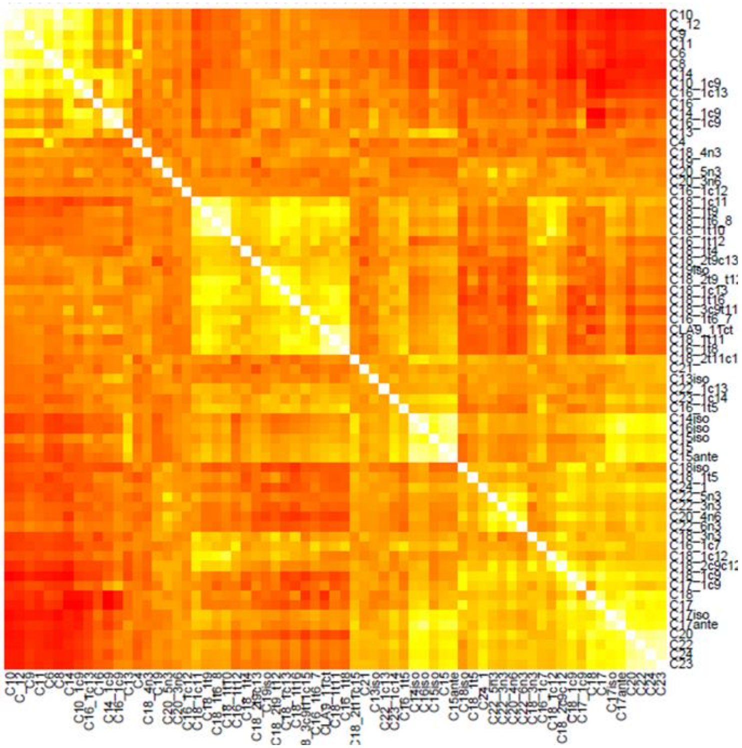
Eigenvalues and eigenvectors



Multivariate system

- ✓ **More variables measured in the same experimental unit**
- ✓ **More variables are considered as dependent simultaneously**

Why multivariate?

- ✓ Scientific discipline (social science, biology, psycometrics)
- ✓ (Nano)technology development
- ✓ Livestock precision farming
- ✓ Big data
- ✓ Omics

Multivariate system

Univariate system

$$y = f(x_1, x_2, \ldots\ldots x_n)$$

Multivariate system

$$y_1, y_2, \ldots. y_n = f(x_1, x_2, \ldots\ldots x_n)$$

Univariate system

Sheep	Milk yield
190	1014
217	1000
528	1410
691	1380
456	1280
545	778
791	835
545	710
528	1274
351	1332
791	824

Sheep	milk	weight	Breed
190	1014	49,2	1
217	1000	55,2	1
528	1410	53,2	2
691	1380	54,4	1
456	1280	51,5	2
545	778	48,3	2
791	835	50,1	2
545	710	52,3	1
528	1274	48,3	1
351	1332	50,2	2
791	824	49,7	1

- ✓ Description
- ✓ Parameter estimation
- ✓ Hypothesis testing

Relevant quantities are scalars

Description

- ✓ Mean, median, standard deviation, percentiles

Parameter estimation

- ✓ Confidence interval of the mean, variance
- ✓ Covariance, correlation

Hypothesis testing

- ✓ ANOVA, F test

In Univariate systems only one dimension is considered

Sheep	Milk yield
190	1014
217	1000
528	1410
691	1380
456	1280
545	778
791	835
545	710
528	1274
351	1332
791	824

Sheep	milk	weight	Breed
190	1014	49,2	1
217	1000	55,2	1
528	1410	53,2	2
691	1380	54,4	1
456	1280	51,5	2
545	778	48,3	2
791	835	50,1	2
545	710	52,3	1
528	1274	48,3	1
351	1332	50,2	2
791	824	49,7	1

In Univariate systems only one dimension is considered

Sheep	milk	Breed
190	1014	1
217	1000	1
528	1410	2
691	1380	1
456	1280	2
545	778	2
791	835	2
545	710	1
528	1274	1
351	1332	2
791	824	1

ANOVA

Source	*SS*	*df*	*MS*	*F*	*Significance*
Between	568980,8	1	568980,75	36,73778621	0,000121791
Within	154876,2	10	15487,61667		
Total	723856,9	11			

From uni to multivariate systems

Sheep	milk
190	1014
217	1000
528	1410
691	1380
456	1280
545	778
791	835
545	710
528	1274
351	1332
791	824

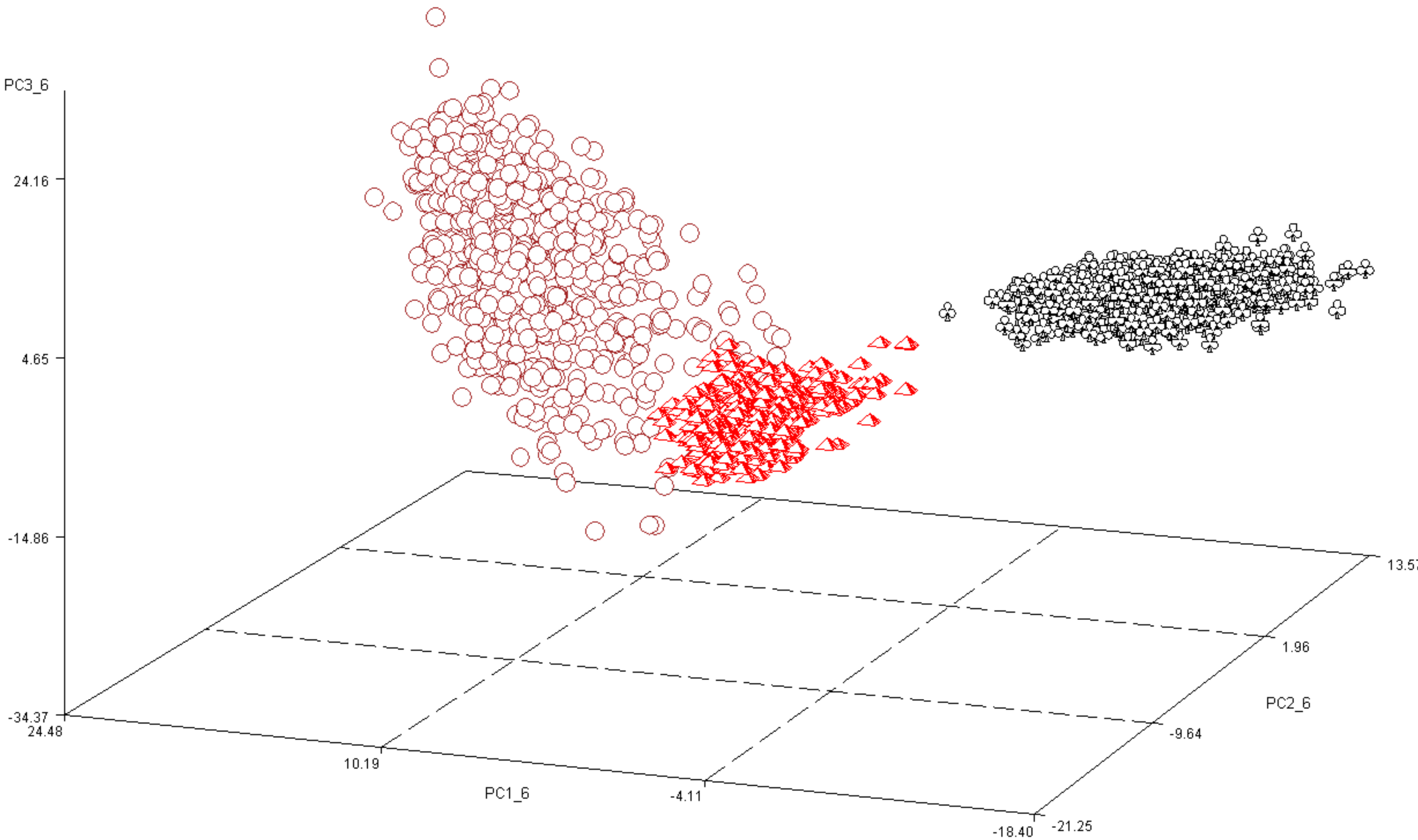
ID	C4	C6	C8	C10	C10:1	C11	C12	C13	iso C14:0	C14	C14-1 cis 9	C15	C15-1	C16
1	1,57	1,71	2,03	7,10	0,23	0,07	4,37	0,11	0,11	11,43	0,20	1,15	0,34	24,74
2	1,50	1,66	1,99	6,97	0,26	0,08	4,33	0,10	0,10	11,35	0,20	1,19	0,32	23,33
3	1,52	1,45	1,57	5,41	0,18	0,05	3,50	0,08	0,13	10,60	0,21	1,30	0,34	24,01
4	1,45	1,33	1,40	4,63	0,16	0,02	2,97	0,05	0,11	10,23	0,19	1,08	0,27	25,01
5	1,42	1,14	1,14	3,76	0,12	0,12	2,69	0,05	0,11	9,48	0,17	0,93	0,32	24,39
6	1,33	1,14	1,17	4,19	0,15	0,04	2,95	0,06	0,12	10,30	0,24	1,00	0,32	27,09
7	1,34	1,28	1,37	5,16	0,23	0,06	3,52	0,09	0,10	11,34	0,29	0,96	0,31	30,98

In multivariate systems more dimensions are considered

- ✓ Variability within each variable
- ✓ Co-variability between different variables

ID	C4	C6	C8	C10	C10:1	C11	C12	C13	iso C14:0	C14	C14-1 cis 9	C15	C15-1	C16
1	1,57	1,71	2,03	7,10	0,23	0,07	4,37	0,11	0,11	11,43	0,20	1,15	0,34	24,74
2	1,50	1,66	1,99	6,97	0,26	0,08	4,33	0,10	0,10	11,35	0,20	1,19	0,32	23,33
3	1,52	1,45	1,57	5,41	0,18	0,05	3,50	0,08	0,13	10,60	0,21	1,30	0,34	24,01
4	1,45	1,33	1,40	4,63	0,16	0,02	2,97	0,05	0,11	10,23	0,19	1,08	0,27	25,01
5	1,42	1,14	1,14	3,76	0,12	0,12	2,69	0,05	0,11	9,48	0,17	0,93	0,32	24,39
6	1,33	1,14	1,17	4,19	0,15	0,04	2,95	0,06	0,12	10,30	0,24	1,00	0,32	27,09
7	1,34	1,28	1,37	5,16	0,23	0,06	3,52	0,09	0,10	11,34	0,29	0,96	0,31	30,98

The multivariate space has p dimensions



Relevant quantities in multivariate systems are vectors and matrices

- ✓ **Vector of means**
- ✓ **Vector of variances**
- ✓ **Variance-covariance matrix**
- ✓ **Correlation matrix**

Main aims of multivariate statistics

- ✓ Reduction of the dimension of the system ($q<p$)
- ✓ Search for sub-structures (Latent variables)
- ✓ Search for relationships of dependence (co-variation)

Multivariate exploratory analysis

Parameter estimation

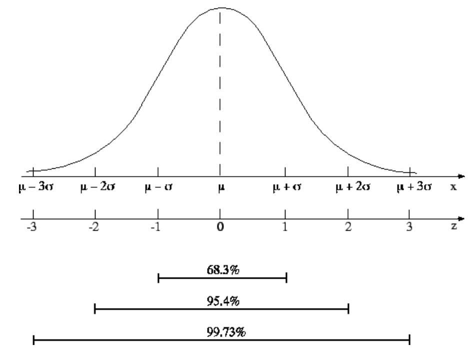
- ✓ Elements of variance-covariance matrix
- ✓ Elements of correlation matrix

Hypothesis testing

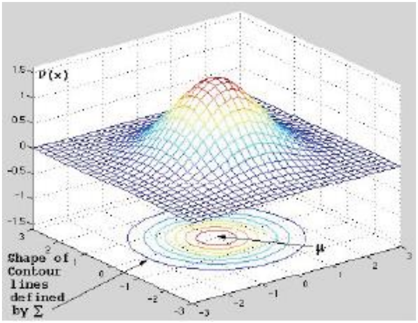
- ✓ MANOVA

The (co)variance matrix

- In a univariate distribution the variance defines the width of the density probability function



- In a multivariate distribution the (co)variance matrix defines the width of the density probability function and the degree of relationships among the different variables



- Given a data matrix $\mathbf{X}_{n,m}$ **with** m variables measured on n experimental units, it is possible to calculate a matrice di variance-covariance matrix of order equal t the number of variables $\boldsymbol{\Sigma}_{m,m}$

The (co)variance matrix S

$$\mathbf{\Sigma}_{m,m} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1j} & \cdots & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22}^2 & & & & & \\ \vdots & \vdots & \ddots & & & & \\ \sigma_{j1} & \sigma_{j2} & & \sigma_{jj}^2 & & & \sigma_{jm} \\ \vdots & \vdots & & & \ddots & & \\ \vdots & \vdots & & & & \ddots & \\ \sigma_{m1} & \sigma_{m2} & & \sigma_{mj} & & & \sigma_{mm} \end{bmatrix}$$

$$s_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

$$s_j^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_{kj} - \bar{x}_j)^2$$

The (co)variance matrix S

NAME	milk_tot_	fat	protein	lactose	CCS
milk_tot_	92131,24	-81,98	-75,70	123,86	-258188,69
fat	-81,98	1,46	0,14	-0,10	661,70
protein	-75,70	0,14	0,29	-0,14	321,95
lactose	123,86	-0,10	-0,14	0,43	-1051,54
CCS	-258188,69	661,70	321,95	-1051,54	10024385,01

Main features of the (co)variance matrix

- ✓ **Symmetric** $(\sigma_{ij} = \sigma_{ji})$

- ✓ **Semi-positive definite, i.e. has non-negative eigenvalues**

- ✓ **Same rank of the data matrix (equal to the number of positive eigenvalues)**

- ✓ **If the data matrix is linearly transformed, the (co)variance matrix has the same rank of the non-transformed matrix**

The correlation matrix R

$$\mathrm{P}_{m,m} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1j} & \cdots & \cdots & \rho_{1m} \\ \rho_{21} & 1 & & & & & \\ \vdots & \vdots & \ddots & & & & \\ \rho_{j1} & \rho_{j2} & & 1 & & & \rho_{jm} \\ \vdots & \vdots & & & \ddots & & \\ \vdots & \vdots & & & & \ddots & \\ \rho_{m1} & \rho_{m2} & & \rho_{mj} & & & 1 \end{bmatrix}$$

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}^2 \cdot \sigma_{jj}^2}}$$

$$\sigma_{ii}^2 = 1$$

The correlation matrix R

NAME	milk_tot_	fat	protein	lactose	CCS	r	k20	a30	pH
milk_tot_	1,00	-0,23	-0,47	0,56	0,00	-0,10	0,01	0,06	-0,28
fat	-0,23	1,00	0,13	-0,12	0,12	-0,13	-0,02	0,08	-0,03
protein	-0,47	0,13	1,00	-0,48	0,00	0,27	-0,09	-0,09	0,32
lactose	0,56	-0,12	-0,48	1,00	-0,12	-0,19	0,01	0,12	-0,57
CCS	0,00	0,12	0,00	-0,12	1,00	0,39	0,38	-0,42	0,27
r	-0,10	-0,13	0,27	-0,19	0,39	1,00	0,66	-0,89	0,36
k20	0,01	-0,02	-0,09	0,01	0,38	0,66	1,00	-0,77	0,24
a30	0,06	0,08	-0,09	0,12	-0,42	-0,89	-0,77	1,00	-0,34
pH	-0,28	-0,03	0,32	-0,57	0,27	0,36	0,24	-0,34	1,00

Main features of the correlation matrix

- ✓ **Symmetric** $(\sigma_{ij} = \sigma_{ji})$
- ✓ **Semi-positive definite, i.e. non-negative eigenvalues**
- ✓ **Same rank of the data matrix (equal to the number of positive eigenvalues)**
- ✓ **Invariant if the data matrix is linearly transformed**
- ✓ **Unit variance on the diagonal** (standardized variabile μ=0; σ^2 =1)
- ✓ **Total variance = tr(R)=number of variables**

(co)variance or correlation matrix?

(Co)variance matrix (S)

- ✓ Real proportion among variables
- ✓ Sensitive to measurement unit
- ✓ The importance of variables is enhanced by their magnitude

Correlation matrix (R)

- ✓ Measurement units are no longer important
- ✓ Variables are standardized
- ✓ Comparable results among different data sets

- ✓ **S** is preferred when variables have the same mesurement unit and similar magnitude
- ✓ In all other cases **R** is preferred

Some R code

```
#-----------------Calculation of variance-covariance matrix S---------------------------------
X=matrix(c(168,184,173,176,176,72,75,58,58,68,30,29,26,26,28),ncol=3)

Xm=apply(X,2,mean)

Xm=diag(Xm)

U=matrix(1,nrow=nrow(X),ncol=ncol(X))

D=X-U%*%Xm

n=nrow(X)

#S= 1/(n-1)XD'D variance-covariance matrix calculation

S=1/(n-1)*t(D)%*%D
```

Vector transformation

Given a column vector $\mathbf{x}_{n,1}$ and a matrix $\mathbf{A}_{m,n}$
The product

$$\mathbf{y=Ax}$$

defines a transformation, creating a vector $\mathbf{y}$ with the same number of rows of the matrix $\mathbf{A}$

i.e. the sistem coordinates are transformed from the n-dimensions of the space of $\mathbf{x}$ to the m-dimensions of the space of $\mathbf{A}$

Eigenvectors and eigenvalues

A= squared (n x n) matrix
λ = scalar (real number)

λ is an **eigenvalue** of **A** if exist a column vector **x**(n,1)

$$\mathbf{Ax} = \lambda\mathbf{x}$$

λ is an **eigenvalue** of **A** corresponding to the **eigenvector** **x**

Eigenvectors and eigenvalues

$$\mathbf{Ax} = \lambda\mathbf{x}$$

is equivalent

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

Solutions $\mathbf{x}\neq 0$ exist only if the matrix

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

equation charateristic of **A**

The set of eigenvalues of **A** is called the ***spectrum*** of **A**

Properties of eigenvalues

- ✓ The product of eigenvalues of a matrix **A** is equal to |**A**|
- ✓ The sum of eigenvalues of **A** is equal to the trace of **A**
- ✓ The eigenvalues of a symmetric matrix with real elements are all real
- ✓ The eigenvalues of a positive-definite matrix are all positive

Properties of eigenvectors

- If λ_1 and λ_2 are the eigenvalues of a symmetric matrix **A** the corresponding eigenvectors $\mathbf{x_1}$ and $\mathbf{x_2}$ are orthogonal

$$\mathbf{Ax_1} = \lambda_1\mathbf{x_1} \qquad \mathbf{Ax_2} = \lambda_2\mathbf{x_2}$$

Decomposition of a matrix A into the product of two or three matrices

Useful for computational reason

- ✓ **Determinant**
- ✓ **Rank**
- ✓ **Ordinary or generalised inverse in a system of linear equation**

Spectral decomposition

For each symetrix matrix **A** there exists an orthogonal matrix **P** so that

$$\mathbf{D=P'AP}$$

D = diagonal matrix of the eigenvalues of **A**
P = matrix of the eigenvectors of **A**

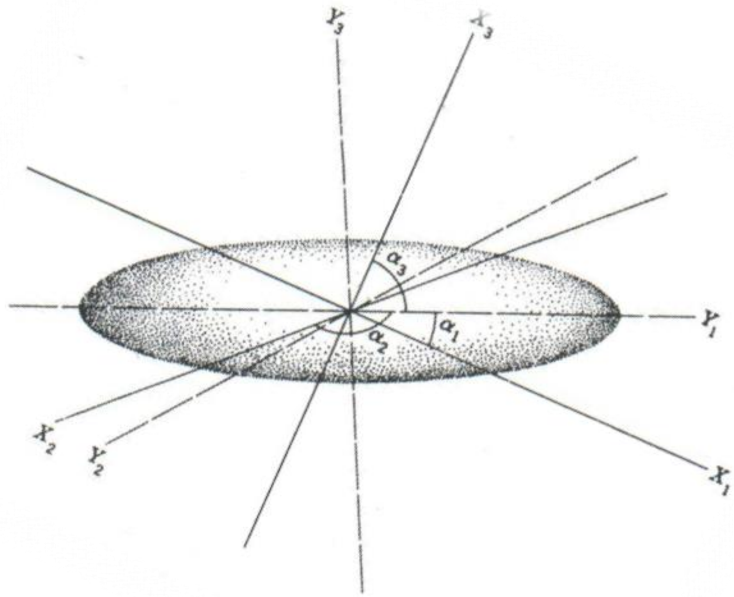
Diagonalisation of the (co)variance matrix

- ✓ The (co)variance matrix **S** can be written in diagonal form through an appropriate change of the reference system

- ✓ The new system coresponds ot the eigenvectors of the (co)variance matrix

Diagonalisation of a squared matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$



$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Similar matrices

- ✓ Similar matrices have the same eigenvalues
- ✓ Eigenvectors of a simmetric matrix with real elements are orthogonal
- ✓ Similar matrices have the same determinant
- ✓ If x and y are normalized then they are defined as orthonormal
- ✓ Similar matrices have the same trace

Some R code

```
#-------------------------- Calculation of correlation matrix R-------------------------

devX=apply(X,2,sd)

Ds=diag(devX)

L=solve(Ds)

Z=D%*%L

# R= Z'Z/(n-1)
R<-1/(n-1)*t(Z)%*%Z


# ---------- R functions for calculating S o R
var(X)
cor(X)
```

Some R code for matrix algebra

```
# creation of a squared matrix A (3 x 3)

A= matrix(c(3,5,4,7,11,8,4,2,6),ncol=3)
A= matrix(c(3,7,4,5,11,2,4,8,6),ncol=3,byrow=T)

#Transposition: function t()
A1=t(A)

# creation of an indentity matrix (4 x 4)
I=diag(4)

# creation of a diagonal matrix (3 x 3) having 2 on the diagonal
p=diag(2,3)

# creation of a diagonal matrix (3 x 3) having the sequence 1-3 on the diagonal
P=diag(1:3)
```

Some R code for matrix algebra

```
#creation of two column vectors (3x1) and their joining for creating a matrix (3x2)
b=matrix(c(4,6,1),ncol=1)
c=matrix(c(1,2,3),ncol=1)

#cbind (join of two or more columns into a matrix)
D=cbind(c,b)

#creation of two row vectors (1x4) and their joining for creating a matrix (2x4)
due
a=-matrix(c(1,2,3,4), ncol=4)

b=matrix(c(4,3,2,1), ncol=4)

#rbind (join of two or more rows into a matrix)
C=rbind(a,b)
```

Some R code for matrix algebra

```
#Sum
A=matrix(c(2,-4,0,1,3,2,2,-1,1),ncol=3)
B=matrix(c(-1,2,3,2,-4,4,1,-1,-2),ncol=3)
C=A+B
C

#Product of a scalar for a matrix
A=matrix(c(15,11,31,23,9,12),ncol=3)
k=3
X=k*A

#Matrix product
A=matrix(c(1,-1,2,0,2,1),ncol=3)
B=matrix(c(6,-1,0,5,1,2,4,-1,0),ncol=3)
C=A%*%B
```

Some R code for matrix algebra

```
#scalar product x'y
x=matrix(c(5,2,1),ncol=1)
y=matrix(c(1,3,6),ncol=1)
x1y=t(x)%*%y

#y'y
y1y=t(y)%*%y

#or
y1y=crossprod(y)

#outer product xy'
xy1=x%*%t(y)

#outer product yy'
Yy1=y%*%t(y)
```

Some R code for matrix algebra

```
# Determinant calculation
A=matrix(c(5,9,5,2,5,4,3,2,1),ncol=3)
DetA=det(A)

#Inverse calculation using solve()
A=matrix(c(5,9,3,2,5,4,3,2,2),ncol=3)
invA=solve(A)
```