ETH Zurich
D-USYS
Institute of Agricultural Sciences

Exam Solutions
Livestock Breeding and Genomics
FS 2017

Peter von Rohr

DATE	*22. December 2017*
BEGIN	*09:15*
END	*11:15*

Name:

Legi-Nr:

Problem	Maximal Number of Points	Number of Points Reached
1	48	
2	22	
3	30	
4	30	
5	29	
Total	159	

Problem 1 Relationship and Inbreeding

We are given the following pedigree

```
##   sire  dam
## 1 <NA> <NA>
## 2 <NA> <NA>
## 3    1    2
## 4    1    3
## 5    4    2
## 6    4    5
```

a) Use the given pedigree to construct the numerator relationship matrix A

36

Solution:

```
(matA <- as.matrix(getA(ped = ped)))
```

```
##        1      2      3    4      5      6
## 1 1.0000 0.0000 0.5000 0.75 0.3750 0.5625
## 2 0.0000 1.0000 0.5000 0.25 0.6250 0.4375
## 3 0.5000 0.5000 1.0000 0.75 0.6250 0.6875
## 4 0.7500 0.2500 0.7500 1.25 0.7500 1.0000
## 5 0.3750 0.6250 0.6250 0.75 1.1250 0.9375
## 6 0.5625 0.4375 0.6875 1.00 0.9375 1.3750
```

b) Which elements of the matrix A contain the inbreeding coefficient F_5 of animal 5? What is the value of the inbreeding coefficient F_5?

8

Solution:

The following elements contain F_5

$$(2,4),(4,2),(5,5)$$

The value of F_5 corresponds to

$$F_5 = 0.125$$

c) Which parents would 5 need to have, such that $F_5 = 0$ would hold?

4

Solution:

Because only 1 and 2 are not related, these would be the only parents with which $F_5 = 0$ would be true.

Problem 2 Variance Components Estimation

We are given the following dataset for the traits body weight and breast circumference for cattle.

Animal	Herd	BodyWeight	BreastCircumference
1	2	669	161
2	1	635	144
3	1	631	151
4	1	632	155
5	1	642	167
6	2	676	165
7	1	645	163
8	2	686	171
9	1	633	138
10	1	641	170

a) What is the value for the estimated resiudal variance for the trait ‘body weight‘, when the following assumptions are met.

12

In a first model, we assume that `body weight` is influenced by the herd leading to the following model

$$y = Xb + e$$

wobei

y	vector of observations for ‘body weight‘
b	vector of herd effects
X	incidence matrix linking observations to herd effects
e	vector with random resiudals with $E[e] = 0$ and $var(e) = I\sigma^2$

The estimates of the herd effects were computed using the function `lm()` in R. The statement for that is given below

```
lm_gewicht <- lm(BodyWeight ~ -1 + Herd, data = dfWtBc)
round(coefficients(lm_gewicht), digits = 1)
```

```
## Herd1 Herd2
##   637   677
```

Your task

Estimate the residual variance σ^2 for the given model for `body weight` with the method based on the residuals. The same value for the residual variance could also be obtained from the Output of the `summary()`-function in R.

Solution:

The vector of residuals e is

```
(vec_residuals <- residuals(lm_gewicht))
```

```
##  1  2  3  4  5  6  7  8  9 10
## -8 -2 -6 -5  5 -1  8  9 -4  4
```

The estimated residual variance corresponds to the squared residuals divided by the degrees of freedom $df_e = n - p$ where n corresponds to the number of observations and p corresponds to the number of herd effects. Therefore $df_e = 10 - 2 = 8$.

The estimated residual variance and the residual standard deviation are

```
(ssq_res <- crossprod(vec_residuals))
```

```
##      [,1]
## [1,]  332
```

```
(n_sigma2_hat <- ssq_res/(n_nr_obs - length(coefficients(lm_gewicht))))
```

```
##      [,1]
## [1,] 41.5
```

```
(n_sigma_hat <- sqrt(n_sigma2_hat))
```

```
##          [,1]
## [1,] 6.442049
```

As a check, we compare the result to the output of R

```
summary(lm_gewicht)
```

```
##
## Call:
## lm(formula = BodyWeight ~ -1 + Herd, data = dfWtBc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8.00  -4.75  -1.50   4.75   9.00
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## Herd1  637.000      2.435   261.6  < 2e-16 ***
## Herd2  677.000      3.719   182.0 9.29e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.442 on 8 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 5.079e+04 on 2 and 8 DF,  p-value: < 2.2e-16
```

b) What is the difference between the estimated residual variance from Problem 2a) and the estimate of the residual variance based on Maximum Likelhood?

6

Your Task

- Describe the difference between the two estimates
- Compute the value of the Maximum-Likelhood-Estimate for the residual variance
- Which of the two estimate is considered to be "better" than the other?

Solution:

- With the Maximum-Likelihood estimate the degrees of freedom p for the estimation of the fixed effects are not considered
- The value of the estimated residual variance is the same as in a), but the factor with which the sums of squares are multiplied is $1/n$ and not $1/(n-p)$. Hence, we get

$$\widehat{\sigma^2_{ML}} = 33.2$$

- The estimate computed in a) based on the residuals is unbiased and is considered better.

c) When the body weight of animal 11 in herd 1 was recorded, the scale broke down. The body weight of animal 11 is estimated such that it is within the range of ±2 standard deviations around the mean body weight for herd 1.

4

Hints

- Use the estimate of the standard deviation computed from Problem 2a)
- If you could not solve Problem 2a), use a value of 10 as an approximation of the estimated residuals standard deviation.

Solution:

```
(n_av_gew_btr1 <- mean(dfWtBc$Gewicht[dfWtBc$Betrieb == 1]))
```

```
## Warning in mean.default(dfWtBc$Gewicht[dfWtBc$Betrieb == 1]): argument is
## not numeric or logical: returning NA
```

```
## [1] NA
```

```
(n_lower <-  n_av_gew_btr1 - 2* n_sigma_hat)
```

```
##      [,1]
## [1,]   NA
```

```
(n_upper <- n_av_gew_btr1 + 2* n_sigma_hat)
```

```
##      [,1]
## [1,]   NA
```

Problem 3 Livestock Breeding

a) What is the difference between livestock population and wildlife populations with respect to the terms ‘selection‘ and ‘mating‘? Please complete the following table.

8

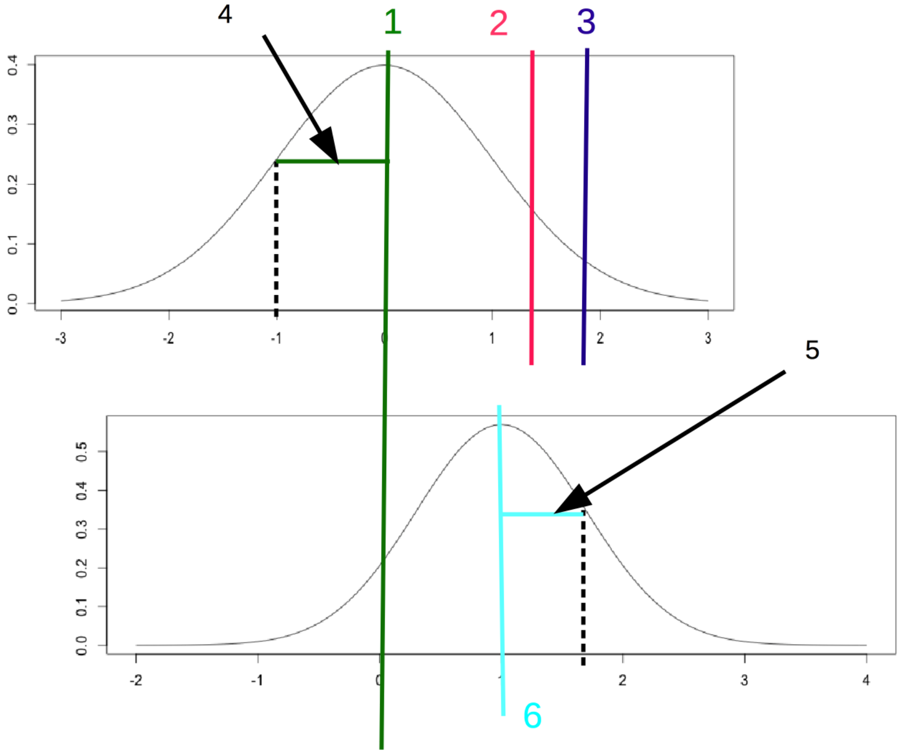
Population	Selection	Mating
Wildlife		
Livestock		

Solution:

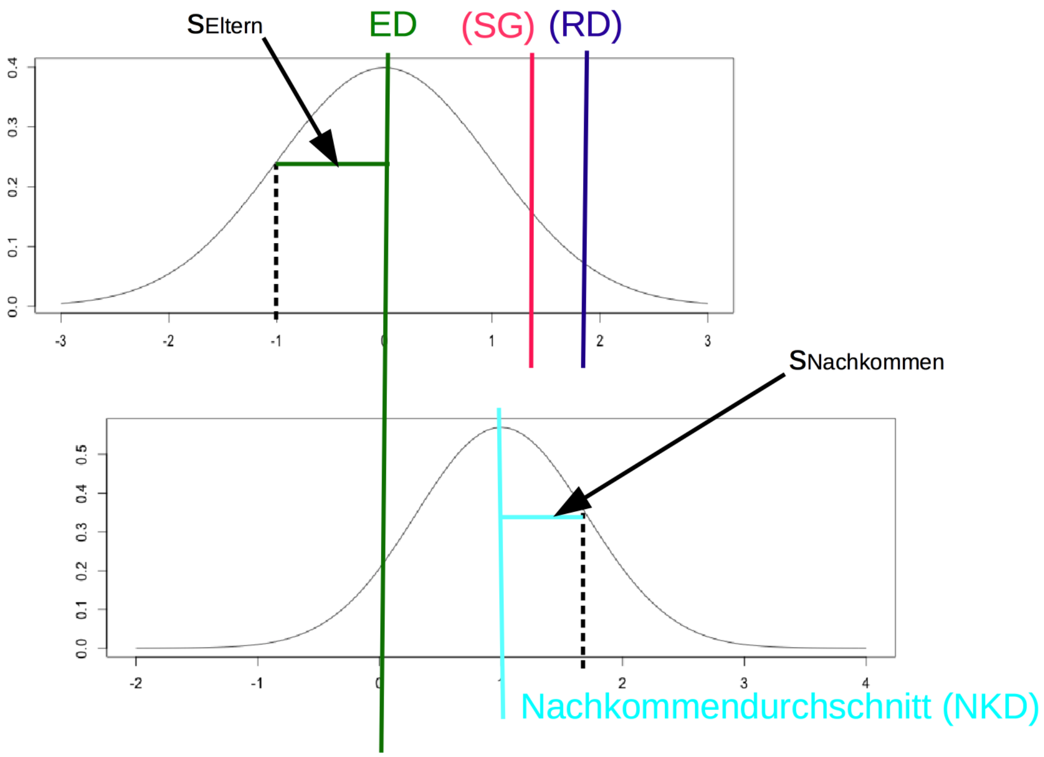
Population	Selektion	Anpaarung
Wildlife	Natural selection based on environment	random mating
Lifestock	artificial and directed selection	planned mating

b) In a selection scheme based on phenotypic observations, the distribution of the parents and the progeny is compared in the following diagram. Please put names on the labels in the following diagram.

12



Solution:



where

Abbreviation	Meaning
ED	parent average
SG	selection threshold
RD	average of selected parents
sEltern	standard deviation of parents
sNachkommen	standard deviation of progeny
NKD	average of progeny

c) In pig breeding the two meat quality traits **tenderness** (ZH) and **juiciness** (SH) should be considered in the aggregate genotype H. The economic values for the two traits are $w_{ZH} = 5$ und $w_{SH} = 1$. Because both traits in the aggregate genotype are difficult to measure, H is estimated using an index I which contains the traits **sheer force** (SK) and **drip loss** (SV) beinhaltet. What is the vector b of index weights which follows from selection index theory?

Assumptions

- Economic values w are given as

$$w = \left[\begin{array}{c} 5 \\ 1 \end{array}\right]$$

- The variance covariance matrix P between the traits `SK` and `SV` in the index is

$$P = \left[\begin{array}{cc} 4 & 0 \\ 0 & 10 \end{array}\right]$$

- The covariance matrix G between the traits in the index and in the aggregate genotype is

$$G = \left[\begin{array}{rr} 1.0 & -0.2 \\ 0.2 & 2.0 \end{array}\right]$$

Your Task

Compute the vector b of index weights.

Solution:

The vector b is computed base on the following index equations

$$Pb = Gw$$

hence we get

$$b = P^{-1}Gw$$

```
(mat_inv_P <- solve(matP))
```

```
##      [,1] [,2]
## [1,] 0.25  0.0
## [2,] 0.00  0.1
```

```
(mat_inv_P %*% matG)
```

```
##      [,1]  [,2]
## [1,] 0.25 -0.05
## [2,] 0.02  0.20
```

$$b = \left[\begin{array}{c} 1.20 \\ 0.30 \end{array}\right]$$

Problem 4 Inbreeding

a) The efficient computation of inbreeding in big pedigrees is baed on the Cholesky-Decomposition of the numerator relationship matrix A. Compute the matrix R which results from the Cholesky-Decomposition for the following pedigree.

25

```
##   sire  dam
## 1 <NA> <NA>
## 2 <NA> <NA>
## 3    2    1
## 4    2    1
## 5    3    4
```

Hint

The Cholesky-Decomposition of the matrix A is

$$A = R * R^T$$

Solution:

The matrix R corresponds to

```
(matR <- t(chol(as.matrix(getA(ped)))))
```

```
##     1   2         3         4         5
## 1 1.0 0.0 0.0000000 0.0000000 0.0000000
## 2 0.0 1.0 0.0000000 0.0000000 0.0000000
## 3 0.5 0.5 0.7071068 0.0000000 0.0000000
## 4 0.5 0.5 0.0000000 0.7071068 0.0000000
## 5 0.5 0.5 0.3535534 0.3535534 0.7071068
```

$$R = \left[\begin{array}{ccccc} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.5000 & 0.5000 & 0.7071 & 0.0000 & 0.0000 \\ 0.5000 & 0.5000 & 0.0000 & 0.7071 & 0.0000 \\ 0.5000 & 0.5000 & 0.3536 & 0.3536 & 0.7071 \end{array}\right]$$

b) Compute the inbreeding coefficients of the five animals in the pedigree of Problem 4a) based on the values in matrix R.

5

Solution:

The inbreeding coefficients are computed based on the sum of the squared values of each row of matrix R

```
(matR2 <- matR*matR)
```

```
##      1    2     3     4   5
## 1 1.00 0.00 0.000 0.000 0.0
## 2 0.00 1.00 0.000 0.000 0.0
## 3 0.25 0.25 0.500 0.000 0.0
## 4 0.25 0.25 0.000 0.500 0.0
## 5 0.25 0.25 0.125 0.125 0.5
```

```
(vec_inb <- apply(matR2, 1, sum) - 1)
```

```
##    1    2    3    4    5
## 0.00 0.00 0.00 0.00 0.25
```

We check with `pedigreemm()`

```
inbreeding(ped = ped)
```

```
## [1] 0.00 0.00 0.00 0.00 0.25
```

Problem 5 Prediction of Breeding Values

Breeding values should be predicted based on the following data set.

Animal	Herd	Observation
1	NA	NA
2	NA	NA
3	B	118
4	A	120
5	A	135
6	B	115

The variances can be taken from the following table.

Component	Value
phenotypic	32
additive genetic	8

a) Predict the breeding value of the 6 animals based on their own performance where the population mean μ corresponds to the average of the phenotypic observation

9

Solution:

The predicted breeding values based on the own performance can be computed as

$$\hat{a}_i = h^2(y_i - \mu)$$

Applying this to the data from the above table, this means

Animal	PBV
1	NA
2	NA
3	-1.00
4	-0.50
5	3.25
6	-1.75

b) Predict the breeding values for the animals in the table above using a BLUP animal model. Set up the model and the resulting mixed model equations. Transfer the information from the data into the model by filling the numeric values into the matrices and the vectors where possible.

16

Solution:

The model corresponds to:

$$y = Xb + Za + e$$

wobei

y	vector of all observations
b	vector of fixed herd effects
X	incidence matrix linking b and y
a	vector of random breeding values
Z	incidence matrix linking a with y
e	vector of random residuals

Mixed model equations

Coefficient matrix M

To set up the coefficient matrix, we need to know the incidence matrices X and Z.

```
(X <- matrix(data = c(0, 1,
                      1, 0,
                      1, 0,
                      0, 1), ncol = 2, byrow = TRUE))
```

```
##      [,1] [,2]
## [1,]    0    1
## [2,]    1    0
## [3,]    1    0
## [4,]    0    1
```

```
(Z <- cbind(matrix(data = 0, nrow = 4, ncol = 2), diag(4)))
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0    0    1    0    0    0
## [2,]    0    0    0    1    0    0
## [3,]    0    0    0    0    1    0
## [4,]    0    0    0    0    0    1
```

The four sub-matrices of M are

1. X^TX

```
(XTX <- t(X) %*% X)
```

```
##      [,1] [,2]
## [1,]    2    0
## [2,]    0    2
```

2. X^TZ

```
(XTZ <- t(X) %*% Z)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0    0    0    1    1    0
## [2,]    0    0    1    0    0    1
```

3. Z^TX

```
(ZTX <- t(XTZ))
```

```
##      [,1] [,2]
## [1,]    0    0
## [2,]    0    0
## [3,]    0    1
## [4,]    1    0
## [5,]    1    0
## [6,]    0    1
```

4. $Z^TZ + A^{-1}\alpha$

```
Ainv <- diag(nNrRecords)
(ZTZAinvAlpha <- t(Z) %*% Z + Ainv * alpha)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    3    0    0    0    0    0
## [2,]    0    3    0    0    0    0
## [3,]    0    0    4    0    0    0
## [4,]    0    0    0    4    0    0
## [5,]    0    0    0    0    4    0
## [6,]    0    0    0    0    0    4
```

Setting up of M

```
(M <- rbind(cbind(XTX, XTZ), cbind(ZTX, ZTZAinvAlpha)))
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    2    0    0    0    0    1    1    0
## [2,]    0    2    0    0    1    0    0    1
## [3,]    0    0    3    0    0    0    0    0
## [4,]    0    0    0    3    0    0    0    0
## [5,]    0    1    0    0    4    0    0    0
## [6,]    1    0    0    0    0    4    0    0
## [7,]    1    0    0    0    0    0    4    0
## [8,]    0    1    0    0    0    0    0    4
```

The right-hand side (rhs)

The vector y is obtained from the data

```
(y <- dfMlrData$Observation[!is.na(dfMlrData$Observation)])
```

```
## [1] 118 120 135 115
```

1. X^Ty

```
(XTy <- t(X) %*% y)
```

```
##      [,1]
## [1,]  255
## [2,]  233
```

2. Z^Ty

```
(ZTy <- t(Z) %*% y)
```

```
##      [,1]
## [1,]    0
## [2,]    0
## [3,]  118
## [4,]  120
## [5,]  135
## [6,]  115
```

Setting up rhs

```
(rhs <-  rbind(XTy,ZTy))
```

```
##      [,1]
## [1,]  255
## [2,]  233
## [3,]    0
## [4,]    0
## [5,]  118
## [6,]  120
## [7,]  135
## [8,]  115
```

Solution

The solution is obtained as $sol = M^{-1} * rhs$

```
(sol <- solve(M, rhs))
```

```
##         [,1]
## [1,] 127.500
## [2,] 116.500
## [3,]   0.000
## [4,]   0.000
## [5,]   0.375
## [6,]  -1.875
## [7,]   1.875
## [8,]  -0.375
```

c) What are the differences when comparing the result from Problem a) and b)?

4

Solution:

- In a) only animals with an own performance get a predicted breeding value. In b) all animals get a predicted breeding value.
- In a) the population mean is used to account for environmental effects, in b) also herd effects are accounted for.