

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

751-7602-00 V
Exam in
Applied Statistical Methods
in Animal Sciences
Summer Semester 2022

Date: 2022-05-30

Name:

Legi-Nr:

Problem	Maximum Number of Points	Number of Points Reached
1	15	
2	6	
3	9	
4	33	
5	30	
Total	93	

Questions in German are in italics

Problem 1: Fixed Linear Effects Model

The following dataset on the slaughter weight and the sex of 12 beef animals is given.

Gegeben ist der folgende Datensatz zum Schlachtgewicht und zum Geschlecht von 12 Fleischrindern.

Animal	Sex	Slaughter Weight
1	female	276.4
2	female	274.6
3	female	285.3
4	female	275.7
5	castrate	304.2
6	castrate	293.4
7	castrate	284.8
8	castrate	296.5
9	male	286.6
10	male	300.9
11	male	290.5
12	male	302.9

The data is available from the address below and can be read by the function `readr::read_csv()`

Die Daten sind unter der nachfolgenden Adresse verfügbar und können mit der Funktion `readr::read_csv()` gelesen werden.

```
## https://charlotte-ngs.github.io/asmss2022/data/asm\_exam\_p01.csv
```

- a) Do an F-test with the data above to answer the question whether the fixed effect of the ‘Sex’ of the animal has any influence at all on the slaughter weight.

Verwenden Sie einen F-Test zur Beantwortung der Frage ob der fixe Effekt des Geschlechts des Tieres überhaupt einen Einfluss auf das Schlachtgewicht hat.

2

Solution

- b) Fit the linear fixed effects model showing the effects of the different levels of ‘Sex’ on ‘Slaughter Weight’. What is the order of the different levels of the factor ‘Sex’ when ordering them according to the size of the effect obtained from the fitted model?

Passen Sie ein lineares fixes Modell an die Daten an, welches den Einfluss des Geschlechts auf das Schlachgewicht zeigt. Wie lautet die Reihenfolge der Effektstufen des Faktors Geschlecht, wenn diese nach der Effektgrösse aus dem geschätzten Modell sortiert werden?

4

Solution

- c) Show how the different effect estimates (Intercept and factor levels of ‘Sex’) are computed from a solution to the least squares normal equations using the data on slaughter weight and sex when treatment contrasts are used.

Zeigen Sie wie die Schätzwerte der verschiedenen Effekte (Achsenabschnitt und die Faktoren des Geschlechtereffekts) aus einer Lösung der Least Squares-Normalgleichungen berechnet werden für die Daten zum Schlachtgewicht und Geschlecht unter der Verwendung von Treatment-Kontrasten.

9

Solution

Problem 2: Linear Regression

The following dataset contains the logarithm of methane emission ($\ln\text{CH}_4$) and the logarithm of dry matter intake ($\ln\text{DMI}$) of 15 cows.

Der folgende Datensatz enthält die logarithmierten Werte der Methanemissionen ($\ln\text{CH}_4$) und der täglichen Futteraufnahme ($\ln\text{DMI}$) für 15 Kühe.

Animal	$\ln\text{DMI}$	$\ln\text{CH}_4$
1	9.51	12.97
2	13.79	20.15
3	10.30	11.70
4	9.35	15.23
5	12.87	17.74
6	1.60	0.85
7	10.77	16.35
8	13.09	19.73
9	6.33	8.79
10	5.38	11.05
11	2.74	3.16
12	9.66	12.79
13	11.15	11.77
14	8.08	16.95
15	4.80	4.70

The data is available from the address below and can be read by the function `readr::read_csv()`

Die Daten sind unter der nachfolgenden Adresse verfügbar und können mit der Funktion `readr::read_csv()` gelesen werden.

```
## https://charlotte-ngs.github.io/asmss2022/data/asm\_exam\_p02.csv
```

a) Fit the linear regression model of 'lCH4' on 'IDMI'.

Passen Sie ein lineares Regressionsmodell von 'lCH4' auf 'IDMI' an.

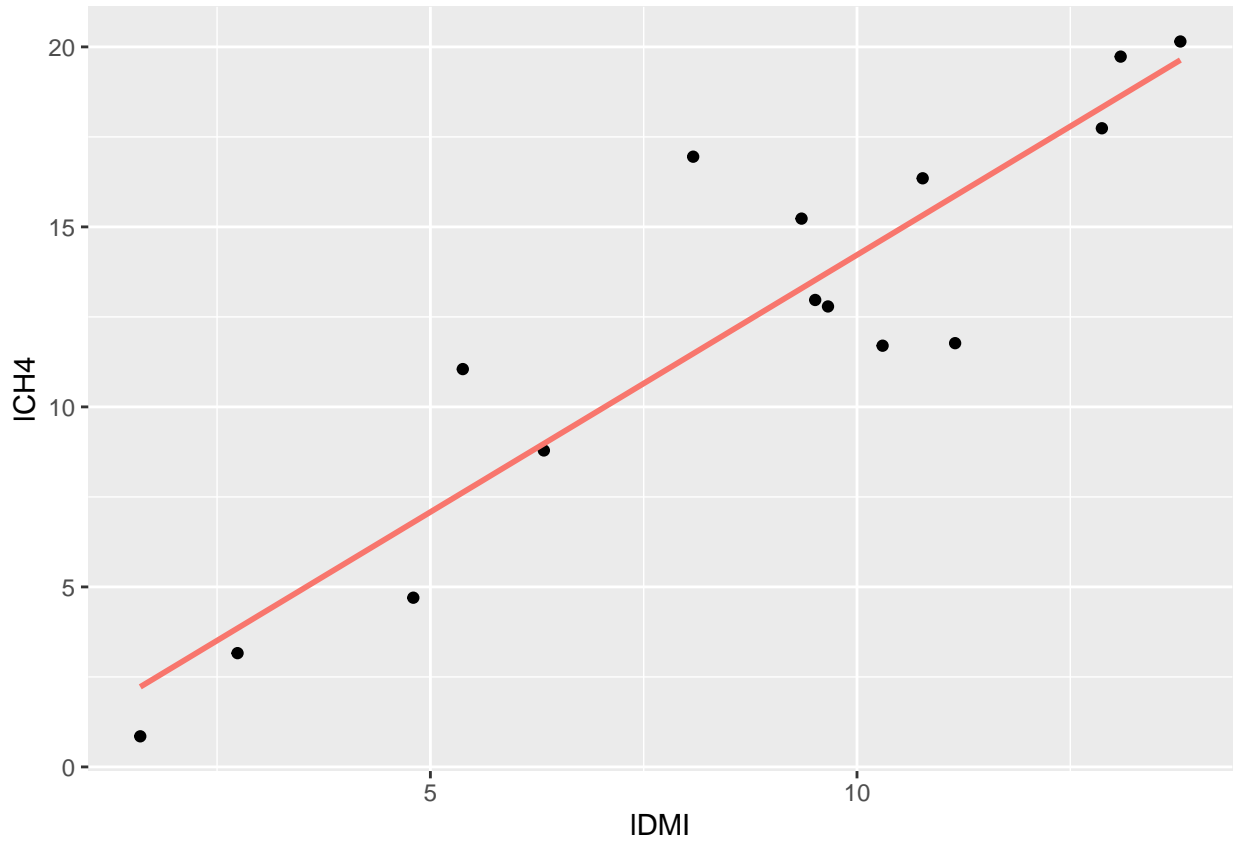
2

Solution

- b) Show in the plot below, the estimates of the model coefficients obtained from the linear regression in Problem 2a. For a selected example observation, show the fitted value and the residual belonging to that selected observation.

Zeigen Sie im nachfolgenden Plot die geschätzten Modellkoeffizienten der linearen Regression aus der Aufgabe 2a). Für einen bestimmten Beobachtungswert zeigen Sie den Modellschätzwert und das Residuum, welches zur ausgewählten Beobachtung gehört.

4



Solution

Problem 3: Model Selection

The following dataset contains fat yield (**fat**) of dairy cows as a response variable. Lactation number (**lact**), days in milk (**dim**) and height of the cow (**hei**) are available as predictor variables.

*Der folgende Datensatz enthält Fettleistung (**fat**) von Milchkühen als eine Zielvariable. Laktationsnummer (**lact**), Laktationslänge in Tagen (**dim**) und Grösse der Kuh (**hei**) sind verfügbar als beschreibende Variablen.*

id	lact	dim	fat	hei
5649	2	301	936	153
5361	1	447	955	150
4312	2	368	615	147
6324	1	294	984	149
6190	2	384	1040	150
5923	2	399	899	148
6187	1	280	1198	149
5653	3	464	903	148
5670	4	305	881	145
5498	1	321	832	147
5907	1	328	1090	150
6299	3	291	745	148
5224	2	295	805	153
5339	3	337	822	145
6187	2	342	1121	147
5795	3	297	1146	148
6255	1	610	1204	146
6507	2	305	794	149
6470	1	323	761	148
5737	4	273	935	146

The data is available from the address below and can be read by the function `readr::read_csv()`

Die Daten sind unter der nachfolgenden Adresse verfügbar und können mit der Funktion `readr::read_csv()` gelesen werden.

```
## https://charlotte-ngs.github.io/asmss2022/data/asm\_exam\_p03.csv
```

- a) Use model selection based on the C_p -value on the above dataset to find the best model. Which predictor variables are included in the best model based on the C_p -value? Which are the parameter estimates of the best model?

Verwenden Sie Modellselektion basierend auf dem C_p -Wert für den oben gegebenen Datensatz. Welche beschreibenden Variablen sind im besten Modell nach C_p -Wert enthalten? Wie lauten die geschätzten Parameter des besten Modells?

6

Solution

- b) Verify the result of the model selection using an analysis of variance (`aov()`) on the full model. Are you getting the same result as shown in Problem 3a?

Verifizieren Sie das Resultat der Modellselektion mit einer Varianzanalyse (`aov()`) auf dem vollen Modell. Erhalten Sie das gleiche Resultat, wie in Aufgabe 3a?

3

Solution

Problem 4: Pedigree-Based BLUP

The dataset shown below shows observations of a trait called P for 6 animals. The phenotypic variance is assumed to be 80. The heritability is 0.2.

Der unten gezeigte Datensatz zeigt Beobachtungen eines Merkmals namens P für 6 Tiere. Die phänotypische Varianz beträgt 80. Die Erbllichkeit ist 0.2.

ID	SIRE	DAM	SEX	P
4	1	3	m	22.4
5	2	3	m	49.2
6	1	NA	m	18.0
7	2	NA	f	11.7
8	5	7	f	27.3
9	5	7	f	28.3

The data is available from the address below and can be read by the function `readr::read_csv()`

Die Daten sind unter der nachfolgenden Adresse verfügbar und können mit der Funktion `readr::read_csv()` gelesen werden.

```
## https://charlotte-ngs.github.io/asmss2022/data/asm\_exam\_p04.csv
```

- a) Use the above shown dataset to predict breeding values using a sire model. In that model include **SEX** as a fixed effect. Specify all model components with expected values and variance-covariance matrices for all random effects in the model. The ratio between residual variance and sire variance can be assumed as 19.

*Schätzen Sie Zuchtwerte mit dem oben gezeigten Datensatz mit einem Vatermodell. In diesem Modell soll **SEX** als fixer Effekt modelliert werden. Geben Sie alle Modellkomponenten an und spezifizieren Sie Erwartungswerte und Varianz-Kovarianzmatrizen für alle zufälligen Effekte im Modell. Das Verhältnis zwischen Restvarianz und Vatervarianz kann angenommen werden als 19.*

15

Solution

- b) Predict breeding values for all animals using an animal model. SEX is modelled as fixed effect. Specify all model components with expected values and variance-covariance matrices for all random effects in the model.

Schätzen Sie Zuchtwerte für alle Tiere mit dem Tiermodell. SEX soll als fixer Effekt modelliert werden. Geben Sie alle Modellkomponenten an und spezifizieren Sie Erwartungswerte und Varianz-Kovarianzmatrizen für alle zufälligen Effekte im Modell.

15

Solution

c) Compare the order of the sires according to the predicted breeding values from Problem 4a and 4b.

Vergleichen Sie die Reihenfolge der Stiere aufgrund der geschätzten Zuchtwerte aus den Aufgaben 4a und 4b.

3

Solution

Problem 5: Genomic Prediction of Breeding Values

The following dataset is used to predict genomic breeding values.

Der nachfolgende Datensatz wird zur Schätzung von genomischen Zuchtwerten verwendet.

ID	SIRE	DAM	SEX	P	SNP1	SNP2	SNP3
5	1	3	m	70.5	0	0	1
6	2	3	f	74.5	1	1	1
7	1	4	m	79.7	0	0	1
8	2	4	f	111.1	2	0	2
9	7	8	f	63.1	1	0	2
10	5	6	f	89.2	1	1	2
11	7	8	f	82.2	1	0	1
12	5	6	f	59.3	1	1	1

The data is available from the address below and can be read by the function `readr::read_csv()`

Die Daten sind unter der nachfolgenden Adresse verfügbar und können mit der Funktion `readr::read_csv()` gelesen werden.

```
## https://charlotte-ngs.github.io/asmss2022/data/asm\_exam\_p05.csv
```


- a) Predict genomic breeding values based on the dataset shown above using a marker effect model. The ratio between residual variance and marker effect is 1.

Schätzen Sie genomische Zuchtwerte aufgrund des oben gezeigten Datensatzes mit einem Markereffektmodell. Das Verhältnis zwischen Restvarianz und Markervarianz beträgt 1.

15

Solution

- b) Predict genomic breeding values based on the dataset shown above using a breeding value-based model. The ratio between residual variance and genomic variance is 1.

Schätzen Sie genomische Zuchtwerte aufgrund des oben gezeigten Datensatzes mit einem Zuchtwertmodell. Das Verhältnis der Restvarianz zur genomischen Varianz beträgt 1.

15

Solution