

Lösung Zettel 2

2023-06-01

Aufgabe 1

Lade zun nächst den Datenrahmen df spacebar aus dem Paket bcogsci.

```
library(bcogsci)
data("df_spacebar")
```

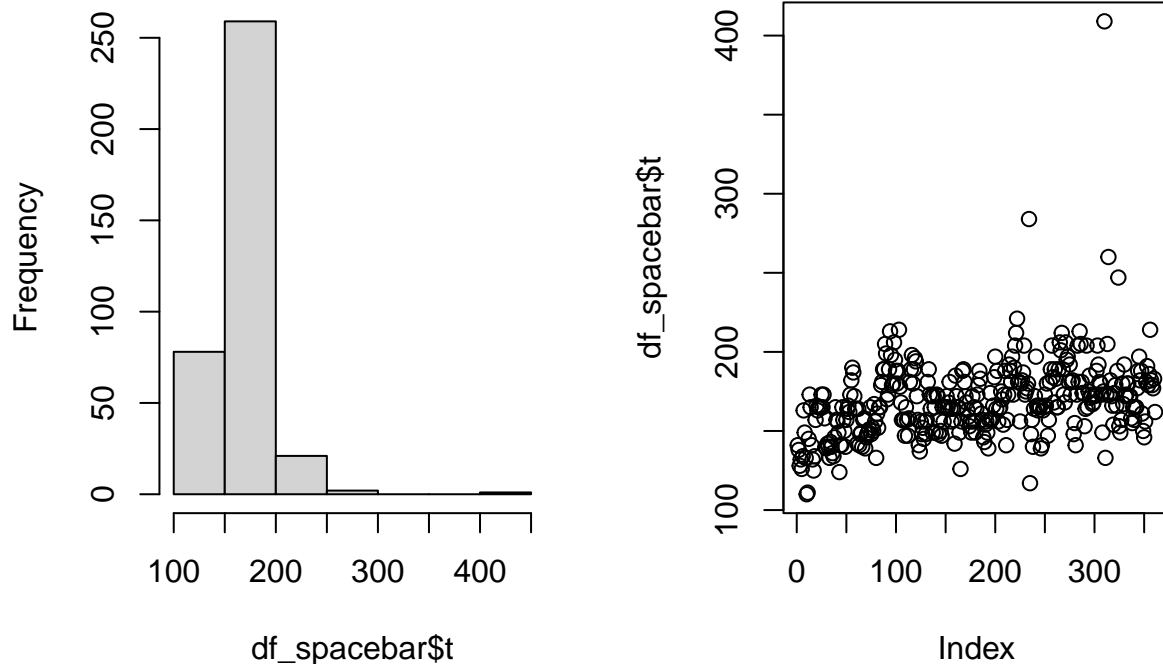
a) Es ist immer eine gute Idee, Daten grafisch darzustellen, bevor man etwas anderes tut; plotte die Daten. (Unsere Response ist "t").

Lösungsansatz:

- Histogram der Response Variable
- Line/Dot-Plot der Daten

```
par(mfrow = c(1, 2)) # zwei Grafiken in einer Figure
hist(df_spacebar$t)
plot(df_spacebar$t)
```

Histogram of df_spacebar\$t



b) Fitte ein Intercept Model auf den Daten mit Hilfe des brms Packages. t sei dabei die Response. Die Angenommene Verteilungsfamilie für t ist die Normalverteilung (`gaussian()`). Unterstelle erst einmal uninformative Prior: Für den Intercept $Uniform(0, 600)$ und Sigma $Uniform(0, 20)$. Setze die Iterationen auf 50 und das warmup auf 25. Konvergiert das Model? Woran erkennst du es? *Hint: Anhand des R Outputs und eines Model Plots.*

```
model_b = brm(t ~ 1,      # Intercept: 1
              family = gaussian(),
              prior =
                c(
                  prior(uniform(0,600), lb = 0, ub = 600, class = Intercept),
                  prior(uniform(0,20), lb = 0, ub = 20, class = sigma)
                ),
              warmup = 25,
              iter = 50,
              data = df_spacebar
            )
```

Lösungsansatz: Model Ergebnisse anschauen

```
model_b
```

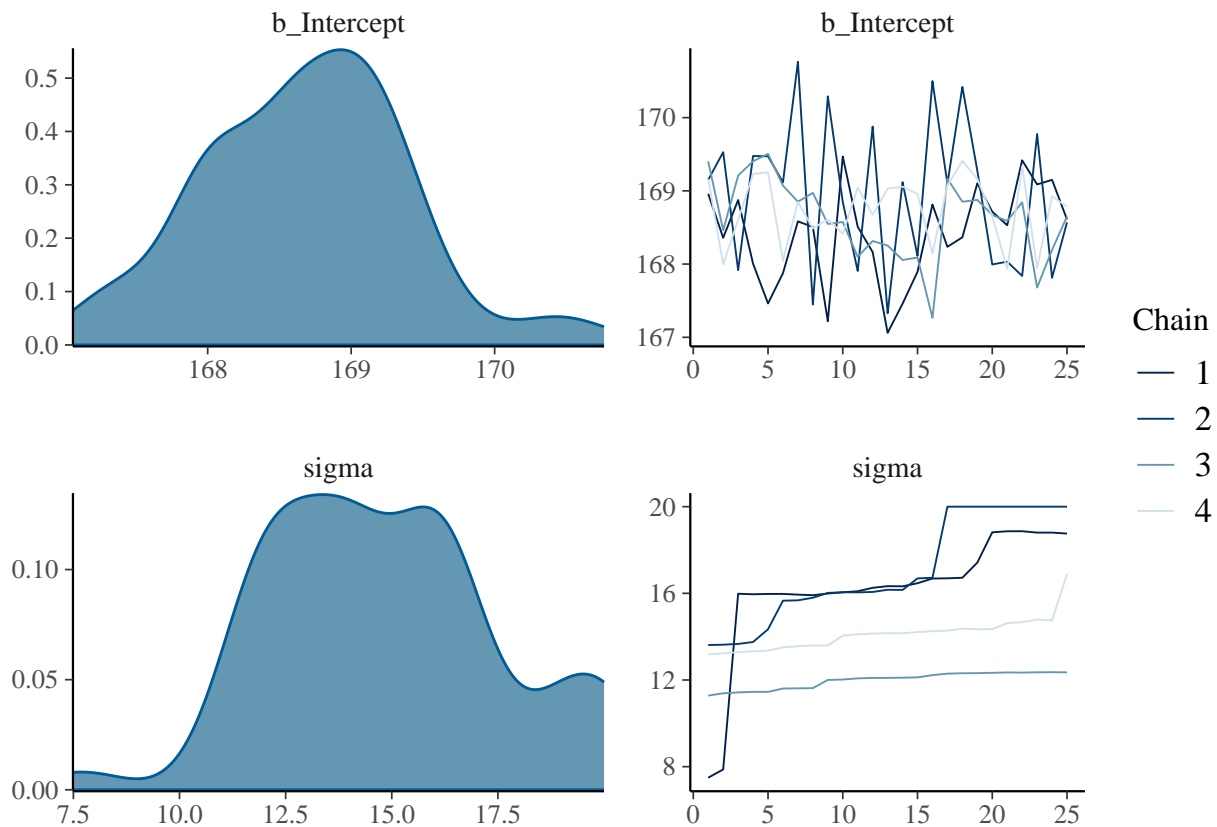
```
## Warning: Parts of the model have not converged (some Rhats are > 1.05). Be
## careful when analysing the results! We recommend running more iterations and/or
```

```
## setting stronger priors.

## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: t ~ 1
## Data: df_spacebar (Number of observations: 361)
## Draws: 4 chains, each with iter = 50; warmup = 25; thin = 1;
## total post-warmup draws = 100
##
## Population-Level Effects:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept  168.68      0.72  167.30  170.36 1.10      73      43
##
## Family Specific Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma    14.83      2.72   11.33   20.00 2.19      7      18
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Analyse des Models

```
plot(model_b)
```



Wir sehen hier ein Problem mit den “Catapillar” Plots. Die Graphen liegen nicht übereinander. Es bedeutet, dass wir hier keine Konvergenz beobachten. Die Anzahl der Iterationen reicht scheinbar nicht aus.

c) Fitte das gleiche Modell, verändere nur die Anzahl der Iterationen auf 2000 und die Warmup auf 1000. Unterstelle für den Intercept einen Prior *Uniform*(0, 60000) und einen Sigma Prior *Uniform*(0, 2000). Konvergiert das Modell? Woran machst du diese Entscheidung fest?

```
model_c = brm(t ~ 1,      # Intercept: 1
              family = gaussian(),
              prior =
                c(
                  prior(uniform(0,600), lb = 0, ub = 600, class = Intercept),
                  prior(uniform(0,20), lb = 0, ub = 20, class = sigma)
                ),
              data = df_spacebar
            )
```

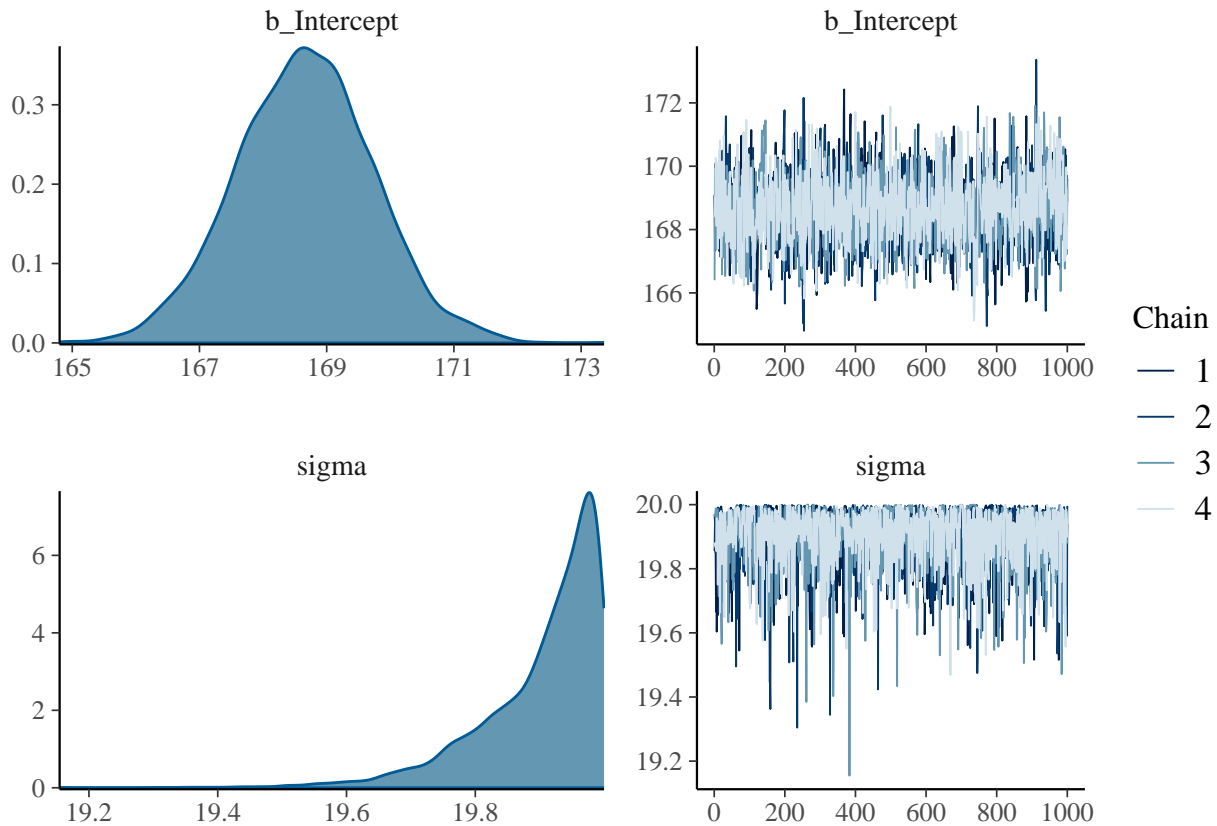
Lösungsansatz: Model Ergebnisse anschauen

```
model_c

## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: t ~ 1
## Data: df_spacebar (Number of observations: 361)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##       total post-warmup draws = 4000
##
## Population-Level Effects:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept   168.65      1.08  166.53  170.82 1.00    2052    1853
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma    19.90      0.10   19.66   20.00 1.00    1677     948
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Analyse des Modells

```
plot(model_c)
```



Wir beobachten Konvergenz! Allerdings sehen wir, dass der Prior für das Sigma vielleicht nicht wirklich gut gewählt worden ist.

d) Wähle unter Berücksichtigung von Normalverteilung der Daten, einen Prior der besser zu den Daten passt. Überlege einen vernünftigen Satz von Prior für μ und σ . Begründe wie du auf diese Auswahl kommst. Es gibt hier keine richtige Antwort! *Die Grafik aus a) kann euch helfen einen Entscheidung zu treffen.*

Lösungsansatz:

- Wir beobachten keine negativen Werte.
-

```
summary(df_spacebar$t)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    110.0   156.0   166.0   168.6   181.0   409.0
```

```
mean(df_spacebar$t); sd(df_spacebar$t)
```

```
## [1] 168.6399
```

```
## [1] 24.9118
```

```
model_d = brm(t ~ 1,      # Intercept: 1
              family = gaussian(),
              prior =
                c(
```

```

        prior(normal(168,30), class = Intercept),
        prior(normal(0,30), class = sigma)
    ),
    data = df_spacebar
)

```

```
model_d
```

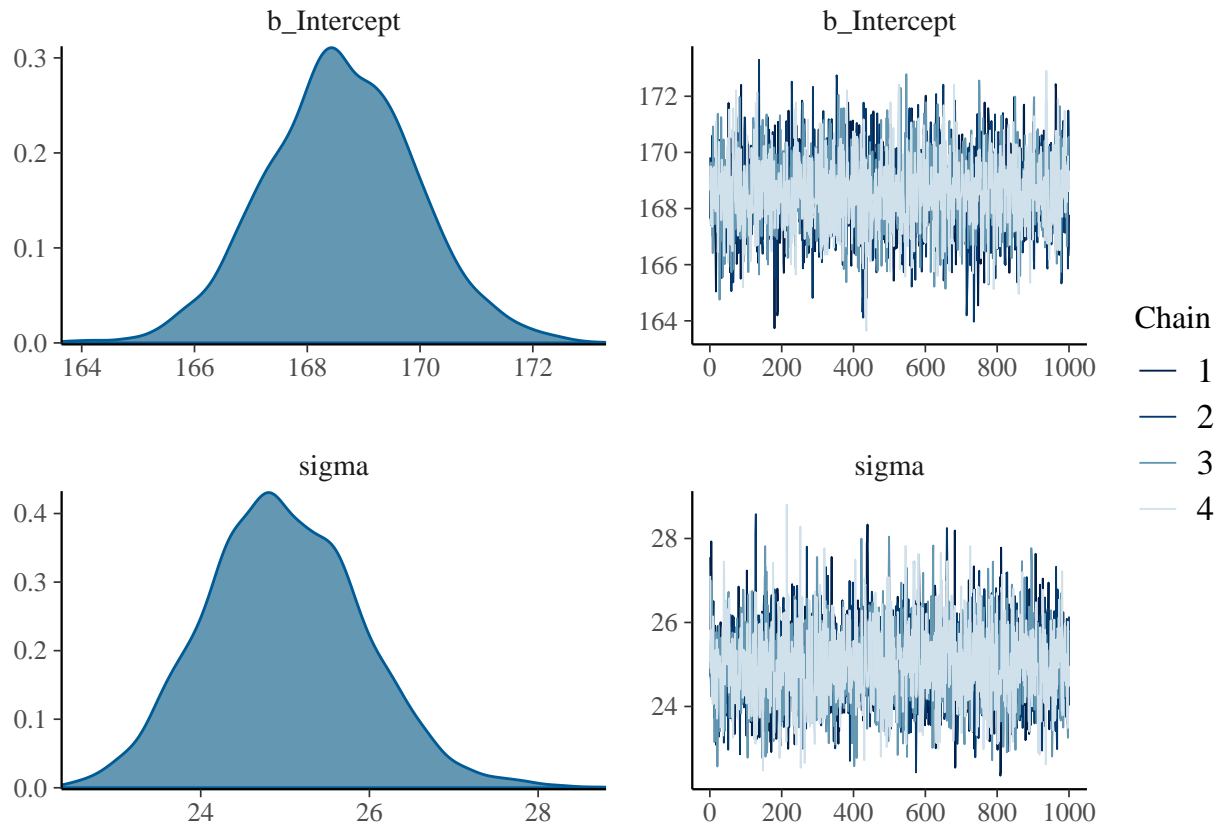
```

## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: t ~ 1
## Data: df_spacebar (Number of observations: 361)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Population-Level Effects:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept  168.60      1.33  165.98  171.25 1.00    2945    2368
##
## Family Specific Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma    24.99      0.93   23.27   26.85 1.00    2965    2475
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).

```

Analyse des Models

```
plot(model_d)
```



e) Nutze die von dir aufgestellten Prior um das neue Modell zu fitten. Wenn du die Modellergebnisse vergleichst was fällt dir auf? *Hint: die `summary()` des Modells ist immer ein guter Startpunkt.*

Lösungsansatz:

f) Kannst du aussagekräftige Priors finden, die den Posterior in spürbarer Weise beeinflussen? (verwende Normalverteilungen für Priors, keine gleichmäßigen Priors) Auch hier gibt es keine richtigen Antworten. Möglicherweise musst du mehrere verschiedene Priors ausprobieren, bevor du den Posterior merklich beeinflussen können. *Wiederholung: Aussagekräftige Prior: Prior die "Wissen" in die Analyse bringen. Z.B. Weißt du vorher schon wo der Mittelwert sein müsste? Wie ist die Varianz?*

Lösungsansatz:

g) Erzeuge auf der Grundlage dieses neuen Priors ein Modell und zeige die Posterior Predictive Verteilungen und plote sie. *Tipp: Das R-Package bayesplot hat nützliche Funktionen.*

Lösungsansatz: