

# ASMNW - Lösung 2

*Peter von Rohr*

*2016-05-01*

## Kontrollfrage 1

Weshalb werden in Schritt 1 der genomischen Selektion geschätzte Zuchtwerte anstelle von phänotypischen Beobachtungen verwendet?

### Lösung

- Für bestimmte Tiere sind phänotypische Beobachtungen nicht verfügbar (Ahnen, ausländische Tiere)
- Bei internationalem Austausch: unterschiedliche nationale Modelle berücksichtigen verschiedene Umweltfaktoren

## Kontrollfrage 2

Welche Eigenschaft der BLUP-Zuchtwerte führt zu der problematischen Reduktion der Varianz der geschätzten Zuchtwerte im Vergleich zur Varianz der phänotypischen Werte?

### Lösung

Da bei BLUP Zuchtwerten gilt, dass

$$\text{cov}(\hat{g}_i, g_i) = \text{var}(\hat{g}_i)$$

folgt, dass

$$\text{var}(\hat{g}_i - g_i) = \text{var}(\hat{g}_i) + \text{var}(g_i) - 2 * \text{cov}(\hat{g}_i, g_i) = \text{var}(g_i) - \text{var}(\hat{g}_i) \geq 0$$

$$\rightarrow \text{var}(g_i) \geq \text{var}(\hat{g}_i)$$

## Kontrollfrage 3

Für Schritt 1 der genomischen Selektion wurden folgende Modelle angeschaut:

|                             |  |
|-----------------------------|--|
| wahre Zuchtwerte            | $\mathbf{g} = \mathbf{1}\mu + \mathbf{M}\mathbf{a} + \epsilon$   |
| phänotypische Beobachtungen | $\mathbf{y} = (\mathbf{1}\mu + \mathbf{X}\mathbf{b}) + \mathbf{M}\mathbf{a} + (\epsilon + \mathbf{e})$   |
| geschätzte Zuchtwerte       | $\hat{\mathbf{g}} = \mathbf{1}\mu + \mathbf{M}\mathbf{a} + (\epsilon + (\hat{\mathbf{g}} - \mathbf{g}))$ |

1. Welche Struktur haben diese Modelle
2. Welche Parameterschätzmethode wird üblicherweise für diese Modelle verwendet
3. Wiese kann diese Parameterschätzmethode für die genomische Selektion nicht verwendet werden

## Lösung

1. Regressionsmodelle
2. Least Squares
3. Anzahl der Parameter ist grösser als die Anzahl der Beobachtungen

## Aufgabe 1: Schätzung der a-Werte aufgrund phänotypischer Leistungen

Der Datensatz `geno_sel_step1.csv` enthält phänotypische Werte in der Kolonne `bwt_sim`. Diese Werte wurden aufgrund von sechs genetischen Loci simuliert. Abgesehen von den genetischen Informationen gibt es auch noch zwei fixe Faktoren, welche mit `sex` und `age` bezeichnet sind. Alle Angaben sind in der gleichen Datei enthalten.

### Ziel

Schätzen Sie für die sechs Loci die entsprechenden a-Werte.

### Ihre Aufgaben

Da es sich nur um sechs Loci handelt, können die a-Werte mit einer Regression geschätzt werden. Folgende Schritte sollten zum Ziel führen.

1. Einlesen der Daten mit

```
dfGenoSelDat <- read.csv2(file = "http://charlotte-ngs.github.io/GELASM/w10/geno_sel_step1.csv",
                          stringsAsFactors = FALSE,
                          row.names = 1)
str(dfGenoSelDat)
```

```
## 'data.frame': 496 obs. of 9 variables:
## $ sex : chr "M" "M" "M" "F" ...
## $ age : int 76 78 67 76 84 77 81 72 67 75 ...
## $ rs3722996 : int 2 3 2 1 3 1 3 3 3 3 ...
## $ rs3671256 : int 2 3 2 1 3 1 3 3 2 3 ...
## $ rs13475745 : int 2 3 2 1 3 1 3 3 2 3 ...
## $ rs6404446 : int 2 3 2 1 3 1 3 3 1 3 ...
## $ mCV23695025: int 2 3 2 1 3 1 3 3 1 3 ...
## $ rs6173215 : int 3 3 2 2 3 1 3 3 1 3 ...
## $ bwt_sim : num 61.9 75.1 51.8 28.2 72.4 ...
```

2. Lineare Regression mit funktion `lm()` wobei die phänotypischen Werte in Kolonne `bwt_sim` gespeichert sind.

```
lmBwtSim <- lm(dfGenoSelDat$bwt_sim ~ ., data = dfGenoSelDat)
```

3. Anzeige der Resultate mit Funktion `summary()`

```
summary(lmBwtSim)
```

```

##
## Call:
## lm(formula = dfGenoSelDat$bwt_sim ~ ., data = dfGenoSelDat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.1634 -1.0341  0.0044  0.9228  4.4683
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.66765    0.82430  -9.302 < 2e-16 ***
## sexM         5.00614    0.13934  35.927 < 2e-16 ***
## age        -0.01365    0.01078  -1.266  0.206
## rs3722996    1.19800    0.19298   6.208 1.15e-09 ***
## rs3671256    3.30766    0.55473   5.963 4.77e-09 ***
## rs13475745   5.13212    0.55524   9.243 < 2e-16 ***
## rs6404446    1.68132    0.35007   4.803 2.09e-06 ***
## mCV23695025  7.17154    0.44844  15.992 < 2e-16 ***
## rs6173215    8.84643    0.34206  25.862 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.545 on 487 degrees of freedom
## Multiple R-squared:  0.9929, Adjusted R-squared:  0.9928
## F-statistic: 8545 on 8 and 487 DF, p-value: < 2.2e-16

```

Wie gross sind nun die a-Werte der sechs Loci?

Die a-Werte entsprechen den einzelnen Schätzwerten für die gentischen Loci. Diese können mit folgender Funktion angezeigt werden:

```
coefficients(lmBwtSim)[4:9]
```

```

##      rs3722996      rs3671256      rs13475745      rs6404446 mCV23695025      rs6173215
##      1.198003       3.307656       5.132123       1.681321       7.171540       8.846433

```