.962	-0.886	1.000	-0.920
SNP2_IB	-0.937	0.962	-0.920	1.000

Cross table of called genotypes:

	0	1	2	NA
0	512	0	0	0
1	1	949	0	0
2	0	0	470	0
NA	0	4	0	3

Displaying correlations





Bibliography

- Johnson & Wichern, (2002) *Applied Multivariate Statistical Analysis*, 5th edition, Prentice Hall, Chapters 2 & 3.
- Mardia, K.V. et al. (1979) *Multivariate Analysis*. Appendix A. Academic press.