

Peter von Rohr

Institute of Agricultural Sciences

D-USYS

ETH Zurich

751-7602-00 V

Exam

Applied Statistical Methods

in Animal Sciences

SS 2020

Date: 25th May 2020

Name: Xenia Klaus

Legi-Nr: 13761044

Problem	Maximum Number of Points	Number of Points Reached
1	14	
2	15	
3	32	
4	19	
Total	80	

*Questions in German are in italics*

## Problem 1: Traditional Breeding versus Genomic Selection

In Livestock breeding three different sources of information are available

1. Performance data
2. Pedigree information
3. Genomic Marker Data

*In der Tierzucht stehen uns drei verschiedene Informationsquellen zur Auswahl*

1. *Leistungsdaten*
2. *Pedigreeinformationen*
3. *Genomische Markerdaten*

- a) Complete the following table indicating which source of information is used in either traditional breeding or in genomic selection.

*Vervollständigen Sie die folgende Tabelle und geben Sie an, welche Informationsquelle in der traditionellen Tierzucht oder der genomischen Selektion verwendet werden.*

**6**

### Solution

Source of Information	Traditional Breeding	Genomic Selection
Performance data		
Pedigree information		
Genomic Marker Data		

- b) In the two step approach of genomic selection the reference population is used to estimate marker effects. Animals with high reliabilities ( $B\%$ ) are included in the reference population. Let us assume that we have a population with 12605 breeding animals. What is the expected number of animals in the reference population, if we require a reliability of at least 59% for animals to be included into the reference population. We assume the reliabilities on the percentage scale to follow a normal distribution with mean 37% and standard deviation of 11%

*Im Zwei-Schritt Verfahren der genomischen Selektion werden die Markereffekte anhand einer Referenzpopulation geschätzt. Tiere mit einem hohen Bestimmtheitsmass ( $B\%$ ) werden in die Referenzpopulation aufgenommen. Wir nehmen an, dass wir eine Population von einer totalen Grösse von 12605 Tieren haben. Was ist die erwartete Anzahl Tiere in der Referenzpopulation, wenn wir ein Bestimmtheitsmass von mindestens 59% für Tiere in der Referenzpopulation verlangen. Wir nehmen an, dass das Bestimmtheitsmass auf der Prozentskala einer Normalverteilung mit Mittelwert 37% und Standardabweichung 11% folgt.*

4

## Solution

- c) To be able to obtain good estimates of marker effects, the minimum size of the reference population is set to 1000 animals. What is the minimum reliability ( $B\%$ ) to get to a reference population of 1000 animals using the same distributional assumptions for the reliabilities as under 1b).

*Damit wir zuverlässige Markereffektschätzungen erhalten müssen wir eine Referenzpopulation von mindestens 1000 Tiere haben. Wie muss die untere Grnze für das Bestimmtheitsmass ( $B\%$ ) festgelegt werden, damit wir eine Referenzpopulation von 1000 Tieren erhalten. Die Annahmen betreffend der Verteilung der Bestimmtheitsmasse können Sie aus der Aufgabe 1b) übernehmen.*

4

## Solution

## Problem 2: Fixed Linear Effect Model

Hyperketonemia (HYK) is a metabolic disorder in cattle characterized by elevated levels of blood ketone bodies. The disorder affects early lactating cows. Ketone bodies such as beta-hydroxy-butyrate (BHB) and acetone (ACE) can be diagnosed in blood, milk and urine. For diagnostic reasons it is interesting to be able to predict ketone bodies such as BHB in the blood based on measurements of BHB and ACE in the milk. In a publication by Chandler et al. (2018) the following regression coefficients were found for cows which were not diagnosed of HYK. In that regression model BHB in blood serum (bBHB) was the response variable and a number of predictors either measured in a milk sample or observed from the cows performance record were used. These predictors are listed in the column entitled by the term **Variable** in the table below.

*Hyperketonemia (HYK) ist eine Stoffwechselstörung beim Rind, welche sich durch erhöhte Werte von Ketonkörpern im Blut manifestiert. Die Störung betrifft vor allem Milchrinder zu Beginn der Laktation. Ketonkörper wie Beta-Hydroxy-Butyrat (BHB) und Aceton (ACE) können im Blut, in der Milch und im Urin diganostiziert werden. Aus Gründen der Diagnostik ist es interessant Ketonkörper, wie z.B. BHB im Blut über Messungen von BHB und ACE in der Milch schätzen zu können. In einer Publikation von Chandler et al. (2018) wurden die folgenden Regressionskoeffizienten für Kühe, welche keine HYK-Diagnose erhielten, gefunden. In diesem Regressionsmodell wurde BHB im Blut (bBHB) als Zielgrösse festgelegt. Verschiedene Milchinhaltsstoffe oder Leistungsparameter der untersuchten Kühe dienten als unabhängige Variablen verwendet.*

Variable	Regression Coefficients (Non-HYK)
Intercept	-2.380
Milk acetone, mmol/L	1.220
Milk protein, %	-0.100
Fat-to-protein ratio	0.240
Production on test day, kg	-0.003
Gestation length, d	0.010

- a) Compute the predicted value of ‘bBHB’ for the three cows which are described in the following table.

*Schätzen Sie die ‘bBHB’-Werte der drei Kühe, welche in der nachfolgenden Tabelle beschrieben sind.*

9

Variable	Observed Values of Cow A	Observed Values of Cow B	Observed Values of Cow C
Milk acetone, mmol/L	0.082	0.103	0.177
Milk protein, %	3.420	3.350	3.300
Milk fat, %	4.300	4.300	4.370
Production on test day, kg	23.700	22.900	23.500
Gestation length, d	276.000	274.000	276.000

## Solution

- b) According to the publication by Chandler et al. (2018), the mean levels of BHB and Aceton found in the blood serum and in the milk are higher for cows with a diagnosed HYK compared to cows which are HYK-free. The different levels are listed in the table below. Is it possible to use the regression coefficients listed in Problem 2a) to predict the levels of blood BHB for cows that have HYK? Please reason about your answer.

*Gemäss der Publikation von Chandler et al. (2018) liegen die mittleren Werte für BHB und Aceton im Blut und in der Milch höher bei den Kühen, bei welchen eine HYK diagnostiziert wurde im Vergleich zu den HYK-freien Kühe. Die Werte sind in der nachfolgenden Tabelle angegeben. Ist es möglich den Blut-BHB-Wert einer auf HYK diagnostizierten Kuh anhand der unter Aufgabe 2a aufgelisteten Regressionskoeffizienten zu schätzen? Bitte begründen Sie Ihre Antwort.*

4

Variable	Non-HYK Cows	HYK Cows
Blood Serum BHB, mmol/L	0.60	2.30
Milk BHB, mmol/L	0.07	0.15
Milk Acetone, mmol/L	0.08	0.69
Fat, %	4.37	5.10
Protein, %	3.36	2.97
Production on test-day, kg	23.90	22.20
Gestation length, d	275.00	282.00

## Solution

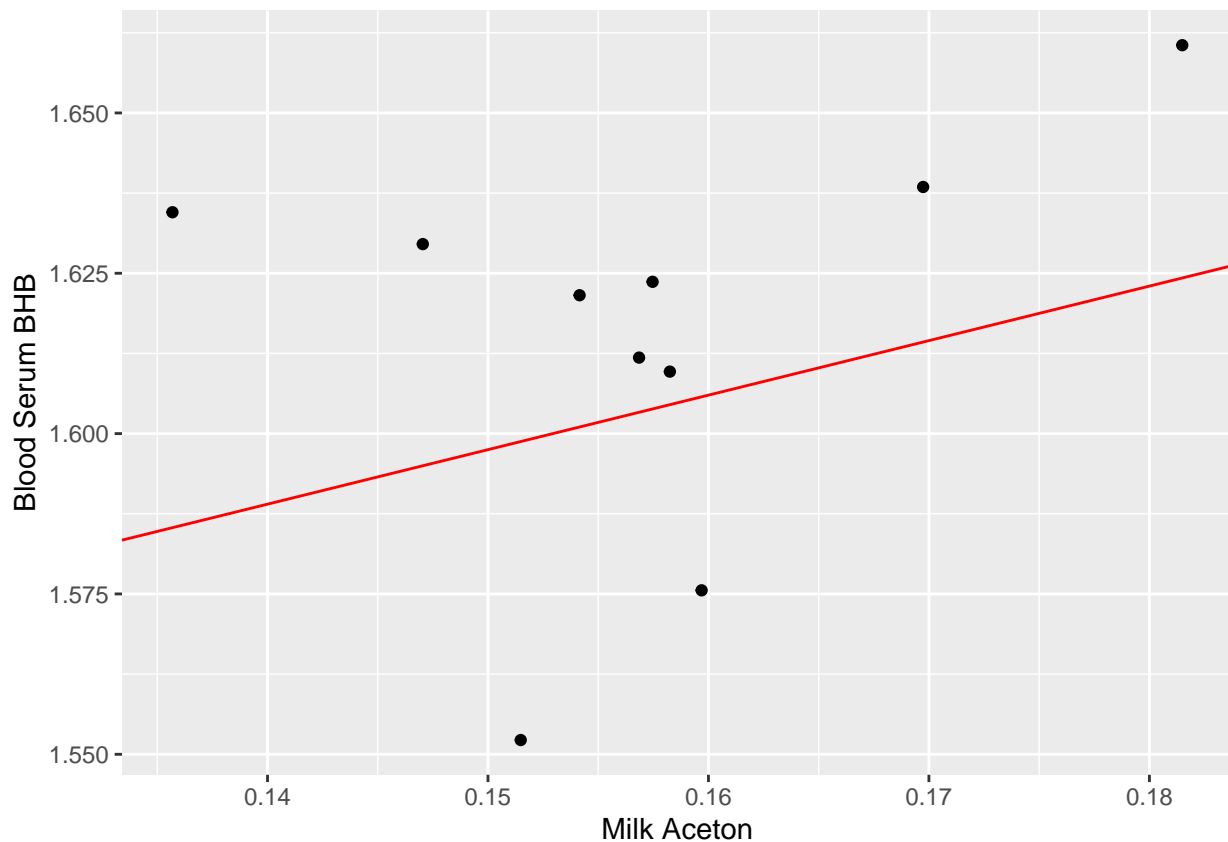
- c) As shown in the table below, there are also breed differences in the levels of keton bodies in blood serum and in the milk. The following tables contains levels of blood-BHB and acetone in the milk for cows of the breeds Holstein and Jersey. When looking only at the intercept of the complete model and at the regression coefficient of milk-acetone for the two breeds the following two plots can be drawn. Please indicate which plot (A and B) belong to which breed.

*Gemäss der folgenden Tabelle gibt es auch Rassenunterschiede bei den Werten der Ketonkörper im Blut und in der Milch. Die folgende Tabelle zeigt Blut-BHB und Milch Aceton für Kühe der Rassen Holstein und Jersey. Schaut man sich nur den Achsenabschnitt und den Regressionskoeffizienten von Milch-Aceton für die zwei Rassen an, dann entstehen die unten gezeigten Plots für die beiden Rassen. Bitte geben Sie an, welches Diagramm (A oder B) zu welcher Rasse gehört.*

2

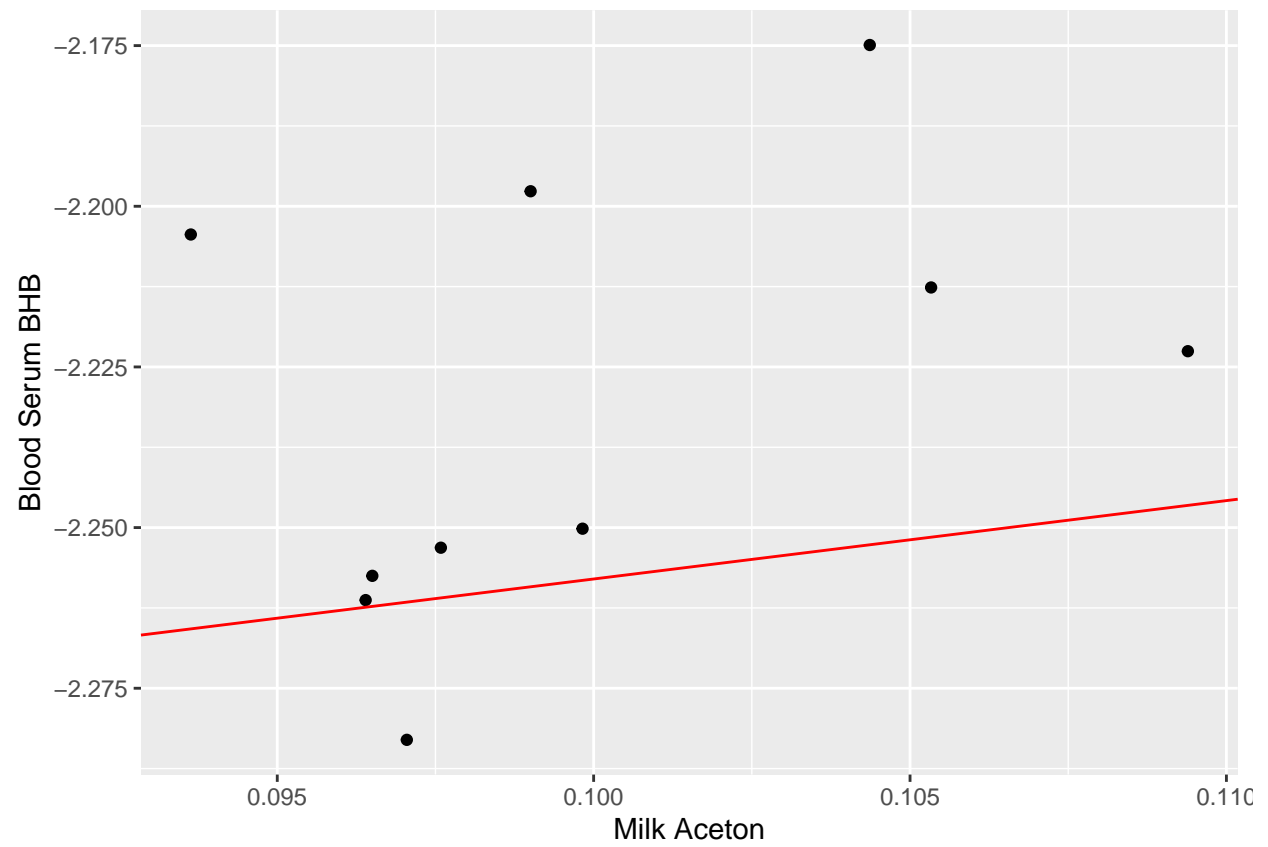
Variable	Holstein	Jersey
Mean Blood Serum BHB, mmol/L	0.800	0.920
Standard deviation Blood Serum BHB, mmol/L	0.030	0.030
Mean Acetone, mmol/L	0.101	0.159
Standard deviation Acetone, mmol/L	0.010	0.010
Intercept	-2.380	1.470
Regression Coefficient	1.220	0.850

Plot A



Plot B





**Solution**

### Problem 3: Genomic BLUP

The dataset shown below on blood-BHB is used to quantify the influence of 5 SNP-loci. The only fixed effect in this model is the herd.

*Der unten gezeigte Datensatz zu BHB im Blut soll verwendet werden um den Einfluss von 5 SNP-Loci zu quantifizieren. Der Betrieb wird als fixer Effekt berücksichtigt.*

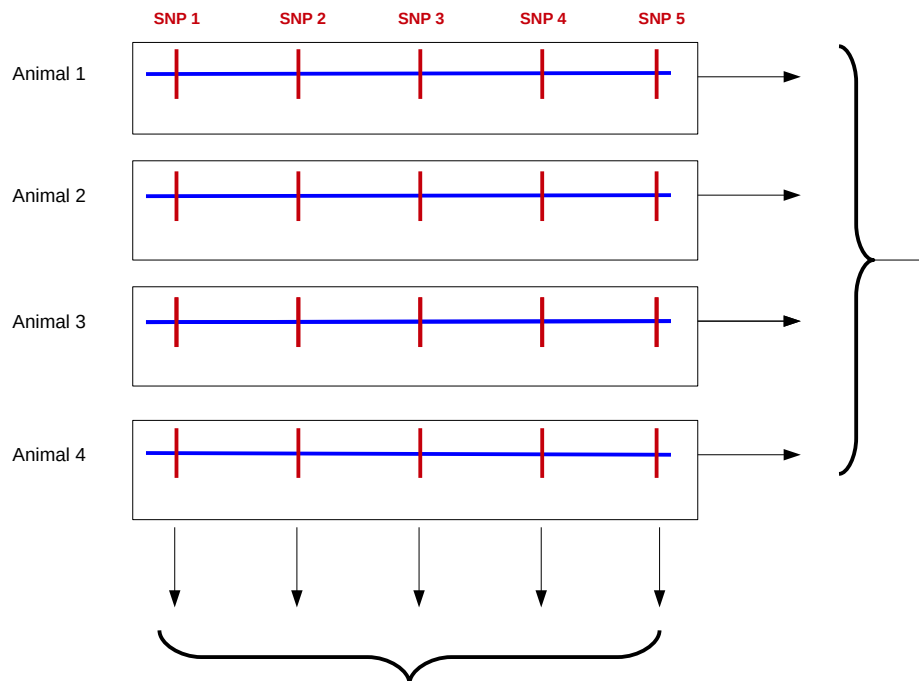
Tier	bBHB	Herd	SNP1	SNP2	SNP3	SNP4	SNP5
1	1.552	1	-1	0	1	1	1
2	1.610	2	1	1	0	1	0
3	1.661	1	-1	0	-1	1	-1
4	1.612	1	1	1	-1	1	0

- a) GBLUP models can either be specified as marker-effect models or as breeding-value models. Please complete the following diagram which illustrates the difference between the two types of models. Enter the SNP-genotypes for all animals at each SNP-position. Use  $G_1$  as the allele with the positive effect.

*GBLUP Modelle können entweder als Marker-Effekt-Modelle oder als Zuchtwertmodelle angegeben werden. Bitte vervollständigen Sie das folgende Diagramm, welches den Unterschied zwischen den beiden Modelltypen erklärt. Bitte notieren Sie die SNP-Genotypen der Tiere an allen SNP-Positionen. Verwenden Sie  $G_1$  als das Allele mit der positiven Wirkung.*

8

### Solution



- b) Use a marker-effect model for the above given dataset. Specify all model components and enter the numbers from the dataset into the model. Setup the mixed model equations to solve for the parameters to be estimated. You can assume a ratio of 1 between the genetic and the residual variance components.

*Verwenden Sie ein Marker-Effekt Modell für den oben gegebenen Datensatz. Geben Sie alle Modellkomponenten an und verwenden Sie die Zahlen des Datensatzes als Information im Modell. Stellen Sie die Mischmodellgleichungen auf für die Lösung nach den zu schätzenden Parameter. Sie können ein Verhältnis von 1 annehmen zwischen der genetischen Varianz und der Varianz der Resteffekte.*

**12**

## **Solution**

- c) Use a breeding-value model for the above given dataset. Specify all model components and enter the numbers from the dataset into the model. Setup the mixed model equations to solve for the parameters to be estimated. You can assume a ratio of 1 between the genetic and the residual variance components.

*Verwenden Sie ein Zuchtwertmodell für den oben gegebenen Datensatz. Geben Sie alle Modellkomponenten an und verwenden Sie die Zahlen des Datensatzes als Information im Modell. Stellen Sie die Mischmodellgleichungen auf für die Lösung nach den zu schätzenden Parameter. Sie können ein Verhältnis von 1 annehmen zwischen der genetischen Varianz und der Varianz der Resteffekte.*

**12**

## **Solution**

## Problem 4: Bayes

- a) In a Bayesian data analysis, we differentiate between known and unknown quantities. Please, separate the relevant quantities for the dataset and the model shown below into known and unknown quantities.

*In einer Bayes'schen Datenanalyse wird zwischen bekannten und unbekannten Grössen unterschieden. Machen Sie die Einteilung für den unten gezeigten Datensatz und das angegebene Modell.*

13

Tier	bBHB	Herd	SNP1	SNP2	SNP3	SNP4	SNP5
1	1.552	1	-1	0	1	1	1
2	1.610	2	1	1	0	1	0
3	1.661	1	-1	0	-1	1	-1
4	1.612	1	1	1	-1	1	0

For the above shown dataset, we assume the following model

*Für den Datensatz nehmen wir das folgende Modell an*

$$y = Xb + Zq + e$$

where: Herd is modelled as a fixed effect ( $b$ ) and SNP-effects ( $q$ ) are taken as random effects.

## Solution

Grösse	bekannt	unbekannt
$y_1$		
$y_2$		
$y_3$		
$y_4$		
$b_{H1}$		
$b_{H2}$		
$X$		
$q_1$		
$q_2$		
$q_3$		
$q_4$		
$q_5$		
$Z$		

- b) In a first step of the Bayesian analysis, the influence of the ‘herd’ on the response variable ‘bBHB’ is to be quantified. This is done by the following junk of R-code. Compute the Bayesian estimate for the intercept and the effects of the two herds, based on the output shown below.

*In einem ersten Schritt der Bayes’schen Analyse soll der Einfluss des Betriebs (‘herd’) auf die Zielgrösse ‘bBHB’ abgeschätzt werden. Dies wird mit dem folgenden R-programm gemacht. Berechnen Sie die Bayes’sche Schätzung für den Achsenabschnitt und die Effekte der beiden Betriebe aufgrund des unten gezeigten Outputs.*

3

```
### # Matrix X as incidence matrix for beta0 and beta1
X <- cbind(1,model_matrix_bhb)
### # y as vector of observations
y <- tbl_herd_bhb_gen_data$bBHB
### # starting values
beta = c(0, 0, 0)
# loop for Gibbs sampler
niter = 10 # number of samples
for (iter in 1:niter) {
  # sampling intercept
  w = y - X[, 2] * beta[2] - X[, 3] * beta[3]
  x = X[, 1]
  xpxi = 1/(t(x) %*% x)
  betaHat = t(x) %*% w * xpxi
  beta[1] = rnorm(1, betaHat, sqrt(xpxi)) # using residual var = 1
  # sampling herd 1
  w = y - X[, 1] * beta[1] - X[, 3] * beta[3]
  x = X[, 2]
  xpxi = 1/(t(x) %*% x)
  betaHat = t(x) %*% w * xpxi
  beta[2] = rnorm(1, betaHat, sqrt(xpxi)) # using residual var = 1
  # sampling herd 2
  w = y - X[, 1] * beta[1] - X[, 2] * beta[2]
  x = X[, 3]
  xpxi = 1/(t(x) %*% x)
  betaHat = t(x) %*% w * xpxi
  beta[3] = rnorm(1, betaHat, sqrt(xpxi)) # using residual var = 1

  # output current sample
  cat("iteration: ", iter,
      " -- intercept: ", beta[1],
      " -- herd1: ", beta[2],
      " -- herd2: ", beta[3], "\n")
}

## iteration: 1 -- intercept: 1.445489 -- herd1: 0.8000551 -- herd2: 3.333651
## iteration: 2 -- intercept: -0.004837851 -- herd1: 0.6981241 -- herd2: 2.561529
## iteration: 3 -- intercept: 1.013742 -- herd1: 0.3023628 -- herd2: 4.52377
## iteration: 4 -- intercept: -0.145773 -- herd1: 2.273671 -- herd2: 0.9943099
## iteration: 5 -- intercept: -0.1065868 -- herd1: 2.753995 -- herd2: 1.529626
## iteration: 6 -- intercept: -1.018371 -- herd1: 2.825402 -- herd2: 1.381579
## iteration: 7 -- intercept: 0.3562579 -- herd1: 0.7712689 -- herd2: 3.142257
## iteration: 8 -- intercept: 1.327032 -- herd1: 0.4034003 -- herd2: 3.676941
## iteration: 9 -- intercept: -0.1259976 -- herd1: 1.140415 -- herd2: 0.2413188
## iteration: 10 -- intercept: 1.287475 -- herd1: -0.1194743 -- herd2: 3.745461
```

## **Solution**

- c) How can the effect of the herds be quantified using the least squares method instead of a Bayesian approach? Please specify the command in R that you would use to get to the least squares estimates of the herd-effects. The data to be analysed is available in a dataframe called `tbl_herd_bhb_gen_data` with a column named 'bBHB' for the response variable and a column 'Herd' with the indicators of the herd.

*Wie können die Betriebseffekte mit der Methode der kleinsten Quadrate anstelle eines Bayes'schen Ansatzes geschätzt werden? Bitte geben sie den R-Befehl, welcher die Betriebseffekte schätzt. Die zu analysierenden Daten sind in einem Datenframe namens `tbl_herd_bhb_gen_data` abgelegt. Der Datenframe enthält eine Kolonne 'bBHB' für die Zielgrösse und eine Kolonnen namens 'Herd' für die Betriebsinformationen.*

**3**

## **Solution**