

Lösungen zu den Aufgaben

1. Aufgabe

Der Datensatz `wo_men` untersucht Geschlechtsunterschiede u.a. in Körpergröße.

Sie können ihn hier herunterladen:

```
wo_men_path <- "https://raw.githubusercontent.com/sebastiansauer/modar/master/datasets/wo_men.csv"
d <- read_csv(wo_men_path)
```

Testen Sie anhand dieses Datensatzes folgende Hypothese:

“Männer haben im Schnitt ein höheres Gewicht als Frauen!”.

Präziser formuliert:

$\mu_M > \mu_F$, wobei μ den Mittelwert angibt (in der Population, auch Erwartungswert genannt).

Geben Sie als Lösung an, welcher Anteil der Postverteilung einen höheren Mittelwert zugunsten der Männer aufweist!

Lösung

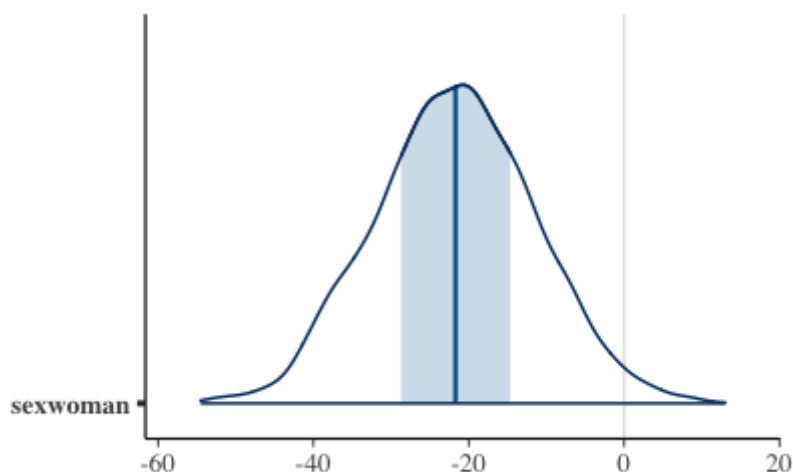
```
lm1 <- stan_glm(height ~ sex, data = d, refresh = 0)
```

Schauen wir uns die Post-Verteilung an:

```
lm1_post <-
  lm1 %>%
  as_tibble()

library(bayesplot)

mcmc_areas(lm1, pars = "sexwoman")
```



Wie man sieht ist die Verteilung des Mittelwertsunterschied ziemlich eindeutig: Nur mit geringer Wahrscheinlichkeit geht das Modell davon aus, dass Frauen einen höheren Mittelwert (in der

Population) aufweisen als Männer. Mit hoher Gewissheit geht das Modell davon aus, dass Männer einen höheren Mittelwert aufweisen (als Frauen).

Lassen wir uns den genauen Anteil ausgeben:

```
lml_post %>%
  count(sexwoman < 0)

## # A tibble: 2 × 2
##   `sexwoman < 0`      n
##   <lgl>             <int>
## 1 FALSE             72
## 2 TRUE             3928
```

Rechnen wir den Anteil noch aus:

```
prop_maenner_schwerer_df <-
  lml_post %>%
  count(sexwoman < 0) %>%
  mutate(prop = n/sum(n))

prop_maenner_schwerer_df

## # A tibble: 2 × 3
##   `sexwoman < 0`      n prop
##   <lgl>             <int> <dbl>
## 1 FALSE             72 0.018
## 2 TRUE            3928 0.982
```

Die Antwort lautet also, laut diesem Modell (mit dieser Simulationsvarianz):

```
prop_maenner_schwerer <- prop_maenner_schwerer_df %>%
  pull(prop)

sol <- prop_maenner_schwerer[2]
```