

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

751-6212-00L V
Solutions to Exam
Applied Genetic Evaluation
SS 2021

Date: 31st May 2020

Name: Firstname Name

Legi-Nr: LegiNr

Problem	Maximum Number of Points	Number of Points Reached
1	17	
2	50	
3	25	
4	21	
Total	113	

Questions in German are in italics

Problem 1: Breeding Program

A beef production herd reported the following economic results with respect to their beef fattening operation.

Ein Rindviehmastbetrieb weist die folgenden wirtschaftlichen Resultate aus dem Mastbereich aus.

Parameter	Value
Weaning Weight (<i>Absetzgewicht</i>)	250 kg
Slaughter Weight (<i>Schlachtgewicht</i>)	550 kg
Dressing Percentage (<i>Schlachtausbeute</i>)	0.6
Average Daily Gain (<i>Tageszunahme</i>)	1 kg/day
Carcass Price (<i>Schlachtkörperpreis</i>)	12.25 SFr / [kg carcass weight]
Subsidies (<i>Subventionen</i>)	2340 SFr / slaughter animal
Direct Costs (<i>Direkte Kosten</i>)	10 SFr / slaughter animal and day
Fixed Costs (<i>Fixe Kosten</i>)	2800 SFr / slaughter animal

- a) Compute the profit per slaughter animal based on the economic parameters given in the table above.

Berechnen Sie den Gewinn pro Schlachttier aufgrund der oben gegebenen wirtschaftlichen Parameter.

6

Solution

Master Solution

The profit (P) per slaughter animal is given by the difference between revenue (R) and costs (C). The revenue corresponds to the sum of the price per slaughter animal plus the subsidies. The price per slaughter animal is the slaughter weight times the slaughter price.

The costs are the sum from fixed costs and direct costs. The direct costs depend on the length of the fattening period which is the difference in slaughter weight minus weaning weight divided by the average daily gain.

```
# revenue
n_sl_price <- n_sla_weight * n_dress * n_carcass_price
n_rev <- n_sl_price + n_sub
# length of fattening period
n_len_fat_period <- (n_sla_weight - n_wean_weight) / n_agd
# direct costs
n_tot_dir_cost <- n_len_fat_period * n_dir_cost
# total costs
n_tot_cost <- n_tot_dir_cost + n_fix_cost
# profit
n_profit <- n_rev - n_tot_cost
```

The profit per slaughter animal is: 582.5 SFr.

- b) Compute the economic value for the traits average daily gain (AGD), fleshiness and fat coverage based on the values given in the table above. For the economic value of fleshiness and fat coverage, the price deductions per class as shown in the table below are assumed.

Berechnen Sie die wirtschaftlichen Gewichte für Tageszunahme, Fleischigkeit und Fettabdeckung basierend auf den oben gemachten Angaben. Für die wirtschaftlichen Gewichte der Fleischigkeit und der Fettabdeckung werden die folgenden Preisabzüge pro Klasse angenommen.

9

Solution

Master Solution

Economic values for the three traits are computed as follows

- ADG: Change of profit caused by an increase of ADG by 0.001 kg/day

```
n_agd_new <- n_agd + n_agd_incr
# length of fattening period
n_len_fat_period_new <- (n_sla_weight - n_wean_weight) / n_agd_new
# direct costs
n_tot_dir_cost_new <- n_len_fat_period_new * n_dir_cost
# total costs
n_tot_cost_new <- n_tot_dir_cost_new + n_fix_cost
# profit
n_profit_new <- n_rev - n_tot_cost_new
# economic value for agd
n_ev_agd <- n_profit_new - n_profit
```

- Fleshiness: The profit is only changed by the change in revue. The revenue is changed by the price deduction.

```
# revenue
n_sl_price_new <- n_sla_weight * n_dress * n_carcass_price * (1-n_price_red_flesh)
n_rev_new <- n_sl_price_new + n_sub
# length of fattening period
n_len_fat_period <- (n_sla_weight - n_wean_weight) / n_agd
# direct costs
n_tot_dir_cost <- n_len_fat_period * n_dir_cost
# total costs
n_tot_cost <- n_tot_dir_cost + n_fix_cost
# profit
n_profit_new <- n_rev_new - n_tot_cost
# economic value for fleshiness
n_ev_fl <- n_profit - n_profit_new
```

- Fat coverage: the same as with fleshiness

```
# revenue
n_sl_price_new <- n_sla_weight * n_dress * n_carcass_price * (1-n_price_red_fat)
n_rev_new <- n_sl_price_new + n_sub
# length of fattening period
n_len_fat_period <- (n_sla_weight - n_wean_weight) / n_agd
# direct costs
n_tot_dir_cost <- n_len_fat_period * n_dir_cost
# total costs
n_tot_cost <- n_tot_dir_cost + n_fix_cost
```

```

# profit
n_profit_new <- n_rev_new - n_tot_cost
# economic value for fleshiness
n_ev_fc <- n_profit - n_profit_new

```

- Result table

Trait	Economic Value
ADG	2.997
Fleshiness	121.275
Fat Coverage	80.850

- c) For two bulls the following predicted breeding values are given. Which of the two bulls would you select based on the index value obtained from the aggregate genotype (H) containing the three traits: average daily gain, fleshiness and fatness?

Für zwei Bullen sind die folgenden geschätzten Zuchtwerte gegeben. Welchen der beiden Bullen würden Sie aufgrund des Indexwertes des Gesamtzuchtwertes (H) mit den drei Merkmalen Tageszunahme, Fleischigkeit und Fettabdeckung auswählen?

2

Trait	Bull 1	Bull 2
Average Daily Gain (g/day)	250.00	20.000
Fleshiness (class)	-0.01	0.020
Fat Coverage (class)	-0.02	0.015

Solution

Master Solution

- Bull 1: Index value

```
(n_idx_val_bull1 <- crossprod(vec_bv_bull1, vec_ev_result))
```

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

```
## [1,] 746.421
```

- Bull 2: Index value

```
(n_idx_val_bull2 <- crossprod(vec_bv_bull2, vec_ev_result))
```

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

```
## [1,] 63.57831
```

- Decision: Bull 1 is better than Bull 2, hence we select Bull 1

Problem 2: Prediction of Breeding Values

The following dataset will be used to predict breeding values using different methods. The dataset is available under the following URL:

Der nachfolgende Datensatz wird für die Schätzung von Zuchtwerten mit verschiedenen Methoden verwendet. Der Datensatz ist unter dem folgenden URL verfügbar:

https://charlotte-ngs.github.io/gelasmss2021/data/gel_exam_p02.csv

Animal	Sire	Dam	Sex	Observation
4	1	2	1	3.38
5	1	2	2	3.94
6	3	5	2	2.81
7	4	6	2	2.51
8	4	7	1	1.59
9	8	6	2	3.47

Hints

- Heritability (*Erblichkeit*) of the given trait is: 0.25
- Phenotypic variance (*Phänotypisch Varianz*): 2.25

- a) Predict the breeding values of the animals given in the table above using the given observation as an own-performance.

Schätzen Sie die Zuchtwerte der Tiere, welche in der Tabelle oben gegeben sind und verwenden Sie die angegebenen Beobachtungen als Eigenleistung.

8

Solution

Master Solution

The predicted breeding value (\hat{u}_i) for animal i using own-performance is computed as

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

The population mean μ is taken as the mean of all observations.

```
n_mu <- mean(tbl_data_p02$Observation)
n_hat_u <- n_h2_imf * (tbl_data_p02$Observation - n_mu)
tbl_result_pbv_p02a <- tibble::tibble(Animal = tbl_data_p02$Animal,
                                     Result = round(n_hat_u, digits = 3))
knitr::kable(tbl_result_pbv_p02a)
```

Animal	Result
4	0.107
5	0.247
6	-0.035
7	-0.110
8	-0.340
9	0.130

- b) Use the following dataset to predict breeding values using a BLUP animal model. Specify all model components, enter all information from the dataset into the model components and specify expected values and variance-covariance matrices for all random components in the model.

Verwenden Sie den folgenden Datensatz zur Schätzung der Zuchtwerte mit einem BLUP-Tiermodell. Spezifizieren Sie alle Modellkomponenten, setzen Sie alle Informationen aus dem Datensatz in die Modellkomponenten ein und geben Sie die Erwartungswerte und die Varianz-Kovarianz-Matrizen der zufälligen Effekte im Modell an.

25

Solution

Master Solution

The BLUP animal model corresponds to the following mixed linear effect model

$$y = Xb + Zu + e$$

where y is the vector of observations, b is the vector of fixed effects, u is the vector of random breeding values, e is the vector of random residuals. The matrices X and Z are known incidence matrices which link the fixed effects and the random breeding values to the observations, respectively.

Inserting the data into the known components of the model leads to

$$y = \begin{bmatrix} 3.38 \\ 3.94 \\ 2.81 \\ 2.51 \\ 1.59 \\ 3.47 \end{bmatrix}$$

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

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

$$b = \begin{bmatrix} b_M \\ b_F \end{bmatrix}$$

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \end{bmatrix}$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

The expected values and the variance-covariance matrices for the random effects are

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

$$\text{var} \begin{bmatrix} e \\ u \\ y \end{bmatrix} = \begin{bmatrix} R & 0 & R \\ 0 & U & UZ^T \\ R & ZU & ZUZ^T + R \end{bmatrix}$$

with $R = I * \sigma_e^2$, $U = A * \sigma_u^2$ and A corresponding to the numerator relationship matrix.

The mixed model equations are given by

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

The solutions to the mixed model equations are computed as follows

```
lambda <- (1 - n_h2_imf) / n_h2_imf
# numerator relationship matrix
ped <- pedigreeemm::pedigree(sire = c(rep(NA, n_nr_founder), tbl_data_p02$Sire),
                             dam = c(rep(NA, n_nr_founder), tbl_data_p02$Dam),
                             label = as.character(1:n_nr_ped_animal))
mat_Ainv <- as.matrix(pedigreeemm::getAinv(ped = ped))
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 solution for the fixed effects are

$$\hat{b} = \begin{bmatrix} 2.5351 \\ 3.2493 \end{bmatrix}$$

$$\hat{u} = \begin{bmatrix} 0.0518 \\ 0.0518 \\ -0.1035 \\ 0.0566 \\ 0.0987 \\ -0.106 \\ -0.172 \\ -0.1568 \\ -0.088 \end{bmatrix}$$

- c) Compare the Reliabilities of the predicted breeding values from a) and b). What is the reason for them being different?

Vergleichen Sie die Bestimmtheitsmasse der unter a) und b) geschätzten Zuchtwerte. Wo liegt der Grund für die unterschiedlichen Werte der Bestimmtheitsmasse?

17

Solution

Master Solution

Problem 3: Model Selection

- a) Model selection can be done using **forward** selection or **backward** selection. Describe the four steps that are needed in both model selection procedures.

*Modellselektion kann mit **Forward** Selektion oder **Backward** Selektion gemacht werden. Beschreiben Sie die vier Schritte, welche es für beide Modellselektionsprozeduren braucht.*

16

Solution

Step	Forward	Backward
1		
2		
3		
4		

Master Solution

Step	Forward	Backward
1	Start with smallest model \mathcal{M}_0	Start with full model \mathcal{M}_0
2	Include predictor reducing RSS the most	Eliminate predictor increasing RSS the least
3	Continue step 2 until all predictors chosen	Continue step 2 until all predictors eliminated
4	From sequence of submodels $\mathcal{M}_0 \subseteq \mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \dots$ choose model with smallest C_p	From sequence of submodels $\mathcal{M}_0 \supseteq \mathcal{M}_1 \supseteq \mathcal{M}_2 \supseteq \dots$ choose model with smallest C_p

- b) We are given the following data set on body weight, breast circumference and shoulder height. Furthermore, the results of fitting the full regression model and a reduced model with only breast circumference as predictor. Compute the Mallow C_p value for both models and decide which of the two models is better based on C_p

Gegeben ist der folgende Datensatz zu Körpergewicht, Brustumfang und Widerristhöhe. Weiter sind auch die Resultate des Fits des vollen Modells und eines reduzierten Modells mit nur Brustumfang als beschreibende Variable gegeben. Berechnen Sie den Mallow C_p -Wert für beide Modelle und begründen Sie, welches der beiden Modelle besser ist.

6

Animal	BreastCircumference	BodyWeight	ShoulderHeight
1	176	471	151
2	177	463	155
3	178	481	158
4	179	470	157
5	179	496	157
6	180	491	151
7	181	518	146
8	182	511	159
9	183	510	149
10	184	541	148

The dataset is available under:

Der Datensatz ist verfügbar unter:

https://charlotte-ngs.github.io/gelasmss2021/data/gel_exam_p03.csv

The results of the full model

Die Resultate des vollen Modells

```
##
## Call:
## lm(formula = BodyWeight ~ BreastCircumference + ShoulderHeight,
##     data = tbl_data_p03)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.278  -7.139   4.814   6.838  11.722
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -798.1362    346.0669  -2.306  0.05448 .
## BreastCircumference     8.0011     1.5203   5.263  0.00117 **
## ShoulderHeight     -0.9541     0.8498  -1.123  0.29859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.91 on 7 degrees of freedom
## Multiple R-squared:  0.8504, Adjusted R-squared:  0.8076
## F-statistic: 19.89 on 2 and 7 DF,  p-value: 0.001296
```

The results of the reduced model

Die Resultate des reduzierten Modells

```
##
## Call:
## lm(formula = BodyWeight ~ BreastCircumference, data = tbl_data_p03)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3941  -6.5525  -0.0673   9.3707  13.2594
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1065.115     255.483   -4.169  0.003126 **
## BreastCircumference     8.673       1.420    6.108  0.000287 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.08 on 8 degrees of freedom
## Multiple R-squared:  0.8234, Adjusted R-squared:  0.8014
## F-statistic: 37.31 on 1 and 8 DF, p-value: 0.000287
```

Solution

Master Solution

Mit `olsrr` erhalten wir

```
olsrr::ols_step_best_subset(lm_bw_bc_sh)
```

```
## Registered S3 methods overwritten by 'car':
##   method             from
##   influence.merMod    lme4
##   cooks.distance.influence.merMod lme4
##   dfbeta.influence.merMod lme4
##   dfbetas.influence.merMod lme4
##
##           Best Subsets Regression
## -----
## Model Index   Predictors
## -----
##      1      BreastCircumference
##      2      BreastCircumference ShoulderHeight
## -----
##
##           Subsets Regression Summary
## -----
## Model   R-Square   Adj. R-Square   Pred R-Square   C(p)   AIC   SBIC   SBC   MSEP   FPE   HSP   APC
## -----
##      1      0.8234      0.8014      0.7137      2.2605  80.2529  52.6278  81.1606  1235.0138  147.3596  17.5428  0.2649
##      2      0.8504      0.8076      0.7148      3.0000  80.5972  54.4225  81.8075  1220.9905  154.6059  19.8213  0.2779
## -----
## 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
```

Das reduzierte Model erhält einen tieferen C_p -Wert und ist somit das bessere Modell.

- c) What is the reason why we are doing model selection, why should we not take always the full model? Please explain the underlying phenomenon in your own words.

Was ist der Grund, dass wir Modellselektion machen? Weshalb können wir nicht einfach immer das volle Modell mit allen erklärenden Variablen zur Anpassung verwenden? Bitte erklären Sie das unterliegende Phänomen in Ihren eigenen Worten.

3

Solution

Master Solution

The problem that makes model selection necessary is called Bias-Variance trade-off. The Bias is reduced, the more predictors we include. But on the other hand, the variance is increased. Hence we have to find the model that balances this trade-off. This is done with a criterion such as the Mallows C_p statistic.

Problem 4: Variance Components Estimation

The following data set with average amounts of methane gas emission per year for three offspring per bull are given in the following table.

Im folgenden Datensatz sind die mittleren Mengen an Methanemissionen pro Jahr für je drei Nachkommen für fünf verschiedene Bullen gegeben.

Offspring	Bull	Methane
1	1	104
2	1	99
3	1	100
4	2	97
5	2	97
6	2	87
7	3	116
8	3	93
9	3	126
10	4	110
11	4	83
12	4	86
13	5	117
14	5	80
15	5	120
16	6	86
17	6	102
18	6	90
19	7	105
20	7	108
21	7	86

The data set is available under:

Der Datensatz ist verfügbar unter:

https://charlotte-ngs.github.io/gelasmss2021/data/gel_exam_p04.csv

- a) Given that we want to reduce the amount of methane gas emitted in cattle using the tools of livestock breeding, we first have to do a variance components estimation. Why is it important to have a certain variability in a given trait and which variance component is important when we want to improve a certain trait with livestock breeding tools?

Wir möchten die Methanemission beim Rind mit den Werkzeugen der Tierzucht senken. Weshalb brauchen wir für die züchterische Bearbeitung eines Merkmals ein gewisse Variabilität im Merkmal und welche Varianzkomponente interessiert uns besonders?

2

Solution

Master Solution

One of the main livestock breeding tools is selection. Selection means that we choose from a given population the best individuals to be parents of the next generation. Hence selection can only happen, when a certain level of variation can be observed in a population. Because livestock breeding means improvement of a population at the genetic level, the genetic variance component or in our case the sire variance component is important.

- b) What is the linear model that would be used to allow us a separation of the different variance components in the given dataset on methane gas emission?

Wie sieht das lineare Modell aus, welches eine Aufteilung der Varianz in ihre verschiedenen Komponenten im Methandatensatz erlaubt?

11

Solution

Master Solution

The model is

$$y_{ij} = \mu + s_i + e_{ij}$$

where

y_{ij}	measurement j of animal i
μ	expected value of y
s_i	deviation of y_{ij} from μ attributed to bull i
e_{ij}	measurement error

The expected values are: $E[s_i] = 0$, $E[e_{ij}] = 0$, $E[y_{ij}] = \mu$

The variances: $\text{var}(s_i) = \sigma_s^2$, $\text{var}(e_{ij}) = \sigma_e^2$, $\text{var}(y_{ij}) = \sigma_s^2 + \sigma_e^2$

- c) Use an analysis of variance (ANOVA) to estimate the variance components given in the above data set.

Schätzen Sie mit einer Varianzanalyse (ANOVA) die Varianzkomponenten aus dem oben angegebenen Datensatz zu den Methanemissionen.

8

Solution

Master Solution

The sires are first converted into factors to prevent R to fit them as covariables.

```
tbl_met$Bull <- as.factor(tbl_met$Bull)
```

Use `aov()` to get the ANOVA-table

```
aov_meth <- aov(Methane ~ Bull, data = tbl_met)
(sy_aov_meth <- summary(aov_meth))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Bull        6  933.6   155.6    0.869  0.541
## Residuals   14 2507.3   179.1
```

From this the estimate of the residual variance corresponds to

```
(n_hat_sigma_e2 <- sy_aov_meth[[1]]$`Mean Sq`[2])
```

```
## [1] 179.0952
```

```
n_nr_dau <- 3
```

```
(n_hat_sigma_s2 <- (sy_aov_meth[[1]]$`Mean Sq`[1] - sy_aov_meth[[1]]$`Mean Sq`[2]) / n_nr_dau)
```

```
## [1] -7.830688
```