ETH Zurich
D-USYS
Institute of Agricultural Sciences

Solutions To Post-Final Exam
Livestock Breeding and Genomics
FS 2018

Peter von Rohr

Date	*14. January 2019*
Begin	*09:15*
End	*11:15*

Name:

Legi-Nr:

Problem	Maximum Number of Points	Number of Points Reached
1	14	
2	19	
3	72	
4	17	
5	69	
Total	191	

Problem 1 Livestock Breeding

Sheep breeders are selecting male animals for rate of gain in body weight. After one year the rams with the highest phenotypic values of the trait `rate of gain in body weight` are selected as sires of the next generation. These sires are mated to the female animals randomly.

a) Do you think this is a good strategy? Please reason about your answer by providing facts from Livestock Breeding. Could you suggest some improvements to the breeding program of the sheep breeders.

8

Solution:

- The strategy is not very good, because parents do not pass phenotypes to their offspring but random samples of their alleles, hence selection should not be based on phenotypic values.
- The measurements are done only at two years of age, which is probably very late. It would be better to have observations earlier in the life of a selection candidate.
- Only male animals are measured and selected, the same should be done on the female side.
- Animals are related, this should be recorded as important information in a pedigree
- Breeding values should be predicted based on a BLUP animal model which uses all information in an optimal way
- Selection decisions should be based for males and females on the ranking according to the predicted breeding values

b) The sheep breeders are selecting for rate of body weight gain until all sires have a rate of gain larger than $0.8 kg$ per day. Then the breeders start to select for wool weight gain. What is the name of this selection strategy for selecting for two traits? Please specify two other selection strategies when multiple traits should be considered in a breeding program. Which strategy is the best when multiple traits should be selected for?

4

Solution:

The described strategy is called `tandem selection`

The two other strategies are

1. Selection based on independent selection thresholds
2. Selection based on dependent selection thresholds which corresponds to a selection index using multivariate BLUP breeding values

- The best strategy is selection based on dependent selection thresholds which corresponds to selection index.

c) The sheep breeders combine the traits ‘rate of weight gain‘ and ‘wool yield‘ into an aggregate genotype. The economic values of the two traits are given below. Let us assume that the two traits are combined into an index where the BLUP breeding values from an animal model are used as information sources. What are the index weights for such an selection index?

2

Trait	Economic Value
rate of weight gain	0.027
wool yield	0.500

Solution:

Because the same traits are used in the aggregate genotype and in the index and because the information sources used in the selection index are based on BLUP predictions, the index weights are the same as the econmic values. Hence

$$b = w = \left[\begin{array}{c} 0.027 \\ 0.50 \end{array}\right]$$

Problem 2 Own Performance and Progeny Records

Given is the following data set

Animal	Sire	Dam	Herd	Observation
4	1	2	1	15.68
5	3	2	1	-10.28
6	3	4	2	12.26
7	6	5	2	14.68

The variance components σ_e^2 (residual) and σ_a^2 (genetic) are

$$\sigma_e^2 = 72$$
$$\sigma_a^2 = 36$$

a) Predict the breeding values and the corresponding accuracies of the predicted breeding values based on the animals own performance. The population mean μ is taken as the average of all observations, $\mu = 8.08$.

8

Solution:

The predicted breeding value $\hat{a}_i$ based on an own performance is

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

where

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} = \frac{36}{36 + 72} = 0.3333333$$

The accuracies of all predicted breeding values correspond to the correlation $r_{a,y}$ between the true breeding value and the information source y.

$$r_{a,y} = \frac{cov(a, y)}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h = \sqrt{0.3333333} = 0.5773503$$

The results are shown in the following table

Animal	Predicted Breeding Value	Accuracy
4	2.532427	0.5773503
5	-6.120906	0.5773503
6	1.392427	0.5773503
7	2.199094	0.5773503

b) For all animals with a performance record some more observations could be collected resulting in the table of observations shown below. Predict the breeding values using the repeated observations, assuming that the repeatability t is 0.6. Indicate for each breeding value the corresponding accuracy.

8

Animal	Sire	Dam	Herd	Observation
4	1	2	1	2.34
4	1	2	1	5.29
4	1	2	1	17.99
5	3	2	1	2.47
5	3	2	1	-10.66
6	3	4	2	0.91
6	3	4	2	9.83
6	3	4	2	-2.37
7	6	5	2	5.07
7	6	5	2	9.03
7	6	5	2	4.70

Solution:

The predicted breeding value $\hat{a}_i$ based on the repeated performances is

$$\hat{a}_i = b(\tilde{y}_i - \mu)$$

where

$$b = \frac{nh^2}{1 + (n-1)t} \text{ and } \mu = 4.05$$

The term $\tilde{y}_i$ corresponds to the mean of the repeated measures of animal i.

First, we are counting the number of repetitions per animal and the mean of the repeated observations

Animal	Mean Observation	Number of Repetition
4	8.540000	3
5	-4.095000	2
6	2.790000	3
7	6.266667	3

Then predicted breeding values are added

Animal	Mean Observation	Number of Repetition	Predicted Breeding Value	Accuracy
4	8.540000	3	2.039	0.67
5	-4.095000	2	-3.396	0.65
6	2.790000	3	-0.575	0.67
7	6.266667	3	1.005	0.67

c) What is the reason for the difference between the accuracies found in Problem 2a and 2b? How many observations per animal are required, if the accuracy should be greater than 0.7 using the same repeatability as shown in 2b?

3

Solution:

- In 2a accuracies only depend on h or h^2.
- In 2b accuracies depend on the number of observations n per animal and on the repeatability.
- The relevant inequality is

$$\sqrt{\frac{nh^2}{1+(n-1)t}} \geq 0.7$$

Squaring both sides of the inequality leads to

$$\frac{nh^2}{1+(n-1)t} \geq 0.7^2$$

Then solving for n leads to

$$\begin{aligned} nh^2 &\geq 0.7^2(1+(n-1)t) = 0.7^2 + 0.7^2nt - 0.7^2t \\ nh^2 - 0.7^2nt &= n(h^2 - 0.7^2t) \geq 0.7^2(1-t) \end{aligned}$$

$$n \geq \frac{0.7^2(1-t)}{h^2 - 0.7^2t} = \frac{0.7^2(1-0.6)}{0.3333333 - 0.7^2 * 0.6} = 4.9830508$$

Hence we must have at least $n = 5$ observations per animal.

Problem 3 BLUP Animal Model

Given is the same data set as in Problem 2a.

Animal	Sire	Dam	Herd	Observation
4	1	2	1	15.68
5	3	2	1	-10.28
6	3	4	2	12.26
7	6	5	2	14.68

The variance components σ_e^2 (residual) and σ_a^2 (genetic) are

$$\sigma_e^2 = 72$$
$$\sigma_a^2 = 36$$

a) Predict the breeding values based on the data given above, using an animal model. Herds are treated as fixed effects and breeding values are modeled as random effects. Please specify all model components including expected values and variances for the random components. Set up the mixed model equations to get solutions for the predicted breeding values.

5

Solution:

Model components

The animal model is defined as

$$y = X\beta + Za + e$$

where		
	y	vector of length 4 of observations
	β	vector of length 2 of fixed herd effects
	X	design matrix linking fixed effects to observations
	a	vector of length 7 of animal breeding values
	Z	design matrix linking animal breeding values to observations
	e	vector of length 4 of random rest effects

The vectors of the model are

$$y = \begin{bmatrix} 15.68 \\ -10.28 \\ 12.26 \\ 14.68 \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{bmatrix}, e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

The design matrices

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, Z = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Expected Values and Variances of Random Components

$$E\left[\begin{array}{c} y \\ a \\ e \end{array}\right] = \left[\begin{array}{c} X\beta \\ 0 \\ 0 \end{array}\right]$$

$$var\left[\begin{array}{c} y \\ a \\ e \end{array}\right] = \left[\begin{array}{ccc} V & ZG & 0 \\ GZ^T & G & 0 \\ 0 & 0 & R \end{array}\right]$$

where $R = I * \sigma_e^2$, $G = A * \sigma_a^2$, $V = ZGZ^T + R$ and A is the numerator relationship matrix between the animals given below

$$A = \left[\begin{array}{ccccccc} 1.0000 & 0.0000 & 0.0000 & 0.5000 & 0.0000 & 0.2500 & 0.1250 \\ 0.0000 & 1.0000 & 0.0000 & 0.5000 & 0.5000 & 0.2500 & 0.3750 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.5000 & 0.5000 & 0.5000 \\ 0.5000 & 0.5000 & 0.0000 & 1.0000 & 0.2500 & 0.5000 & 0.3750 \\ 0.0000 & 0.5000 & 0.5000 & 0.2500 & 1.0000 & 0.3750 & 0.6875 \\ 0.2500 & 0.2500 & 0.5000 & 0.5000 & 0.3750 & 1.0000 & 0.6875 \\ 0.1250 & 0.3750 & 0.5000 & 0.3750 & 0.6875 & 0.6875 & 1.1875 \end{array}\right]$$

Mixed Model Equations

MME for the animal model are

$$\left[\begin{array}{cc} X^TX & X^TZ \\ Z^TX & Z^TZ + \lambda * A^{-1} \end{array}\right] \left[\begin{array}{c} \widehat{\beta} \\ \widehat{a} \end{array}\right] = \left[\begin{array}{c} X^Ty \\ Z^Ty \end{array}\right]$$

The single components of MME are

$$X^TX = \left[\begin{array}{cc} 2 & 0 \\ 0 & 2 \end{array}\right], X^TZ = \left[\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right], Z^TZ = \left[\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right], \lambda = \frac{\sigma_e^2}{\sigma_s^2} = \frac{72}{36} = 2$$

$$A^{-1} = \left[\begin{array}{rrrrrrr} 1.5000 & 0.5000 & 0.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.5000 & 2.0000 & 0.5000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.5000 & 2.0000 & 0.5000 & -1.0000 & -1.0000 & 0.0000 \\ -1.0000 & -1.0000 & 0.5000 & 2.5000 & 0.0000 & -1.0000 & 0.0000 \\ 0.0000 & -1.0000 & -1.0000 & 0.0000 & 2.5000 & 0.5000 & -1.0000 \\ 0.0000 & 0.0000 & -1.0000 & -1.0000 & 0.5000 & 2.5000 & -1.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1.0000 & -1.0000 & 2.0000 \end{array}\right]$$

Putting all mixed model equations together leads to

$$\begin{bmatrix} 2.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 2.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 \\ 0.00 & 0.00 & 3.00 & 1.00 & 0.00 & -2.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 4.00 & 1.00 & -2.00 & -2.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 4.00 & 1.00 & -2.00 & -2.00 & 0.00 \\ 1.00 & 0.00 & -2.00 & -2.00 & 1.00 & 6.00 & 0.00 & -2.00 & 0.00 \\ 1.00 & 0.00 & 0.00 & -2.00 & -2.00 & 0.00 & 6.00 & 1.00 & -2.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & -2.00 & -2.00 & 1.00 & 6.00 & -2.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & -2.00 & -2.00 & 5.00 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{a}_1 \\ \widehat{a}_2 \\ \widehat{a}_3 \\ \widehat{a}_4 \\ \widehat{a}_5 \\ \widehat{a}_6 \\ \widehat{a}_7 \end{bmatrix} = \begin{bmatrix} 5.40 \\ 26.94 \\ 0.00 \\ 0.00 \\ 0.00 \\ 15.68 \\ -10.28 \\ 12.26 \\ 14.68 \end{bmatrix}$$

The solution

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{a}_1 \\ \widehat{a}_2 \\ \widehat{a}_3 \\ \widehat{a}_4 \\ \widehat{a}_5 \\ \widehat{a}_6 \\ \widehat{a}_7 \end{bmatrix} = \begin{bmatrix} 2.6120 \\ 13.8315 \\ 2.2800 \\ 0.1173 \\ -2.3973 \\ 3.4787 \\ -3.3026 \\ 0.3060 \\ -1.0290 \end{bmatrix}$$

b) Compute the 95% confidence interval of the predicted breeding values from 3a

21

Solution:

The confidence interval is based on the standard error of prediction (SEP) which is the square-root of the prediction error variance (PEV). PEV depends on the inverse of the coefficient matrix of the mixed model equations. The confidence interval is based on the assumption of the predicted breeding values being normal distributed around the current predictions with SEP corresponding to the standard deviation. The value α is given in the task as

$$\alpha = 1 - 0.95 = 0.05$$

Animal	PBV	SEP	Lower Bound	Upper Bound
1	2.2799941	5.857380	-9.200261	13.760249
2	0.1173421	5.990181	-11.623197	11.857881
3	-2.3973361	5.861034	-13.884751	9.090079
4	3.4786621	5.682078	-7.658006	14.615331
5	-3.3026490	5.642383	-14.361515	7.756217
6	0.3059789	5.933630	-11.323723	11.935681
7	-1.0289669	6.356862	-13.488187	11.430253

Problem 4 Quantitative Genetics

We are given a dataset where a trait of interest is mostly determined by a single genetic locus. The locus is bi-allelic with alleles G_1 and G_2. G_1 is the favorite allele which increases the enzyme level and it has a frequency of 0.2. For reasons of simplicity, we assume that the Hardy-Weinberg equilibrium holds.

We take the mean levels of the trait values per genotype are taken as the characteristic values for a given genotype. This results in

Genotype	Mean Level
G_1G_1	11.23
G_1G_2	10.14
G_2G_2	7.66

After re-scaling the levels for each genotype such that the origin is exactly between the two homozygous genotypes G_1G_1 and G_2G_2, we get the following genotypic values for the three genotypes.

Genotype	Genotypic Value
G_1G_1	1.785
G_1G_2	0.695
G_2G_2	-1.785

a) Assuming a single locus model, what are the values of a and d for the trait levels at locus G and what is the population mean μ?

5

Solution:

The genotypic values given in the last table of the problem corresponds to the following symbolic values

Genotype	Genotypic Value
G_1G_1	$V_{11} = a$
G_1G_2	$V_{12} = d$
G_2G_2	$V_{22} = -a$

Hence

$$a = 1.785$$

and

$$d = 0.695$$

The population mean is defined as

$$\mu = V_{11} * f(G_1G_1) + V_{12} * f(G_1G_2) + V_{22} * f(G_2G_2)$$

where $f(G_iG_j)$ stands for the frequency of genotype G_iG_j. From Hardy-Weinberg we know that

Genotype	Genotype Frequency
G_1G_1	$f(G_1G_1) = p^2$
G_1G_2	$f(G_1G_2) = 2pq$
G_2G_2	$f(G_2G_2) = q^2$

where $p = f(G_1)$ the allele frequency of G_1 which is given in the task and is $p = 0.2$. Because the locus G is bi-allelic, we get $q = f(G_2) = 1 - p = 0.8$.

Using all this, we get the population mean μ as

$$\mu = a * p^2 + d * 2pq + (-a) * q^2 = (p - q)a + 2pqd = (0.2 - 0.8) * 1.785 + 2 * 0.2 * 0.8 * 0.695 = -0.8486$$

b) Compute the breeding values and the dominance deviations for all three genotypes

6

Solution:

All breeding values depend on the allele substitution effect α which corresponds to

$$\alpha = a + (q - p)d = 1.785 + (0.8 - 0.2) * 0.695 = 2.202$$

Breeding values are computed according to the following table

Genotype	Breeding Value
G_1G_1	$2q\alpha = 3.5232$
G_1G_2	$(q - p)\alpha = 1.3212$
G_2G_2	$-2p\alpha = -0.8808$

The dominance deviation is

Genotype	Dominance Deviation
G_1G_1	$-2q^2d = -0.889600000000001$
G_1G_2	$2pqd = 0.2224$
G_2G_2	$-2p^2d = -0.0556$

c) Compute the genetic additive variance and the dominance variance.

2

Solution:

The additive genetic variance corresponds to

$$\sigma_A^2 = 2pq\alpha^2 = 2 * 0.2 * 0.8 * 2.202^2 = 1.5516173$$

$$\sigma_D^2 = (2pqd)^2 = (2 * 0.2 * 0.8 * 0.695)^2 = 0.0494618$$

d) When two heterozygous parents have offspring, these offspring individuals can have genotypes G_1G_1, G_1G_2 or G_2G_2. When comparing the breeding values of the offspring with different genotypes, we realize that the breeding values are different. Compute all pairwise differences between the breeding values of all offspring genotypes. What is the name of the specific effect that appears in the differences of the breeding values?

4

Solution:

$$\begin{aligned}
BV_{11} - BV_{12} &= 2q\alpha - (q-p)\alpha = \alpha \\
BV_{12} - BV_{22} &= (q-p)\alpha - (-2p\alpha) = \alpha \\
BV_{11} - BV_{22} &= 2q\alpha - (-2p\alpha) = 2\alpha
\end{aligned}$$

The effect α appears in all differences and is called **allele substitution** effect.

Problem 5 Relationship and Inbreeding

Given is the following pedigree

Animal	Sire	Dam
ANTONIA	SCHAERZ	PHILADELPHIA
PHILADELPHIA	PAUER	VERA
VERA	VITGOR	PRISCA
PRISCA	HARRY	REIA

a) Compute the numerator relationship matrix for the above given pedigree.

48

Solution:

First we have to re-order the pedigree such that parents are always before progeny. This results in

Animal	Sire	Dam
PRISCA	HARRY	REIA
VERA	VITGOR	PRISCA
PHILADELPHIA	PAUER	VERA
ANTONIA	SCHAERZ	PHILADELPHIA

Recoding the pedigree to numers with the following mapping

Name	Code
HARRY	1
REIA	2
VITGOR	3
PAUER	4
SCHAERZ	5
PRISCA	6
VERA	7
PHILADELPHIA	8
ANTONIA	9

Applying the above shown mapping

Animal	Sire	Dam
6	1	2
7	3	6
8	4	7
9	5	8

Extending the pedigree with founders

Animal	Sire	Dam
1	NA	NA

Animal	Sire	Dam
2	NA	NA
3	NA	NA
4	NA	NA
5	NA	NA
6	1	2
7	3	6
8	4	7
9	5	8

This pedigree can be used to compute the numerator relationship matrix

$$A = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.2500 & 0.1250 & 0.0625 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.2500 & 0.1250 & 0.0625 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.2500 & 0.1250 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.2500 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 \\ 0.5000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.5000 & 0.2500 & 0.1250 \\ 0.2500 & 0.2500 & 0.5000 & 0.0000 & 0.0000 & 0.5000 & 1.0000 & 0.5000 & 0.2500 \\ 0.1250 & 0.1250 & 0.2500 & 0.5000 & 0.0000 & 0.2500 & 0.5000 & 1.0000 & 0.5000 \\ 0.0625 & 0.0625 & 0.1250 & 0.2500 & 0.5000 & 0.1250 & 0.2500 & 0.5000 & 1.0000 \end{bmatrix}$$

b) Compute the inverse numerator relationship matrix A^{-1} from the above given pedigree.

21

Solution:

$$A^{-1} = \left[\begin{array}{rrrrrrrrr}
1.5000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.5000 & 1.5000 & 0.0000 & 0.0000 & 0.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 1.5000 & 0.0000 & 0.0000 & 0.5000 & -1.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.5000 & 0.0000 & 0.0000 & 0.5000 & -1.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.5000 & 0.0000 & 0.0000 & 0.5000 & -1.0000 \\
-1.0000 & -1.0000 & 0.5000 & 0.0000 & 0.0000 & 2.5000 & -1.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & -1.0000 & 0.5000 & 0.0000 & -1.0000 & 2.5000 & -1.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & -1.0000 & 0.5000 & 0.0000 & -1.0000 & 2.5000 & -1.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & -1.0000 & 0.0000 & 0.0000 & -1.0000 & 2.0000
\end{array}\right]$$