ETH Zürich
D-USYS
Institut für Agrarwissenschaften

Lösungen zur Prüfung Züchtungslehre HS 2016

Peter von Rohr

Datum	*23. Dezember 2016*
Beginn	*09:15 Uhr*
Ende	*11:15 Uhr*

Name:

Legi-Nr:

Aufgabe	Maximale Punktzahl	Erreichte Punktzahl
1	52	
2	17	
3	24	
4	30	
5	25	
Total	148	

Aufgabe 1: Verwandtschaft und Inzucht (52)

Gegeben ist das folgende Pedigree.

```
  sire  dam
1 <NA> <NA>
2 <NA> <NA>
3 <NA> <NA>
4    1    2
5    1    4
6    1    4
```

a) Stellen Sie die additiv genetische Verwandtschaftsmatrix A auf

36

Lösung:

$$A = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.500 & 0.750 & 0.750 \\ 0.000 & 1.000 & 0.000 & 0.500 & 0.250 & 0.250 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ 0.500 & 0.500 & 0.000 & 1.000 & 0.750 & 0.750 \\ 0.750 & 0.250 & 0.000 & 0.750 & 1.250 & 0.750 \\ 0.750 & 0.250 & 0.000 & 0.750 & 0.750 & 1.250 \end{bmatrix}$$

b) Welche Tiere im gezeigten Pedigree sind ingezüchtet und wie gross sind die Inzuchtkoeffizienten F_X? Bitte vervollständigen Sie die nachfolgende Tabelle.

12

Tier ID	ingezüchtet (ja/nein)	Inzuchtkoeffizient F_X
1		
2		
3		
4		
5		
6		

Lösung:

Tier Id	ingezuechtet (ja/nein)	Inzuchtkoeffizient
1	nein	0.00
2	nein	0.00
3	nein	0.00
4	nein	0.00
5	ja	0.25
6	ja	0.25

c) In der Paarungsplanung geht es oft darum Inzucht zu verhindern. Wenn wir einen Paarungspartner für Kuh 6 im oben gezeigten Pedigree suchen, kommen die Stiere 1, 3 oder 5 in Frage. Wählen Sie den Paarungspartner so aus, dass das Kalb aus der Paarung nicht ingezüchtet ist. Begründen Sie Ihre Auswahl des Paarungspartners.

4

Lösung:

Der Paarungspartner muss so gewählt werden, dass er nicht mit Kuh 6 verwandt ist. Dies ist nur für den Paarungspartner 3 der Fall. Als Kontrolle können wir die das Pedigree erweitern und die Inzuchtkoeffizienten nochmals rechnen.

```
  sire  dam
1 <NA> <NA>
2 <NA> <NA>
3 <NA> <NA>
4    1    2
5    1    4
6    1    4
7    3    6
```

Die Inzuchtkoeffizienten der Tiere lauten:

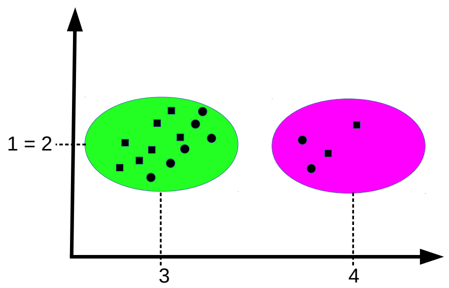
```
[1] 0.00 0.00 0.00 0.00 0.25 0.25 0.00
```

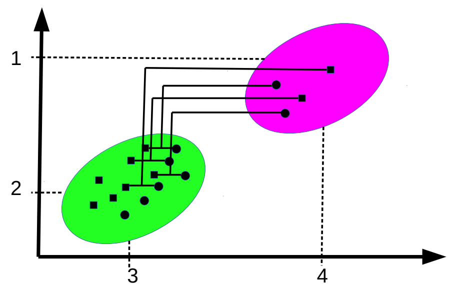
Wobei der Inzuchtkoeffizient von Tier 7 ist gleich: 0

Aufgabe 2: Selektion und Selektionsindex (17)

a) Was veranschaulichen die zwei folgenden Diagramme. Benennen Sie die Punkte 1 bis 4

8





Lösung:

In den beiden Diagrammen wird der Unterschied zwischen gerichteter Selektion in Zuchtpopulationen und Vermehrung in Wildtier oder Produktions- und Vermehrungspopulationen dargestellt.

Die vier Punkte bedeuten

1. Nachkommendurchschnitt
2. Elterndurchschnitt
3. Eltern
4. Nachkommen

b) Die Theorie des Selektionsindexes zeigt auf, wie der Gesamtzuchtwert H aufgrund von verfügbaren Informationsquellen geschätzt werden kann. Dabei werden die Informationsquellen durch einen Vektor x repräsentiert und mit einem Vektor b von unbekannten Indexgewichten zu einem Index I zusammengefasst. Das Ziel ist nun den Vektor b so zu bestimmen, dass I den Gesamtzuchtwert H möglichst genau schätzt. Das heisst, die Fehlervarianz $var(H - I)$ soll minimal sein.

Wie lauten die Indexgleichungen zur Bestimmung der Indexgewichte b, aufgrund der Anforderung der minimalen Fehlervarianz? Benennen Sie die Komponenten in den Indexgleichungen.

5

Lösung:
Die Indexgleichungen lauten

$$Pb = Gv$$

mit	P	Covarianzmatrix zwischen den Informationsquellen x
	b	Vektor der unbekannten Indexgewichte
	G	Covarianzmatrix zwischen Informationsquellen und Merkmalen im Gesamtzuchtwert
	v	Vektor der wirtschaftlichen Gewichte

c) Eine Fleischrinderzuchtorganisation entwirft ein neues Zuchtziel mit den zwei Merkmalen Geburtsgewicht (`GBG`) und Absetzgewicht (`ABG`). Die gleichen Merkmale wie im Zuchtziel, werden bei den Selektionskandidaten erhoben und stehen in Form von geschätzten Zuchtwerten als Informationsquellen zur Verfügung. Die wirtschaftlichen Gewichte der beiden Merkmale im Zuchtziel betragen 3.00 Fr/kg für `GBG` und 2.50 Fr/kg für `ABG`. Wie lauten die Gewichtungsfaktoren für einen Index aus den Informationsquellen `GBG` und `ABG`?

4

Lösung:
Da die gleichen Merkmale im Zuchtziel und im Index sind, dann gilt $b = v$ und somit ist

$$b = \left[\begin{array}{c} 3.0 \\ 2.5 \end{array}\right]$$

Aufgabe 3: Varianzanalyse (24)

Aufgrund des folgenden Datensatzes aus der Pflanzenzucht für das Merkmal `Stengellänge` sollen Varianzkomponenten geschätzt werden.

Pflanze	Sorte	Messperson	Stengellaenge
1	S1	A	13.00
2	S1	A	11.30
3	S1	B	12.20
4	S2	A	10.90
5	S2	C	20.70
6	S2	C	21.20
7	S3	C	13.80
8	S3	A	19.00
9	S3	B	12.40
10	S3	A	17.40

a) In einer ersten Analyse soll nur der Einfluss der Sorte auf die Stengellänge betrachtet werden. Dabei soll Sorte als fixer Effekt modelliert werden. Schätzen Sie die Restvarianz σ_e^2 aufgrund der Residuen für das folgende Modell.

10

Das Modell mit der Sorte als fixen Effekt.

$$y = Xb + e$$

mit	y	Vektor der gemessenen Stengellängen
	b	Vektor der fixen Sorteneffekte
	X	Inzidenzmatrix für b
	e	Resteffekte

Wir nehmen an, dass die Reste unabhängig sind und somit gilt, dass $var(e) = I * \sigma_e^2$ ist. Die geschätzten Sorteneffekte aus dem oben gezeigten Regressionsmodell lauten

```
> lmRegModelSorte <- lm(Stengellaenge ~ -1 + Sorte, data = dfMlrData)
> round(coefficients(lmRegModelSorte), digits = 2)
```

```
SorteS1 SorteS2 SorteS3
  12.17   17.60   15.65
```

Lösung:
Der Residuenvektor lautet

```
> vecRes <- residuals(lmRegModelSorte)
> print(round(vecRes, digits = 2))
```

```
   1     2     3     4     5     6     7     8     9    10 
 0.83 -0.87  0.03 -6.70  3.10  3.60 -1.85  3.35 -3.25  1.75 
```

Die geschätzte Restvarianz beträgt somit

```
> nRestVarEst <- crossprod(vecRes)/(nrow(dfMlrData) - nNrSorte)
> cat("Restvarianz: ", round(nRestVarEst, digits = 2))
```

```
Restvarianz:  13.88
```

```
> cat("Reststandardabweichung: ", round(sqrt(nRestVarEst), digits = 2))
```

```
Reststandardabweichung:  3.73
```

Als Vergleich dazu, was `lm()` liefert

```
> summary(lmRegModelSorte)
```

```
Call:
lm(formula = Stengellaenge ~ -1 + Sorte, data = dfMlrData)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.7000 -1.6042  0.4333  2.7625  3.6000 

Coefficients:
        Estimate Std. Error t value Pr(>|t|)    
SorteS1   12.167      2.151   5.656  0.00077 ***
SorteS2   17.600      2.151   8.182 7.90e-05 ***
SorteS3   15.650      1.863   8.401 6.66e-05 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 3.726 on 7 degrees of freedom
Multiple R-squared:  0.9603,	Adjusted R-squared:  0.9433 
F-statistic:  56.5 on 3 and 7 DF,  p-value: 2.847e-05
```

b) Schätzen Sie für den gegebenen Datensatz die Restvarianz mit Maximum Likelihood. Wo liegt der Unterschied zur Schätzung aus Aufgabe a)? Weshalb wird die Schätzung der Restvarianz aufgrund der Residuen als "glaubwürdiger" betrachtet?

6

Lösung:
Die Maximum Likelihood Schätzung der Restvarianz ist gleich, wie die Schätzung aufgrund der Residuen bis auf den Faktor vor der Summe der quadrierten Residuen. Dieser lautet bei Maximum Likelihood $1/n$ und bei der Residuen-Methode $1/(n-p)$. Da der Faktor bei ML immer kleiner ist als bei der Residuen-Methode wird die Restvarianz mit ML tendenziell immer unterschätzt.

Somit beträgt der ML-Schätzwert für die Restvarianz

```
> nRestVarEstML <- crossprod(vecRes)/nrow(dfMlrData)
> cat("ML-Schaetzung der Restvarianz: ", round(nRestVarEstML, digits = 2))

ML-Schaetzung der Restvarianz:  9.72

> cat("ML-Schaetzung der Reststandardabweichung: ",
+     round(sqrt(nRestVarEstML), digits = 2))

ML-Schaetzung der Reststandardabweichung:  3.12
```

Die Schätzung der Restvarianz mit der Residuen-Methode wird als glaubwürdiger betrachtet, da diese erwartungstreu ist.

c) Der Einfluss der Messperson soll in einer separaten Analyse untersucht werden, wobei die Messperson als zufälliger Effekt ins Modell einfliessen soll. Schätzen Sie die Varianzkomponenten des zufälligen Effektes der Messperson und des Resteffektes mit Hilfe einer Varianzanalyse. Folgendes Modell soll den Daten zu Grunde liegen.

8

$$y = 1\mu + Zm + e$$

mit	y	Vektor der gemessenen Stengellängen
	μ	allgemeines Mittel
	1	Inzidenzvektor für μ
	m	Vektor der zufälligen Effekte der Messperson
	Z	Inzidenzmatrix für m
	e	Vektor der Resteffekte

Die Varianzen der zufälligen Effekte haben die folgende Struktur:

$$var(m) = I * \sigma_m^2 \quad \text{und} \quad var(e) = I * \sigma_e^2$$

wobei I für die Einheitsmatrix steht.

Die Summenquadrate (SSQ) für die Messpersonen und die Residuen und die Freiheitsgrade (DF) sind in der folgenden Tabelle zusammengefasst.

Effekt	DF	SSQ
Messperson	2	54.694
Residuen	7	88.175

Lösung:

```
> nMsqResMessPers <- nSsqResMessPers/nDfResidual
```

Somit ist die Schätzung der Restvarianz:

$$88.175/7 = 12.596$$

Für die Varianz der Effekte der Messpersonen erhalten wir folgende Schätzung:

```
> nMsqMessPers <- nSSqFitMessPers / nDfMessPers
> nEstVarMessPers <- (nMsqMessPers - nMsqResMessPers)/nAnzObs
```

Die Schätzung der Varianz des Effektes der Messperson lautet:

$$(27.347 - 12.596/10) = 1.475$$

Aufgabe 4: Zuchtwertschätzung (30)

Der folgende Datensatz soll für die Vorhersage von Zuchtwerten verwendet werden.

Tochter	Herde	Vater	Leistung
4	1	C	112.00
5	1	A	105.00
6	2	B	118.00
7	1	B	120.00
8	2	C	135.00
9	2	A	115.00

Die Varianzen sind aus der folgenden Tabelle ersichtlich

Komponenten	Symbol	Wert
phaentypisch	σ_p^2	128.00
genetisch additiv	σ_a^2	32.00

a) Schätzen Sie die Zuchtwerte der Väter aufgrund der Mittelwerte ihrer Nachkommen. Wir nehmen an, dass das allgemeine Populationsmittel $\mu = 117.5$ beträgt.

6

Lösung:
Der geschätzte Zuchtwert $\hat{a}_s$ eines Vaters s aufgrund der Mittelwerte $\tilde{y}_s$ seiner Nachkommen lautet

$$\hat{a}_s = \frac{2n}{n+k}(\tilde{y}_s - \mu)$$

wobei n für die Anzahl Beobachtungen steht und $k = \frac{4-h^2}{h^2} = 15$ ist.
Der Faktor vor der Differenz zwischen $\tilde{y}_s$ und μ ist:

$$\frac{2n}{n+k} = 0.57$$

Somit sind die Zuchtwerte der drei Väter:

Vater	Zuchtwert
A	-4.29
B	0.86
C	3.43

b) Wie lautet das BLUP-Vatermodell in Matrix-Vektor-Schreibweise für den in dieser Aufgabe gezeigten Datensatz. Benennen Sie die einzelnen Modellkomponenten. Geben Sie Erwartungswerte und die Varianzen für alle zufälligen Effekte im Modell an.

12

Lösung:
Das allgemeine BLUP-Vatermodell lautet:

$$y = Xb + Zs + e$$

mit	y	Vektor der phänotypischen Leistungen, entspricht der Kolonne Leistung im Datensatz
	b	Vektor der fixen Herden-Effekte
	X	Inzidenzmatrix verknüpft b mit y
	s	Vektor der zufälligen Vatereffekte
	Z	Inzidenzmatrix verknüpft s mit y
	e	Vektor der Resteffekte

Die Erwartungswerte und die Varianzen für die zufälligen Teile des Modells lauten:

$$E[s] = 0 \quad \text{und} \quad E[e] = 0$$

Somit ist

$$E[y] = Xb$$

.

Die Varianzen von e und s sind definiert als

$$var(s) = G = A * \sigma_s^2 = I * \sigma_s^2 \quad \text{und} \quad var(e) = R = I * \sigma_e^2$$

Da die Väter nicht miteinander verwandt sind, ist in diesem Bespiel $A = I$. Die Covarianzen zwischen s und e werden auf 0 gesetzt. Daraus können wir die Covarianzmatrix von y berechnen als

$$var(y) = ZGZ^T + R = ZZ^T * \sigma_s^2 + I * \sigma_e^2$$

c) Stellen Sie die Mischmodellgleichungen für das BLUP-Vatermodell für den in dieser Aufgabe angegebenen Datensatz an.

12

Lösung:
Die Komponenten der Mischmodellgleichungen des Vatermodells lauten:

- Varianzverhältnis

$$\alpha = \frac{\sigma_e^2}{\sigma_s^2} = \frac{120}{8} = 15$$

- Koeffizientenmatrix

$$X^T X = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

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

$$Z^T Z + I * \alpha = \begin{bmatrix} 17 & 0 & 2 \\ 0 & 17 & 0 \\ 2 & 0 & 17 \end{bmatrix}$$

- Lösungsvektor

$$\begin{bmatrix} \hat{b}_{Herde1} \\ \hat{b}_{Herde2} \\ \hat{s}_{VaterA} \\ \hat{s}_{VaterB} \\ \hat{s}_{VaterC} \end{bmatrix}$$

- Rechte Handseite

$$X^T y = \begin{bmatrix} 345.00 \\ 347.00 \end{bmatrix}$$

$$Z^T y = \begin{bmatrix} 238.00 \\ 240.00 \\ 238.00 \end{bmatrix}$$

- Zusammengesetzt

$$\begin{bmatrix} 3.00 & 2.00 & 1.00 & 0.00 & 1.00 \\ 2.00 & 3.00 & 1.00 & 0.00 & 1.00 \\ 1.00 & 1.00 & 17.00 & 0.00 & 2.00 \\ 0.00 & 0.00 & 0.00 & 17.00 & 0.00 \\ 1.00 & 1.00 & 2.00 & 0.00 & 17.00 \end{bmatrix} \begin{bmatrix} \hat{b}_{Herde1} \\ \hat{b}_{Herde2} \\ \hat{s}_{VaterA} \\ \hat{s}_{VaterB} \\ \hat{s}_{VaterC} \end{bmatrix} = \begin{bmatrix} 345.00 \\ 347.00 \\ 238.00 \\ 240.00 \\ 238.00 \end{bmatrix}$$

Aufgabe 5: Inverse der Verwandtschaftsmatrix (25)

Für das folgende Pedigree soll die Inverse A^{-1} der genetisch additiven Verwandtschaftsmatrix aufgestellt werden.

```
  sire  dam
1 <NA> <NA>
2 <NA> <NA>
3    1    2
4    1 <NA>
5 <NA>    3
```

25

Lösung:

```
> print(as.matrix(getAInv(ped = ped)))
```

```
           1    2          3          4          5
1  1.8333333  0.5 -1.0000000 -0.6666667  0.0000000
2  0.5000000  1.5 -1.0000000  0.0000000  0.0000000
3 -1.0000000 -1.0  2.3333333  0.0000000 -0.6666667
4 -0.6666667  0.0  0.0000000  1.3333333  0.0000000
5  0.0000000  0.0 -0.6666667  0.0000000  1.3333333
```