

# 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 sind im Schnitt größer als Frauen!”

Präziser formuliert:

$\mu_M > \mu_F$ , wobei  $\mu$  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

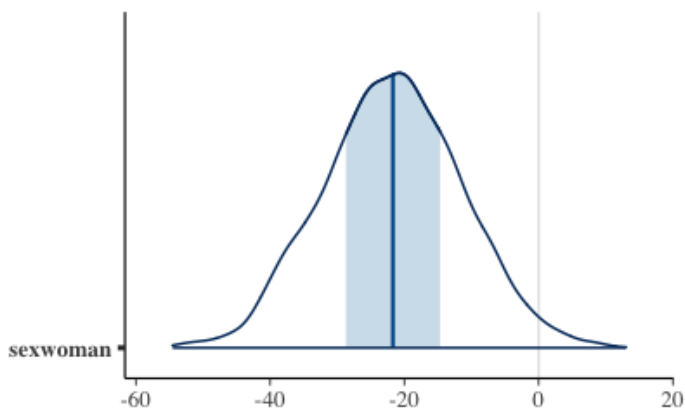
```
lml <- stan_glm(height ~ sex, data = d, refresh = 0)
```

Schauen wir uns die Post-Verteilung an:

```
lml_post <-
  lml %>%
  as_tibble()

library(bayesplot)

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



Wie man sieht ist die Verteilung des Mittelwertsunterschieds 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 gegebener Simulationsvarianz):

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

sol <- prop_maenner_schwerer[2]
sol

## [1] 0.982
```

Wir resümieren: Die Wahrscheinlichkeit, dass  $\mu_M > \mu_F$  liegt bei 0.98, laut unserem Modell.

## 2. 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  $H$ :

“Männer sind im Schnitt größer als Frauen!”

Präziser formuliert:

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

Über die Hypothese soll anhand folgender Entscheidungsregel  $R$  entschieden werden:

- $\$Pr(H|D) > .95$   $H$  \$ annehmen
- $\$Pr(H|D) < .05$   $H$  \$ verwerfen
- ansonsten: *keine Entscheidung* bzgl.  $H$  aufgrund uneindeutiger Evidenz

Zur Erinnerung:  $Pr(H|D)$  benennt die Posterior-Verteilung.

## Lösung

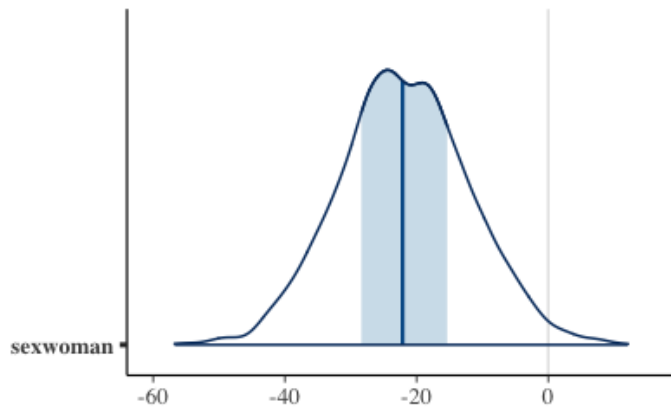
```
lml <- stan_glm(height ~ sex, data = d, refresh = 0)
```

Schauen wir uns die Post-Verteilung an:

```
lml_post <-
  lml %>%
  as_tibble()
```

```
library(bayesplot)

mcmc_areas(lml, 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           58
## 2 TRUE          3942
```

Mann kann den Anteil natürlich einfach aus der Tabelle ablesen, aber man kann ihn sich auch von R ausgeben lassen; das tut hier nicht wirklich Not. Es würde reichen, wenn Sie den Wert einfach hinschreiben.

Also rechnen wir den Anteil 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           58  0.0145
## 2 TRUE          3942  0.986
```

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

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

sol <- prop_maenner_schwerer[2]

if (sol > .95) entscheidung <- "H annehmen"
if (sol < .05) entscheidung <- "H verwerfen"
if (sol > .05 & sol < .95) entscheidung <- "keine Entscheidung"
```

Wir resümieren:

Auf Basis der Evidenz treffen wir folgende Entscheidung: H annehmen.

Hinweis: Oft wird folgende Entscheidungsregel  $R_2$  verwendet:

```
if (sol > .95) entscheidung <- "H annehmen"
if (sol < .95) entscheidung <- "H verwerfen"
```

In Worten: Bei  $R_2$  verhält man sich so: Nur wenn wir ziemlich sicher sind, nehmen wir  $H$  an, ansonsten verwerfen wir  $H$ . Wir sind also eher dem Verwerfen der  $H$  zugeneigt, mithin also der Nullhypothese zugeneigt. Das nennt man manchmal eine *skeptische Haltung*.

Nachtrag:

Vielleicht ist es einfacher, die Grenzen des 95%-PI zu benennen:

```
posterior_interval(lm1, pars = "sexwoman", prob = .95)

##                2.5%          97.5%
## sexwoman -41.03284 -2.502507
```

Die Ausgabe zeigt, dass innerhalb der Grenzen des 95%-PI die Null nicht enthalten ist. Das ist gleichbedeutend mit der Aussage, dass wir die Null mit einer Sicherheit von mind. 95% ausschließen können (laut unserem Modell).

- a.  $H$  annehmen
- b.  $H$  verwerfen
- c. Keine Entscheidung möglich

### 3. 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  $H$ :

“Männer sind im Schnitt größer als Frauen!”

Präziser formuliert:

$X_M > X_F$ , wobei  $X$  die Größe *einer zufällige gezogenen Person* angibt (in der Population, auch Erwartungswert genannt).  $M$  steht für Männer,  $F$  für Frauen.

Über die Hypothese soll anhand folgender Entscheidungsregel  $R$  entschieden werden:

- $\Pr(p|D, H) > .95$   $H$  annehmen
- $\Pr(p|D, H) < .05$   $H$  verwerfen
- ansonsten: *keine Entscheidung* bzgl.  $H$  aufgrund uneindeutiger Evidenz

Dabei benennt  $\Pr(p|D, H)$  die Wahrscheinlichkeit auf Basis der PPV, dass ein beliebiger Mann ein höheres Gewicht hat als eine beliebige Frau (d.h. zufällig gezogen).

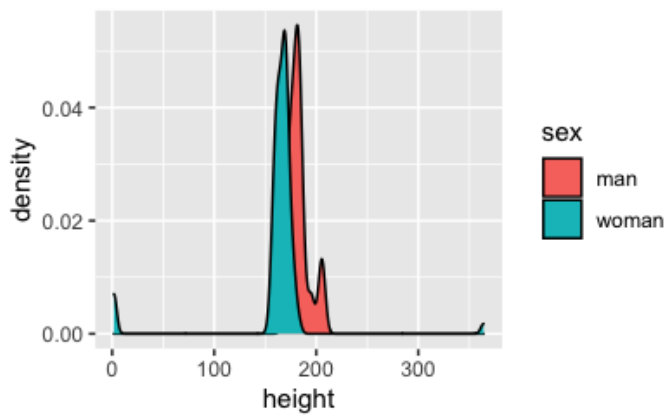
Geben Sie die gesuchte Wahrscheinlichkeit als Lösung an!

### Lösung

Betrachten wir zunächst die Daten:

```
d %>%
  ggplot(aes(x = height, fill = sex)) +
  geom_density()

## Warning: Removed 1 rows containing non-finite values (stat_density).
```



Wie man sieht, scheint es einige “komische” Werte zu geben. Besser wir entfernen diese vorab.

```
d2 <-
  d %>%
  filter(between(height, 150, 210))

lm2 <- stan_glm(height ~ sex, data = d2, refresh = 0)
```

Berechnen wir dann die PPV. Dabei interessieren uns zwei Ausprägungen, nämlich `woman` und `man`. Wir brauchen also nur diese zwei Beobachtungen:

```
lm2_ppv <-
  posterior_predict(lm2, newdata = tibble(sex = c("woman", "man"))) %>%
  as_tibble()
```

Schauen wir uns die PPV an:

```
lm2_ppv %>%
  head()

## # A tibble: 6 × 2
##   `1`   `2`
##   <ppd> <ppd>
## 1 176.6743 169.8805
## 2 167.9193 197.3932
## 3 161.6888 178.5250
## 4 163.9002 180.9727
## 5 169.4206 191.5362
## 6 159.1413 174.6664
```

Die Tabelle `lm1_ppv` hat 4000 Zeilen, das ist jeweils eine Stichprobe aus der PPV. Die Tabelle hat 2 Spalten, das sind die zwei Beobachtungen `woman` und `man`, Personen also.

Jetzt schauen wir für jede der 4000 Stichproben, wie oft der Mann größer war:

```
lm2_ppv <-
  lm2_ppv %>%
  mutate(mann_ist_groesser = `1` < `2`)

lm2_ppv %>%
  head()

## # A tibble: 6 × 3
##   `1`   `2`   mann_ist_groesser
##   <ppd> <ppd>   <lgl>
## 1 176.6743 169.8805 FALSE
## 2 167.9193 197.3932 TRUE
## 3 161.6888 178.5250 TRUE
## 4 163.9002 180.9727 TRUE
## 5 169.4206 191.5362 TRUE
## 6 159.1413 174.6664 TRUE
```

Und dann zählen wir aus, wie oft in den 4000 Stichproben der Mann größer war, also wie oft `mann_ist_groesser == TRUE` gilt.

```
lm2_ppv %>%
  summarise(prop_mann_ist_groesser = mean(mann_ist_groesser))

## # A tibble: 1 × 1
##   prop_mann_ist_groesser
##               <dbl>
## 1                   0.942
```

Vielleicht wundern Sie sich, dass wir hier den Befehl `mean` verwendet haben. Praktischerweise kann `mean()` auch logische Variablen verarbeiten. `mann_ist_groesser` ist eine logische Variable (Typ `lgl`, logical); die Variable hat nur die Werte `TRUE` und `FALSE`. Dabei zählt `mean()` für jedes `TRUE` eine 1 und für jedes `FALSE` eine 0. Diese Summe wird durch die Anzahl der Zeilen, also  $n$ , geteilt: Voilà, der Mittelwert.

Wie man sieht ist die Verteilung des Mittelwertsunterschieds in der PPV ziemlich eindeutig: Nur mit geringer Wahrscheinlichkeit geht das Modell davon aus, dass eine *beliebige* (zufällig gezogene) Frau einen höheren Mittelwert (in der Population) aufweisen als ein beliebiger Mann. Anders gesagt: Mit hoher Gewissheit geht das Modell davon aus, dass ein zufällig gezogener Mann einen höheren Mittelwert aufweist (als eine zufällig gezogene Frau).

Lassen wir uns den genauen Anteil ausgeben:

```
anteil_beliebiger_mann_ist_groesser <-
  lm2_ppv %>%
    summarise(prop_mann_ist_groesser = mean(mann_ist_groesser)) %>%
    as.numeric()

anteil_beliebiger_mann_ist_groesser

## [1] 0.94225
```

Da die Tabelle nur aus einer Zelle besteht, wandeln wir sie mit `as.numeric()` in einen numerischen Wert um, also eine einfache Zahl in diesem Fall.

Mann kann den Anteil natürlich einfach aus der Tabelle ablesen, aber man kann ihn sich auch von R ausgeben lassen; das tut hier nicht wirklich Not. Es würde reichen, wenn Sie den Wert einfach hinschreiben.

Die Antwort (solution, `sol`) lautet also, laut diesem Modell (mit gegebener Simulationsvarianz):

```
sol <- anteil_beliebiger_mann_ist_groesser
round(sol, 2)

## [1] 0.94
```

## 4. 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  $H$ :

“Männer sind im Schnitt größer als Frauen!”

Präziser formuliert:

$d = \mu_M > \mu_F + c$ , wobei  $\mu$  den Mittelwert angibt (in der Population, auch Erwartungswert genannt) und  $c$  einen *minimalen Unterschied praktischer Relevanz* benennt.

Nehmen Sie an:  $c = 5$ . Das heißt: Wir gehen davon aus, dass Männer mehr als 5cm größer sind als Frauen im Durchschnitt.

Über die Hypothese soll anhand folgender Entscheidungsregel  $R$  entschieden werden:

- Liegt die Verteilung von  $d$  mit mind. 95% im ROPE, verwerfen wir  $H$ .
- Liegt die Verteilung von  $d$  mit mind. 95% außerhalb des ROPE, verwerfen wir  $H$ .
- Andernfalls behalten wir uns eine Entscheidung vor, aufgrund uneindeutiger Evidenz.

In Worten sagt  $R$  etwa: “Wenn von der Post-Verteilung der Großteil (sagen wir 95%) außerhalb des ROPE liegt, dann ist der Unterschied bedeutsam. Wenn der größte Teil der Post-Verteilung innerhalb des ROPE liegt, ist der Unterschied vernachlässigbar. Ansonsten ist die Datenlage unklar”.

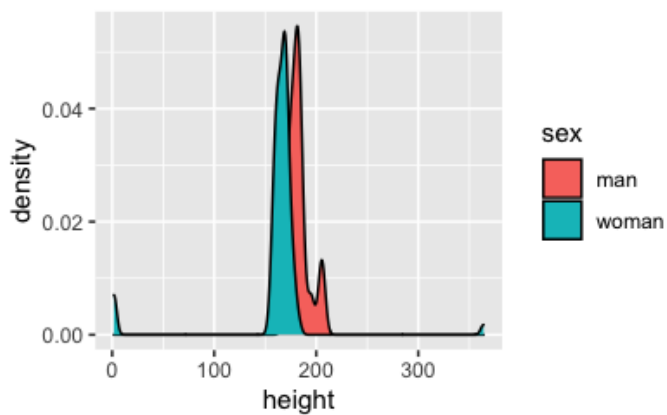
Geben Sie den Anteil der Post-Verteilung an, der außerhalb des Rope liegt!

## Lösung

Betrachten wir zunächst die Daten:

```
d %>%
  ggplot(aes(x = height, fill = sex)) +
  geom_density()

## Warning: Removed 1 rows containing non-finite values (stat_density).
```

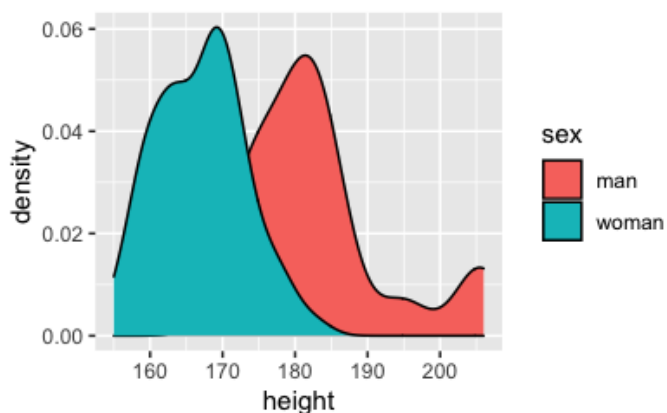


Wie man sieht, scheint es einige “komische” Werte zu geben. Besser wir entfernen diese vorab.

```
d2 <-
  d %>%
  filter(between(height, 150, 210))
```

Hat's was gebracht:

```
d2 %>%
  ggplot(aes(x = height, fill = sex)) +
  geom_density()
```



Einigermaßen ok.

```
lm2 <- stan_glm(height ~ sex, data = d2, refresh = 0)
```

Berechnen wir dann das ROPE:

```
library(bayestestR)
lm2_rope <- rope(lm2, range = c(-5, 5))

lm2_rope

## # Proportion of samples inside the ROPE [-5.00, 5.00]:
##
## Parameter | inside ROPE
## -----
## (Intercept) | 0.00 %
## sexwoman | 0.00 %

plot(lm2_rope)

## Error: Failed at retrieving data :( Please provide original model or data through the `data` argument
```

Wie man sieht, kann man die Hypothese  $H_0$ , dass der Unterschied zwischen den mittleren Größen von Frauen und Männern vernachlässigbar klein ist, *verwerfen*.

Wir nehmen also die  $H$  an, dass Männer im Mittel mehr als 5cm größer sind als Frauen.

Die Antwort (solution, `sol`) lautet also, laut diesem Modell (mit gegebener Simulationsvarianz):

```
sol <- 1
```

## 5. 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  $H$ :

“Männer sind im Schnitt größer als Frauen!”

Diese Aussage soll wie folgt verstanden werden:

Der Anteil erklärter Varianz  $R^2$  in einem entsprechenden Modell soll größer als ein kritischer Wert  $c$ , der die Grenze zu einem vernachlässigbar kleinen Effekt kennzeichnet.

Nehmen Sie an:  $c = .10$ .

Das Geschlecht soll also mindestens 10% der Gesamtvarianz erklären.

Geben Sie an, ob der Median des (Bayes-)  $R^2$ -Koeffizienten den Wert  $c$  erreicht, ob also gilt  $R^2 \geq c$ .

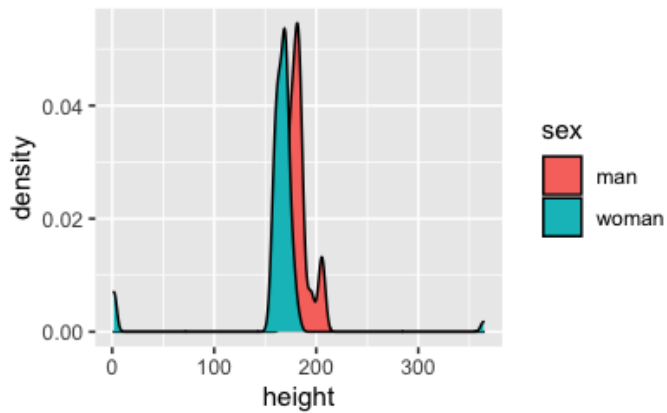
## Lösung

Betrachten wir zunächst die Daten:

```
d %>%
  ggplot(aes(x = height, fill = sex)) +
  geom_density()

## Warning: Removed 1 rows containing non-finite values (stat_density).
```



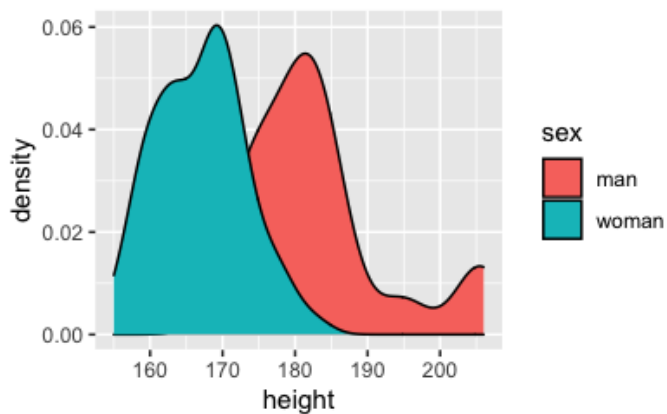


Wie man sieht, scheint es einige “komische” Werte zu geben. Besser wir entfernen diese vorab.

```
d2 <-
  d %>%
  filter(between(height, 150, 210))
```

Hat's was gebracht?

```
d2 %>%
  ggplot(aes(x = height, fill = sex)) +
  geom_density()
```



Naja, einigermaßen besser.

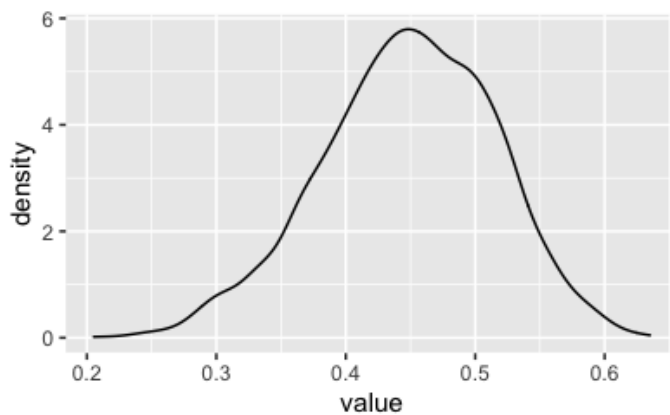
```
lm2 <- stan_glm(height ~ sex, data = d2, refresh = 0)

lm2_r2 <- bayes_R2(lm2) %>%
  as_tibble()
```

Da die meisten Funktionen aus `rstanarm` keine Tabellen zurückliefern, aber mit `as_tibble()` leicht in eine Tabelle (ein “Tibble”) umgewandelt werden können, haben wir das hier (und an anderen Stellen) gemacht.

Wieder bekommen wir viele Stichproben aus der Post-Verteilung:

```
lm2_r2 %>%
  ggplot(aes(x = value)) +
  geom_density()
```



Rechnen wir noch flux die Quantile für 2.5% und 97.5% aus:

```
lm2_r2 %>%
  summarise(qs = quantile(value, probs = c(.025, .975)))

## # A tibble: 2 × 1
##   qs
##   <dbl>
## 1 0.305
## 2 0.570
```

Unser  $R^2$  liegt also mit hoher Wahrscheinlichkeit in diesem Bereich (behauptet zumindest unser Golem).

Aber es wurde nach dem Median gefragt:

```
lm2_r2 %>%
  summarise(r2_md = median(value))

## # A tibble: 1 × 1
##   r2_md
##   <dbl>
## 1 0.451
```

Das Geschlecht erklärt also fast die Hälfte der Varianz in der Größe der Gesamt-Stichprobe (bzw. auch in der Population).

Das ist mal ein *superstarker* Effekt. Die allermeisten Effekte, die man in der Psychologie findet, sind viel kleiner.

Merken Sie sich mal den Wert von  $R^2$ , als Vergleich, wenn Sie andere Werte von sehen.

Die Antwort (solution, `sol`) lautet also, laut diesem Modell (mit gegebener Simulationsvarianz):

```
sol <- lm2_r2 %>%
  summarise(r2_md = median(value)) %>%
  as.numeric()

sol

## [1] 0.4505247
```

## 6. 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  $H$ :

“Männer sind im Schnitt größer als Frauen!”

Formaler:

$d > 0$  mit  $d = \mu_M - \mu_F$ .

Wenn man zwei Gruppen (zwei Populationen) vergleicht, kann man einfach den Unterschied in den Mittelwerten,  $d$ , berechnen. Das macht meist Sinn, gerade wenn die Einheiten gut verständlich sind, wie es bei der Körpergröße der Fall ist. Bei einem psychometrischen Merkmal, sagen wir “Smartphone-Abhängigkeit”, das vielleicht auf einer Skale von 4 bis 44 gemessen wird, sind die Einheiten nicht wirklich gut interpretierbar. In diesem Fall macht man gerne Folgendes: Man berechnet den Mittelwertsunterschied und schaut dann, wie groß dieser Wert im Verhältnis zur Streuung (SD) der Variable ist.

Ein Beispiel zur Verdeutlichung: Der Unterschied  $d$  betrage 10cm. Die Streuung (sd) betrage 5cm. Dann ist der Unterschied genau zwei Streuungseinheiten.

Diesen Koeffizienten nennt man *Cohens d*.

Er ist so definiert:

$$C_d = \frac{d}{sd}$$

Anmerkung: Um die Streuung zu berechnen, gibt es ein paar Varianten. Meist nimmt man den Mittelwert der Streuungen (SDs) der beiden Gruppen, ungewichtet von der Gruppengröße. Das nennt man auch die “gepoolte SD” (pooled sd).

Nehmen Sie an:  $C_d = .5$ .

Damit formulieren wir unsere Hypothese  $H$  neu:

$H : C_d > .5$ .

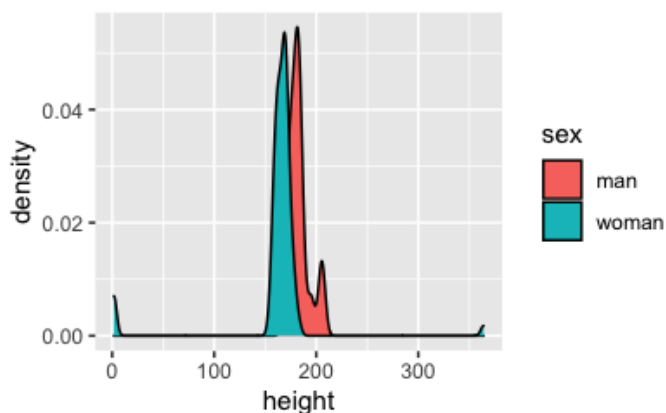
Das Unterschied in der mittleren Körpergröße soll also, laut  $H$ , mindestens eine halbe SD-Einheit ausmachen.

Geben Sie an einen Schätzwert für  $C_d$  für obige Hypothese an! Nutzen Sie den Median der entsprechenden Post-Verteilung!

## Lösung

Betrachten wir zunächst die Daten:

```
d %>%  
  ggplot(aes(x = height, fill = sex)) +  
  geom_density()  
  
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

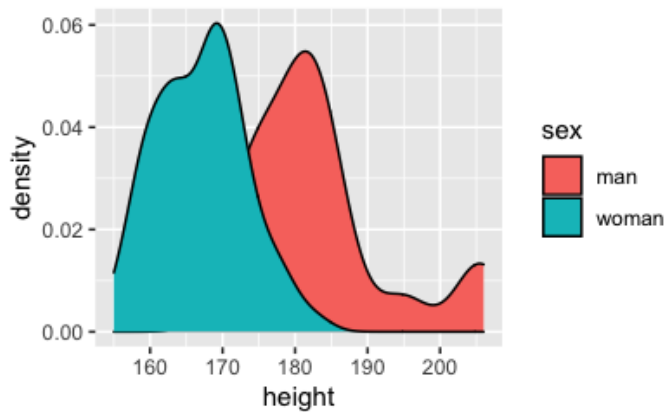


Wie man sieht, scheint es einige “komische” Werte zu geben. Besser wir entfernen diese vorab.

```
d2 <-
  d %>%
  filter(between(height, 150, 210))
```

Hat's was gebracht?

```
d2 %>%
  ggplot(aes(x = height, fill = sex)) +
  geom_density()
```



Naja, einigermaßen besser.

```
lm2 <- stan_glm(height ~ sex, data = d2, refresh = 0)
```

```
lm2_post <-
  lm2 %>%
  as_tibble()
```

```
lm2_post %>%
  slice_head(n=5)
```

```
## # A tibble: 5 × 3
##   `(Intercept)` sexwoman sigma
##   <dbl>         <dbl> <dbl>
## 1      184.      -16.9  7.04
## 2      184.      -17.5  7.04
## 3      184.      -17.9  7.12
## 4      181.      -14.0  6.90
## 5      185.      -17.8  6.77
```

Da die meisten Funktionen aus `rstanarm` keine Tabellen zurückliefern, aber mit `as_tibble()` leicht in eine Tabelle (ein "Tibble") umgewandelt werden können, haben wir das hier (und an anderen Stellen) gemacht.

Jetzt noch ein kleines (vielleicht nerviges) Detail: Wir berechnen wir noch die gepoolte SD,  $sd_p$ , also den Mittelwert der Streuung in den beiden Geschlechtern:

```
d2 %>%
  group_by(sex) %>%
  summarise(height_sd = sd(height))
```

```
## # A tibble: 2 × 2
##   sex   height_sd
##   <chr>      <dbl>
## 1 man         9.96
## 2 woman        6.11
```

Also:

```
sd_m = 9.96
sd_f = 6.11
```

Die gepoolte Varianz berechnet sich dann so:

```
sd_pooled = sqrt((sd_m^2 + sd_f^2) / 2)
sd_pooled
```

```
## [1] 8.262376
```

Also etwa der Mittelwert von `sd_m` und `sd_f`.

Jetzt können wir  $C_d$  berechnen.

Zur Erinnerung: Die Differenz zwischen den Gruppen wird durch die Spalte `sexwoman` angegeben:

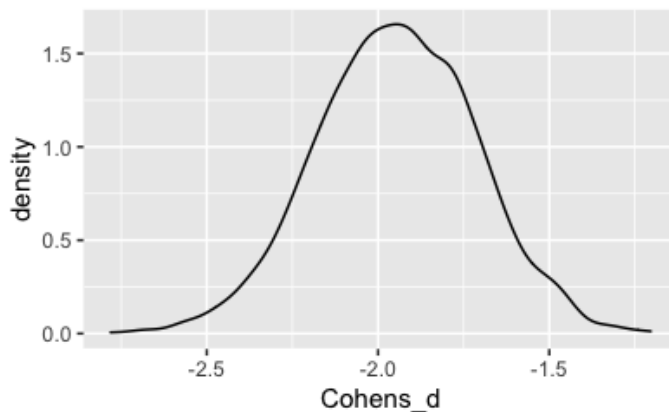
```
lm2_post %>%
  slice_head(n=2)

## # A tibble: 2 × 3
##   `(Intercept)` sexwoman sigma
##   <dbl> <dbl> <dbl>
## 1    184.   -16.9  7.04
## 2    184.   -17.5  7.04

lm2_post <-
  lm2_post %>%
  mutate(Cohens_d = sexwoman / sd_pooled)
```

Hier ist die Verteilung von  $C_d$ :

```
lm2_post %>%
  ggplot(aes(x = Cohens_d)) +
  geom_density()
```



Gefragt war nach dem Median:

```
lm2_post %>%
  summarise(Cohens_d_md = median(Cohens_d))

## # A tibble: 1 × 1
##   Cohens_d_md
##   <dbl>
## 1    -1.95
```

Der mittlere Größenunterschied beträgt also ca. 2 SD-Einheiten.

Das ist mal ein *superstarker* Effekt. Die allermeisten Effekte, die man in der Psychologie findet, sind viel kleiner.

Merken Sie sich mal diesen Wert von  $C_d$ , als Vergleich, wenn Sie andere Werte von sehen.

Die Antwort (solution, `sol`) lautet also, laut diesem Modell (mit gegebener Simulationsvarianz):

```
sol <- -1.95
sol

## [1] -1.95
```