Solutions To Exam

Livestock Breeding and Genomics

FS 2020

Peter von Rohr

Date	18. December 2020
Begin	09:15
End	11:15

Name:

Legi-Nr:

Problem	Maximum Number of Points	Number of Points Reached
1	45	
2	40	
3	18	
4	16	
5	10	
Total	129	

Problem 1 Numerator Relationship Matrix and Inbreeding

We are given the following pedigree.

Das folgende Pedigree ist gegeben.

ID	Sex	Sire	Dam
5	M	3	2
6	F	1	4
7	M	5	6
8	F	5	6
9	M	7	8
10	M	9	8

a) Compute the additive genetic relationship matrix A for the above pedigree.

   Berechnen Sie die additiv-genetische Verwandtschaftsmatrix A für das oben angegebene Pedigree

20

Solution:

The numerator relationship matrix is computed using `pedigreemm::getA()`. In a first step, we have to extend the pedigree to contain the founder animals in the ID-column.

```r
vec_founder_sire <- setdiff(tbl_ped_nrm$Sire, tbl_ped_nrm$ID)
n_nr_founder_sire <- length(vec_founder_sire)
vec_founder_dam <- setdiff(tbl_ped_nrm$Dam, tbl_ped_nrm$ID)
n_nr_founder_dam <- length(vec_founder_dam)

# check that founder_sire and founder_dam are not the same
if (length(intersect(vec_founder_sire, vec_founder_dam)) != 0)
  stop(" * ERROR: Founder sires and founder dams are not exclusive")
tbl_ped_nrm_ext <- dplyr::bind_rows(
  tibble::tibble(ID = vec_founder_sire[order(vec_founder_sire)],
                 Sex = rep('M', n_nr_founder_sire),
                 Sire = rep(NA, n_nr_founder_sire),
                 Dam  = rep(NA, n_nr_founder_sire)),
  tibble::tibble(ID = vec_founder_dam[order(vec_founder_dam)],
                 Sex = rep('F', n_nr_founder_dam),
                 Sire = rep(NA, n_nr_founder_dam),
                 Dam  = rep(NA, n_nr_founder_dam)),
  tbl_ped_nrm)

ped_nrm <- pedigreemm::pedigree(sire = tbl_ped_nrm_ext$Sire,
                                dam = tbl_ped_nrm_ext$Dam,
```

```
                                        label = as.character(tbl_ped_nrm_ext$ID))
(matA_nrm <- as.matrix(pedigreemm::getA(ped = ped_nrm)))
```

```
##       1    3    2    4   5   6     7     8    9     10
## 1  1.00 0.00 0.00 0.00 0.0 0.5 0.250 0.250 0.25 0.250
## 3  0.00 1.00 0.00 0.00 0.5 0.0 0.250 0.250 0.25 0.250
## 2  0.00 0.00 1.00 0.00 0.5 0.0 0.250 0.250 0.25 0.250
## 4  0.00 0.00 0.00 1.00 0.0 0.5 0.250 0.250 0.25 0.250
## 5  0.00 0.50 0.50 0.00 1.0 0.0 0.500 0.500 0.50 0.500
## 6  0.50 0.00 0.00 0.50 0.0 1.0 0.500 0.500 0.50 0.500
## 7  0.25 0.25 0.25 0.25 0.5 0.5 1.000 0.500 0.75 0.625
## 8  0.25 0.25 0.25 0.25 0.5 0.5 0.500 1.000 0.75 0.875
## 9  0.25 0.25 0.25 0.25 0.5 0.5 0.750 0.750 1.25 1.000
## 10 0.25 0.25 0.25 0.25 0.5 0.5 0.625 0.875 1.00 1.375
```

∎

b) Compute the inbreeding coefficients of all animals in the given pedigree. Complete the following table and indicate which of the animals are inbred.

Berechnen Sie die Inzuchtkoeffizienten aller Tiere im gegebenen Pedigree. Vervollständigen Sie die folgende Tabelle und geben an, welche Tiere ingezüchtet sind.

20

ID	Inbreeding Coefficient	Animal Inbred (TRUE/FALSE)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Solution:

The numeric solution is

```r
vec_inbr_yn <- ifelse(diag(matA_nrm) > 1, "TRUE", "FALSE")
vec_inbr_coef <- diag(matA_nrm) - 1
# tibble for table
(tbl_inb_sol <- tibble::tibble(ID = tbl_ped_nrm_ext$ID,
                               `Inbreeding Coefficient` = round(vec_inbr_coef, digits = 3),
                               `Animal Inbred (TRUE/FALSE)` = vec_inbr_yn))
```

```
## # A tibble: 10 x 3
##       ID `Inbreeding Coefficient` `Animal Inbred (TRUE/FALSE)`
##    <dbl>                    <dbl> <chr>
## 1     1                        0  FALSE
## 2     3                        0  FALSE
## 3     2                        0  FALSE
## 4     4                        0  FALSE
## 5     5                        0  FALSE
## 6     6                        0  FALSE
## 7     7                        0  FALSE
## 8     8                        0  FALSE
## 9     9                     0.25  TRUE
```

4

```
## 10     10                          0.375 TRUE
```

■

c) Assume that dam 8 and sire 9 are mated. What is the inbreeding coefficient of their offspring?

Wir nehmen an, dass die Mutter 8 mit dem Vater 9 angepaart wird. Wie gross ist der Inzuchtkoeffizient des Nachkommens aus dieser Paarung?

5

Solution:

The inbreeding coefficient of the offspring of dam 8 and sire 9 is half of the relationship between the parents.

```
0.5 * matA_nrm[sire_id, dam_id]
```

```
## [1] 0.375
```

∎

Problem 2 Prediction of Breeding Values

The following dataset is used to predict breeding values.

Der folgende Datensatz soll für die Schätzung von Zuchtwerten verwendet werden.

ID	Sire	Dam	Herd	Phenotypic Observation
4	1	2	Planta	7.53
5	3	2	Moos	8.48
6	4	5	Moos	0.26
7	4	5	Moos	6.60
8	6	7	Moos	2.44

Assumptions

- the error variance is: $\sigma_e^2 = 28.8$
- the heritability is: $h^2 = 0.2$
- the genetic additive variance is: $\sigma_u^2 = 7.2$
- the population mean μ can be taken as the mean of the given phenotypic observations

Annahmen

- *die Restvarianz beträgt: $\sigma_e^2 = 28.8$*
- *die Heritabilität beträgt: $h^2 = 0.2$*
- *die genetisch-additive Varianz beträgt: $\sigma_u^2 = 7.2$*
- *das Populationsmittel μ kann als Mittelwert der phänotypischen Beobachtungen angenommen werden.*

a) Use the regression method to predict breeding values based on own-performance records for the animals in the table given above.

   Verwenden Sie die Regressionsmethode zur Schätzung der Zuchtwerte basierend auf der Eigenleistung der Tiere in der obigen Tabelle.

10

Solution:

According to the regression method, the predicted breeding value ($\hat{u}_i$) for animal i is

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

where y_i is the phenotypic observation of animal i, h^2 is the heritability and μ is the population mean.

```
vec_y <- tbl_data_pbv$`Phenotypic Observation`
n_mu <- mean(vec_y)
vec_uhat <- n_h2 * (vec_y - n_mu)
```

ID	PBV
4	0.4936
5	0.6836
6	-0.9604
7	0.3076
8	-0.5244

∎

b) Use a BLUP animal model to predict the breeding values for all animals in the pedigree based on the data given in the table above. Specify the model to predict breeding values, name all model components, compute the expected values and the variance-covariance matrices for all random model components. Insert the information from the above table into the model components where possible. Set up the mixed model equations and compute the solutions for the estimates of fixed effects and for the predicted breeding values.

Verwenden Sie ein BLUP Tiermodell zur Schätzung der Zuchtwerte aller Tiere im Pedigree basierend auf den Daten in der obigen Tabelle. Spezifizieren Sie das Modell für die Schätzung der Zuchtwerte, benennen Sie alle Modellkomponenten, berechnen Sie die Erwartungswerte und die Varianz-Kovarianz Matrizen aller zufälligen Effekte im Modell. Setzen Sie die verfügbaren Information aus dem Datensatz in die Modellkomponenten ein. Stellen Sie die Mischmodellgleichungen auf und berechnen Sie die Schätzungen der fixen Effekte und der Zuchtwerte.

30

Solution:

Model

The BLUP animal model corresponds to the following mixed-effects model

$$y = Xb + Zu + e$$

where y is the vector of observations, b is the vector of fixed effects for the two herds, u is the vector of breeding values for all animals in the pedigree and e is the vector of random residuals. The design matrices X and Z link the fixed effects and the breeding values to the observations, respectively.

Expected Values and Variance-Covariance Matrices

Expected values and variance-covariance matrices of the random components y, u and e

$$E \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \end{bmatrix}, \; var \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} ZGZ^T + R & ZG & 0 \\ GZ^T & G & 0 \\ 0 & 0 & R \end{bmatrix}$$

Informations in Model Components

The model vectors are

$$y = \begin{bmatrix} 7.53 \\ 8.48 \\ 0.26 \\ 6.6 \\ 2.44 \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \; u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \end{bmatrix}, \; e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

The model matrices

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

Mixed Model Equations

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} * \lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

Solution

```
# nrm A
tbl_ped_pbv_ext <- rmdexam::extend_pedigree(ptbl_ped =
                  tbl_data_pbv[,c('ID', 'Sire', 'Dam')])
ped_pbv <- pedigreemm::pedigree(sire  = tbl_ped_pbv_ext$Sire,
                                dam   = tbl_ped_pbv_ext$Dam,
                                label = as.character(tbl_ped_pbv_ext$ID))
mat_Ainv_pbv <- as.matrix(pedigreemm::getAInv(ped = ped_pbv))
lambda <- n_var_e / n_var_u

# coefficient matrix
mat_xtx <- crossprod(mat_X)
mat_xtz <- crossprod(mat_X, mat_Z)
mat_ztx <- crossprod(mat_Z, mat_X)
mat_ztz_ainv_lambda <- crossprod(mat_Z) + lambda * mat_Ainv_pbv
mat_coef <- rbind(cbind(mat_xtx,mat_xtz), cbind(mat_ztx,mat_ztz_ainv_lambda))

# rhs
mat_xty <- crossprod(mat_X, vec_y)
```

```
mat_zty <- crossprod(mat_Z, vec_y)
mat_rhs <- rbind(mat_xty, mat_zty)

# solutions
(mat_sol <- solve(mat_coef, mat_rhs))
```

```
##             [,1]
##     4.575806e+00
##     7.864645e+00
## 2 -2.567391e-16
## 1 -2.230968e-01
## 3  2.230968e-01
## 4 -3.346452e-01
## 5  3.346452e-01
## 6 -5.753190e-01
## 7  1.291254e-01
## 8 -4.116737e-01
```

■

Problem 3 Genomics

Given is the following data set of SNP-Genotyping results.

Gegeben sind die Genotypisierungsresultate in der nachfolgenden Tabelle.

Animal	SNPLGH	SNPFS2	Observation
1	-1	0	12.7
2	1	1	46.0
3	1	1	32.7
4	0	0	19.3
5	-1	0	14.8
6	0	-1	6.7
7	-1	-1	2.4
8	1	0	33.0
9	1	0	29.4
10	0	-1	4.9
11	1	-1	14.4
12	0	0	19.5
13	1	0	25.6
14	-1	1	19.1
15	1	0	25.6

a) Use a marker effect model to estimate the fixed effects of both markers on the observation. Please specify the fixed-effect model that you use, name all the model components and insert the information from the data into the components where possible.

   Verwenden Sie eine Marker-Effekt Modell zur Schätzung der fixen Effekte der beiden Marker auf die Beobachtung. Bitte spezifizieren Sie das fixe Modell, benennen Sie alle Modellkomponenten und fügen Sie die Daten aus der obigen Tabelle in die Modellkomponenten ein, wo das möglich ist.

12

Solution:

Model

The fixed effect model to estimate the marker effects is given by

$$y = Xb + e$$

where y is the vector of observations, b is the vector of fixed effects containing intercept and effect of both markers, e is the vector of random error terms.

Information in Model

The model vectors are

$$
y = \begin{bmatrix} 12.7 \\ 46 \\ 32.7 \\ 19.3 \\ 14.8 \\ 6.7 \\ 2.4 \\ 33 \\ 29.4 \\ 4.9 \\ 14.4 \\ 19.5 \\ 25.6 \\ 19.1 \\ 25.6 \end{bmatrix}, \ b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \ e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \end{bmatrix}
$$

Marker Effects

The marker effects are estimated using `lm()`

```
vec_mrk_names <- grep('SNP', colnames(tbl_data_gnm), value = TRUE)
lm_mrk_eff <- lm(as.formula(paste0('Observation ~ ',
              paste0(vec_mrk_names, collapse = ' + '), collapse = '')),
              data = tbl_data_gnm)
summary(lm_mrk_eff)
```

```
##
## Call:
## lm(formula = as.formula(paste0("Observation ~ ", paste0(vec_mrk_names,
##     collapse = " + "), collapse = "")), data = tbl_data_gnm)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4098 -1.9313 -0.2904  2.0989  6.8902
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.590      1.028  19.060 2.45e-10 ***
## SNPLGH         7.941      1.202   6.604 2.52e-05 ***
## SNPFS2        11.578      1.473   7.863 4.49e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

13

```
## Residual standard error: 3.839 on 12 degrees of freedom
## Multiple R-squared:  0.9108, Adjusted R-squared:  0.896
## F-statistic: 61.28 on 2 and 12 DF,  p-value: 5.03e-07
```

∎

b) Predict the direct genomic breeding values for all animals of the dataset using the marker effects estimated in Task a).

Schätzen Sie die direkt-genomischen Zuchtwerte für alle Tiere im Datensatz unter Verwendung der aus Aufgabe a) geschätzten Markereffekte.

6

Solution:

Direct genomic breeding values $\hat{g}_i$ for animal i corresponds to the sums of the marker effects, hence

$$\hat{g}_i = x_i^T \cdot \hat{b}$$

where $\hat{b}$ is the vector of estimated marker effects and x_i^T is the row i of the design matrix X corresponding to animal i.

```
vec_mrk_eff <- coefficients(lm_mrk_eff)
n_intercept_pos <- which(names(vec_mrk_eff) == '(Intercept)')
if (length(n_intercept_pos) > 0) vec_mrk_eff <- vec_mrk_eff[-n_intercept_pos]
vec_snp_names <- names(vec_mrk_eff)
mat_geno_gnm <- as.matrix(tbl_data_gnm[, vec_snp_names])
(vec_hat_g <- crossprod(t(mat_geno_gnm), vec_mrk_eff))
```

```
##              [,1]
##  [1,]  -7.940943
##  [2,]  19.519434
##  [3,]  19.519434
##  [4,]   0.000000
##  [5,]  -7.940943
##  [6,] -11.578491
##  [7,] -19.519434
##  [8,]   7.940943
##  [9,]   7.940943
## [10,] -11.578491
## [11,]  -3.637547
## [12,]   0.000000
## [13,]   7.940943
## [14,]   3.637547
## [15,]   7.940943
```

∎

Problem 4 Variance and Inbreeding

In dairy cattle semen and embryos of the breeds Brown Swiss and Holstein are often imported from North America. For the Brown Swiss breed, the North American population is based on 280 female animals. The following assumptions can be made.

- Although, cattle does not follow the same mode of inheritance as the organism shown in the lecture, the computations as shown in the lecture can be used as an approximation.
- In contrast to the size N of the subpopulations that was assumed to be the number of individuals, here N is the number of female animals in a given subpopulation.

Samen und Embryos der Rassen Brown Swiss und Holstein werden oft aus Nordamerika importiert. Für die Rasse Brown Swiss basiert die Nordamerikanische Population auf 280 weiblichen Tieren. Die folgenden Annahmen können getroffen werden.

- *Obwohl das Rind nicht den gleichen Vererbungsmodus zeigt, wie die Organismen, welche in den Vorlesungsunterlagen verwendet wurden, können die dort eingeführten Berechnungen der Inzucht als Annäherungen verwendet werden.*
- *Die Populationsgrösse N hier entspricht der Anzahl weiblichen Tiere.*

a) What is the expected level of inbreeding (F) of imported semen in the Brown Swiss breed after 16 generations?

   Welches ist der erwartet Wert der Inzucht (F) von importiertem Samen in der Rasse Brown Swiss nach 16 Generationen?

4

Solution:

The level of inbreeding F_t after t generations is computed as

$$F_t = 1 - (1 - \Delta F)^t$$

with $\Delta F = 1/(2N)$ and N being the number of female animals in the base population.

```
(inb_level_ft <- 1 - (1 - 1/(2*n_nr_base_animals))^n_nr_gen)
```

```
## [1] 0.02819195
```

∎

b) Compute the between-line, the within-line and the total genetic variance for a single additive Locus where the difference between the homozygous genotypes is 50, the allele frequency $p = 0.2$ and the level of inbreeding is 0.01.

Berechnen Sie die Innerhalb-Linie, Zwischen-Linie und die totale genetische Varianz für einen additiven Lokus, wobei die Differenz der homozygoten Genotypen 50 entspricht, die Allelefrequenz $p = 0.2$ ist und der Inzuchtwert 0.01 ist.

10

Solution:

The variance according to the sources are computed as

Source	Variance
Between lines	$2FV_G$
Within lines	$(1-F)V_G$
Total	$(1+F)V_G$

with

$$V_G = 2pqa^2 = 2 * 0.2 * 0.8 * 625 = 200$$

Inserting the values, we get

Source	Variance	Results
Between lines	$2FV_G$	4
Within lines	$(1-F)V_G$	198
Total	$(1+F)V_G$	202

∎

c) After how many generations is the expected inbreeding depression bigger than 0.2 in a population with $N = 280$ animals. The following assumptions can be made

Nach wie vielen Generationen ist die erwartete Inzuchtdepression grösser als 0.2 in einer Population von $N = 280$ Tieren. Die folgenden Annahmen können getroffen werden.

2

- single bi-allelic locus
- minor allele frequency $p = 0.2$
- dominance deviation $d = 30$
- *einzelner Locus mit zwei Allelen*
- *Frequenz des seltenen Alleles $p = 0.2$*
- *Dominanzabweichung $d = 30$*

Solution:

Inbreeding depression is computed as

$$M_0 - M_F = 2dp(1 - p)F = 0.2$$

hence

$$F = \frac{M_0 - M_F}{2dp(1 - p)} = \frac{0.2}{2 * 30 * 0.2 * (1 - 0.2)} = 0.0208333$$

From

$$F_t = 1 - (1 - \Delta F)^t$$

we get

$$(1 - \Delta F)^t = 1 - F_t$$

and

$$t = \frac{log(1 - F_t)}{log(1 - \Delta F)} = 11.7793793$$

Hence after 12 generations the inbreeding depression is bigger than 0.2

■

Problem 5 Variance Components

We are given the following dataset for the trait live weight (`LiveWeight`) for cattle.

Der folgende Datensatz umfasst das Merkmal Lebendgewicht ('LiveWeight') von Rindern.

Animal	Farm	LiveWeight
1	3	613
2	1	621
3	2	630
4	3	614
5	1	629
6	2	611
7	3	612
8	1	614
9	2	606
10	1	621
11	2	621
12	1	623
13	2	608
14	2	603
15	1	589
16	1	599
17	1	610
18	2	595
19	3	612
20	3	616

a) Compute the estimate of the error variance σ_e^2 from the residuals of the fixed linear model specified below.

Schätzen Sie die Fehlervarianz σ_e^2 basierend auf den Residuen des folgenden fixen Modells.

5

The fixed linear model that is used is

$$y = Xb + e$$

where y is the vector of all live weight values, b is the vector of the effects caused by the different farms. The fixed linear model is specified in R by

```
tbl_data_anova$Farm <- as.factor(tbl_data_anova$Farm)
lm_lweight <- lm(LiveWeight ~ 0 + Farm, data = tbl_data_anova)
```

The resulting effects of the farms are

```
(vec_coef_lweight <- coefficients(lm_lweight))
```

```
##    Farm1    Farm2    Farm3
## 613.2500 610.5714 613.4000
```

Solution:

The esimate of the error variance is computed based on the resiudals. The residuals can be obtained by the function `residuals()` in R.

```
vec_res <- residuals(lm_lweight)
ssq_res <- crossprod(vec_res)
(n_est_res_var <- ssq_res / (nrow(tbl_data_anova)-length(vec_coef_lweight)))
```

```
##          [,1]
## [1,] 122.9655
```

The error standard deviation is

```
(n_est_res_sd <- sqrt(n_est_res_var))
```

```
##          [,1]
## [1,] 11.08898
```

∎

b) Verify your result from task a) with the output of the 'summary()'-function applied to the result of the 'lm()'-function

Verifizieren Sie das Resultat aus Aufgabe a) anhand des Outputs der 'summary'-Funktion angewendet auf das Resultat der 'lm()'-Funktion

5

Solution:

From the task, we have the result object of the `lm()`-function which is called `lm_lweight`. Applying the `summary()`-method leads to

```
summary(lm_lweight)
```

```
##
## Call:
## lm(formula = LiveWeight ~ 0 + Farm, data = tbl_data_anova)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.2500  -3.5804   0.0143   7.7500  19.4286
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## Farm1   613.250      3.921   156.4   <2e-16 ***
## Farm2   610.571      4.191   145.7   <2e-16 ***
## Farm3   613.400      4.959   123.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.09 on 17 degrees of freedom
## Multiple R-squared:  0.9997, Adjusted R-squared:  0.9997
## F-statistic: 2.033e+04 on 3 and 17 DF,  p-value: < 2.2e-16
```

The number next to `Residual standard error:` corresponds to the estimated value of the error standard deviation.

∎