

# **Lösungen zur Prüfung Angewandte Statistische Methoden in den Nutzwissenschaften FS 2016**

Peter von Rohr

DATUM    *30. Mai 2016*  
BEGINN    *08:00 Uhr*  
ENDE       *08:45 Uhr*

Name:

Legi-Nr:

Aufgabe	Maximale Punktzahl	Erreichte Punktzahl
1	10	
2	13	
3	6	
4	6	
Total	35	

## Aufgabe 1: Modellierung vor und nach Einführung der Genomischen Selektion

- a) Wo liegen die Unterschiede im Bezug auf die Modellierung von Tierzuchtdate vor und nach der Einführung der genomischen Selektion (GS) im Bezug auf die folgendne Punkte?

6

Punkt	vor GS	nach GS
Informationsquellen		
statistisches Modell		
genetisches Modell		

- b) In der genomischen Selektion werden häufig geschätzte BLUP-Zuchtwerte aus einem Tiermodell als Beobachtungen verwendet.
- Nennen sie je einen Vorteil und einen Nachteil der Verwendung von BLUP Zuchtwerten als Beobachtungen
  - Welches Verfahren wird verwendet, um den Nachteil von der Verwendung von BLUP-Zuchtwerten als Beobachtungen, zu beheben und nach welchem Prinzip funktioniert dieses Verfahren?

## Aufgabe 2: Lineare Regression

Gegeben sind die folgenden Resultate einer linearen Regression

Call:

```
lm(formula = y ~ snp1 + snp2, data = dfSnpData)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.6819	-2.9583	0.1485	2.7452	8.4649

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.9661	1.6819	0.574	0.570
snp1	-2.3806	1.0970	-2.170	0.039 *
snp2	6.5272	0.9994	6.531	5.28e-07 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.999 on 27 degrees of freedom

Multiple R-squared: 0.6179, Adjusted R-squared: 0.5896

F-statistic: 21.83 on 2 and 27 DF, p-value: 2.286e-06

a) Aus welchen Komponenten besteht das lineare Modell?

b) Wie sieht das Modell aus, welches zu den oben gezeigten Resultaten geführt hat?

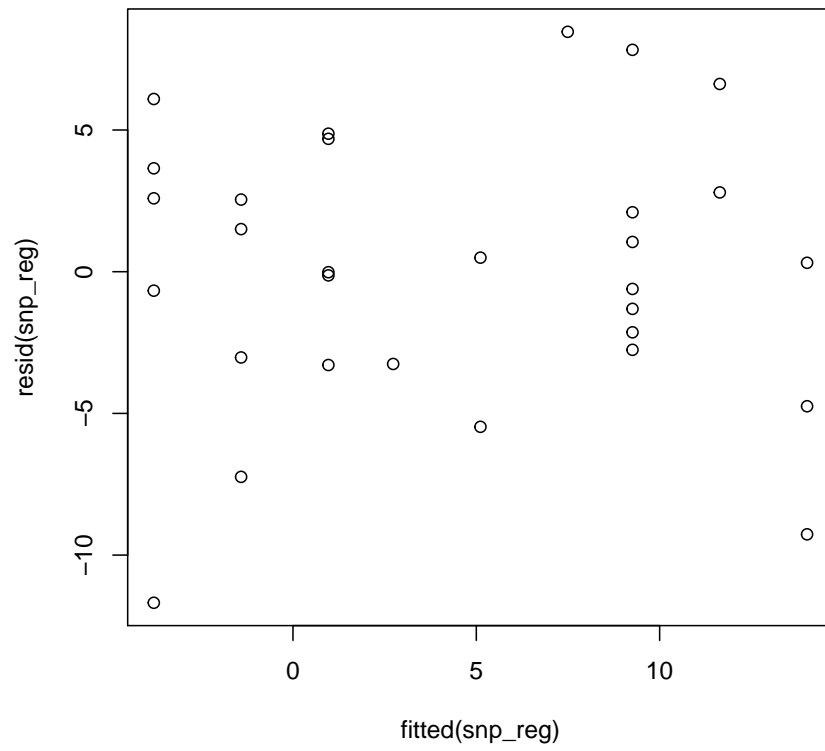
**5**

- c) Berechnen Sie aus den oben gezeigten Resultaten das Vertrauensintervall für die erklärende Variable `snp1`. Wie gross ist die Irrtumswahrscheinlichkeit für dieses Vertrauensintervall?

**3**

d) Wie heisst der folgende Plot und wozu kann dieser Plot verwendet werden?

2



**Zusatz:** Durch welches Statement wird der oben gezeigte Plot in R erzeugt?

**2**



### **Aufgabe 3: LASSO**

- a) Was bedeutet die Abkürzung LASSO?

1

- b) Sobald in einem linearen Modell die Anzahl Parameter grösser ist als die Anzahl Beobachtungen können wir Least Squares nicht verwenden. Was sind in einem solchen Fall Alternativen zu Least Squares?

**3**

- c) Wie unterscheiden sich die Schätzer durch Least Squares vom Schätzer durch LASSO und wie wird die Selektion der Variablen erreicht?

**2**

## **Aufgabe 4: Bayes**

- a) In welche Kategorien unterteilen Bayesianer die Komponenten eines Modells?

**2**

- b) Worauf basieren Schätzungen in der Bayes'schen Statistik, aus welchen Komponente besteht das gesuchte Objekt und wie wird dieses berechnet?

**3**

- c) Angenommen, Sie haben vor der Schätzung eines Parameters keine Information über den Parameter. Wie lassen Sie diese Tatsache in einer Bayes'schen Analyse einfließen?

**1**