

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

751-7602-00 V

Exam in

Applied Statistical Methods

in Animal Sciences

Spring Semester 2024

Date: 2024-05-27

Name:

Legi-Nr:

Problem	Maximum Number of Points	Number of Points Reached
1	22	
2	18	
3	12	
4	16	
5	40	
Total	108	

Questions in German are in italics

Problem 1: Linear Regression

According to Figure 1 of (Chen et al. 2023) the concentration of citrate in milk between days in milk (DIM) 20 and 50 can be modeled by a linear regression. The dataset shown below contains measurements of 20 cows with their respective days in milk.

Gemäss Abbildung 1 von (Chen et al. 2023) kann die Zitratkonzentration in der Milch zwischen DIM 20 und 50 mit einer linearen Regression modelliert werden. Der nachfolgend gezeigte Datensatz enthält Beobachtungen von 20 Kühen.

Cow	DIM	Citrate
1	22	9.10
2	22	9.23
3	22	9.10
4	24	8.68
5	24	9.10
6	24	9.14
7	25	8.69
8	26	8.85
9	27	8.93
10	29	8.67
11	29	8.54
12	30	9.34
13	31	8.63
14	33	8.66
15	35	8.78
16	36	8.29
17	42	8.86
18	44	8.34
19	44	8.72
20	46	8.26

The dataset is available from

https://charlotte-ngs.github.io/asmasss2024/data/asm_exam_2024_p01.csv

- a) Write down the linear regression model in matrix-vector notation. List the meaning of all vectors and matrices used in the model. Furthermore, insert all information known from the dataset into the known components of the dataset.

Geben Sie das lineare Regressionmodell als Gleichung in Matrix-Vektor-Notation an. Spezifizieren Sie alle verwendeten Vektoren und Matrizen im Modell. Die bekannten Grössen aus dem Datensatz soll in die bekannten Komponenten des Modells eingesetzt werden.

10

Solution

b) Fit a linear regression model with Citrate as response variable using the 'lm()' function in R. Specify the estimates obtained by the model fit for the following components

- slope
- intercept
- residual variance

Passen Sie ein lineares Regressionsmodell mit Zitratkonzentration als Zielgrösse an. Verwenden Sie dazu die Funktion 'lm()' in R. Geben Sie die erhaltenen Schätzwerte für die folgenden Komponenten an

- *die Steigung,*
- *der Achsenabschnitt und*
- *die Varianz der Residuen*

6

Solution

c) Which of the two plots reflects the modelling results found under Problem 1b)? If possible insert the following components into the correct plot below

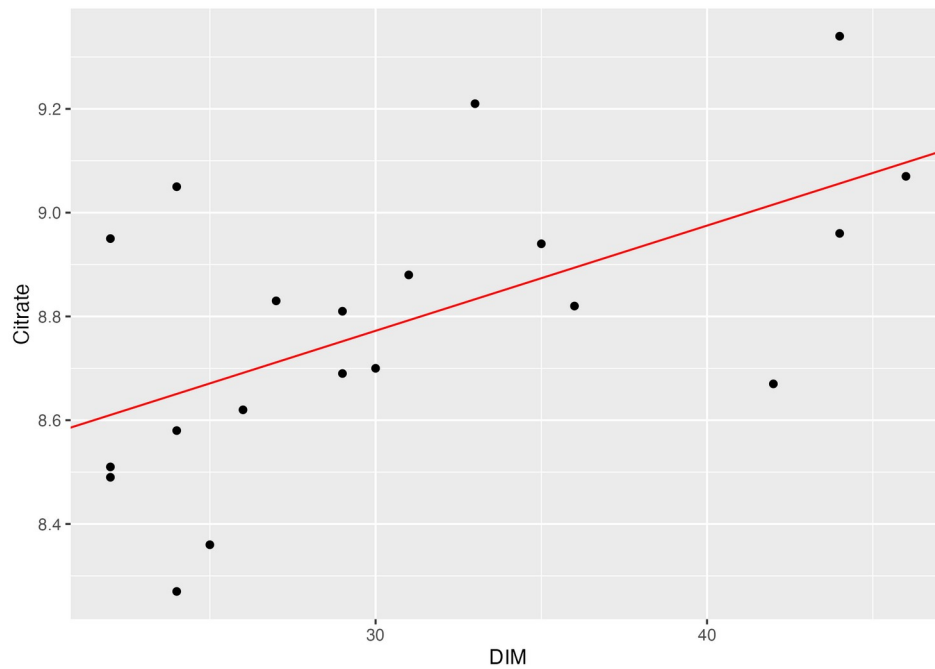
- slope
- intercept
- residual of the data point the the highest citrate concentration

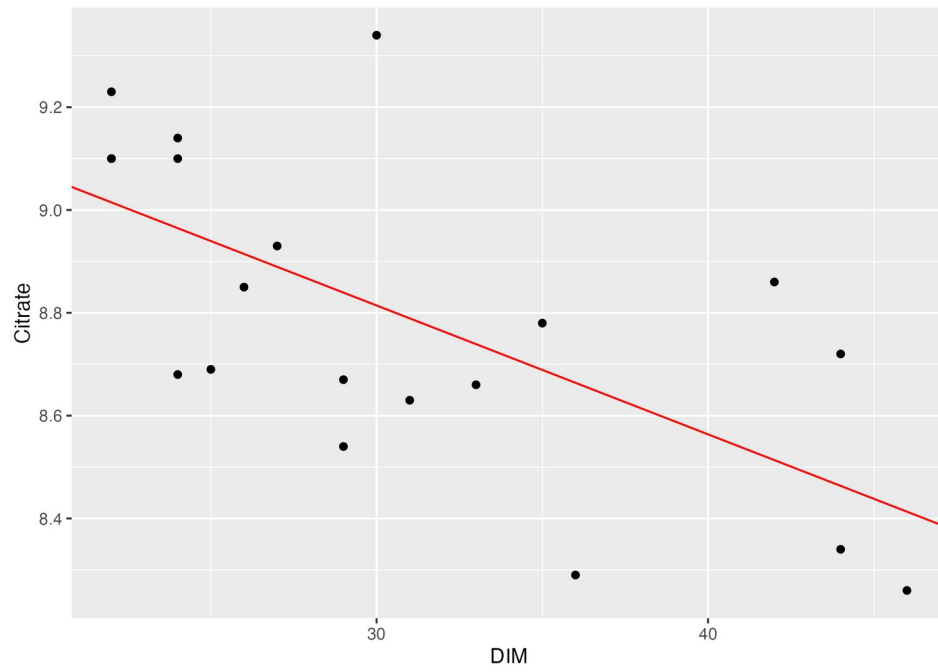
Welcher der beiden nachfolgend gezeigten Plots spiegelt die unter 1b) gefundenen Modellresultate wieder?

- Steigung
- Achsenabschnitt
- Residuum des Datenpunktes mit der höchsten Zitratkonzentration

4

Solution





- d) Citrate can be used as biomarker for a negative energy balance (NEB). NEB can be problematic in very early stages of the lactation. Hence there is a certain demand in predicting citrate concentration in milk for days in milk between 1 and 10. Is it possible to use the modelling results found in 1b) to predict citrate concentration for the time span of DIM between 1 and 10. If yes, how are those citrate concentrations predicted? If no, what is the reason that this is not possible?

Zitratkonzentration wird als Biomarker für eine negative Energiebilanz (NEB) verwendet. Eine NEB kann zu Beginn einer Laktation zu Problemen führen. Somit besteht ein Bedürfnis die Zitratkonzentration in der Milch für die Zeitperiode von DIM zwischen 1 und 10 vorherzusagen. Ist es möglich die Zitratkonzentration für die Zeitperiode von DIM zwischen 1 und 10 mit den Ergebnissen aus 1b) vorherzusagen? Falls ja, wie würde eine solche Vorhersage gemacht werden? Falls nein, was ist der Grund weshalb eine Vorhersage nicht möglich ist?

2

Solution

Problem 2: Fixed Linear Effects Model

In the dataset below, the influence of the breed on the citrate concentration in milk is investigated.

Im nachfolgenden Datensatz soll der Einfluss der Rasse auf die Zitratkonzentration in der Milch untersucht werden.

Cow	DIM	Breed	Citrate
1	22	BS	8.71
2	22	HO	9.99
3	22	BS	8.84
4	24	BS	8.87
5	24	BS	9.03
6	24	BS	8.57
7	25	BS	9.25
8	26	HO	9.22
9	27	HO	9.57
10	29	BS	8.66
11	29	BS	8.80
12	30	BS	8.80
13	31	HO	9.67
14	33	HO	9.21
15	35	HO	9.63
16	36	HO	8.83
17	42	HO	8.76
18	44	HO	8.95
19	44	BS	8.67
20	46	BS	8.56

The dataset is available from

https://charlotte-ngs.github.io/asmasss2024/data/asm_exam_2024_p02.csv

- a) Fit a linear fixed effects model to the dataset shown above. Use ‘Citrate’ as response variable, ‘DIM’ as regression predictor and ‘Breed’ as fixed effect.

Passen Sie ein lineares Modell mit fixen Effekten an den oben gezeigten Datensatz an. Verwenden Sie Citrate als Zielgrösse, DIM als Regressionsvariable und Breed als fixen Effekt.

6

Solution

b) What are the estimates for the following parameters based on the modelling results of 2a)

- regression slope
- group mean of Citrate for animals of breed BS
- group mean of Citrate for animals of breed HO

Wie lauten die Schätzwerte für die folgenden Parameter aufgrund der Modellresultate

- *Steigung der Regressionsgeraden*
- *Gruppenmittel der Tiere der Rasse BS*
- *Gruppenmittel der Tiere der Rasse HO*

6

Solution

- c) What are the expected value for citrate concentration for animals of breed BS and HO at DIM 30?
Wie gross sind die erwarteten Zitratkonzentrationen für Tiere der Rasse BS und HO bei DIM 30?

6

Solution

Problem 3: Interactions

Use the following dataset to fit a fixed linear model assuming that the regression of **Citrate** on **DIM** shows an interaction with the **Breed** factor.

Verwenden Sie den folgenden Datensatz für die Anpassung eines fixen linearen Modells unter der Annahme, dass die Regression von Citrate auf DIM eine Interaktion mit dem Faktor Breed zeigt.

Cow	DIM	Breed	Citrate
1	22	BS	8.21
2	22	HO	9.51
3	22	BS	8.34
4	24	BS	8.37
5	24	BS	8.53
6	24	BS	8.07
7	25	BS	8.75
8	26	HO	8.74
9	27	HO	9.10
10	29	BS	8.16
11	29	BS	8.30
12	30	BS	8.30
13	31	HO	9.20
14	33	HO	8.75
15	35	HO	9.16
16	36	HO	8.36
17	42	HO	8.30
18	44	HO	8.49
19	44	BS	8.17
20	46	BS	8.06

The dataset is available from

https://charlotte-ngs.github.io/asmasss2024/data/asm_exam_2024_p03.csv

- a) Fit a fixed linear effects model including an interaction term between ‘DIM’ and ‘Breed’.

Passen Sie ein fixes lineares Modell an unter Berücksichtigung der Interaktion zwischen DIM und Breed.

6

Solution

- b) What is the expected difference in citrate concentration in the milk for animals of both breeds ('BS' and 'HO') between DIM 30 and 40?

Was beträgt der erwartete Unterschied in der Zitratkonzentration in der Milch für Tiere der beiden Rassen BS und HO zwischen DIM 30 und 40?

6

Solution

Problem 4: Contrasts

The dataset below is used to fit a fixed linear effects model with **Citrate** as response and **Breed** as fixed effect.

Der nachfolgende Datensatz wird für die Anpassung eines fixen linearen Modells mit Citrate als Zielgrösse und Breed als fixen Effekt verwendet

Cow	Breed	Citrate
1	BS	8.71
2	HO	9.99
3	BS	8.84
4	BS	8.87
5	BS	9.03
6	BS	8.57
7	BS	9.25
8	HO	9.22
9	HO	9.57
10	BS	8.66
11	BS	8.80
12	BS	8.80
13	HO	9.67
14	HO	9.21
15	HO	9.63
16	HO	8.83
17	HO	8.76
18	HO	8.95
19	BS	8.67
20	BS	8.56

The dataset is available from

https://charlotte-ngs.github.io/asmasss2024/data/asm_exam_2024_p04.csv

- a) Use ‘treatment’ contrasts to fit a fixed linear model with ‘Citrate’ as response and ‘Breed’ as fixed effect. Verify the obtained ‘Breed’-effect estimates using appropriate group means of ‘Citrate’ and solutions of the least-squares normal equations.

Verwenden Sie ‘treatment’ Kontraste für die Anpassung des fixen linearen Modells mit ‘Citrate’ als Zielgrösse und ‘Breed’ als fixen Effekt. Verifizieren Sie die erhaltenen Schätzungen der ‘Breed’-Effekte anhand der passenden Gruppen-Mittel und anhand von Lösungen der Least-Squares Normalgleichungen.

8

Solution

- b) Use 'sum' contrasts to fit a fixed linear model with 'Citrate' as response and 'Breed' as fixed effect. Verify the obtained 'Breed'-effect estimates using appropriate group means of 'Citrate' and solutions of the least-squares normal equations.

Verwenden Sie 'sum' Kontraste für die Anpassung des fixen linearen Modells mit 'Citrate' als Zielgrösse und 'Breed' als fixen Effekt. Verifizieren Sie die erhaltenen Schätzungen der 'Breed'-Effekte anhand der passenden Gruppen-Mittel und anhand von Lösungen der Least-Squares Normalgleichungen.

8

Solution

Problem 5: Linear Mixed Effects Models

Use the dataset below to fit a linear mixed effects model. In this model **Citrate** is the response and **DIM** and **Breed** are regression variables and fixed effects, respectively. You do not have to include any interaction effects.

Verwenden Sie den nachfolgenden Datensatz um ein lineares gemischtes Modell anzupassen. In diesem Modell sei 'Citrate' die Zielgrösse und 'DIM' die Regressionvariable und 'Breed' ein fixer Effekt. Interaktionseffekte müssen nicht berücksichtigt werden.

ID	Sire	Dam	DIM	Breed	Citrate
5	1	NA	22	BS	9.30
6	1	NA	22	HO	9.26
7	1	5	22	BS	8.92
8	1	5	24	BS	9.30
9	1	6	24	BS	9.03
10	1	6	24	BS	8.82
11	2	7	25	BS	9.09
12	2	7	26	HO	9.04
13	2	8	27	HO	9.71
14	2	8	29	BS	8.52
15	2	9	29	BS	9.18
16	3	9	30	BS	8.85
17	3	10	31	HO	9.04
18	3	10	33	HO	9.04
19	3	11	35	HO	9.44
20	3	11	36	HO	9.18
21	3	12	42	HO	8.88
22	4	12	44	HO	8.96
23	4	13	44	BS	8.74
24	4	13	46	BS	8.52

The dataset is available from

https://charlotte-ngs.github.io/asmasss2024/data/asm_exam_2024_p05.csv

- a) Use a sire model to predict breeding values for all sires. Please specify the model in the form of an equation in matrix-vector notation and explain the meaning of all model components. Write down the expected values and the variance-covariance matrices for all random effects in the model. Put all the information from the data into the model components consisting of response variable, regression and fixed effects, random effects, and their associated design matrices. Construct the mixed model equations and provide solutions to these equations. Extract predicted breeding values for all sires and estimates for the regression and fixed effects from these solutions.

The following assumptions can be made

- the ratio λ_s between residual variance and sire variance is given by $\lambda_s = \sigma_e^2/\sigma_s^2 = 7.11$
- the sires are assumed to be unrelated

Verwenden Sie ein Vatermodell zur Schätzung der Zuchtwerte aller männlichen Tiere. Bitte geben Sie die Modellgleichung in Matrix-Vektor-Schreibweise an und erläutern Sie die Bedeutung aller Modellkomponenten. Schreiben Sie die Erwartungswerte und die Varianz-Kovarianzmatrizen aller zufälligen Effekte im Modell auf. Verwenden Sie alle Informationen aus den Daten für die genaue Spezifikation der Modellkomponenten, wie zum Beispiel Zielgrösse, Regressionseffekte, fixe Effekte, zufällige Effekte und die zugehörigen Designmatrizen. Stellen Sie die Mixed-Model-Gleichungen auf und lösen Sie diese Gleichungen. Extrahieren Sie die geschätzten Zuchtwerte aller männlichen Tiere und die Schätzungen für die Regressionseffekte und für die fixen Effekte aus den Lösungen.

Die folgenden Annahmen können getroffen werden:

- das Verhältnis (λ_s) zwischen der Residuenvarianz und der Vätervarianz beträgt $\lambda_s = \sigma_e^2/\sigma_s^2 = 7.11$
- die Väter sind untereinander nicht verwandt

20

Solution

- b) Use an animal model to predict breeding values for all animals in the dataset, including sires and dams. Please specify the model in the form of an equation in matrix-vector notation and explain the meaning of all model components. Write down the expected values and the variance-covariance matrices for all random effects in the model. Put all the information from the data into the model components consisting of response variable, regression and fixed effects, random effects, and their associated design matrices. Construct the mixed model equations and provide solutions to these equations. Extract predicted breeding values for all animals and estimates for the regression and fixed effects from these solutions.

The following assumptions can be made

- the ratio λ_u between residual variance and additive genetic variance is given by $\lambda_u = \sigma_e^2 / \sigma_u^2 = 1.78$

Verwenden Sie ein Tiermodell zur Schätzung der Zuchtwerte aller Tiere im Datensatz inklusive von Vätern und Müttern. Bitte geben Sie die Modellgleichung in Matrix-Vektor-Schreibweise an und erläutern Sie die Bedeutung aller Modellkomponenten. Schreiben Sie die Erwartungswerte und die Varianz-Kovarianzmatrizen aller zufälligen Effekte im Modell auf. Verwenden Sie alle Informationen aus den Daten für die genaue Spezifikation der Modellkomponenten, wie zum Beispiel Zielgrösse, Regressionseffekte, fixe Effekte, zufällige Effekte und die zugehörigen Designmatrizen. Stellen Sie die Mixed-Model-Gleichungen auf und lösen Sie diese Gleichungen. Extrahieren Sie die geschätzten Zuchtwerte aller Tiere und die Schätzungen für die Regressionseffekte und für die fixen Effekte aus den Lösungen.

Die folgenden Annahmen können getroffen werden:

- das Verhältnis (λ_u) zwischen der Residuenvarianz und der additiv-genetischen Varianz beträgt $\lambda_u = \sigma_e^2 / \sigma_u^2 = 1.78$

20

Solution