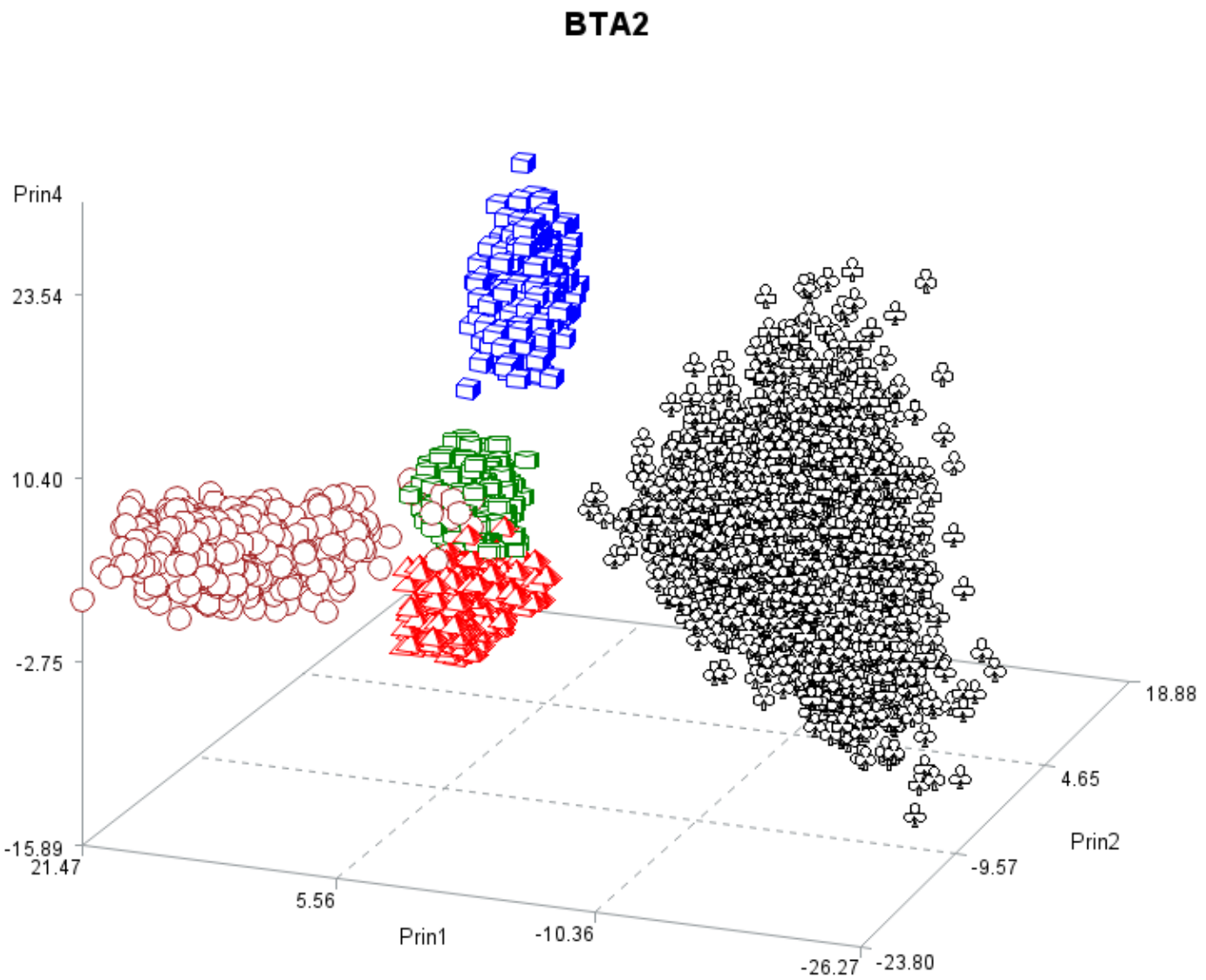
Factor analysis



Main aims of multivariate statistics

- ✓ **Reduction of system dimensions**
- ✓ **Search for latent structures**
- ✓ **Analysis of dependence relatioships**

Search for latent structures

- ✓ **Definition of covariance structure**
- ✓ **New variables with technical or biological meaning**
- ✓ **Latent structure that regulates the covariance of the system**

Search for latent structures

PCA is aimed at reducing the dimensions of the system

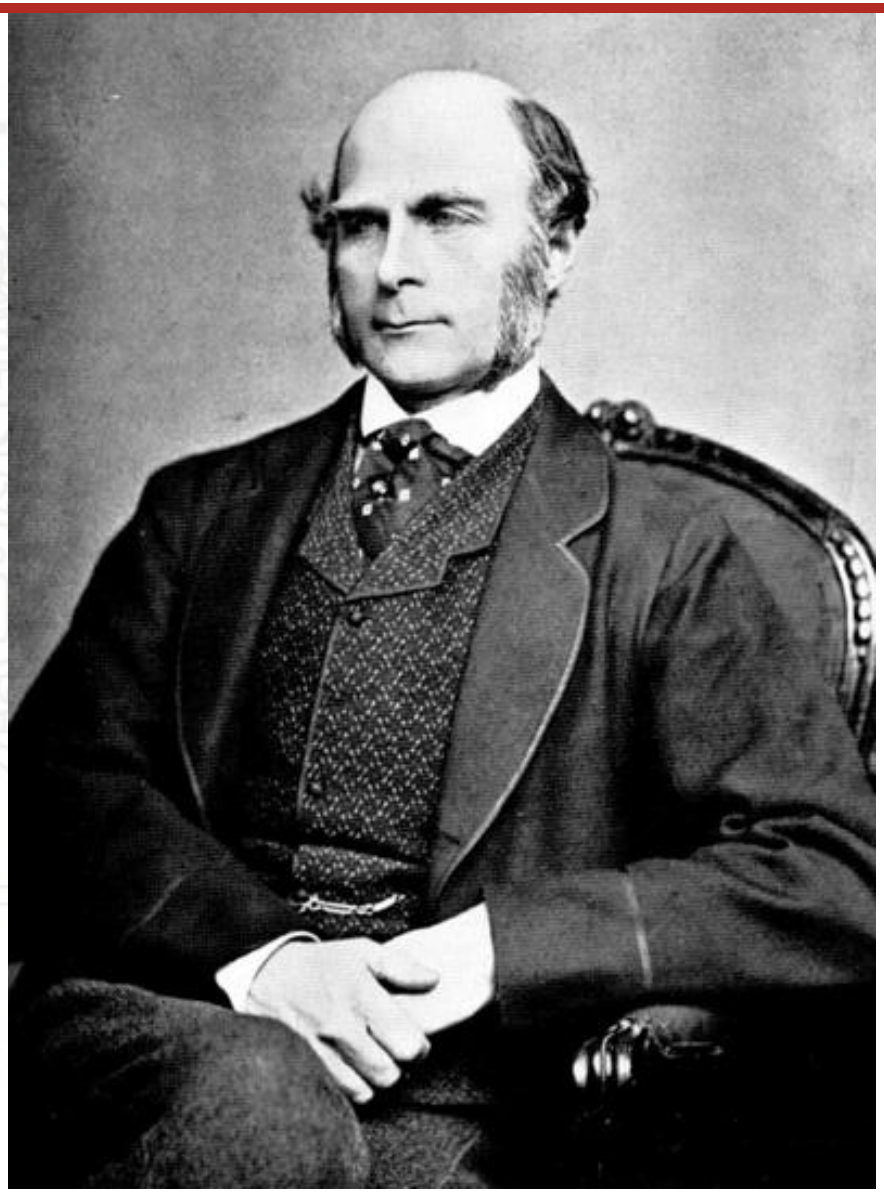
	milk1	milk2	milk3	milk4	milk5	milk6	milk7	milk8
milk1	1.00000	0.72719	0.62818	0.58538	0.52913	0.45254	0.33354	0.20025
milk2	0.72719	1.00000	0.84702	0.74284	0.72049	0.62749	0.45795	0.31355
milk3	0.62818	0.84702	1.00000	0.84072	0.79685	0.71108	0.57151	0.44148
milk4	0.58538	0.74284	0.84072	1.00000	0.85651	0.75159	0.62385	0.51051
milk5	0.52913	0.72049	0.79685	0.85651	1.00000	0.84531	0.71791	0.62356
milk6	0.45254	0.62749	0.71108	0.75159	0.84531	1.00000	0.82266	0.73125
milk7	0.33354	0.45795	0.57151	0.62385	0.71791	0.82266	1.00000	0.82216
milk8	0.20025	0.31355	0.44148	0.51051	0.62356	0.73125	0.82216	1.00000

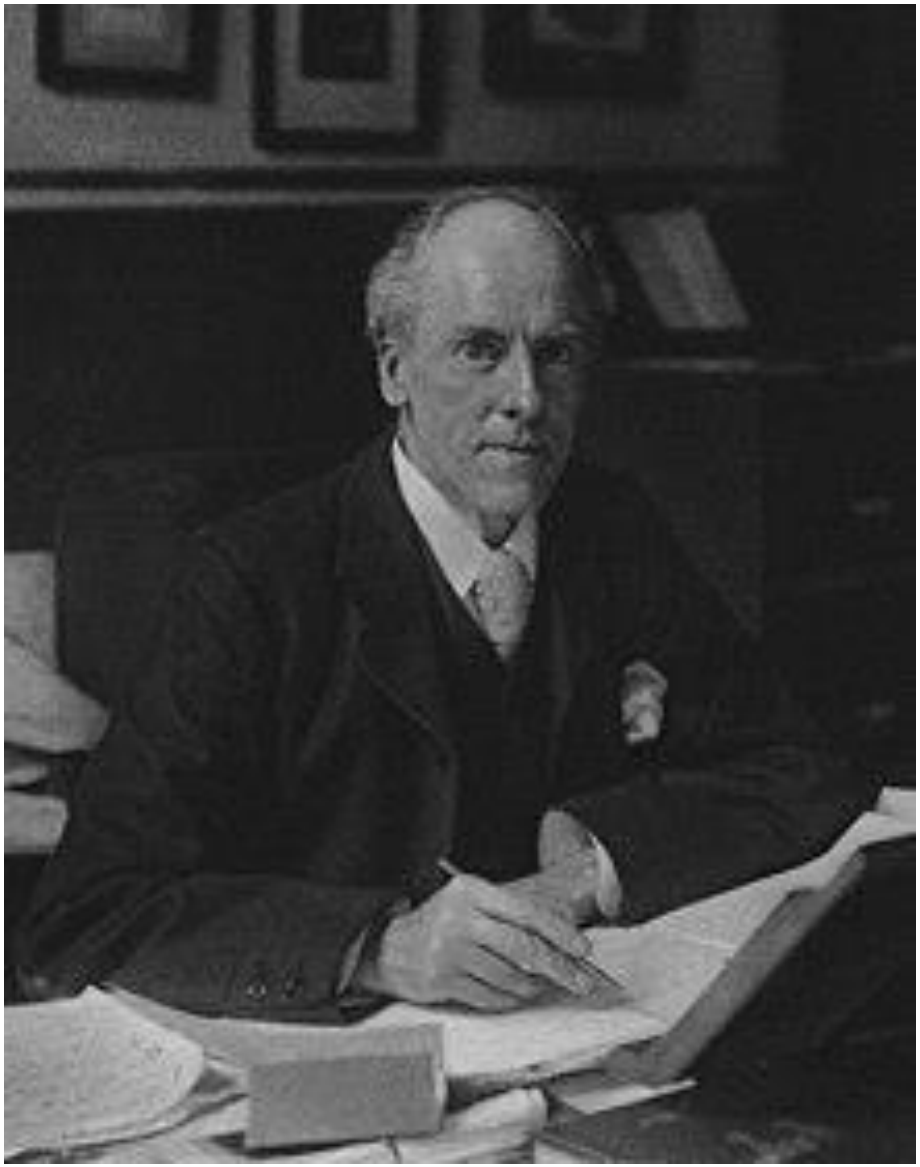
Factor analysis fits a theoretical model of the structure of the correlation matrix

Factor Analysis

- ✓ **Proposed by Galton at the end of the 800 ad further developed by Pearson and Spearman**
- ✓ **Intepretation of learning tests**
- ✓ **Theoretical assumption on the covariance structure of a multivariate system**

Factor Analysis





Theoretical assumptions

Variability of each X_i in a multivariate system

- ✓ Quota shared with other variables

COMMUNALITY

- ✓ Quota specific for each variable

UNIQUENESS

The factorial model of the covariance

The (co)variance matrix of a system could be decomposed

$$S = BB' + \Psi$$

S = (Co)variance matrix of original data

B = (Co)variance between original data and latent factors

Ψ = Specific variance matrix

The factorial model

The original variables are represented as linear functions of a smaller number (q) of latent common factor variables and of a single latent specific variable:

$$X_1 = b_{11}Y_1 + b_{12}Y_2 + \ldots\ldots\ldots\ldots + b_{1q}Y_q + e_1$$
$$X_2 = b_{21}Y_1 + b_{22}Y_2 + \ldots\ldots\ldots\ldots + b_{2q}Y_q + e_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots$$
$$X_p = b_{p1}Y_1 + b_{p2}Y2_+ \ldots\ldots\ldots\ldots + b_{pq}Y_q + e_p$$

The factorial model

$$X_1 = b_{11}Y_1 + b_{12}Y_2 + \ldots\ldots\ldots\ldots + b_{1q}Y_q + e_1$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Y_j = *j*-th common factor,

b_{ij} = factor loading, i.e. parameter that reflects the importance of the *j*-th factor in the composition of the della *i*-th original variable (factorial load o loading)

e = specific casual variable for X_i

Difference between factor analysis and PCA

The j-th principal component

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \ldots + a_{pj}X_p$$

The factor model

$$X_p = b_{p1}Y_1 + b_{p2}Y_2 + \ldots\ldots\ldots\ldots + b_{pq}Y_q + e_p$$

Factors are called common because generates the covariances among the original variables

The factorial model in matrix terms

$$\begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \dots & \dots & \dots & \dots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_q \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_p \end{bmatrix}$$

$$\mathbf{X = BY + e}$$

The matrix of factor loadings

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots\cdots & b_{1p} \\ b_{21} & b_{22} & \cdots\cdots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \cdots\cdots & b_{pp} \end{bmatrix}_{(pxp)}$$

b are correlations between factors and original variables

B allows for interpreting the factorial structure

The factorial model of the covariance

Diagonal elements of **BB'**

$$\sigma_i^2 - \Psi_i = \sum_{j=1}^{q} b_{ij}^2$$

Are the **communality** of original variables

Called also h^2

Factor analysis

- ✓ Estimation of loading matrix **B**
- ✓ Estimation of uniqueness matrix **Ψ**
- ✓ Different methods
- ✓ Constraint of the communalieties

Factor analysis:

The principal component method

Similar to PCA

Differences in **R** structure

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \\ r_{p1} & r_{p2} & & 1 \end{bmatrix} \qquad \ddot{\mathbf{R}} = \begin{bmatrix} h_1^2 & r_{12} & \dots & r_{1p} \\ r_{21} & h_2^2 & \dots & r_{2p} \\ \dots & \dots & \dots & \\ r_{p1} & r_{p2} & & h_p^2 \end{bmatrix}$$

Communalities have to be estimated a priori

Factor analysis: ML method

- ✓ Iterative
- ✓ Test for establishing the number of factors to be extracted
- ✓ **In PCA the factorisation of R is unique**
- ✓ **In factor analysis no**

Factor rotation:

The concept of simple structure of B

- ✓ Each factor should have very large loadings for few variables e many close to zero for the others
- ✓ A variable should have a large loading only for one factor and close to zero for the other factors
- ✓ Variables with common origin, type etc. should be associated to the same factor
- ✓ Loading matrix **B** could be modified trough the **rotation**

Factor rotation: Analytical aspects

Factor rotation consists in the right multplication of the loading matrix **B** by an orthogonal matrix **T**

$$\mathbf{S} = \mathbf{BB'} + \mathbf{\Psi}$$

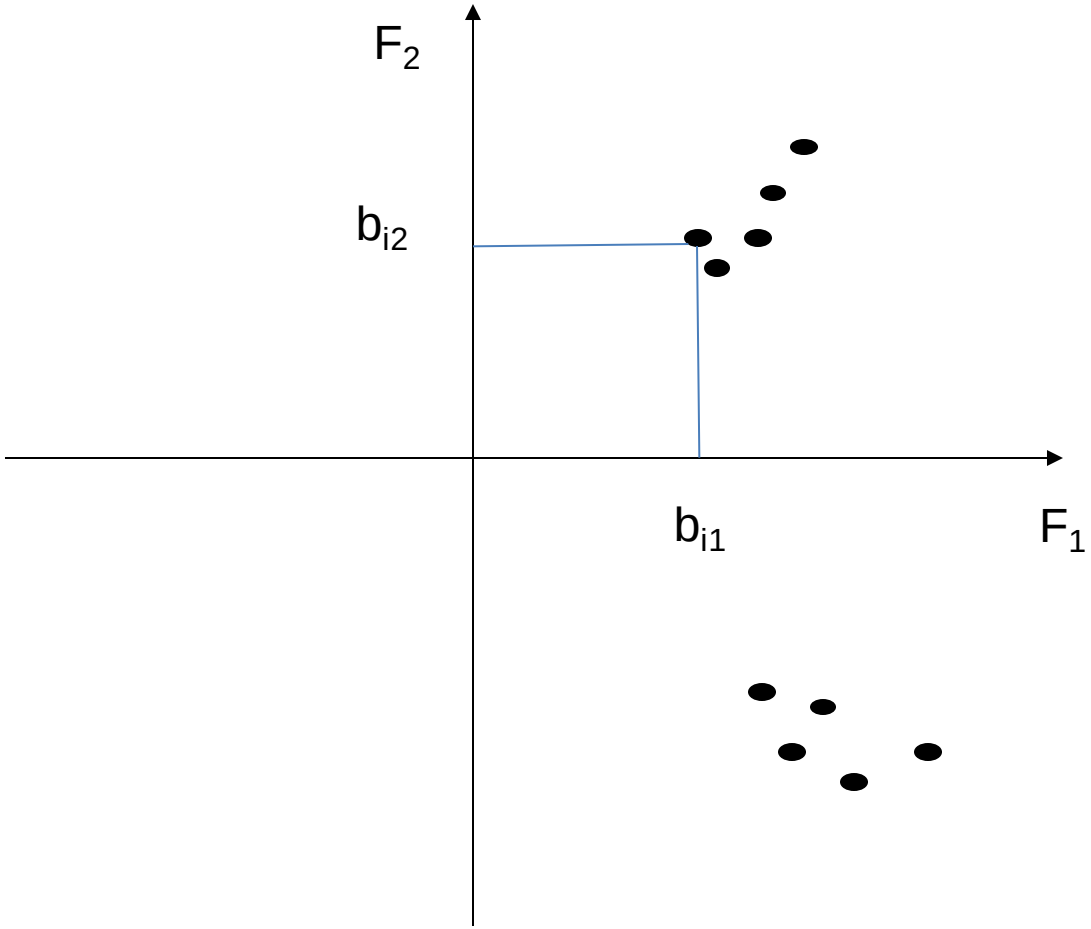
Right-multiplying **B** by **T**

$$\mathbf{S} = \mathbf{BT(BT)'} + \mathbf{\Psi} = \mathbf{BTT'B'} + \mathbf{\Psi} = \mathbf{BB'} + \mathbf{\Psi}$$

Factor rotation: Geometrical aspects

- ✓ Factorial space with q dimensions
- ✓ Axis are the factors
- ✓ The dots are the variables
- ✓ Coordinates on factorial axis are the correlations of the variables with the factors (loadings)
- ✓ Factor rotation is equivalent to a rogid rotation of the axis in the q dimensons of the factorial space

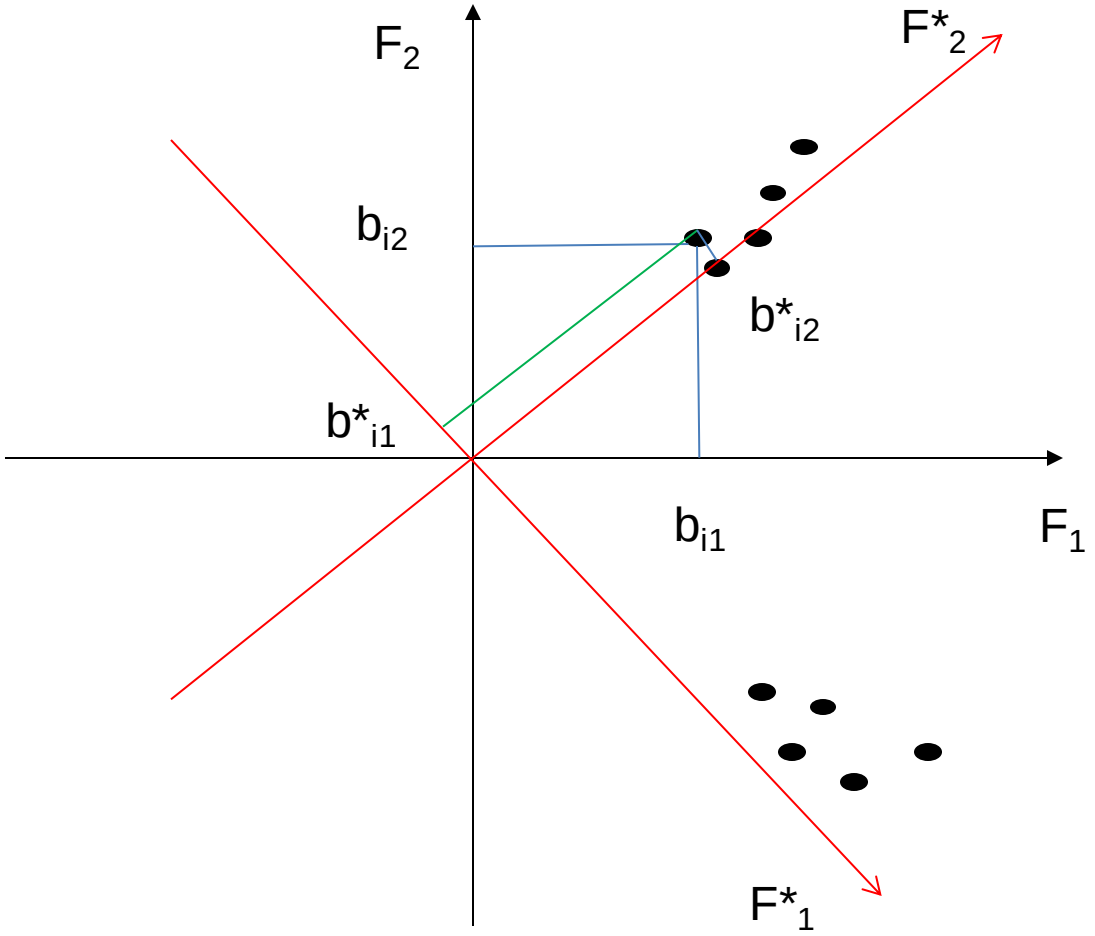
Factor rotation: Geometrical aspects



•Space defined by two factors (F_1 e F_2)

uniss
UNIVERSITÀ DEGLI STUDI DI SASSARI
agraria
DIPARTIMENTO DI AGRARIA

Factor rotation: Geometrical aspects



•Factors are orthogonally rotated (F^*_1 e F^*_2)

uniss
UNIVERSITÀ DEGLI STUDI DI SASSARI
agraria
DIPARTIMENTO DI AGRARIA

Methods of Factor rotation

Orthogonal

VARIMAX = maximisation of squared weight sum within each colomn of **B**

QUARTIMAX = maximisation of squared weight sum within each row of **B**

Oblique

OBLIMIN, **PROMAX**

Steps of Factor analysis

1. Assessment of data suitability
2. Evaluation of number of factors to extract
3. Factor extraction
4. Rotation and intepretation
5. Evaluation of goodness of extracted factors
6. Factor score calculation

Assessment of data suitability

Factor analysis assumes the existence of a latent structure that generates the covariance among observed variables

X_1 *and* X_2 r_{x1x2}

A third variable,
X_k,

could explain part of the relationships between X_1 *e* X_2

$R_{x1x2,k}$ = *partial correlation between* X_1 *and* X_2 *given* X_k

Comparison between Pearson and partial correlations

	MILK1	MILK2	MILK3	MILK4	MILK5	MILK6
MILK1	*	0.85	0.78	0.71	0.61	0.44
MILK2	0.53	*	0.88	0.81	0.71	0.54
MILK3	0.14	0.42	*	0.87	0.79	0.61
MILK4	0.03	0.13	0.40	*	0.86	0.69
MILK5	-0.01	0.01	0.13	0.43	*	0.79
MILK6	-0.08	-0.00	-0.00	0.07	0.54	*

Comparison between Pearson and partial correlations

If Pearson correlations are high and partial correlations are small the existence of a latent structure can be hypothesized

The data set could be considered suitable for factor analysis

An index to assess suitability for factor Analysis is the Kaiser Measure of Sample Adequacy MSA

MSA>0,75 for good suitability

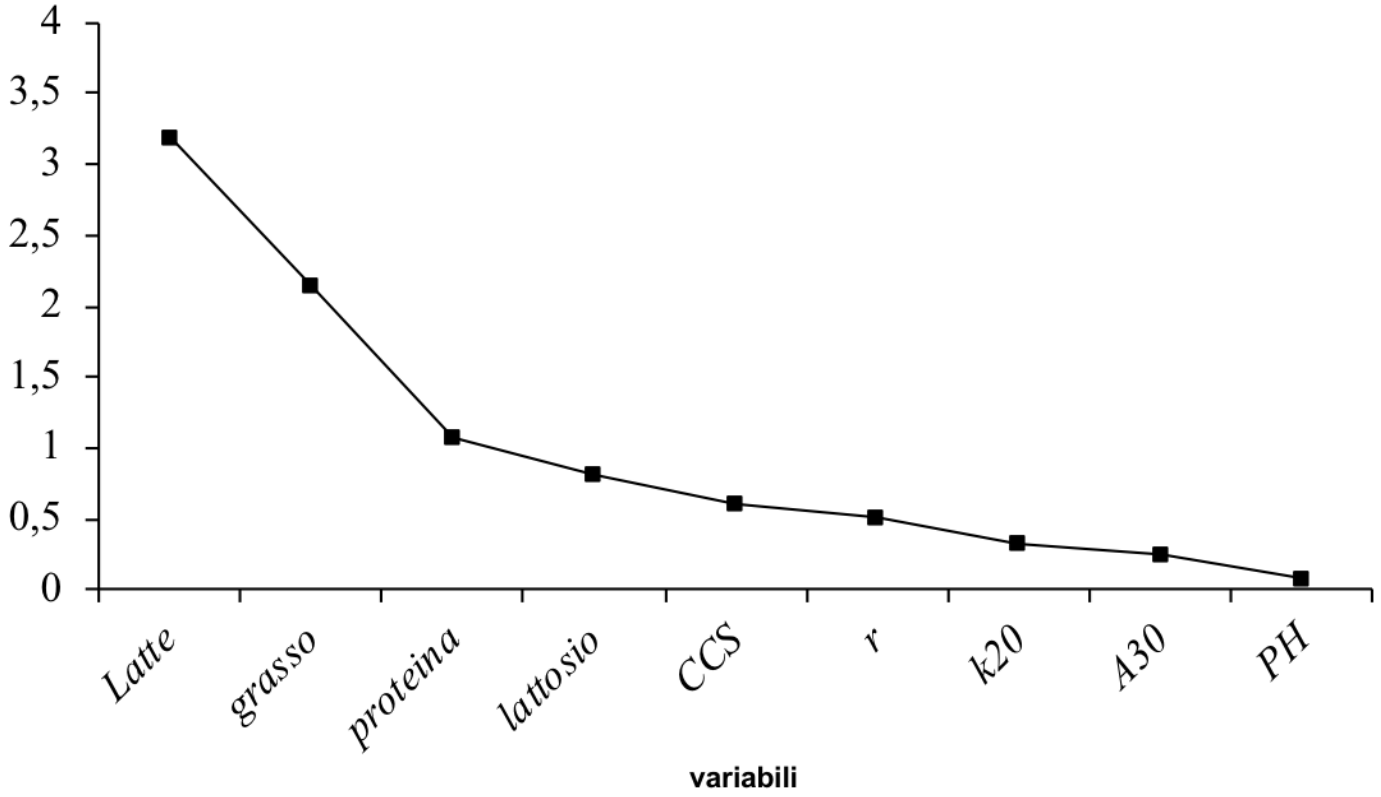
Steps of Factor analysis

1. Assessment of data suitability
2. Evaluation of number of factors to extract
3. Factor extraction
4. Rotation and intepretation
5. Evaluation of goodness of extracted factors
6. Factor score calculation

Evaluation of number of factors to extract

- ✓ **Relevant literature**
- ✓ **Amount of explained variance**
- ✓ **Factor with eigenvalues>1**
- ✓ **Statistical test**

Evaluation of number of factors to extract



Steps of Factor analysis

1. Assessment of data suitability
2. Evaluation of number of factors to extract
3. Factor extraction
4. Rotation and intepretation
5. Evaluation of goodness of extracted factors
6. Factor score calculation

Factor rotation and intepretation

	Factor1	Factor2	Factor3	Factor1	Factor2	Factor3
Milk	-0,36	0,71	-0,09	0,04	0,76	-0,25
Fat	0,02	-0,32	0,88	-0,04	-0,11	0,93
Protein	0,39	-0,64	-0,14	-0,01	-0,76	0,01
Lactose	-0,50	0,70	0,11	-0,06	0,86	-0,04
SCC	0,55	0,22	0,41	0,63	-0,02	0,35
R	0,86	0,28	-0,12	0,87	-0,24	-0,18
k20	0,70	0,51	0,13	0,87	0,08	0,02
a30	-0,85	-0,41	0,02	-0,93	0,10	0,11
pH	0,65	-0,32	-0,21	0,37	-0,64	-0,13

Evaluation of goodness of extracted factors

Factors generate the correlation matrix through the loading matrix **B**

$$\mathbf{R} = \mathbf{BB'} + \mathbf{\Psi}$$

Thus

$$\mathbf{R - BB'} = \mathbf{\Psi}$$

$$\mathbf{\Psi} = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Psi p \end{bmatrix}$$

Evaluation of goodness of extracted factors

R

$$\begin{bmatrix} 1 & 0.83 & 0.78 & 0.73 & 0.68 & 0.61 & 0.50 \\ & 1 & 0.89 & 0.85 & 0.79 & 0.72 & 0.62 \\ & & 1 & 0.90 & 0.86 & 0.79 & 0.68 \\ & & & 1 & 0.90 & 0.84 & 0.74 \\ & & & & 1 & 0.89 & 0.80 \\ & & & & & 1 & 0.86 \\ symm & & & & & & 1 \end{bmatrix} =$$

BB'

$$\begin{bmatrix} 0.73 & 0.81 & 0.79 & 0.75 & 0.68 & 0.60 & 0.49 \\ & 0.90 & 0.89 & 0.86 & 0.80 & 0.72 & 0.61 \\ & & 0.91 & 0.89 & 0.86 & 0.79 & 0.68 \\ & & & 0.89 & 0.88 & 0.84 & 0.75 \\ & & & & 0.90 & 0.89 & 0.81 \\ & & & & & 0.91 & 0.85 \\ symm & & & & & & 0.80 \end{bmatrix}$$

Ψ

$$\begin{bmatrix} 0.27 & -0.03 & -0.01 & -0.02 & 0.00 & 0.01 & 0.01 \\ & 0.10 & 0.00 & -0.01 & -0.0 & 0.00 & 0.01 \\ & & 0.09 & 0.01 & 0.00 & 0.00 & 0.00 \\ & & & 0.11 & 0.02 & 0.00 & -0.01 \\ & & & & 0.10 & 0.00 & -0.01 \\ & & & & & 0.09 & 0.01 \\ symm & & & & & & 0.20 \end{bmatrix}$$

Factor score calculations

Common factor *scores* (**f'**) could be calculated as:

$$\mathbf{f}' = \mathbf{x}'[(\mathbf{B}\mathbf{B}' + \Psi)^{-1} \cdot \mathbf{B}]$$

Scores could the be used as new phenotypes for further elaborations:

Some examples of use of factor analysis in animal science

- ✓ Analysis of differences between genomic and chromosomal correlation matrices
- ✓ Derivation of indicators of milk quality
- ✓ Analysis of correlation structure of milk fatty acid composition

Factor analysis and milk quality

J. Dairy Sci. 95:7346–7354
http://dx.doi.org/10.3168/jds.2012-5546
© American Dairy Science Association®, 2012.

Use of multivariate factor analysis to define new indicator variables for milk composition and coagulation properties in Brown Swiss cows

N. P. P. Macciotta,*[1] A. Cecchinato,† M. Mele,‡ and G. Bittante†
*Dipartimento di Agraria, Sezione Scienze Zootecniche, Università di Sassari, Via de Nicola 9, 07100 Sassari, Italy
†Dipartimento di Agronomia, Animali, Alimenti, Risorse Naturali e Ambiente, Università di Padova, Viale dell'Università 16, 35020 Legnaro, Padova, Italy
‡Dipartimento di Agronomia e Gestione dell'Agroecosistema, Sezione di Scienze Zootecniche, Università di Pisa, Via S. Michele degli Scalzi, 2, 56124 Pisa, Italy

J. Dairy Sci. 99:4547–4557
http://dx.doi.org/10.3168/jds.2015-10589
© American Dairy Science Association®, 2016.

Derivation of multivariate indices of milk composition, coagulation properties, and individual cheese yield in dairy sheep

M. G. Manca,* J. Serdino,* G. Gaspa,* P. Urgeghe,* I. Ibba,† M. Contu,† P. Fresi,‡ and N. P. P. Macciotta*[1]
*Dipartimento di Agraria, Università di Sassari, 07100 Sassari, Italy
†Associazione Regionale Allevatori della Sardegna, 09128 Cagliari, Italy
‡Associazione Nazionale della Pastorizia, 00161 Rome, Italy

Factor analysis and milk quality

- ✓ Variables describing milk chemical e tecnological properties
- ✓ Complex correlation pattern
- ✓ Conflicting results in literature
- ✓ Possible existence of a latent structure

Factor analysis and milk quality

Table 2. Pearson (above the diagonal) and partial (under the diagonal) correlations among milk composition traits and milk coagulation properties.

	RCT	a_{30}	k_{20}	ILCY	FP	PP	LAC	SCS	CAS	FRE	Ph	NaCl
RCT	*	-0.40	0.66	0.34	0.11	0.28	-0.27	0.40	0.27	-0.06	0.61	0.29
a_{30}	-0.02	*	-0.70	-0.30	-0.17	0.02	0.25	-0.23	0.03	0.19	-0.15	-0.23
k_{20}	0.47	-0.56	*	0.28	0.09	0.00	-0.31	0.34	-0.01	-0.20	0.36	0.36
ILCY	-0.01	-0.16	0.05	*	0.55	0.42	-0.29	0.30	0.42	0.07	0.21	0.16
FP	-0.08	-0.05	-0.06	0.45	*	0.46	-0.49	0.26	0.48	0.09	-0.06	0.17
PP	-0.08	-0.03	-0.01	0.05	-0.26	*	-0.43	0.25	0.99	0.14	-0.05	0.13
LAC	0.03	0.03	-0.07	0.26	-0.72	-0.02	*	-0.41	-0.39	0.40	0.04	-0.84
SCS	0.07	0.00	0.07	0.01	0.15	0.01	0.12	*	0.24	-0.09	0.26	0.41
CAS	0.13	0.05	-0.02	-0.00	0.21	0.99	-0.08	0.01	*	0.17	-0.06	0.07
FRE	-0.02	0.03	0.02	-0.17	0.57	-0.00	0.77	-0.11	0.08	*	0.03	-0.26
Ph	0.55	0.14	-0.03	0.14	0.07	0.02	0.15	0.09	-0.04	-0.08	*	0.09
NaCl	0.08	0.06	-0.01	0.23	-0.58	0.20	-0.92	0.21	-0.30	0.66	0.09	*

Factor analysis and milk quality

Trait	Factor 1	Factor 2	Factor 3	Factor4	Communality
Fat percentage	**0.72**	-0.12	-0.18	-0.34	0.68
Protein percentage	**0.91**	-0.09	0.12	0.15	0.88
Casein percentage	**0.92**	-0.03	0.11	0.14	0.89
Lactose percentage	-0.42	**0.87**	-0.02	0.14	0.95
Somatic Cell Score	0.29	-0.42	0.47	-0.09	0.49
Freezing point	0.28	0.64	0.08	0.14	0.52
pH	-0.09	0.08	**0.86**	-0.08	0.77
NaCl	0.10	**-0.86**	0.19	-0.08	0.79
Individual cheese yield	0.62	0.03	0.17	-0.48	0.64
Rennet coagulation time	0.20	-0.14	**0.82**	-0.30	0.83
Curd firming time	-0.04	-0.26	0.49	**-0.70**	0.80
Curd firmness	0.01	0.16	-0.15	**0.87**	0.82
Variance explained (%)	25	19	17	15	

Factor analysis and milk quality

Variable	Factor1	Factor 2	Factor 3	Factor 4	Communality
TS	**0.814**	0.006	-0.136	0.378	0.82
FP	**0.526**	-0.241	0.120	-0.403	0.51
PP	**0.940**	0.075	-0.080	-0.093	0.90
CAS	**0.962**	0.009	-0.090	0.047	0.93
LAC	0.074	-0.159	-0.169	**0.818**	0.73
SCC	-0.011	0.101	0.008	**-0.651**	0.43
SH	0.347	-0.106	**-0.779**	0.171	0.77
pH	0.082	0.110	**0.900**	-0.045	0.83
RCT	0.138	**0.743**	0.034	0.095	0.58
K20	-0.021	**0.667**	0.020	-0.208	0.49
A30	0.148	**-0.729**	-0.193	0.166	0.62
Eigenvalues	0.29	0.19	**0.12**	0.10	

Factor analysis and milk quality

Table 5. Features of the marginal posterior distribution of heritability for the investigated traits

Trait	h^2			
	PM[1]	LB95%[2]	UB95%[3]	*P* [4]
Composition factor	0.203	0.081	0.398	94.40
Coagulation factor	0.227	0.106	0.413	98.25
Acidity factor	0.102	0.050	0.190	52.77
Health factor	0.137	0.041	0.288	73.99
Milk yield, kg	0.085	0.010	0.231	39.81

R code for performing multivariate factor analysis on genomic and chromosomal correlation matrices

```
DGV=read.table("…….genomic.csv", header=T, sep=",")

DCV=read.table(":……../BTA14.csv", header=T, sep=",")

G_GEN=as.matrix(DGV[ ,2:15])
G_CHR=as.matrix(DCV[ ,2:15])
#calculation of correlation matrix
R=cor(G_GEN)                                    "
R

#calculation of Kaiser Measure of Sampling Adequacy (Kaiser MSA)
#the package "psych"  is required

library(psych)
KMO(R)
```

R code for performing multivariate factor analysis on milk quality data

```
sheep=read.table("/Users/nicolomacciotta/OneDrive - Università degli Studi di Sassari/Corso Saragozza/Presentations/files/Sheep milk quality.csv", header=T, sep=";" dec=",")
milk=sheep[ ,1:14]
R=cor(milk)

#calculation of Kaiser Measure of Sampling Adequacy (Kaiser MSA)
#the package "psych"  is required
library(psych)                                          "
KMO(R)

Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = R)
Overall MSA =  0.64
MSA for each item =
   resa     a30     fat protein lactose  Casein Urea_FT   pH_FT    NaCl   C18_1
   0.72    0.72    0.76    0.55    0.67    0.56    0.42    0.72    0.52    0.67
  C18_3     CLA       R   latte
   0.68    0.58    0.80    0.78
```

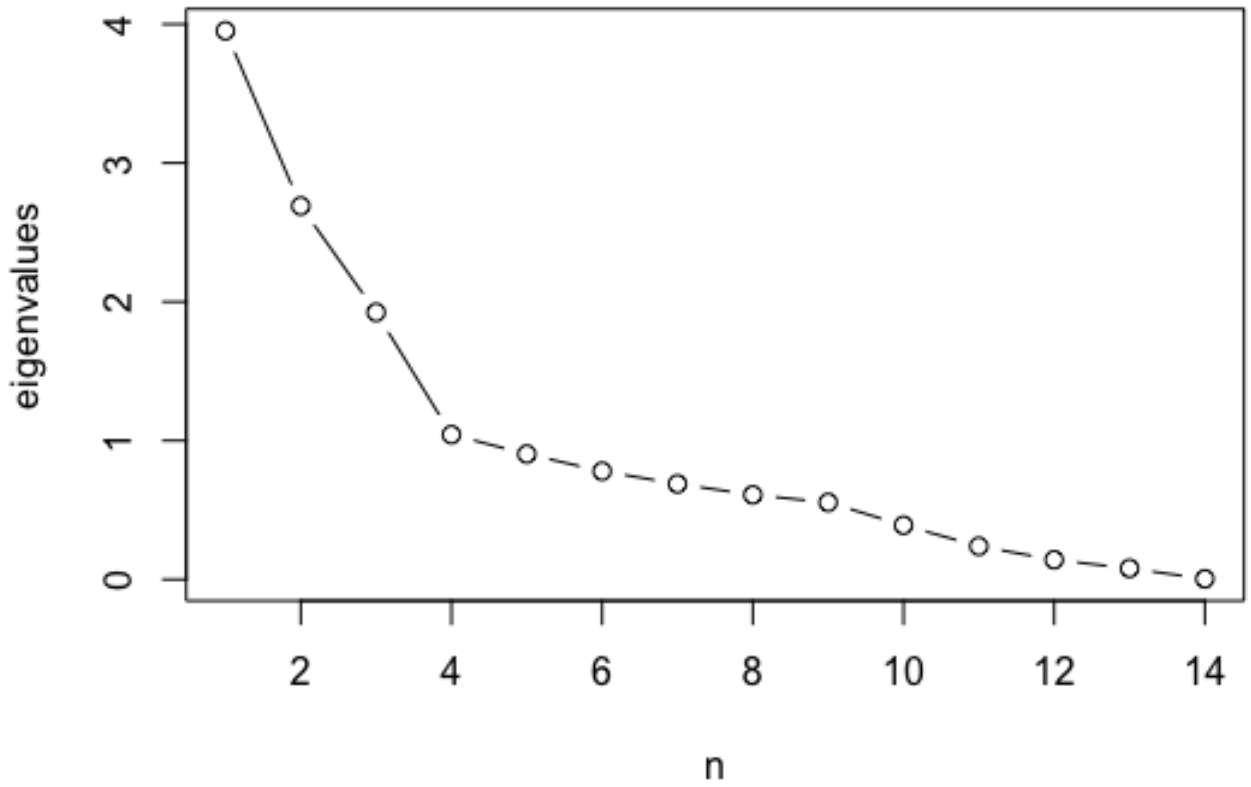
R code for performing multivariate factor analysis on milk quality data

```
#assessment of number of factors to retain
eigenvalues=eigen(R)$values
eigenvalues

var_tot=sum(diag(R))
proportion=eigenvalues/var_tot
proportion
```

```
eigenvalues
 [1] 3.952196392 2.690115698 1.924126251 1.042984767 0.903268209 0.779620769
 [7] 0.688344862 0.609335791 0.554840267 0.390365213 0.239109506 0.141737359
[13] 0.079624499 0.004330416
>
> var_tot=sum(diag(R))
> proportion=eigenvalues/var_tot
> proportion
 [1] 0.2822997423 0.1921511213 0.1374375894 0.0744989119 0.0645191578
 [6] 0.0556871978 0.0491674902 0.0435239851 0.0396314476 0.0278832295
[11] 0.0170792504 0.0101240971 0.0056874642 0.0003093154
```

R code for performing multivariate factor analysis on milk quality data

```
n=c(1:nrow(R))
plot(n,eigenvalues, type="b")
```



R code for performing multivariate factor analysis on milk quality data

```
#factor extraction - package "rela" is required
library(rela)
Factors=factanal(milk, factors=3, rotation="varimax")
Factors
```

```
Uniquenesses:
   resa     a30     fat protein lactose  Casein Urea_FT   pH_FT    NaCl   C18_1
  0.793   0.933   0.510   0.007   0.227   0.005   0.961   0.535   0.008   0.216

 C18_3     CLA       R   latte
  0.769   0.104   0.660   0.778
```

R code for performing multivariate factor analysis on milk quality data

Loadings:

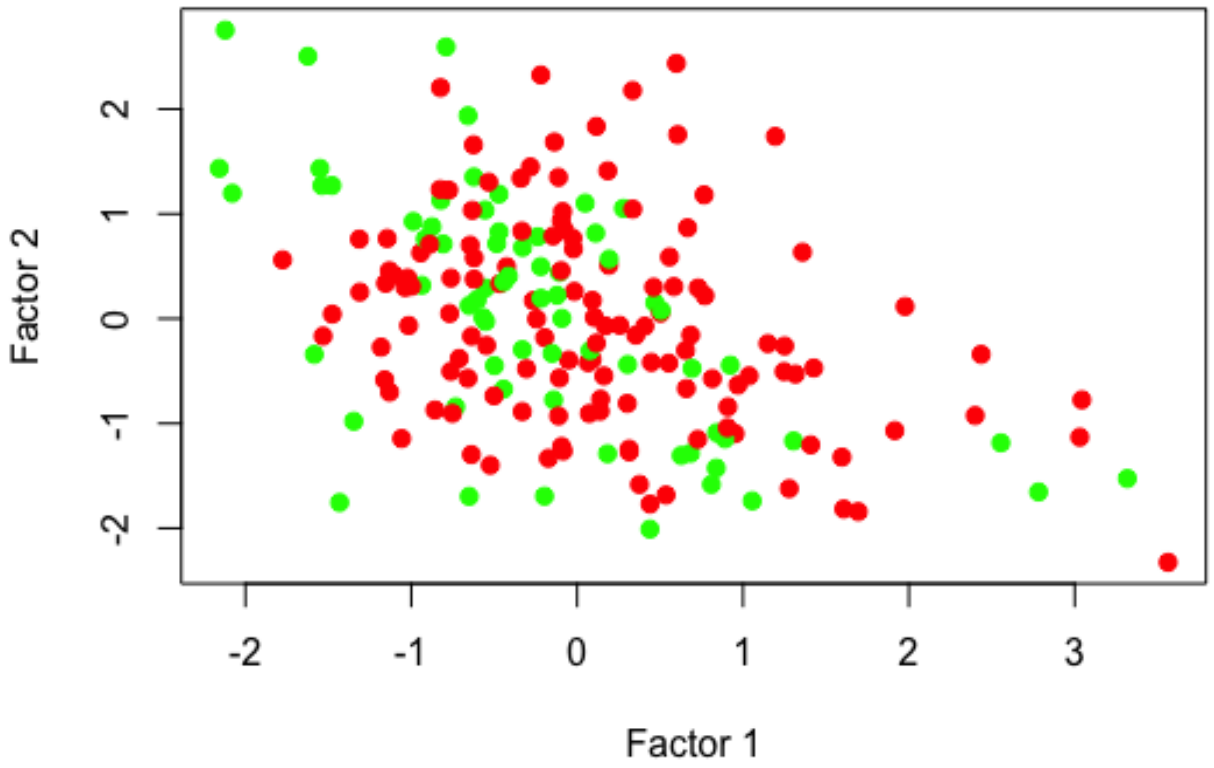
	Factor1	Factor2	Factor3
resa	0.341	-0.167	0.105
a30	0.280		
fat	0.275	-0.532	
protein	1.077	0.216	
lactose	-0.194	0.369	0.706
Casein	1.052	0.156	
Urea_FT		0.154	
pH_FT		-0.220	0.644
NaCl			-0.990
C18_1		0.830	0.264
C18_3	0.238	0.527	-0.178
CLA	0.319	1.054	-0.105
R	0.436	-0.105	0.301
latte	-0.238	0.302	

	Factor1	Factor2	Factor3
SS loadings	2.994	2.774	2.124
Proportion Var	0.214	0.198	0.152
Cumulative Var	0.214	0.412	0.564

uniss UNIVERSITÀ DEGLI STUDI DI SASSARI agraria DIPARTIMENTO DI AGRARIA

R code for performing multivariate factor analysis on milk quality data

```
#production of scores
Scores=factanal(milk, factors=3, rotation="promax",scores="Bartlett")$scores

plot(Scores[ ,1], Scores[ ,2],xlab="Factor 1",ylab="Factor 2",
col=c("red","green","blue")[unclass(sheep$zona)],cex=1, pch=19)
```



Calculation of the residual matrix

$$\mathbf{R = BB' + \Psi}$$

B=Factors$loadings
BB1=B%*%t(B)
Psi=cor(milk)-BB1
round(Psi,2)

resa	0.84	-0.06	0.15	0.07	-0.06	0.06	-0.12	0.18	0.00	-0.07	-0.14	-0.13	0.16	-0.06
a30	-0.06	0.91	0.07	-0.07	-0.06	-0.05	0.05	0.08	-0.01	-0.06	0.00	-0.16	0.01	-0.03
fat	0.15	0.07	0.64	0.23	-0.26	0.23	0.00	-0.03	0.03	-0.05	-0.16	-0.03	0.11	-0.17
protein	0.07	-0.07	0.23	-0.21	-0.10	-0.18	-0.06	0.14	-0.06	-0.39	-0.30	-0.56	0.02	-0.10
lactose	-0.06	-0.06	-0.26	-0.10	0.33	-0.11	-0.01	-0.02	-0.03	0.08	0.07	0.12	-0.13	0.12
Casein	0.06	-0.05	0.23	-0.18	-0.11	-0.13	-0.07	0.15	-0.05	-0.37	-0.27	-0.52	0.05	-0.13
Urea_FT	-0.12	0.05	0.00	-0.06	-0.01	-0.07	0.97	-0.21	-0.01	0.10	0.03	-0.03	-0.03	0.03
pH_FT	0.18	0.08	-0.03	0.14	-0.02	0.15	-0.21	0.53	0.01	0.03	-0.12	0.10	0.32	-0.06
NaCl	0.00	-0.01	0.03	-0.06	-0.03	-0.05	-0.01	0.01	0.02	-0.07	-0.05	-0.09	-0.01	-0.02
C18_1	-0.07	-0.06	-0.05	-0.39	0.08	-0.37	0.10	0.03	-0.07	0.24	-0.12	-0.08	-0.08	0.05
C18_3	-0.14	0.00	-0.16	-0.30	0.07	-0.27	0.03	-0.12	-0.05	-0.12	0.63	-0.25	-0.25	0.16
CLA	-0.13	-0.16	-0.03	-0.56	0.12	-0.52	-0.03	0.10	-0.09	-0.08	-0.25	-0.22	-0.12	0.05
R	0.16	0.01	0.11	0.02	-0.13	0.05	-0.03	0.32	-0.01	-0.08	-0.25	-0.12	0.71	-0.13
latte	-0.06	-0.03	-0.17	-0.10	0.12	-0.13	0.03	-0.06	-0.02	0.05	0.16	0.05	-0.13	0.85

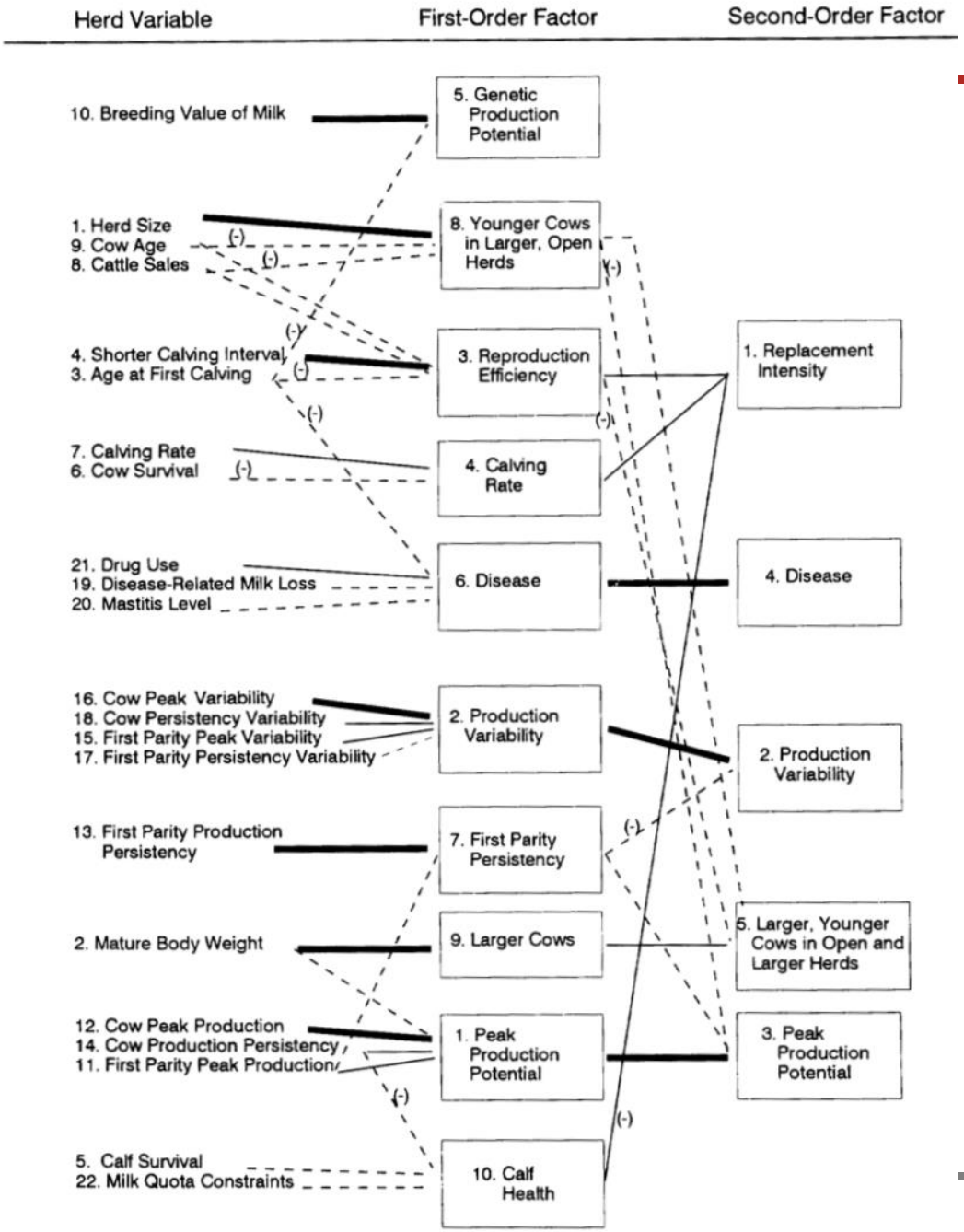
Factor analysis and farm characteristics

Dairy Herd Management Types Assessed from Indicators of Health, Reproduction, Replacement, and Milk Production

CARSTEN ENEVOLDSEN,[1] JENS HINDHEDE, and TROELS KRISTENSEN
Danish Institute of Animal Science, Research Centre Foulum,
DK-8830 Tjele, Denmark

TABLE 1. Five, 25, 50, 75, and 95 percentiles[1] of 22 variables describing 111 Danish dairy herds.

Variable	Percentile				
	5	25	50	75	95
1 Herd size, cow years	34	46	59	71	118
2 Mature BW, kg[2]	431	605	623	625	625
3 Age at first calving, d	754	798	823	862	939
4 Calving interval, median d to calving or culling	412	393	375	361	346
5 Calf survival, % alive at 6 mo postpartum	77	83	88	92	95
6 Cow survival, % in herd at 12 wk postpartum	88	93	95	96	98
7 Calving rate, calves born per 100 cow years	96	105	111	119	128
8 Livestock sales, no.[3]	−23	0	2	10	25
9 Cow age, mean lactation number	1.9	2.2	2.3	2.5	2.9
10 Breeding value for milk, index[4]	−290	−124	−20	79	241
11 First parity peak milk, kg of ECM[5]	18	20	23	25	27
12 Cow peak milk, kg of ECM[5] (parity ≥3)	25	28	31	33	37
13 First parity persistency, kg of ECM[6]	0.5	1.3	1.9	2.4	3.6
14 Cow persistency, kg of ECM[6] (parity ≥3)	2.7	4.2	4.9	5.7	6.9
15 First parity peak variability (CV), %	6	10	12	16	23
16 Cow peak variability (CV), % (parity ≥3)	6	8	10	13	19
17 First parity persistency variability, 75 to 25 percentile	0.5	1.0	1.0	2.0	2.5
18 Cow persistency variability, 75 to 25 percentile	0.0	1.0	2.0	2.5	4.0
19 Disease-related milk loss, %[7]	4	5	7	9	11
20 Mastitis level, mean annual SCC	163	236	304	354	414
21 Drug use, administrations per 100 calvings	4	29	63	99	145
22 Milk quota constraints, composite score[8]	−0.9	−0.4	0.0	0.4	1.0


Herd Variable
First-Order Factor
Second-Order Factor
10. Breeding Value of Milk
5. Genetic Production Potential
1. Herd Size
9. Cow Age
8. Cattle Sales
8. Younger Cows in Larger, Open Herds
4. Shorter Calving Interval
3. Age at First Calving
3. Reproduction Efficiency
1. Replacement Intensity
7. Calving Rate
6. Cow Survival
4. Calving Rate
21. Drug Use
19. Disease-Related Milk Loss
20. Mastitis Level
6. Disease
4. Disease
16. Cow Peak Variability
18. Cow Persistency Variability
15. First Parity Peak Variability
17. First Parity Persistency Variability
2. Production Variability
2. Production Variability
13. First Parity Production Persistency
7. First Parity Persistency
2. Mature Body Weight
9. Larger Cows
5. Larger, Younger Cows in Open and Larger Herds
12. Cow Peak Production
14. Cow Production Persistency
11. First Parity Peak Production
1. Peak Production Potential
3. Peak Production Potential
5. Calf Survival
22. Milk Quota Constraints
10. Calf Health
(-)

Factor analysis and farm characteristics

MANAGEMENT TYPES

TABLE 9. Final regression model of total annual milk production in 111 Danish dairy herds. Effects of five second-order factors and mean number of cows.

Variable[1]	df	Regression coefficient	SE	*P*
1 Replacement intensity	1	21,020	10,279	0.0435
2 Production variability	1	15,039	7224	0.0399
3 Potential peak production	1	−11,319	7881	0.1541
5 Larger cows in younger, open, and larger herds	1	2775	3961	0.4851
6 Cows in the herd, no.	1	6595	152	0.0001

Factor analysis and genomic matrices

J. Dairy Sci. 92:2931–2946
doi:10.3168/jds.2008-1762
© American Dairy Science Association, 2009.

Distribution and location of genetic effects for dairy traits

J. B. Cole,*[1] P. M. VanRaden,* J. R. O'Connell,† C. P. Van Tassell,*‡ T. S. Sonstegard,‡ R. D. Schnabel,§ J. F. Taylor,§ and G. R. Wiggans*
*Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD 20705-2350

Table 7. Correlations among BTA18[1]-specific (above the diagonal) and genome-wide (below the diagonal) sire estimated breeding values for net merit (NM), longevity (PL), sire (SCE) and daughter calving ease (DCE), stature (STAT), strength (STR), body depth (BDep), and rump width (RWid)

Item	NM	PL	SCE	DCE	STAT	STR	BDep	RWid
NM		0.88	−0.64	−0.67	−0.44	−0.53	−0.60	−0.46
PL	0.71		−0.72	−0.71	−0.48	−0.62	−0.67	−0.55
SCE	−0.32	−0.31		0.81	0.69	0.79	0.79	0.79
DCE	−0.56	−0.44	0.58		0.57	0.73	0.75	0.69
STAT	0.16	0.00	0.25	−0.13		0.82	0.83	0.78
STR	0.02	−0.09	0.28	0.03	0.72		0.95	0.90
BDep	−0.01	−0.18	0.30	0.05	0.78	0.91		0.88
RWid	0.11	−0.02	0.23	−0.07	0.68	0.71	0.72	

[1]BTA18 = *Bos taurus* autosome 18.

uniss UNIVERSITÀ DEGLI STUDI DI SASSARI agraria DIPARTIMENTO DI AGRARIA

Correlation between chromosomal values

- **Correlation structures may differ between chromosomes**
- **Differences may indicate changes in the genetic control of a group of traits**
- **Identification of genes with small additive effects that affect several traits**

How to test differences betwen matrices?

- ✓ **Issue in population and evolutionary genetics**
- ✓ **Different aspects**
- ✓ **PC are used to explain the variance of a system**
- ✓ **The covariance should be investigated**

Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Dissection of genomic correlation matrices of US Holsteins using multivariate factor analysis

N.P.P. Macciotta[1], C. Dimauro[1], D.J. Null[2], G. Gaspa[1], M. Cellesi[1] & J.B. Cole[2]

- Genomic correlation matrices calculated for all BTAs for US Holstein
- Based on genomic evaluations of May 2012 (n. 182,233)
- 31 traits (productive, functional)

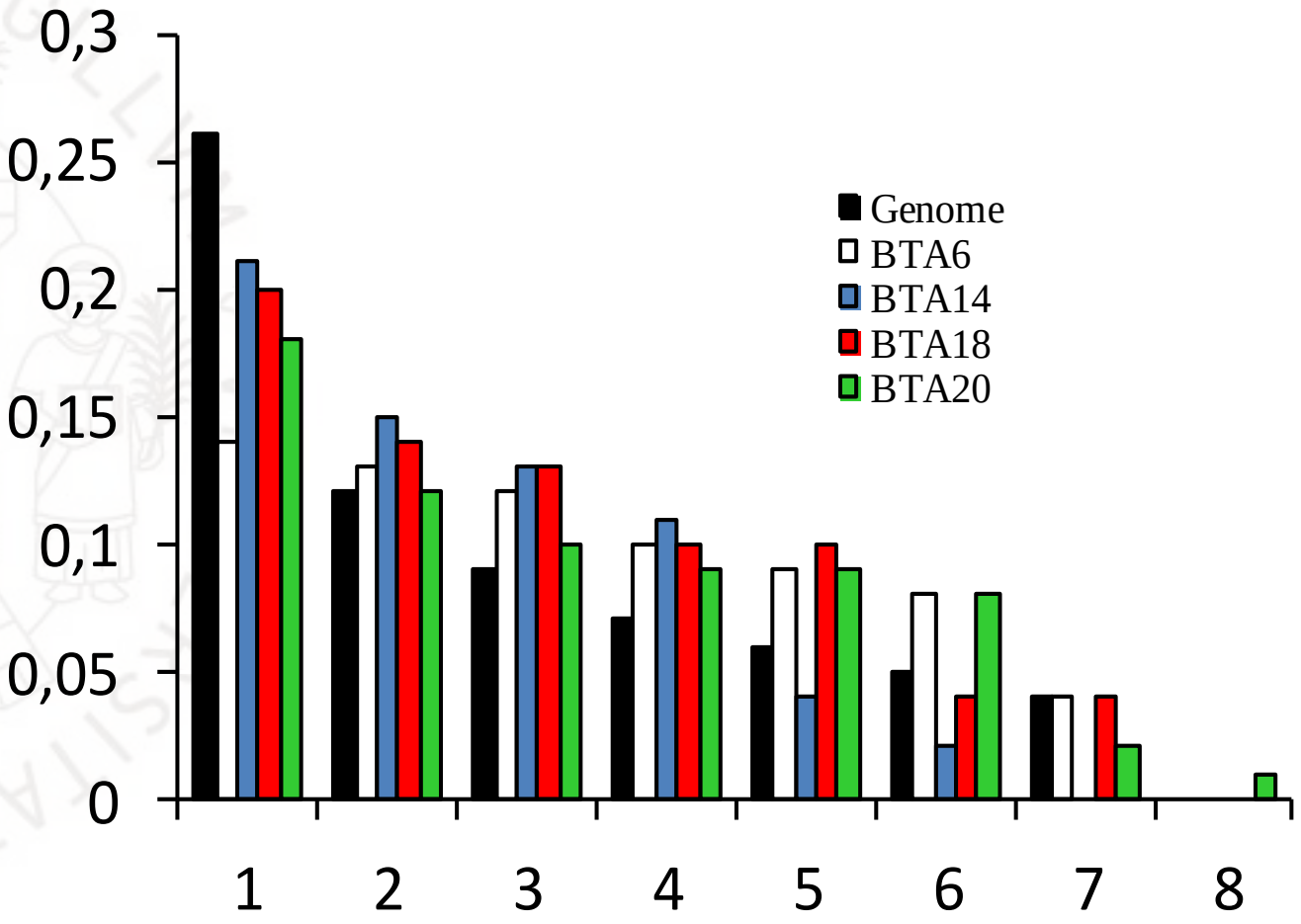
The basic idea

Differences in the correlation pattern among genomic values of different traits

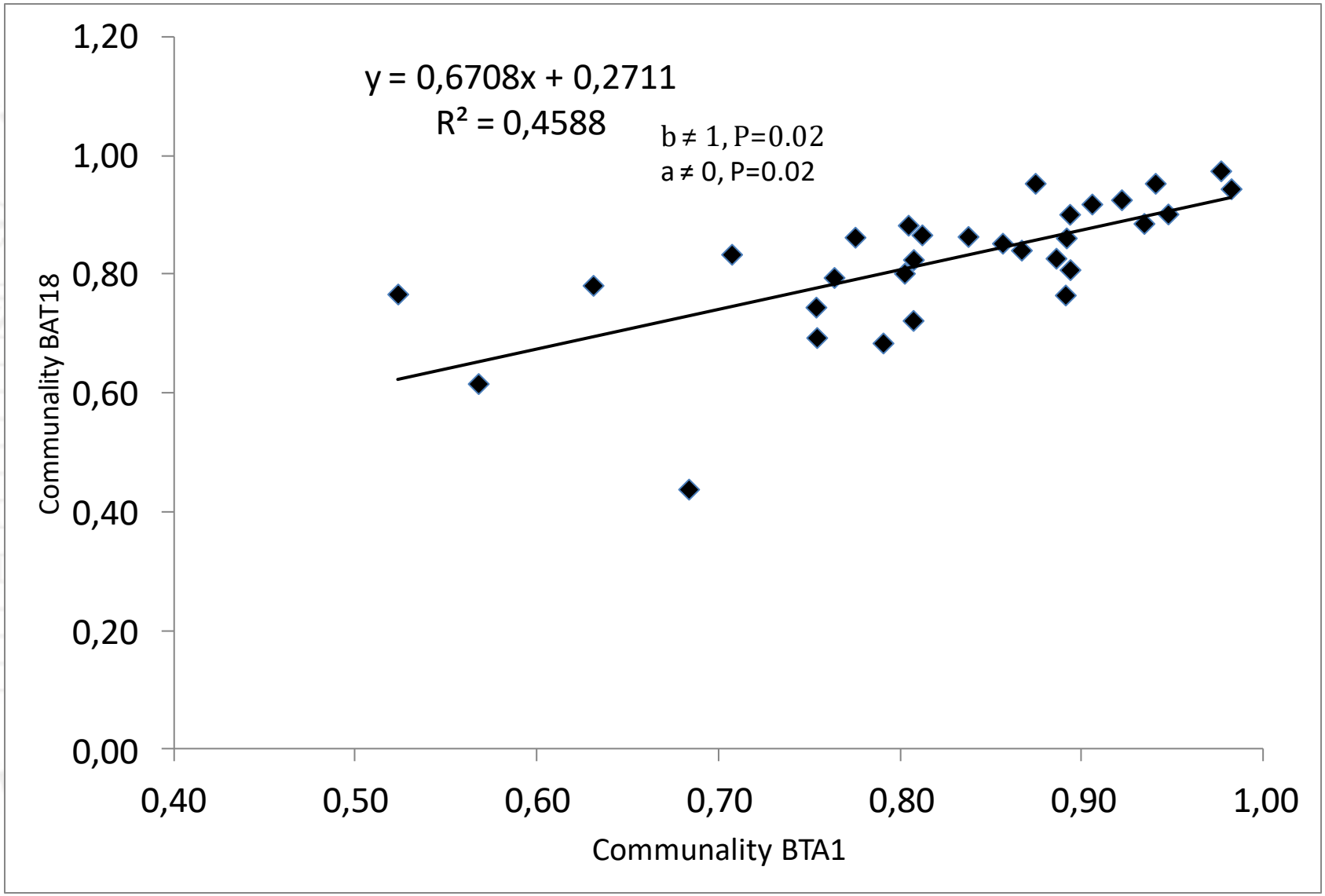
Changes in the quota of variance shared by all variables (**communality)**

Variation in the latent structure that controls the (co)variance of the system

Eigenvalue plot comparison



How to test differences?



How to test differences?

Table 9 Regression analysis of communalities extracted from the genomic correlation matrix on those extracted from the different chromosome matrices

BTA	Intercept	p^1	Slope	p^2
6	0.32 ± 0.09	0.01	0.66 ± 0.10	0.02
14	0.30 ± 0.10	0.02	0.68 ± 0.12	0.03
18	0.51 ± 0.03	<0.001	0.48 ± 0.03	<0.001
20	0.41 ± 0.01	<0.001	0.58 ± 0.02	<0.001
3	−0.12 ± 0.13	0.390	1.12 ± 0.16	0.453

Trait	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Milk	0.29	0.14	**0.89**	0.02	0.03	−0.17	−0.28
Fat	0.30	0.20	**0.66**	0.06	0.04	**0.65**	0.03
Protein	0.31	0.23	**0.90**	0.05	0.03	−0.03	0.20
Fat percentage	0.04	0.07	−0.20	0.04	0.01	**0.92**	0.33
Protein percentage	0.01	0.14	−0.09	0.05	−0.01	0.29	**0.94**
Net merit	0.36	**0.75**	0.47	0.13	−0.04	0.24	0.07
Productive life	0.22	**0.92**	0.10	0.09	−0.10	0.04	−0.02
Somatic cell score	−0.16	**−0.64**	0.11	−0.09	−0.07	−0.11	0.04
Daughter pregnancy rate	−0.22	**0.71**	−0.30	0.03	0.00	−0.09	0.10
Sire calving ease	0.13	−0.42	−0.16	0.01	0.19	−0.01	−0.05
Daughter calving ease	−0.26	−0.48	−0.17	−0.08	0.03	−0.02	0.00
Sire stillbirth	0.13	−0.33	−0.05	0.00	0.09	0.02	−0.04
Daughter stillbirth	−0.15	−0.40	−0.13	−0.07	−0.01	0.00	0.01
Final score	**0.93**	0.09	0.12	0.23	0.24	0.07	0.01
Stature	**0.72**	−0.17	0.09	0.22	0.46	0.02	0.04
Strength	0.41	−0.12	0.08	0.26	**0.86**	0.04	0.05
Dairy form	**0.75**	−0.29	0.34	0.04	0.00	0.10	−0.06
Foot angle	0.52	0.08	0.05	**0.69**	0.27	0.04	0.06
Rear legs (side view)	0.24	−0.14	0.06	−0.58	−0.13	0.02	0.01
Body depth	0.58	−0.28	0.14	0.20	**0.67**	0.08	0.01
Rump angle	−0.06	0.02	0.11	−0.02	0.08	−0.02	−0.06
Rump width	**0.65**	−0.14	0.11	0.11	0.50	0.04	0.04
Fore udder attachment	**0.85**	0.27	−0.06	0.11	0.17	0.06	−0.01
Rear udder height	**0.88**	0.11	0.15	0.16	0.08	0.06	−0.02
Udder depth	**0.73**	0.34	−0.21	0.08	0.09	0.00	0.03
Udder cleft	**0.81**	0.02	0.09	0.06	0.06	0.01	0.00
Front teat placement	**0.63**	0.16	0.14	−0.03	0.04	0.03	0.02
Teat length	0.00	−0.24	−0.03	0.10	0.24	−0.04	−0.06
Rear legs (rear view)	0.53	0.10	0.07	**0.76**	0.11	0.06	0.04
Feet and legs	**0.65**	0.13	0.07	**0.73**	0.05	0.07	0.05
Rear teat placement	**0.62**	0.01	0.14	−0.04	0.00	0.01	0.01
Variance explained (%)	0.26	0.12	0.09	0.07	0.06	0.05	0.04

Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

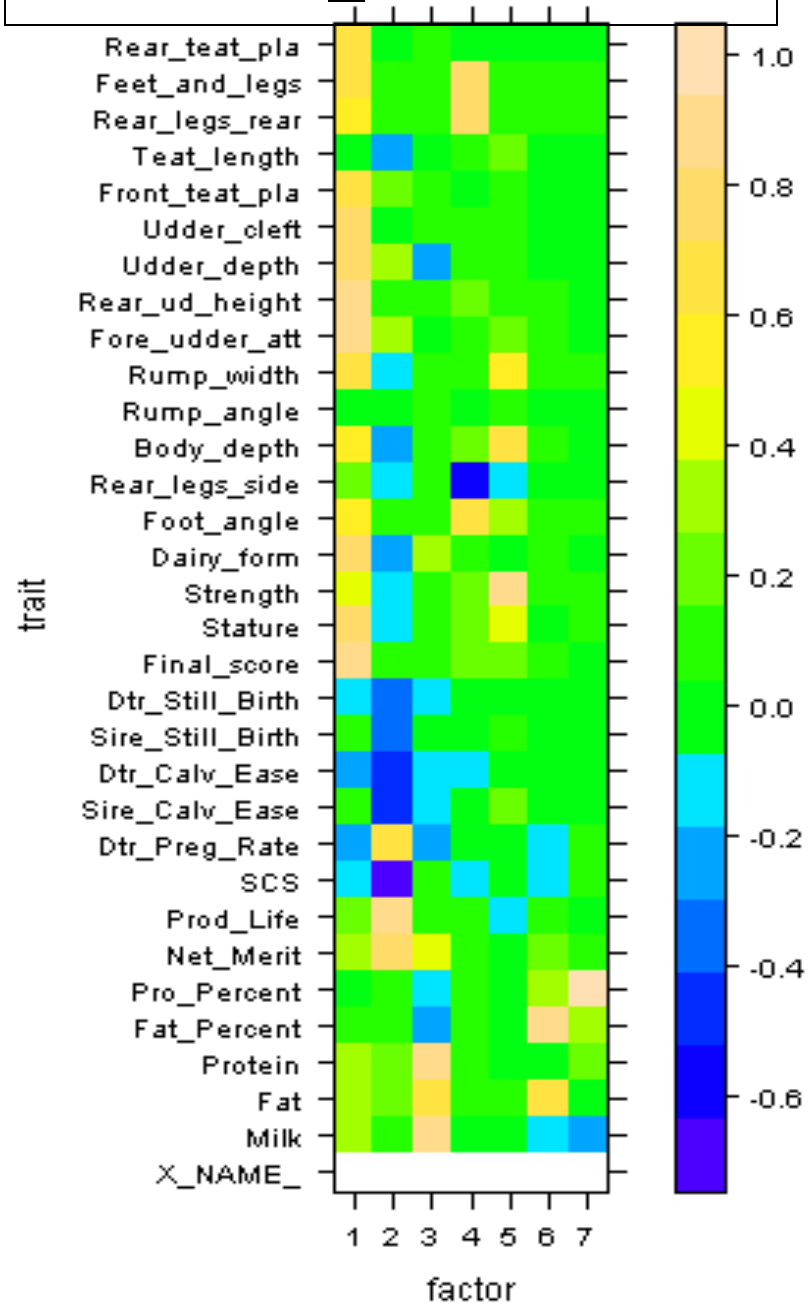
G_GEN
US Holsteins

BTA18	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Milk	−0.08	0.04	0.05	0.01	**0.95**	−0.20	−0.20
Fat	0.18	−0.09	−0.01	−0.07	**0.80**	0.55	0.01
Protein	−0.16	0.05	0.13	−0.04	**0.94**	−0.09	0.26
Fat percentage	0.31	−0.15	−0.07	−0.10	−0.10	**0.89**	0.23
Protein percentage	−0.17	0.01	0.18	−0.11	0.05	0.20	**0.94**
Net merit	−0.31	0.13	**0.78**	0.22	0.46	0.09	0.11
Productive life	−0.41	0.13	**0.83**	0.26	0.15	−0.03	0.09
Somatic cell score	0.00	−0.12	**−0.71**	−0.10	0.01	0.00	0.00
Daughter pregnancy rate	−0.22	−0.08	**0.82**	0.25	−0.13	0.00	0.06
Sire calving ease	**0.72**	0.01	−0.41	0.12	−0.04	0.12	−0.01
Daughter calving ease	0.46	−0.09	−0.50	0.09	−0.17	0.08	−0.12
Sire stillbirth	**0.69**	0.11	−0.31	0.19	−0.01	0.17	−0.02
Daughter stillbirth	0.38	0.08	−0.44	−0.07	−0.26	0.08	−0.10
Final score	0.47	**0.69**	0.22	0.46	0.02	0.00	−0.08
Stature	**0.83**	0.21	−0.03	0.19	−0.06	−0.05	−0.11
Strength	**0.96**	0.01	−0.08	0.09	−0.03	0.09	0.02
Dairy form	0.34	0.37	−0.36	0.11	0.28	−0.06	−0.22
Foot angle	0.52	0.32	0.10	**0.67**	−0.10	−0.03	0.00
Rear legs (side view)	−0.03	−0.04	−0.10	−0.46	−0.03	0.08	0.07
Body depth	**0.93**	0.05	−0.23	0.09	−0.01	0.10	−0.06
Rump angle	−0.37	0.05	−0.14	−0.24	0.11	−0.23	−0.05
Rump width	**0.84**	0.21	−0.07	0.17	−0.05	0.05	−0.06
Fore udder attachment	0.30	**0.67**	0.44	0.33	−0.10	0.01	−0.01
Rear udder height	0.04	**0.71**	0.22	0.36	0.07	−0.12	−0.08
Udder depth	0.17	0.51	0.51	0.27	−0.25	−0.08	−0.11
Udder cleft	0.04	**0.85**	0.14	0.18	0.03	−0.09	0.06
Front teat placement	0.01	**0.81**	−0.06	0.04	−0.02	0.00	0.00
Teat length	0.44	−0.27	−0.13	−0.17	0.17	−0.05	−0.10
Rear legs (rear view)	0.29	0.26	0.16	**0.84**	−0.01	0.11	−0.03
Feet and legs	0.13	0.33	0.16	**0.91**	−0.04	0.00	0.01
Rear teat placement	−0.03	**0.84**	−0.17	0.00	0.07	0.00	0.04
Variance explained (%)	0.20	0.14	0.13	0.10	0.10	0.04	0.04

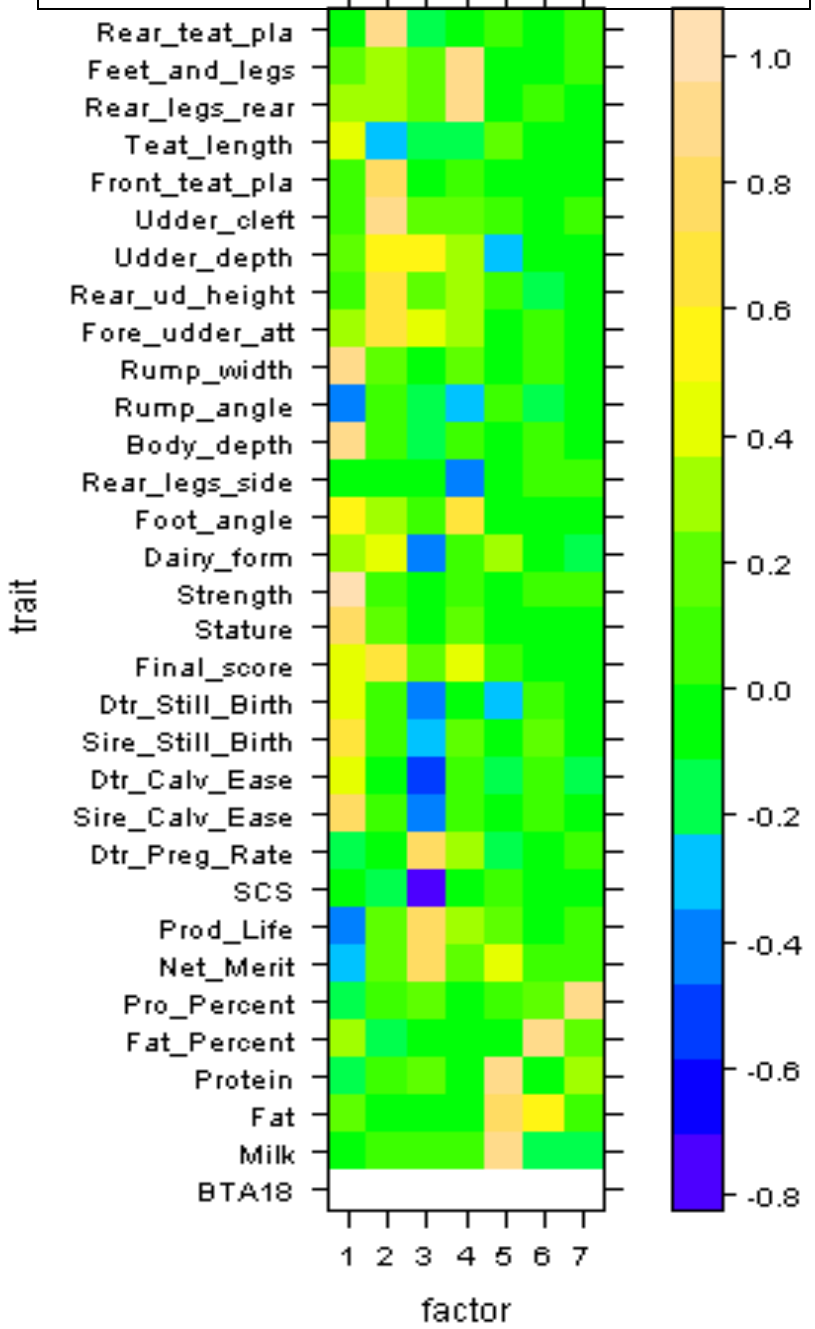
Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

G_CHROM BTA18
US Holsteins

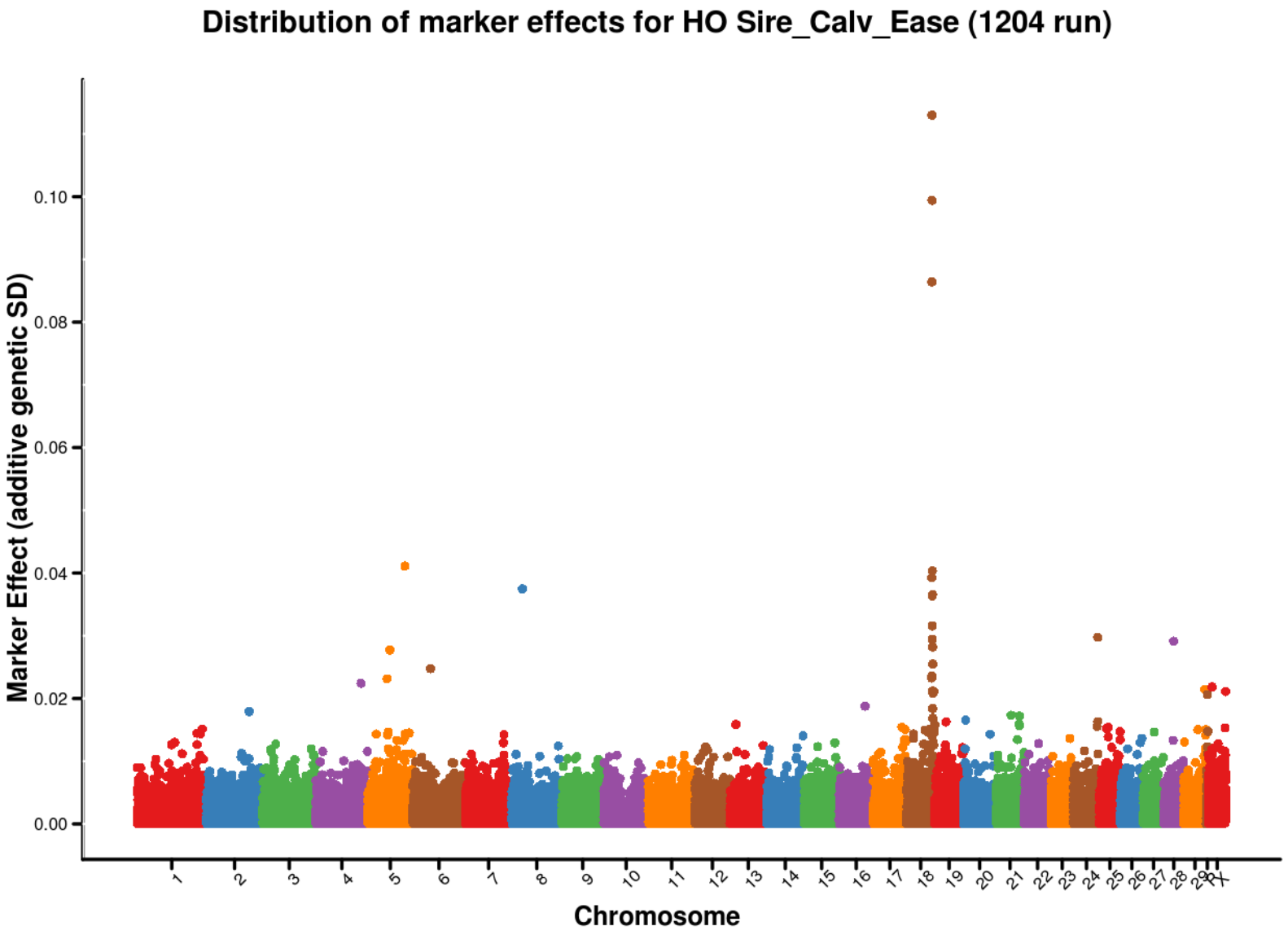
G_GEN



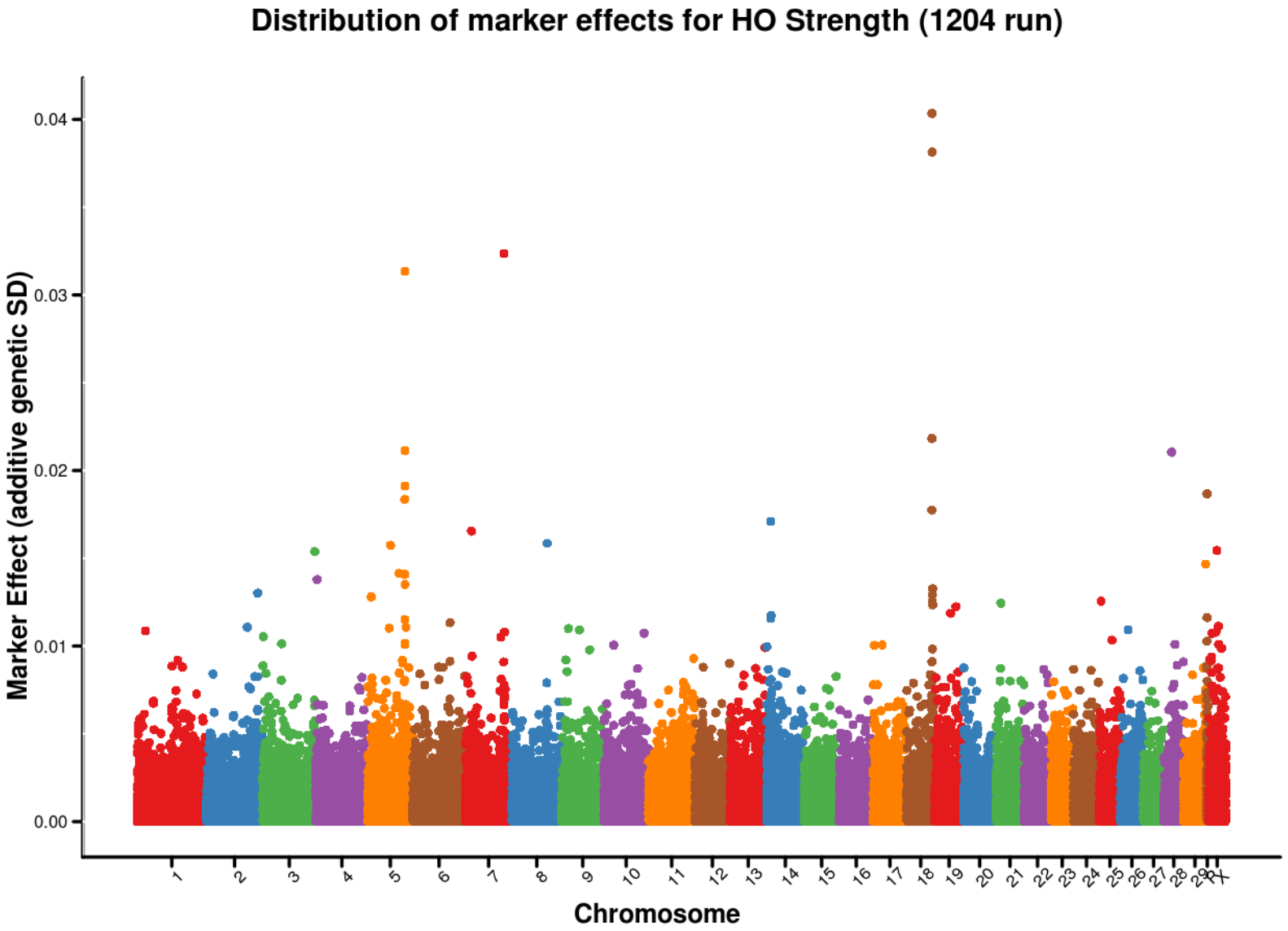
G_CHROM BTA18



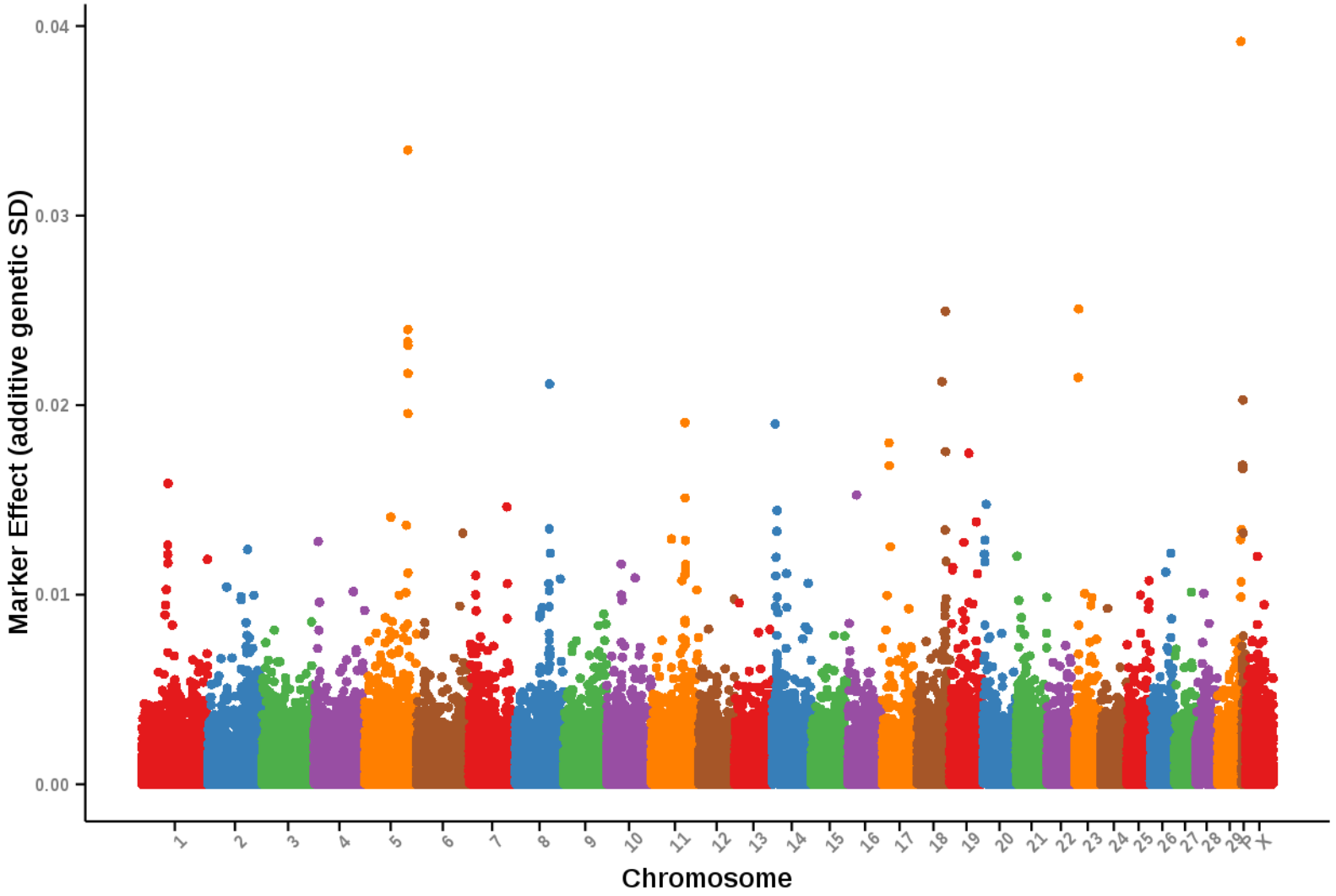
http://aipl.arsusda.gov/Report_Data/Marker_Effects/marker_effects.cfm



http://aipl.arsusda.gov/Report_Data/Marker_Effects/marker_effects.cfm



Distribution of marker effects for HO Stature (1508 run)



What happened later?

ARS-BFGL-NGS-109285 at 57,895,121 Mb on BTA18

Signal consistent in HD data

Intronic to a putative CD33-related Siglec gene

This region is gene-rich

Many Siglecs involved in leptin signaling

Also affects gestation length

Maltecca, C., Gray, K. A., Weigel, K. A., Cassady, J. P., Ashwell, M. 2011. A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. Animal Genetics 42:1365-2052.

Courtesy by John Cole

R code for performing multivariate factor analysis on genomic and chromosomal correlation matrices

```
DGV=read.table("…….genomic.csv", header=T, sep=",")

DCV=read.table(":……../BTA14.csv", header=T, sep=",")

G_GEN=as.matrix(DGV[ ,2:15])
G_CHR=as.matrix(DCV[ ,2:15])
#calculation of correlation matrix
R=cor(G_GEN)                                    "
R

#calculation of Kaiser Measure of Sampling Adequacy (Kaiser MSA)
#the package "psych"  is required

library(psych)
KMO(R)
```

R code for performing multivariate factor analysis on genomic and chromosomal correlation matrices

```
#assesment of number of factors to retain
eigenvalues=eigen(R)$values
eigenvalues

var_tot=sum(diag(R))
proportion=eigenvalues/var_tot
proportion
n=c(1:14)
plot(n,eigenvalues, type="b")                    "

#factor extraction
#package "rela" is required
library(rela)

Factors_GEN=factanal(G_GEN, factors=4, rotation="varimax")
Factors_GEN

Factors_CHR=factanal(G_CHR, nfactors=5, rotation="varimax")
Factors_CHR
```

Loadings:

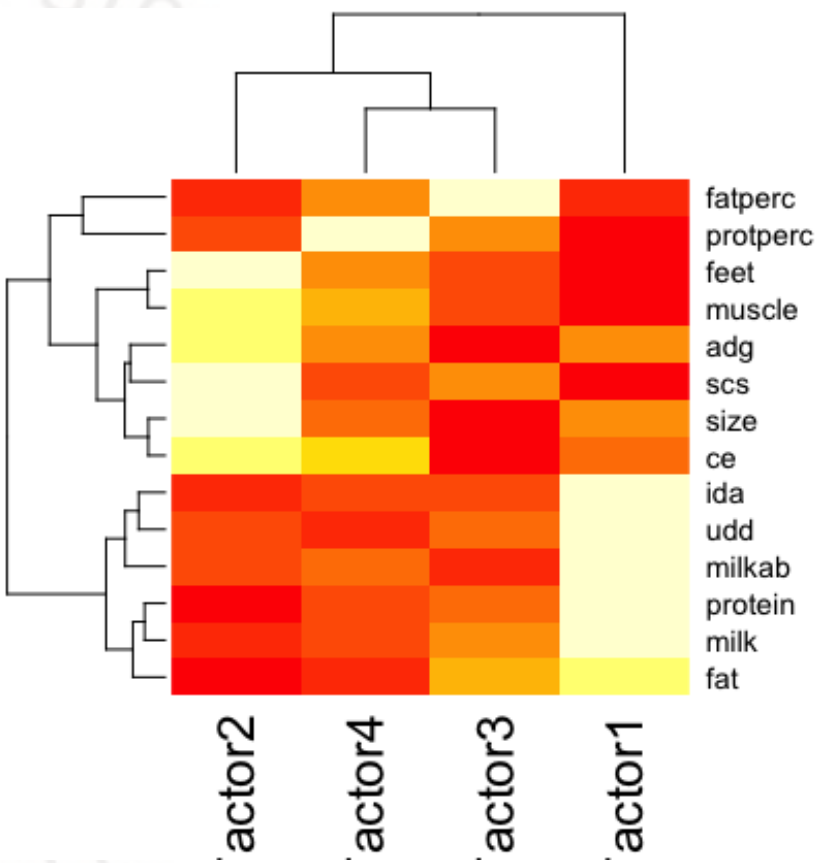
	Factor1	Factor2	Factor3	Factor4
milk	0.867	-0.409	-0.281	
fat	0.848	-0.375	0.330	-0.170
protein	0.908	-0.407		
adg	0.449	-0.226		
milkab	0.562			
fatperc		0.955	0.283	
protperc	-0.132	0.127	0.297	0.935
size		0.227	-0.200	
feet	-0.290	0.737	-0.110	0.138
muscle	-0.470	0.696		0.202
ida	0.995			
udd	0.752		0.111	-0.149
scs	-0.131	0.447		
ce		-0.164		

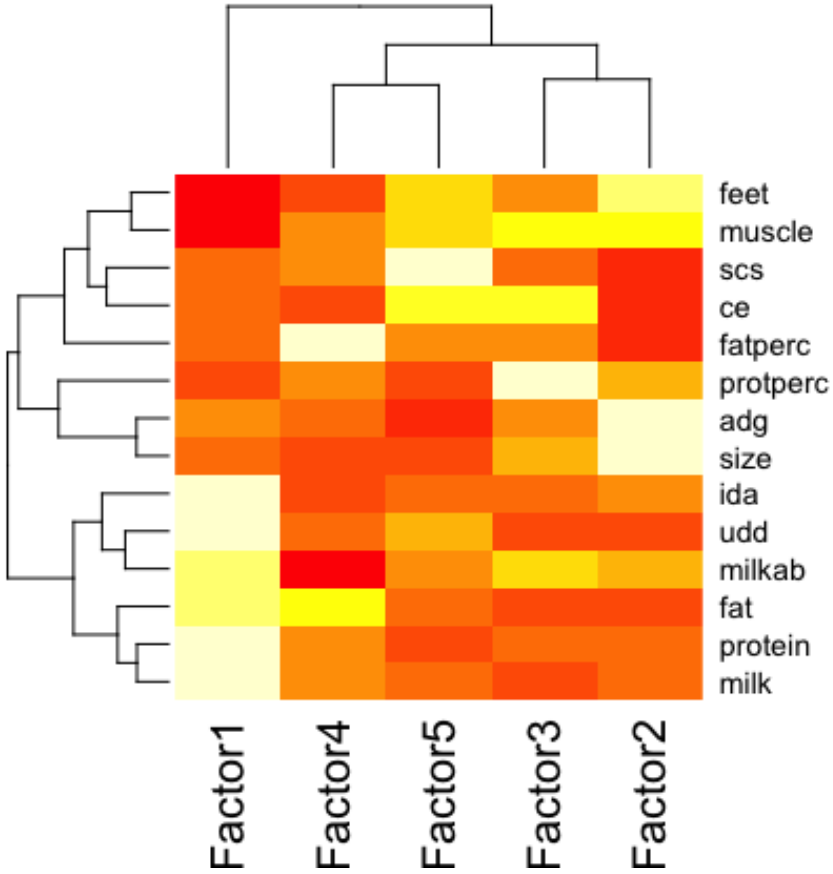
	Factor1	Factor2	Factor3	Factor4
SS loadings	4.520	1.982	1.276	1.170
Proportion Var	0.323	0.142	0.091	0.084
Cumulative Var	0.323	0.464	0.556	0.639

Heat maps

```
#heat map for representing factor patterns
matrix_GEN=as.matrix(factors_GEN$loadings)
matrix_CHR=as.matrix(factors_CHR$loadings)

heatmap(matrix_GEN, symm=F)
heatmap(matrix_CHR, symm=F)

heatmap(matrix_GEN, symm=F, col=c("green","red"))

heatmap(matrix_CHR, symm=F, col=c("green","red"))

heatmap(matrix_GEN, symm=F, col=gray.colors(100,start=1,end=0))
```

Heat maps


fatperc
protperc
feet
muscle
adg
scs
size
ce
ida
udd
milkab
protein
milk
fat
actor2
actor4
actor3
actor1


feet
muscle
scs
ce
fatperc
protperc
adg
size
ida
udd
milkab
fat
protein
milk
Factor1
Factor4
Factor5
Factor3
Factor2