Peter von Rohr
Institute of Agricultural Sciences
D-USYS
ETH Zurich

751-6212-00L V
Solutions to Exam
Applied Genetic Evaluation
SS 2020

Date: 25th May 2020

Name: Firstname Name

Legi-Nr: LegiNr

Problem	Maximum Number of Points	Number of Points Reached
1	24	
2	5	
3	16	
4	17	
Total	62	

Questions in German are in italics

Problem 1: Breeding Program

a) What are the constituent parts of a breeding program?

Wie lauten die Bestandteile eines Zuchtprogramms?

6

Solution

1. Breeding goal
2. Performance Tests
3. Prediction of Breeding Values
4. Reproduction Technologies
5. Selection and Mating
6. Selection Response

b) The following table lists the characteristic variables of two breeding programs. Which quantity do you use to compare the success of two breeding programs? Which of the two breeding programs shown below is better according to the quantity of comparison?

Die nachfolgende Tabelle zeigt die charakteristischen Kennzahlen von zwei Zuchtprogrammen. Welche Grösse wird verwendet um den Erfolg von Zuchtprogrammen zu vergleichen? Welches der beiden unten beschriebenen Zuchtprogramme ist erfolgreicher gemessen an der verwendeten Vergleichsgrösse?

4

Variable	Breeding Program 1	Breeding Program 2
Selection intensity	1.63	1.63
Accuracy of predicted breeding values	0.89	0.59
Genetic standard deviation	262.00	262.00
Generationinterval (in years)	10.00	2.00

Solution

The comparison is done based on `selection response` per year R which is defined as

$$R = \frac{i * r_{u,\hat{u}} * \sigma_g}{L}$$

- Breeding Program 1:

```
sel_r1 <- sel_int * acc_bv1 * gen_sd / gen_int1
cat(" * Selection response for breeding program 1: ", sel_r1, "\n")
```

```
##  * Selection response for breeding program 1:  38.00834
```

*Breeding Program 2:

```
sel_r2 <- sel_int * acc_bv2 * gen_sd / gen_int2
cat(" * Selection response for breeding program 2: ", sel_r2, "\n")
```

```
##  * Selection response for breeding program 2:  125.9827
```

- Comparison

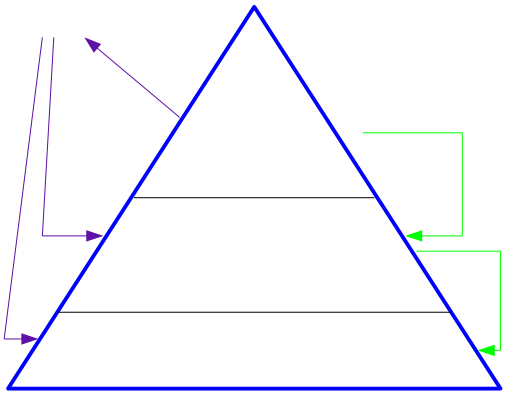
```
##  * Breeding program 2 is better than breeding program 1.
```

c) There are two basic schemes of a breeding program. These schemes are shown below. Please complete the diagrams shown below. In which species are the two schemes implemented?

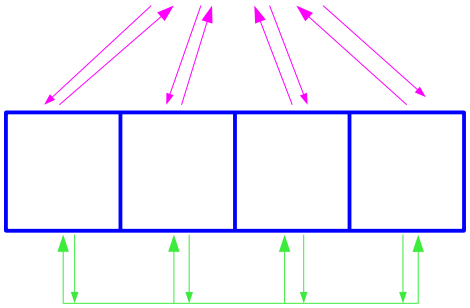
Es gibt zwei grundsätzliche Schemas eines Zuchtprogramms. Diese Schemas sind unten gezeigt. Bitte vervollständigen Sie die nachfolgenden Darstellungen. In welchen Tierarten werden die beiden Schemas umgesetzt?

14

Name of the scheme:

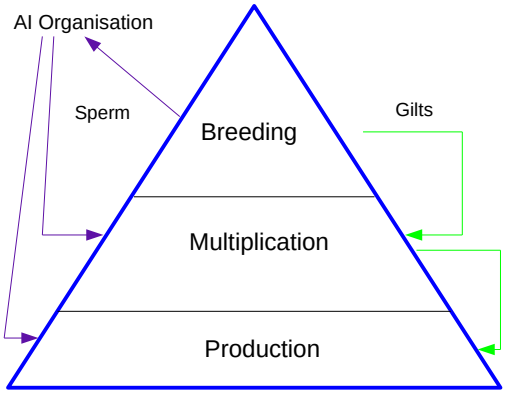


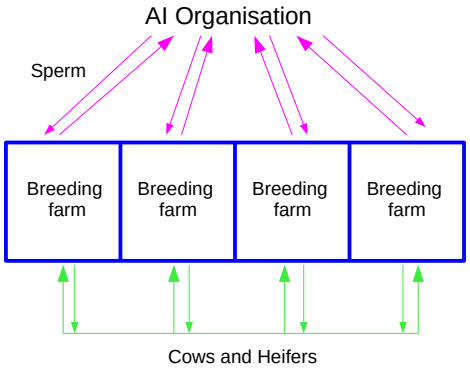
Name of the scheme:



Solution

The solution





- Left: hierarchical and right: monolithic
- Left: pigs and right: cattle / sheep / goats / horse

Problem 2: Economic Value

In a pig production farm that operates as a multiplier and has its own fatting operation, the revenus and the costs for producing a slaughter pig are listed below.

In einem Schweineproduktionsbetrieb mit Ferkelvermehrung und eigener Ausmast, die Erlöse und die Kosten für die Produktion eines Schlachtschweins sind unten aufgelistet.

Variable	Amount in Swiss Francs
Revenu per kg slaughter weight	4.900
Net cost per kg slaughter weight	3.571

a) Compute the profit per slaughter pig based on the information given in the above table. We assume that the slaughter weight per pig is $80kg$ per slaughter pig.

Berechnen Sie den Gewinn pro Mastschwein basierend auf den oben angegebenen Informationen. Dabei betrage das Schlachtgewicht $80kg$ *pro Schlachtschwein.*

2

Solution

The profit per slaughter pig is computed as

```
sl_pig_profit <- slaughter_weight * (slaughter_price - net_cost)
cat(" * Profit per slaughter pig: ", round(sl_pig_profit, digits = 2), "\n")
```

```
##  * Profit per slaughter pig:  106.32
```

b) The number of piglets born alive is an important economic trait for the pig producing farm. Using the trait 'number of piglets born alive' in the breeding program allowed the farm to increase the average of trait by 0.1 piglets. This changed the costs as shown in the table below. Compute the economic value for the trait 'number of piglets born alive' using the profit computed in Problem 2a as base situation and the profit in the new situation with the incremented level of the trait 'number of piglets born alive'.

Die Anzahl lebend geborener Ferkel ist ein wirtschaftlich wichtiges Merkmal in der Schweineproduktion. Dieses Merkmal wird deshalb im Zuchtprogramm verwendet und hat dem Produktionsbetrieb erlaubt die mittlere Anzahl an lebend geborenen Ferkeln um 0.1 Ferkel zu erhöhen. Dadurch verändern sich die Kosten, wie in der unten gezeigten Tabelle aufgeführt. Berechnen Sie das wirtschaftliche Gewicht für das Merkmal Anzahl lebend geborener Ferkel. Verwenden Sie den Gewinn, den Sie in der Aufgabe 2a) berechnet haben als Basissituation. Der Gewinn in der Situation mit dem erhöhten Mittelwert des Merkmals Anzahl lebend geborener Ferkel kann aus den Zahlen aus der nachfolgenden Tabelle berechnet werden.

3

Variable	Amount in Swiss Francs
Revenu per kg slaughter weight	4.900
Net cost per kg slaughter weight	3.557

Solution

The profit per slaughter pig with increased level of number of piglets born

```
sl_pig_profit_incr <- slaughter_weight * (slaughter_price - net_cost_incr)
cat(" * Profit per slaughter pig: ", round(sl_pig_profit_incr, digits = 2), "\n")
```

```
##  * Profit per slaughter pig:  107.44
```

The economic value

```
ec_val <- (sl_pig_profit_incr - sl_pig_profit) / n_incr_npa
cat (" * Economic value per piglet: ", round(ec_val, digits = 2), "\n")
```

```
##  * Economic value per piglet:  11.2
```

Problem 3: Model Selection

In the search of a diagnostic tool to prevent Hyperketonemia (HYK), keton bodies in the milk (aceton and BHB) are measured as predictor variables. These predictor variables are to be used to quantify changes in the response variable which is `Blood Serum BHB`. The following table shows the measurements of 12 cows.

Auf der Suche nach einem Diagnosewerkzeug zur Verhinderung von Hyperketonemia (HYK) wurden Ketonkörper (Aceton und BHB) in der Milch als unabhängige Variablen gemessen. Diese sollen verwendet werden um die Variation in der Zielgrösse BHB im Blut zu quantifizieren. Die folgende Tabelle zeigt die Messwerte von 12 Kühen.

Milk Aceton	Milk BHB	Blood Serum BHB
0.150	0.920	1.569
0.151	0.887	1.631
0.183	0.888	1.611
0.160	0.888	1.585
0.154	0.878	1.586
0.155	0.890	1.553
0.164	0.875	1.574
0.152	0.881	1.534
0.145	0.907	1.553
0.165	0.933	1.601
0.149	0.901	1.583
0.159	0.889	1.649

a) Specify the multiple linear regression model with 'Blood Serum BHB' as response variable. Quantify the effects of the two predictor variables 'Milk Aceton' and 'Milk BHB' on the response variable. What is the expected change of 'Blood Serum BHB' when the response variables 'Milk Aceton' and 'Milk BHB' change by one unit?

Stellen Sie das multiple lineare Regressionsmodell mit Blut-BHB als Zielgrösse auf. Quantifizieren Sie die Effekte der unabhängigen Variablen Milch-Aceton und Milch-BHB auf die Zielgrösse. Welche Änderung erwarten Sie im Blut-BHB-Wert, wenn die unabhängigen Variablen Milch-Aceton und Milch-BHB um eine Einheit ändern?

10

Solution

The multiple linear regression model for `Blood Serum BHB` on `Milk Aceton` and `Milk BHB` can be specified as

$$y = X * b + e$$

where y is the vector of `Blood Serum BHB`, b is the vector consisting of the intercept and the effects of `Milk Aceton` and `Milk BHB`, e is the vector of random residuals. The design matrix X has dimension 12×3 and consists of a column of ones. The second and the third column are the measurement values of `Milk Aceton` and `Milk BHB`.

The effects are quantified by the output of the following

```
lm_bbhb <- lm(`Blood Serum BHB` ~ `Milk BHB` + `Milk Aceton`, data = tbl_jer_reg)
summary(lm_bbhb)
```

```
##
## Call:
## lm(formula = `Blood Serum BHB` ~ `Milk BHB` + `Milk Aceton`,
```

```
##     data = tbl_jer_reg)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.044100 -0.018399 -0.005143  0.005997  0.061490
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.31970    0.57133   2.310   0.0462 *
## `Milk BHB`     0.07648    0.59138   0.129   0.8999
## `Milk Aceton`  1.25676    1.01817   1.234   0.2483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.034 on 9 degrees of freedom
## Multiple R-squared:  0.1448, Adjusted R-squared:  -0.04522
## F-statistic: 0.762 on 2 and 9 DF,  p-value: 0.4946
```

When `Milk BHB` changes by one unit, `Blood Serum BHB` is expected to change by 0.0764764 units. When `Milk Aceton` changes by one unit, then `Blood Serum BHB` is expected to change by 1.2567635 units.

b) The full model where the response variable is described by all the predictor variables is not necessarily the best model. Please specify two model selection techniques to find the best model. Give a short description of the listed model selection techniques.

Das volle Modell in welchem alle unabhängigen Variablen verwendet werden um die Zielgrösse zu beschreiben ist nicht in allen Fällen das beste Modell. Nennen Sie zwei Modellselektionsstrategien und beschreiben sie diese kurz.

4

Solution

1. Forward selection: Start with an empty model. Include those predictors that decrease the resiudal sums of squares the most. Do this until there are not predictor variables available. From the sequence of models take that with the smallest C_p value as the best model.
2. Backwards elimination: Start with the full model. Remove the predictor that increases the residual sum of squares the least. Do this until there are no predictor variables to be excluded. From the sequence of models take that with the smallest C_p value as the best model.

c) Use one of the model selection strategies and determine the best model for the ‘Blood-BHB‘ data.

Verwenden Sie eine Modellselektionsstrategie und bestimmen Sie das beste Modell für den Blut-BHB Datensatz.

2

Solution

Using `olsrr` to do the model selection.

```
tbl_jer_reg_ms <- tibble::tibble(MilkAceton = vec_milk_ace_jer,
                                 MilkBHB = vec_milk_bhb_jer,
                                 BloodSerumBHB = vec_blood_bhb_jer)

lm_bhb_full <- lm(BloodSerumBHB ~ MilkBHB + MilkAceton, data = tbl_jer_reg_ms)
olsrr::ols_step_best_subset(lm_bhb_full)
```

```
##      Best Subsets Regression
## ---------------------------------
## Model Index    Predictors
## ---------------------------------
##      1         MilkAceton
##      2         MilkBHB MilkAceton
## ---------------------------------
##
##                                                    Subsets Regression Summary
## ------------------------------------------------------------------------------------------------------------
##                         Adj.        Pred
## Model    R-Square    R-Square    R-Square     C(p)       AIC         SBIC        SBC         MSEP
## ------------------------------------------------------------------------------------------------------------
##   1        0.1432      0.0576     -0.1015    1.0167    -44.5265    -77.4765    -43.0717    0.0125
##   2        0.1448     -0.0452      -0.183    3.0000    -42.5487    -74.8255    -40.6091    0.0141
## ------------------------------------------------------------------------------------------------------------
## AIC: Akaike Information Criteria
##  SBIC: Sawa's Bayesian Information Criteria
##  SBC: Schwarz Bayesian Criteria
##  MSEP: Estimated error of prediction, assuming multivariate normality
##  FPE: Final Prediction Error
##  HSP: Hocking's Sp
##  APC: Amemiya Prediction Criteria
```

Problem 4: Genetic Evaluation

The trait `Blood Serum BHB` is used to do a genetic analysis. From the dataset in Problem 3, we selected 4 animals. For those animals pedigree information was added. The new dataset is shown below.

Für das Merkmal Blut-BHB soll eine genetische Analyse durchgeführt werden. Vom Datensatz in Aufgabe 3 wurden 4 Tiere ausgewählt. Für diese Tiere wurden die Abstammungsdaten zum Datensatz hinzugefügt. Der resultierende Datensatz ist in der nachfolgenden Tabelle gezeigt.

Animal	Sire	Dam	Milk Aceton	Milk BHB	Blood Serum BHB
3	1	2	0.183	0.888	1.611
4	1	NA	0.155	0.890	1.553
5	3	4	0.145	0.907	1.553
6	3	5	0.164	0.875	1.574

a) Use a BLUP animal model to predict breeding values for all the animals given in the dataset shown above. Specify all model components, insert all numeric information into the model components, and write down expected values and variance-covariance matrices for all random effects. As fixed effects you can take the components from the best model found in Problem 3c. If you did not solve Problem 3c, just use all covariables given in the dataset as fixed effects.

Verwenden Sie ein BLUP-Tiermodell zur Schätzung der Zuchtwerte aller Tiere, welche im oben gezeigten Datensatz gegeben sind. Spezifizieren Sie alle Modellkomponenten, setzen Sie die numerischen Informationen vom Datensatz in die Modellkomponenten ein und notieren Sie die Erwartungswerte und Varianz-Kovarianzmatrizen für alle zufälligen Effekte im Modell. Als fixe Effekte können Sie die Komponenten des besten Modells aus Aufgabe 3c übernehmen. Falls Sie Aufgabe 3c nicht gelöst haben, dann verwenden Sie alle Kovariablen im Datensatz als fixe Effekte.

15

Solution

1. BLUP-animal model:

$$y = Xb + Zu + e$$

2. Model components

- Vector y of observations

$$y = \begin{bmatrix} 1.611 \\ 1.553 \\ 1.553 \\ 1.574 \end{bmatrix}$$

- Vector b of fixed effects

$$b = \begin{bmatrix} b_0 \\ b_{ACE} \end{bmatrix}$$

with b_0 corresponding to the intercept and b_{ACE} to the effect of `Milk Aceton`.

- Design matrix X

$$X = \begin{bmatrix} 1 & 0.183 \\ 1 & 0.155 \\ 1 & 0.145 \\ 1 & 0.164 \end{bmatrix}$$

- The random vector u of breeding values

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{bmatrix}$$

- The design matrix Z

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

- The vector e of random residuals

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

3. Expected values

$$E \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \end{bmatrix}$$

4. Variance-Covariance Matrices

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

with $R = I * \sigma_e^2$, $G = A * \sigma_u^2$, σ_e^2 corresponding to the residual variance component and σ_u^2 is the genetic-additive variance.

5. Mixed Model Equations The following mixed model equations are used to solve for the unknown estimated fixed effects ($\hat{b}$) and the unknown predicted breeding values ($\hat{u}$).

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

with $\lambda = \sigma_e^2 / \sigma_u^2$. Because the variance components are not specified, we assume λ to be equal to 1. The matrix A corresponds to the numerator relationship matrix. We start by computing its inverse using the package `pedigreemm`.

```
n_nr_ani_ped <- n_nr_founder+n_nr_ani
ped <- pedigreemm::pedigree(sire = c(rep(NA, n_nr_founder), tbl_gen$Sire),
                            dam = c(rep(NA, n_nr_founder), tbl_gen$Dam),
                            label = as.character(1:n_nr_ani_ped))
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
```

```
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
mat_Ainv <- as.matrix(pedigreemm::getAInv(ped = ped))
```

With that, the mixed model equations can be constructed and solved for the estimated fixed effects ($\hat{b}$) and the predicted breeding values ($\hat{u}$).

```
mat_XTX <- crossprod(mat_X)
mat_XTZ <- crossprod(mat_X, mat_Z)
mat_ZTX <- crossprod(mat_Z, mat_X)
mat_ZTZAinvlambda <- crossprod(mat_Z) + mat_Ainv * lambda
mat_coef <- rbind(cbind(mat_XTX, mat_XTZ), cbind(mat_ZTX, mat_ZTZAinvlambda))
mat_rhs <- rbind(crossprod(mat_X, vec_y), crossprod(mat_Z, vec_y))
mat_sol <- solve(mat_coef, mat_rhs)
```

The solutions for the estimated fixed effects are

$$\hat{b} = \begin{bmatrix} 1.3163 \\ 1.5808 \end{bmatrix}$$

The solutions for the predicted breeding values are

$$\hat{u} = \begin{bmatrix} 0 \\ 0.0018 \\ 0.0027 \\ -0.0027 \\ 0.0021 \\ 0.0011 \end{bmatrix}$$

b) Animals 6 and 4 have two offspring which do not have any performance records yet. What type of information is required to determine which of the two fullsibs has the better genetic potential?

Die Tiere 6 und 4 haben zwei Nachkommen, welche noch keine Leistungen haben. Welche Informationen sind notwendig, damit wir bestimmen können, welches der beiden Vollgeschwister das bessere genetische Potential hat?

2

Solution

Animals with records and the two fullsibs must be genotyped and based on this, genomic breeding values can be predicted. Because fullsibs do not have the same genotypes, their genomic breeding value is different and can be used to find the animal with the better genetic potential.