

Forschungsfragen mit metrischer AV

Kapitel 6

Gliederung

1. Einleitung: Wissenschaft als Gerechtigkeitsprojekt
2. Teil 1: Eine binäre UV
3. Teil 2: Eine metrische plus eine nominale UV
4. Teil 3: Eine nominale UV mit mehreren Stufen
5. Teil 4: Mehrere metrische UV
6. Hinweise

Einleitung

Wissenschaft als Gerechtigkeitsprojekt

Meinungen als Grundlage der Konfliktlösung

- "Ich find Masken doof!"
- "Impfen ist schädlich!"
- "Corona gibt's gar nicht!"



- "Ich find Masken gut!"
- "Impfen ist nützlich!"
- "Corona ist gefährlich!"



Meinungen kennen kein richtig und kein falsch. Konflikte können auf Basis von Meinungen nur schwer gelöst werden.

Fakten als Grundlage der Konfliktlösung

Wissenschaft produziert Fakten.

Da Fakten universell sind (sein können), ist Wissenschaft potenziell ein Weg zur Konfliktlösung.

Warum helfen Fakten bei Konflikten?

- Fakten sind neutral gegenüber Personen.
- Fakten bieten daher eine Chance zur fairen Einigung.

Wann ist ein Fakt ein Fakt?

- Fakten müssen vor allem nachprüfbar sein (Daten, Analyse und Bericht müssen offen zugänglich sein).

Datenlage spricht zugunsten der Covid19-Impfung

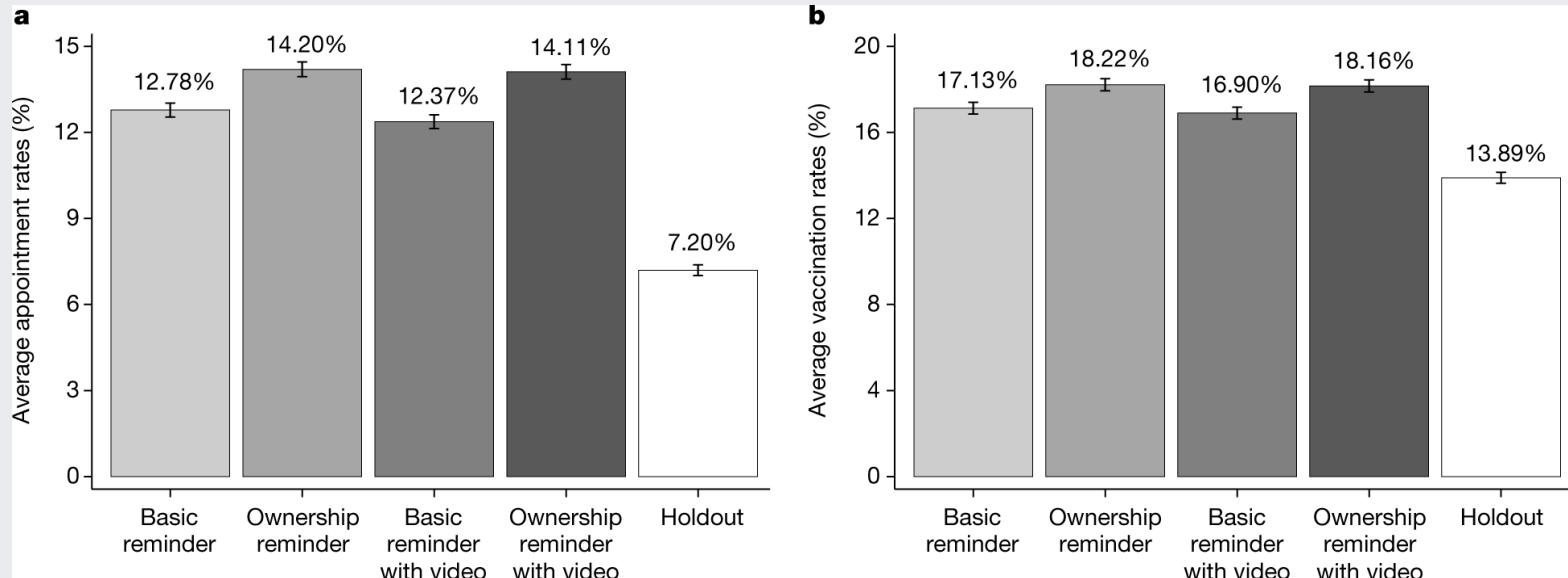
The effectiveness of full messenger RNA (mRNA) vaccination (≥ 14 days after the second dose) was 89% (95% confidence interval [CI], 87 to 91) against laboratory-confirmed SARS-CoV-2 infection leading to hospitalization, 90% (95% CI, 86 to 93) against infection leading to an ICU admission, and 91% (95% CI, 89 to 93) against infection leading to an emergency department or urgent care clinic visit.

Thompson, Stenehjem, Grannis, et al. (2021); vgl. auch Nasreen, Chung, He, et al. (2021), Pormohammad, Zarei, Ghorbani, Mohammadi, Razizadeh, et al. (2021)

Zwei Anforderungen an die Qualität von Studien:

1. *handwerklich gut*: z.B. vergleichbare Gruppen, genaue Messinstrumente
2. *bescheiden*: die Forschungsfrage wird nur dann selbstbewusst beantwortet, wenn es die handwerkliche Qualität der Studie zulässt. Gibt es eine Vielzahl weiterer Studien mit abweichenden Ergebnissen, wird dies bei der Beantwortung der Forschungsfrage berücksichtigt.

Psychologische Intervention zur Erhöhung der Impfquote



Quelle/Volltext

Dai, Saccardo, Han, Roh, Raja, et al. (2021)

Was heißt "ist effektiv"?

Nasreen, Chung, He, et al. (2021) definieren *effectivity*, e , so:

$$e = 1 - C; C = \frac{n_{vacc|pos}}{n_{vacc|neg}}$$

- C nennt man das *Chancenverhältnis (odds ratio)*, es beschreibt einen Bruchterm: $\frac{x}{y}$.
- $n_{vacc|pos}$: Anzahl der geimpften Personen unter allen Personen mit positiver Corona-Diagnose
- $n_{vacc|neg}$: Anzahl der geimpften Personen unter allen Personen mit negativer Corona-Diagnose

Beispiel: Von den 100 Personen mit *positiver* Corona-Diagnose sind 10 geimpft, $n_{vacc|pos} = 10$. Von den 100 Personen mit *negativer* Corona-Diagnose sind 90 geimpft, $n_{vacc|neg} = 90$

$$C = \frac{10}{90} = \frac{1}{9}; e = 1 - \frac{1}{9} = \frac{8}{9} \approx 0.88$$

In diesem Beispiel liegt die Effektivität e bei knapp 90%.

Arten von Forschungsfragen

Deskriktiv (beschreibend)

- Wie stark ist der (lineare) Zusammenhang r von Größe und Gewicht?
- Wie stark ist der (lineare) Zusammenhang b von Lernzeit und Note?
- Bevorzugen unsere Kunden Webshop A oder B?

Prädiktiv (prognostisch, vorhersagend)

- Wie schwer ist ein deutscher Mann der Größe 1,80m im Schnitt?
- Welche Note kann man erwarten, wenn man nichts für die Klausur lernt?
- Wieviel wird ein Kunde ausgeben, wenn er sich in dieser Variante des Webshops aufhält?

Präskriptiv (erklärend, kausal)

- Ist Größe eine Ursache von Gewicht (bei deutschen Männern)?
- Wenn ich 100 Stunden lerne, welche Note schreibe ich dann?
- Hat die Art des Webshops einen Einfluss auf unseren Umsatz?

Metrische AV

- Wir konzentrieren uns im Folgenden auf Regressionsmodelle mit *metrischer* AV.
- Für die UV(s) sind nominale und metrische Skalenniveaus erlaubt.
- Modelle mit mehreren UV (und mehreren Stufen an UV) sind erlaubt.

Teil 1

Eine binäre UV

Forschungsfrage

Hintergrund:

Eine Psychologin, die im öffentlichen Dienst arbeitet, versucht herauszufinden, warum einige Kinder intelligenter sind als andere. Dazu wurden in einer aufwändigen Studie die Intelligenz vieler Kinder gemessen. Zusätzliche wurden verschiedene Korrelate erhoben, in der Hoffnung, "Risikofaktoren" für geringere Intelligenz zu entdecken.

Forschungsfrage:

Unterscheidet sich der mittlere IQ-Wert (kid_score) von Kindern in Abhängigkeit davon, ob ihre jeweilige Mutter über einen Schlusabschluss (mom_hs) verfügt? (ceteris paribus)

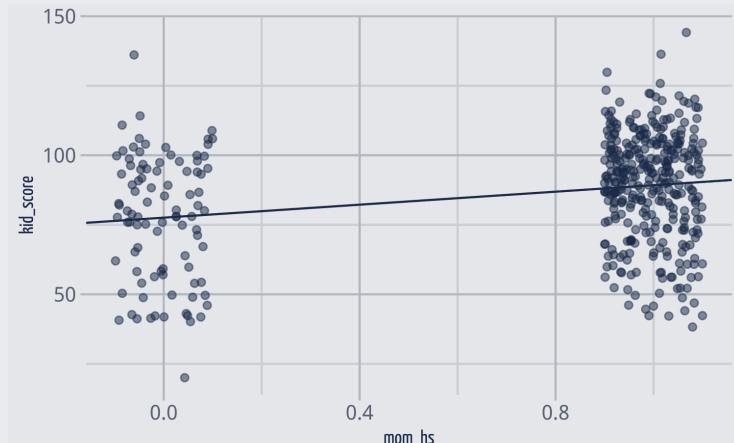
IQ von Kindern, binärer Prädiktor

```
library(rstanarm)
data("kidiq") # Paket rstanarm
m10.1 <- stan_glm(
  kid_score ~ mom_hs,
  data = kidiq)
```

Alternativ können Sie die Daten [hier](#) herunterladen.

```
coef(m10.1)
```

```
## (Intercept)      mom_hs
##    77.56016     11.70580
```



Interpretation von m10.1

m10.1: kid_score = 78 + 12*mom_hs + error

- Der *Achsensabschnitt* (intercept, α) ist der mittlere (bzw. vorhergesagte) IQ-Wert von Kindern, deren Mütter über keinen Schulabschluss verfügen:

kid_score = 78 + 0*12 + error

- Das *Regressionsgewicht* (slope, β) ist der Unterschied im IQ-Wert von Kindern mit Mütter mit Schulabschluss (im Vergleich zum IQ-Wert von Kindern mit Mütter ohne Schulabschluss). Dieser Unterschied entspricht der Steigung der Regressionsgerade.

kid_score = 78 + 1*12 + error = 90 + error

- Die *Streuung* (σ) der IQ-Wert um ihren Mittelwert zeigt, wie genau die Schätzung (Vorhersage) ist bzw. wie stark Prädiktor (UV) und Kriterium (AV) zusammenhängen.

m10.1 als Mittelwertsdifferenz

- UV: binär (zweistufig nominal/kategorial)
- AV: metrisch (quantitativ)

```
kidiq %>%
  group_by(mom_hs) %>%
  summarise(kid_score_avg =
             mean(kid_score))
```

```
## # A tibble: 2 × 2
##   mom_hs     kid_score_avg
##   <int>        <dbl>
## 1     0        77.5
## 2     1        89.3
```

- In der klassischen Statistik untersucht man diese Datensituation mit einem *t-Test*.
- Der t-Test ist ein inferenzstatistisches Verfahren, dass prüft, ob die Mittelwertsdifferenz (in der Population) μ_d Null ist: $\mu_d = 0$.
- In der Bayes-Statistik betrachtet man dazu die Posteriori-Verteilung (z.B. mit 95%PI).

Der mittlere (average, avg) IQ-Wert unterscheidet sich um ca. 12 Punkte (89.4 - 77.6), zugunsten der Kinder von Müttern mit Abschluss. Allerdings gibt es viel Streuung um die Mittelwerte herum.

Antwort auf die Forschungsfrage, m10.1

```
m10.1_post <-  
  m10.1 %>%  
  as_tibble()  
  
dim(m10.1_post)  
  
## [1] 4000      3
```

Stichprobe aus der Post-Verteilung		
(Intercept)	mom_hs	sigma
77.1	12.3	20.3
75.9	12.3	19.8
76.8	12.0	20.3
79.2	11.4	20.5
76.8	12.0	19.6

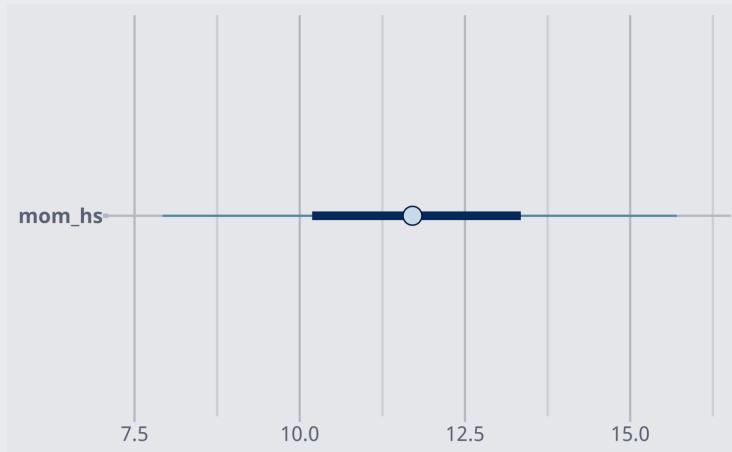
```
pi_mom_hs <-  
  m10.1_post %>%  
  summarise(  
    pi_95 = quantile(  
      mom_hs,  
      c(.025, .975)))
```

Mit 95% Wahrscheinlichkeit liegt der Unterschied im mittleren IQ-Wert zwischen Kindern von Müttern mit bzw. ohne Schulabschluss im Bereich von 7 bis 14 IQ-Punkten, laut unserem Modell: 95%PI : [7, 14].

Die Hypothese, dass es keinen Unterschied oder einen Unterschied in die andere Richtung geben sollte, ist vor diesem Hintergrund als unwahrscheinlich abzulehnen.

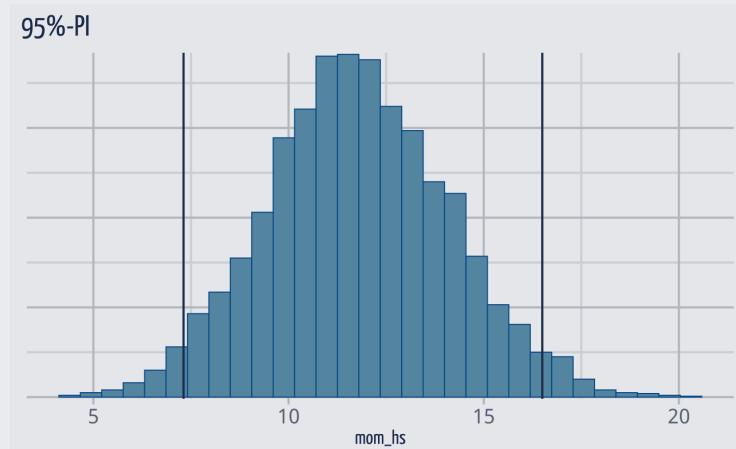
Visualisierung der Mittelwertsdifferenz

```
library(bayesplot)
plot(m10.1,
      pars = "mom_hs")
```

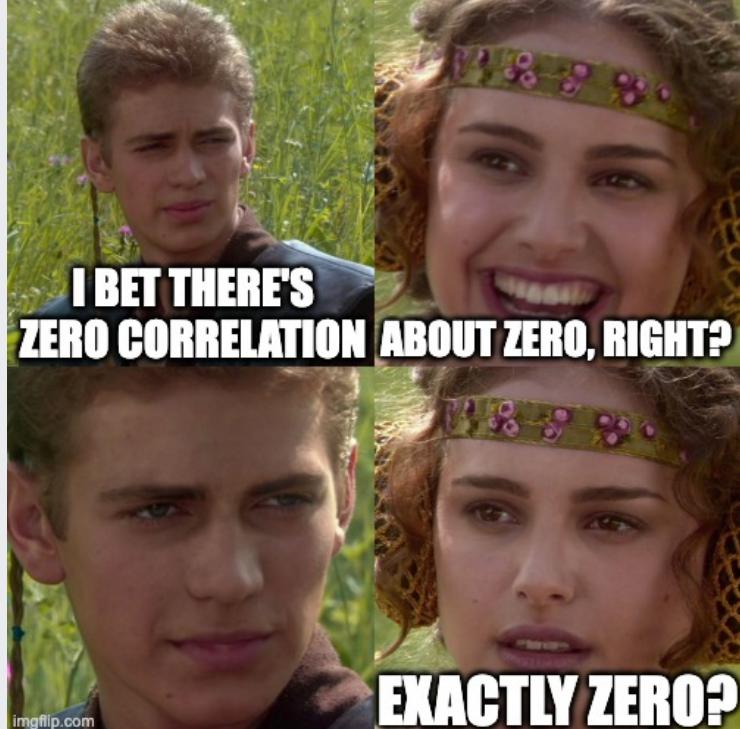


Im Standard werden 50%- und 90%-PI gezeigt.

```
plot(m10.1,
      pars = "mom_hs",
      plotfun = "mcmc_hist") +
geom_vline(
  xintercept =
  pi_mom_hs$pi_95) +
labs(
  title = "95%-PI"
)
```



Nullhypotesen sind praktisch immer falsch



Quelle: Imgflip Meme Generator

We do not generally use null hypothesis significance testing in our own work. In the fields in which we work, we do not generally think null hypotheses can be true: in social science and public health, just about every treatment one might consider will have *some* effect, and no comparison or regression coefficient of interest will be exactly zero. We do not find it particularly helpful to formulate and test null hypotheses that we know ahead of time cannot be true.

Gelman, Hill, and Vehtari (2021)

Alternativen zu Nullhypotesen

- Nullhypotesen, H_0 , sind z.B.: $\rho = 0, \rho_1 = rho_2, \mu_1 = \mu_2, \mu = 0, \beta_1 = 0$.
- Nullhypotesen zu testen, ist sehr verbreitet.
- Ein Grund ist, dass in der Frequentistischen Statistik keine andere Art von Hypothesentest möglich ist.
- Ein anderer Grund ist vermutlich, ... wir haben es schon immer so gemacht.
- Alternativen zum Testen von Nullhypotesen:
 - Posteriori-Intervalle (PI oder HDI) berichten
 - Rope-Konzept ([Kruschke, 2018](#))
 - Wahrscheinlichkeit von inhaltlich bedeutsamen Hypothesen quantifizieren.
 - Wahrscheinlichkeit quantifizieren, dass der Effekt ein positives bzw. ein negatives Vorzeichen hat.

Teil 2

Eine metrische plus eine nominale UV

Forschungsfrage

Wie stark ist der Zusammenhang von jeweils Schulabschluss der Mutter (mom_hs) und IQ der Mutter (mom_iq) auf den IQ des Kindes (kid_score) ?

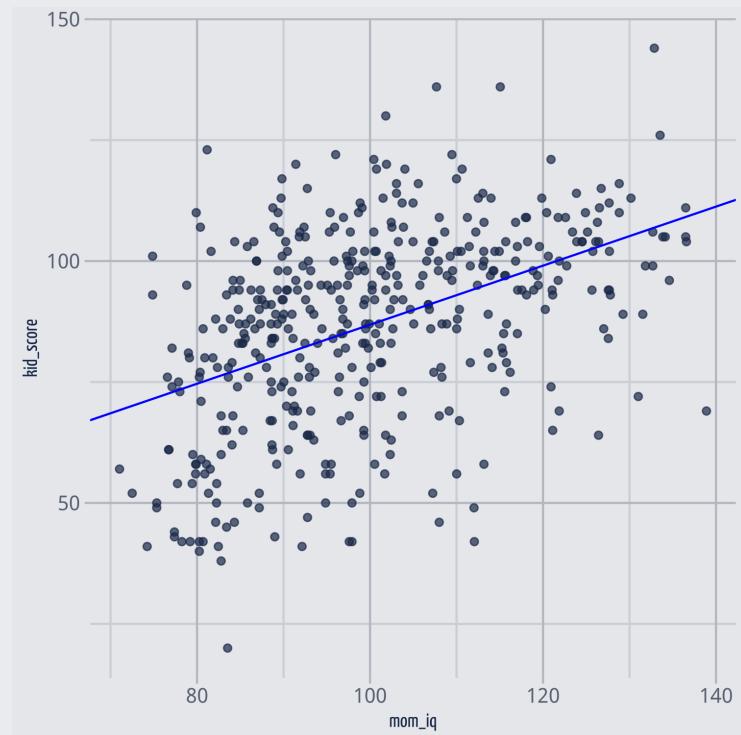
```
library(rstatix)
data("kidiq") # Paket rstanarm, alternativ über CSV einlesen
get_summary_stats(kidiq)
```

variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
kid_score	434	20.00	144.00	90.00	74.00	102.00	28.00	19.27	86.80	20.41	0.98	1.93
mom_age	434	17.00	29.00	23.00	21.00	25.00	4.00	2.96	22.79	2.70	0.13	0.26
mom_hs	434	0.00	1.00	1.00	1.00	1.00	0.00	0.00	0.79	0.41	0.02	0.04
mom_iq	434	71.04	138.89	97.92	88.66	110.27	21.61	15.89	100.00	15.00	0.72	1.42

Datenquelle

Ein metrischer Prädiktor

$$\text{kid_score} = 26 + 0.6 * \text{mom_iq} + \text{error}$$



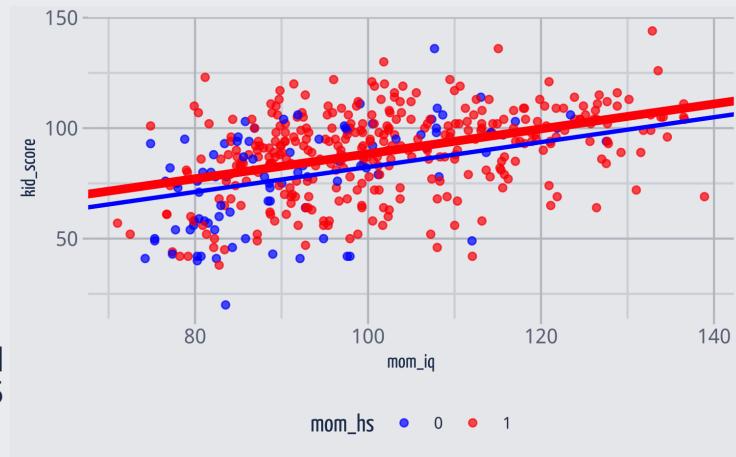
- Die Linie zeigt die vorhergesagten IQ-Werte der Kinder für verschiedene IQ-Werte der Mütter.
- Vergleicht man Teilpopulationen von Müttern mit mittleren Unterschied von einem IQ-Punkt, so findet man 0.6 IQ-Punkte Unterschied bei ihren Kindern *im Durchschnitt*.
- Der Achsenabschnitt hilft uns nicht weiter, da es keine Menschen mit einem IQ von 0 gibt.

Beide Prädiktoren, m10.3

m10.3: kid_score = 26 + mom_hs + 0.6*mom_iq + error

```
m10.3 <-  
  stan_glm(  
    kid_score ~ mom_hs + mom_iq,  
    refresh = 0,  
    data = kidiq)  
coef(m10.3)
```

```
## (Intercept)      mom_hs      mom_iq  
## 25.8222602    5.9211910    0.5642005
```

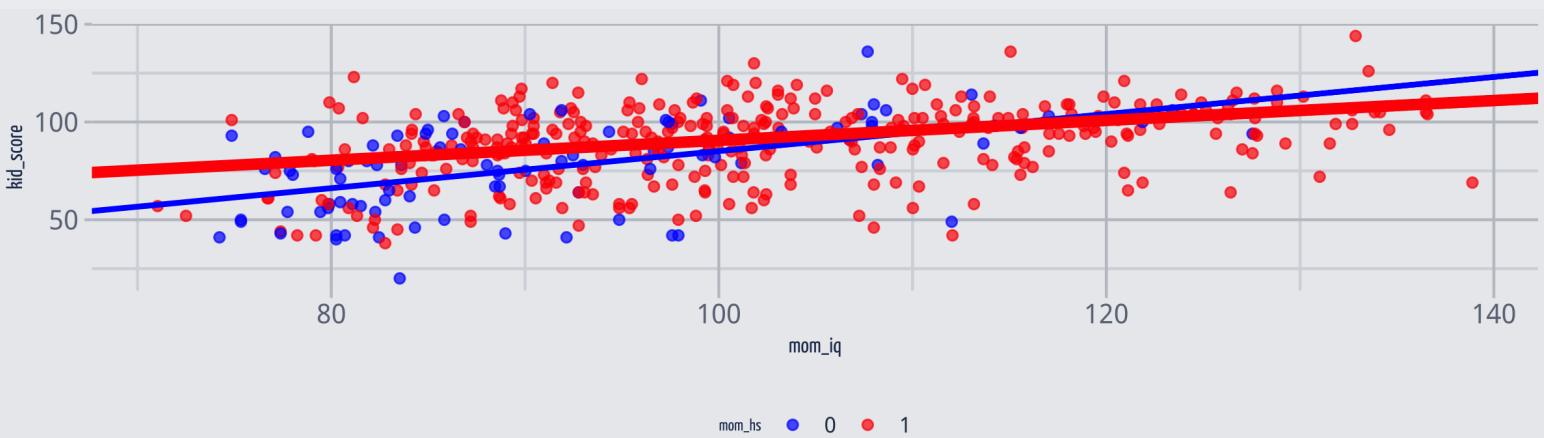


- *Achsenabschnitt:* Hat das Kind eine Mutter mit einem IQ von 0 und ohne Schulabschluss, dann schätzt das Modell den IQ-Wert des Kindes auf 26.
- *Koeffizient zum mütterlichen Schulabschluss:* Vergleicht man Kinder von Müttern gleicher Intelligenz, aber mit Unterschied im Schulabschluss, so sagt das Modell einen Unterschied von 6 Punkten im IQ voraus.
- *Koeffizient zur müttlichen IQ:* Vergleicht man Kinder von Müttern mit gleichem Wert im Schulabschluss, aber mit 1 IQ-Punkt Unterschied, so sagt das Modell einen Unterschied von 0.6 IQ-Punkten bei den Kindern voraus.

Interaktion

- In m10.3 hat das Modell die Regressionsgeraden gezwungen, parallel zu sein.
- Betrachtet man das Streudiagramm, so sieht man, dass nicht-parallele Geraden besser passen.
- Sind die Regressionsgeraden nicht parallel, so spricht man von einer Interaktion (synonym: Interaktionseffekt, Moderation).
- Liegt eine Interaktion vor, so unterscheidet sich also die Steigung in den Gruppen.

```
m10.4 <-  
  stan_glm(kid_score ~ mom_hs + mom_iq + mom_hs:mom_iq,  
            data = kidiq, refresh = 0)
```



Interpretation von m10.4

- Achsenabschnitt: IQ-Schätzwerte für Kinder mit Mütter ohne Abschluss und mit einem IQ von 0. Kaum zu interpretieren.
- mom_hs: Unterschied der IQ-Schätzwerte zwischen Kindern mit Mutter ohne bzw. mit Schulabschluss und jeweils mit einem IQ von 0. Puh.
- mom_iq: Unterschied der IQ-Schätzwerte zwischen Kindern mit Müttern, die sich um einen IQ-Punkt unterscheiden aber jeweils ohne Schulabschluss.
- Interaktion: Der Unterschied in den Steigungen der Regressiongeraden, also der Unterschied des Koeffizienten für mom_iq zwischen Mütter mit bzw. ohne Schulabschluss.

mom_hs=0:

$$\begin{aligned}\text{kid_score} &= -11 + 51 \cdot 0 + 1.1 \cdot \text{mom_iq} + 0.5 \cdot 0 \cdot \text{mom_iq} \\ &= -11 + 1.1 \cdot \text{mom_iq}\end{aligned}$$

mom_hs=1:

$$\begin{aligned}\text{kid_score} &= -11 + 51 \cdot 1 + 1.1 \cdot \text{mom_iq} + 0.5 \cdot 1 \cdot \text{mom_iq} \\ &= 40 + 0.6 \cdot \text{mom_iq}\end{aligned}$$

(Gelman, Hill, and Vehtari, 2021, Kap. 10.3)

Nach der Interpretation von 20 unzentrierten Koeffizienten



via GIPHY

Zentrieren von Prädiktoren

- Unter *Zentrieren* (to center) versteht man das Bilden der Differenz eines Messwerts zu seinem Mittelwert.
- Zentrierte Werte geben also an, wie weit ein Messwert vom mittleren (typischen) Messwert entfernt ist.
- Mit zentrierten Werten ist eine Regression einfacher zu interpretieren.

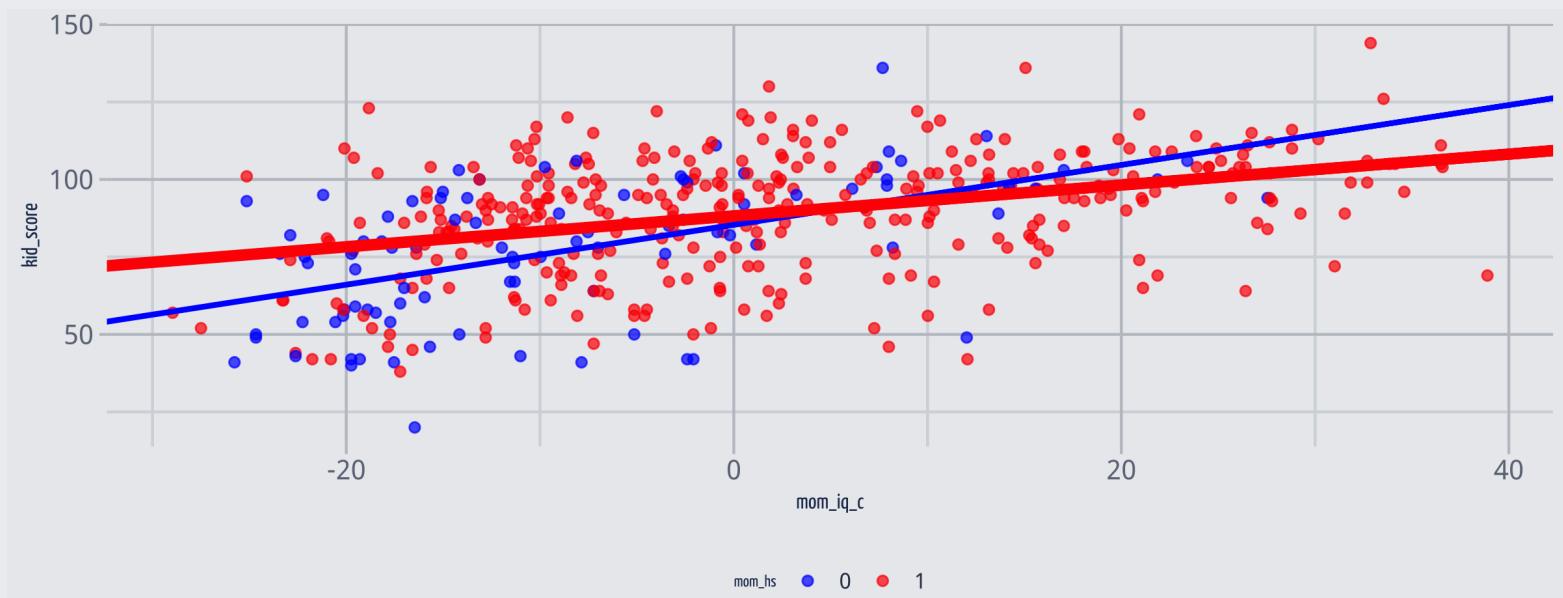
```
kidiq <-  
  kidiq %>%  
  mutate(mom_iq_c = mom_iq - mean(mom_iq),  
         mom_hs_c = mom_hs - mean(mom_hs))
```

```
m10.5 <- stan_glm(kid_score ~ mom_hs + mom_iq_c + mom_hs:mom_iq_c,  
                     data = kidiq,  
                     refresh = 0)  
coef(m10.5)
```

```
##      (Intercept)          mom_hs  
##      85.3986939     2.8648038  
##      mom_iq_c mom_hs:mom_iq_c  
##      0.9668627    -0.4837153
```

Interpretation von m10.5

- Der Achsenabschnitt (Intercept) gibt den geschätzten IQ des Kindes an, wenn man eine Mutter *mittlerer* Intelligenz und *ohne* Schulabschluss betrachtet.
- mom_hs gibt den Unterschied im geschätzten IQ des Kindes an, wenn man Mütter mittlerer Intelligenz aber mit bzw. ohne Schlusabschluss vergleicht.
- mom_iq_c gibt den Unterschied im geschätzten IQ des Kindes an, wenn man Mütter ohne Schlusabschluss aber mit einem IQ-Punkt Unterschied vergleicht.
- mom_hs : mom_iq_c gibt den Unterschied in den Koeffizienten für mom_iq_c an zwischen den beiden Gruppen von mom_hs.



Zentrieren ändert nichts an den Vorhersagen

m10.4:

```
new <- tibble(mom_hs = 0, mom_iq = mean(kidiq$mom_iq))
pred_new <- posterior_predict(m10.4, newdata = new)
mean(pred_new)
```

```
## [1] 85.35087
```

m10.5:

```
new <- tibble(mom_hs = 0, mom_iq_c = 0)
pred_new <- posterior_predict(m10.5, newdata = new)
mean(pred_new)
```

```
## [1] 85.65857
```

Auch die Streuungen der vorhergesagten Werte unterscheiden sich nicht (wirklich): $\sigma_{m10.4} = 18$; $\sigma_{m10.5} = 18$.

Das Zentrieren ändert auch nicht die Regressionskoeffizienten, da die Streuungen der Prädiktoren nicht verändert wurden.

Perzentilintervalle aus der Posteriori-Verteilung

```
posterior_interval(m10.5)
```

```
##                                5%      95%
## (Intercept)     81.5411087 88.9792978
## mom_hs        -1.0468886  6.9571692
## mom_iq_c       0.7224591  1.2092989
## mom_hs:mom_iq_c -0.7465962 -0.2097615
## sigma          17.0058126 19.0427958
```

Highest Density (Posterior) Intervalle (HDI oder HDPI) kann man sich komfortabel ausgeben lassen mit `hdi(m10.5)` aus dem Paket `bayestestR`.

```
hdi(m10.5)
```

Im Falle symmetrischer Posteriori-Verteilungen (wie hier) kommen beide Arten von Intervallen zu gleichen Ergebnissen.

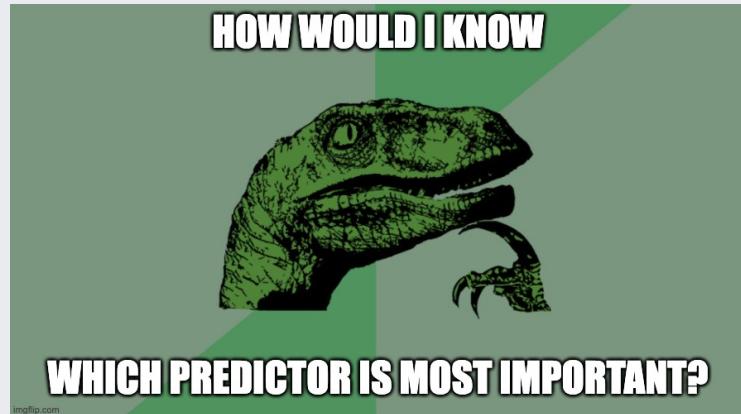
Beantworten der Forschungsfrage

Das Modell zeigt keine Belege, dass sich die mittlere Intelligenz von Kindern bei Müttern mit bzw. ohne Schulabschluss unterscheidet (95%PI: [-2.0, 7.8]). Hingegen fand sich ein Effekt der mütterlichen Intelligenz; pro Punkt Unterschied in mittlerlichem IQ fand sich ein Unterschied von 0.7 bis 1.3 IQ-Punkte (95%PI). Außerdem fand sich ein Beleg, dass der Zusammenhang des IQ zwischen Mutter und Kind durch den Schulabschluss moderiert wird: Bei Mütter mit Schulabschluss war der Zusammenhang zwischen Mutter-IQ und Kind-IQ geringer (95%PI: [-0.80, -0.17]).



Das Modell macht *keine* kausalen Aussagen. Es werden lediglich Unterschiede bzw. Zusammenhänge beschrieben.

Relevanz von Prädiktoren



from Imgflip Meme Generator

- Das Problem ist, dass die Prädiktoren auf verschiedenen Skalen gemessen wurden, so dass sie nicht direkt vergleichbar sind.
- Man könnte
 - die Skalierungen der Prädiktoren anpassen
 - Vorhersagegüte verschiedener Modelle vergleichen (Modell mit einem vs. Modell mit beiden Prädiktoren)
 - ...
- Dazu später mehr 😊

Betrachten wir m10.3 noch einmal.

Welcher Prädiktor, mom_hs oder mom_iq ist wichtiger im Sinne von stärker mit AV kid_score verbunden?

Parameter	95% PI

(Intercept)	[14.04, 37.31]
mom_hs	[1.58, 10.34]
mom_iq	[0.44, 0.69]

Teil 3

Eine nominale UV mit mehreren Stufen

Forschungsfrage

Hintergrund:

Nach Ihrem Studium haben Sie bei einem aufstrebenden Online-Händler angeheuert. Sie sind für alles zuständig, was nach Wissenschaft oder Analyse aussieht oder sonst den Anschein erweckt, kompliziert zu sein. Aus irgendwelchen Gründen liegt Ihre Chefin Diamanten, so dass Ihre erste Aufgabe darin besteht, Diamantenpreise zu analysieren. Das Ziel Ihrer Chefin liegt darin, zu erkennen, was ein Preis ist, für den ein Diamant gut verkauft werden kann. Kennt man diesen Preis, kann man sich auf die Suche machen, ob man den Diamant irgendwo günstiger find. Wenn ja, macht man Gewinn. Gutes Geschäftsmodell, meint Ihre Chefin.

Unterscheiden sich mittlere Diamantenpreise in Abhängigkeit von ihrer Schliffart? (Datensatz diamonds)

Alle Mittelwerte sind gleich (?)

- Formal: $\mu_1 = \mu_2 = \dots = \mu_k$ mit k verschiedenen Gruppen von Schliffart.
- Hypothesen, die keinen (Null) Unterschied zwischen Gruppen oder keinen Zusammenhang zwischen Variablen postulieren, kann man als *Nullhypotesen* bezeichnen.
- Moment. Dass sich *alle* Mittelwerte um 0,00000000 unterscheiden, ist wohl nicht zu vermuten. Wer glaubt sowas? 🤔 Daher ist die bessere Forschungsfrage:

Wie sehr unterscheiden sich mittlere Diamantenpreise in Abhängigkeit von ihrer Schliffart?

Oder wir prüfen die Hypothese, ob die Mittelwerte "praktisch" gleich sind, also sich "kaum" unterscheiden. Der Grenzwert für "praktisch gleich" bzw. "kaum unterschiedlich" ist subjektiv.

Erster Blick in den Datensatz diamonds

Datenquelle, Beschreibung des Datensatzes

```
diamonds_url <- "https://tinyurl.com/up84jp5c"
```

```
set.seed(42) # Zufallszahlen für `sample()` festlegen
diamonds <-
  read_csv(diamonds_url) %>%
  sample_n(1000) %>% # um etwas Rechenzeit zu sparen
  select(-1) # 1. Spalte nur laufende Nummer
```

```
diamonds %>%
  select(price, cut) %>%
  group_by(cut) %>%
  # nehmen wir die robusten Statistiken, da Preis sehr schief ist:
  get_summary_stats(type = "robust")
```

Ein Überblick über die metrischen Variablen

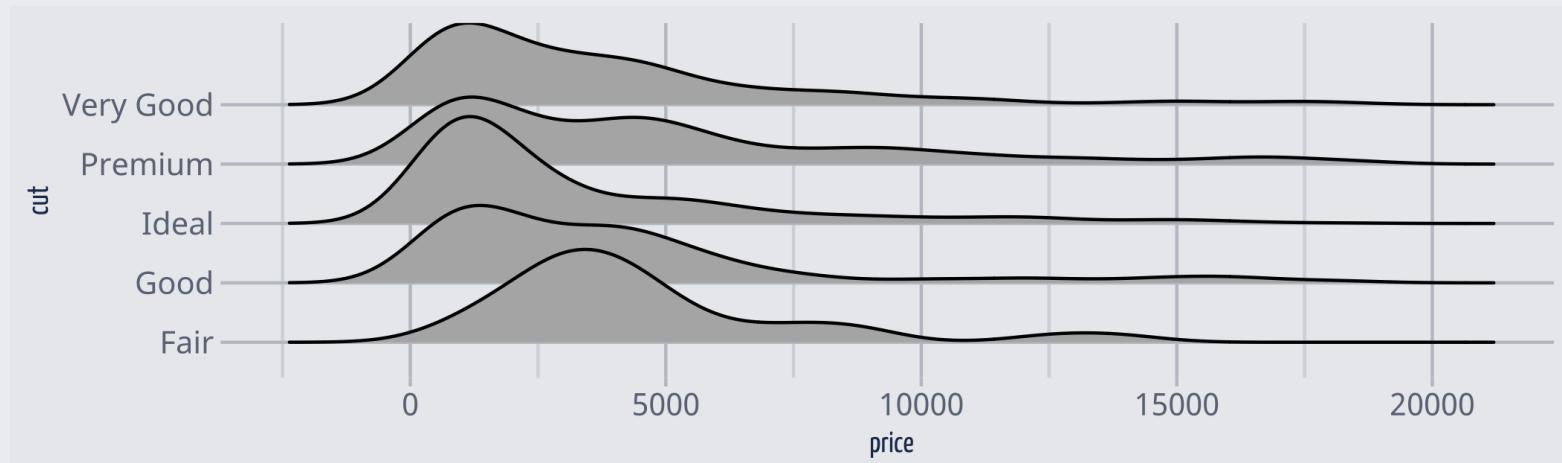
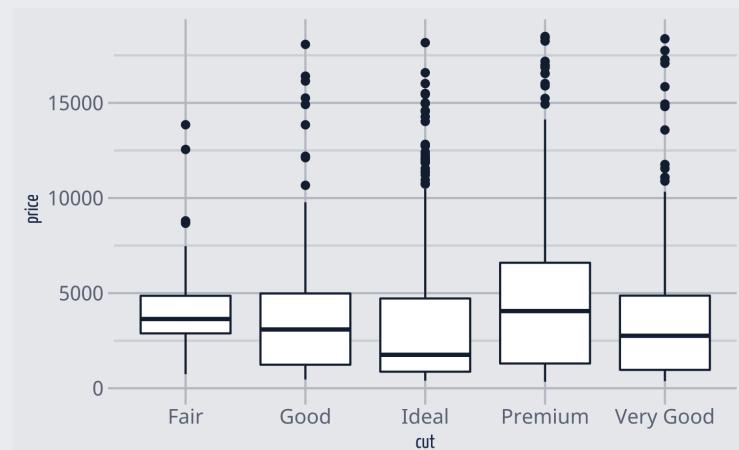
... aufgeteilt in die Stufen von cut:

cut	variable	n	median	iqr
Fair	price	33.0	3,640.0	1,976.0
Good	price	102.0	3,089.5	3,744.8
Ideal	price	407.0	1,759.0	3,849.5
Premium	price	230.0	4,057.5	5,295.2
Very Good	price	228.0	2,759.5	3,902.5

Was fällt Ihnen auf?

Visualisierung (EDA)

cut	price_md	price_iqr
Fair	3,640	1,976
Good	3,090	3,745
Ideal	1,759	3,850
Premium	4,058	5,295
Very Good	2,760	3,902



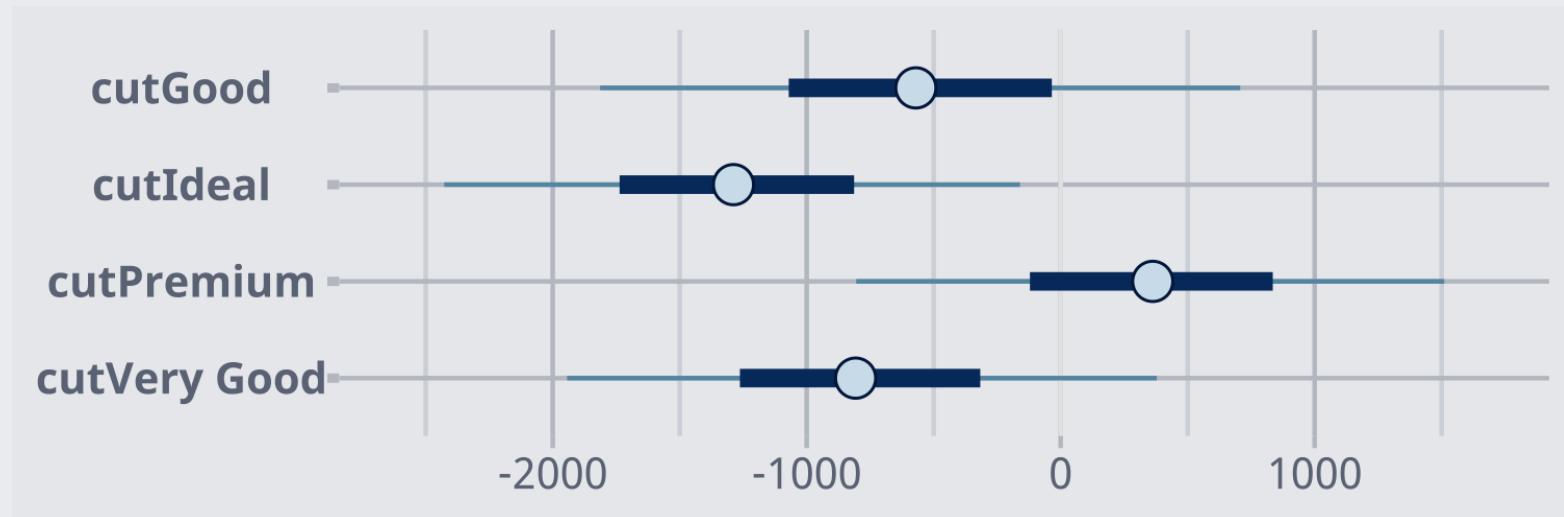
Mittlere Preisunterschiede in der Population

```
options(mc.cores = parallel::detectCores()) # Turbo einschalten  
  
m10.6 <- stan_glm(price ~ cut, data = diamonds, refresh = 0)  
# refresh=0 unterdrückt Ausgabe der Posteriori-Stichproben
```

```
## stan_glm  
## family: gaussian [identity]  
## formula: price ~ cut  
## observations: 1000  
## predictors: 5  
## -----  
##           Median  MAD_SD  
## (Intercept)  4571.7  675.1  
## cutGood     -570.2  777.2  
## cutIdeal    -1288.3  688.1  
## cutPremium   362.5  709.8  
## cutVery Good -807.4  706.3  
##  
## Auxiliary parameter(s):  
##           Median  MAD_SD  
## sigma 3795.0   82.4  
##  
## -----  
## * For help interpreting the printed output see ?print.stanreg  
## * For info on the priors used see ?prior_summary.stanreg
```

Interpretation von m10.6

- cut hat fünf verschiedene Werte (Stufen, Faktorstufen, Ausprägungen), aber es werden nur vier angezeigt.
- Die fünfte (Fair, nicht ausgegeben) ist die *Vergleichs- oder Referenzkategorie* (baseline) und ist im Achsenabschnitt ausgedrückt.
- Die Koeffizienten für cut geben jeweils den Unterschied zur Vergleichskategorie wieder.
- Diamanten der Schliffart Fair haben laut Modell einen mittleren Preis von ca. 4300\$.
- Diamanten der Schliffart Good sind laut Modell im Mittel gut 400\$ billiger als Diamanten der Schliffart Fair, etc.



Schäzbereiche (PI) für die Modellparameter

```
m10.6_post <-  
  m10.6 %>%  
  as_tibble()  
  
grenzen <- c(0.025, 0.975)  
  
m10.6_post %>%  
  summarise(pi_intercept = quantile(`(Intercept)` , probs = grenzen),  
            cutGood = quantile(cutGood, probs = grenzen),  
            cutIdeal = quantile(cutIdeal, probs = grenzen),  
            cutPremium = quantile(cutPremium, probs = grenzen),  
            `cutVery Good` = quantile(`cutVery Good` , probs = grenzen))  
  
## # A tibble: 2 × 5  
##   pi_intercept cutGood cutIdeal  
##       <dbl>    <dbl>    <dbl>  
## 1      3314.   -2040.   -2630.  
## 2      5837.    949.     18.0  
## # ... with 2 more variables:  
## #   cutPremium <dbl>,  
## #   cutVery Good <dbl>
```

Variablennamen, die nach R-Rechtschreiberegeln verboten sind, wie `cutVery Good` oder `(Intercept)`, müssen mit Backticks angeführt werden:

Schäzbereiche (PI) ausgeben lassen, komfortabel

Einfacher bekommt man die gleiche Ausgabe z.B. so:

```
posterior_interval(m10.6,  
                    prob = .89)
```

```
##                      5.5%      94.5%  
## (Intercept)    3528.7861 5617.8655  
## cutGood        -1781.6511  663.9931  
## cutIdeal       -2385.1021 -201.8007  
## cutPremium     -765.0572 1468.4457  
## cutVery Good  -1915.6286  352.0496  
## sigma          3667.2492 3937.5775
```

Hilfeseite für diese Funktion

95%-HDI für die Regressionskoeffizienten

Da die Forschungsfrage nur auf die Regressionskoeffizienten, nicht auf σ abzielt, brauchen wir uns σ auch nicht en Detail anschauen.

Das HDI kann man komfortabel z.B. so bekommen:

```
bayestestR::hdi(m10.6)

## Highest Density Interval
##
## Parameter | 95% HDI
## -----
## (Intercept) | [ 3299.64, 5816.35]
## cutGood | [-2040.26, 952.51]
## cutIdeal | [-2645.03, -3.72]
## cutPremium | [-965.18, 1732.27]
## cutVery Good | [-2102.22, 626.49]
```

Wie man sieht, sind die Intervallgrenzen des HDI ähnlich zu denen des PI.

Glauben wir jetzt an Preisunterschiede?

... zwischen den Preis-Mittelwerten in der Population?

- Teilweise, denn einige Schätzintervalle (für die Preisunterschiede) waren im Modell m10.6 weit von der Null entfernt, andere nicht.
- Auf Basis unseres Modells schließen wir also (mit hoher Sicherheit) aus, dass *alle* Preise im Mittelwert *exakt* identisch sind.
- Ehrlicherweise hätte sowieso niemand geglaubt, dass die *exakte Nullhypothese* $\mu_1 = \mu_2 = \dots = \mu_k$ bis in die letzte Dezimale gilt. Anders gesagt: Die Wahrscheinlichkeit eines bestimmten Wertes einer stetigen Zufallsvariable ist praktisch Null.
- Aber: Viele Forscheris prüfen gerne die Nullhypothese, daher taucht der Begriff hier auf.
- Das Verfahren der Frequentistischen Statistik, um die Nullhypothese $\mu_1 = \mu_2 = \dots = \mu_k$ zu testen, nennt man *Varianzanalyse* (analysis of variance, kurz ANOVA).
- In der Bayes-Statistik nutzt man - wie immer - primär die Post-Verteilung, um Fragen der Inferenz (z.B. Gruppenunterschiede dieser Art) zu inferenzstatistisch zu beurteilen.

Modellkoeffizienten von m10.6

Die Regressionsoeffizienten pro Stufen von cut entsprechen den Mittelwerten \hat{y}_i aus der Posteriori-Verteilung. Mit `coef(m10.6)` kann man sie sich bequem ausgeben lassen.

```
coef(m10.6)
```

```
##   (Intercept)      cutGood      cutIdeal
##   4571.7415     -570.2364    -1288.3183
##   cutPremium  cutVery Good
##   362.4788     -807.3535
```

- $\hat{y}_{Fair} = 4572$
- $\hat{y}_{Good} = -570$
- etc.

Wechsel der Referenzkategorie

- cut ist eine nominale Variable, da passt in R der Typ factor (Faktor) am besten. Aktuell ist der Typ noch character (Text):

```
diamonds <- diamonds %>%  
  mutate(cut = factor(cut))
```

- Im Standard sortiert R die Faktorstufen alphabetisch, aber man kann die Reihenfolge ändern.

```
levels(diamonds$cut)
```

```
## [1] "Fair"      "Good"      "Ideal"  
## [4] "Premium"   "Very Good"
```

Setzen wir Ideal als Referenzkategorie und lassen die restliche Reihenfolge, wie sie ist:

```
library(forcats)  
diamonds <- diamonds %>%  
  mutate(cut = factor(cut),  
        cut = fct_relevel(cut, "Ideal"))
```

Wechsel der Referenzkategorie ändert nichts Wesentliches am Modell

```
m10.6a <- stan_glm(price ~ cut, data = diamonds, refresh = 0)
m10.6a
```

```
## stan_glm
##   family:      gaussian [identity]
##   formula:     price ~ cut
##   observations: 1000
##   predictors:  5
## -----
##           Median MAD_SD
## (Intercept) 3280.3 187.5
## cutFair     1268.9 671.4
## cutGood      725.4 404.6
## cutPremium   1645.8 305.9
## cutVery Good  485.8 309.8
##
## Auxiliary parameter(s):
##           Median MAD_SD
## sigma 3798.3  86.9
##
## -----
## * For help interpreting the printed output see ?print.stanreg
```

Modellgüte mit R^2 bestimmen

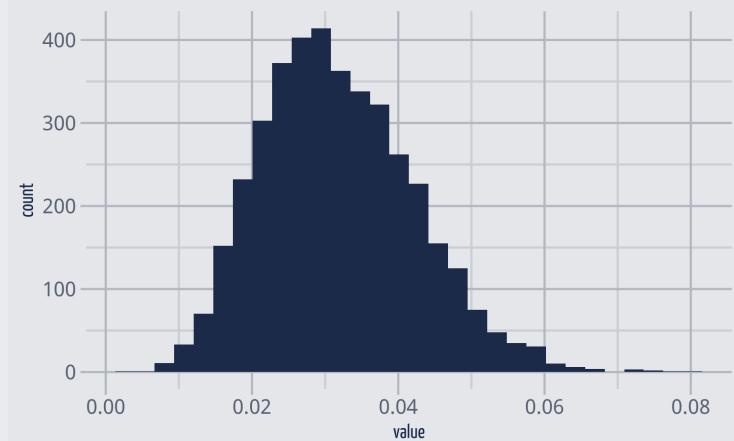
- R^2 gibt den Anteil der Gesamtvarianz (der AV) an, den das Modell erklärt.
- Höhere Wert von R^2 bedeuten, dass das Modell die Daten besser erklärt.
- R^2 wird normalerweise auf Basis eines Punktschätzers definiert.
- Solch eine Definition lässt aber viel Information - über die Ungewissheit der Schätzung - außen vor.
- Daher ist es wünschenswert, diese Information in R^2 einfließen zu lassen: *Bayes-R-Quadrat*.
- mit `bayes_r2()` kann man sich die Verteilung berechnen lassen.

```
m10.6_r2 <- bayes_R2(m10.6)  
median(m10.6_r2)
```

```
## [1] 0.0308118
```

```
IQR(m10.6_r2)
```

```
## [1] 0.01445986
```



Definition vom "klassischen" R^2

- Wie genau sind die Vorhersagen des Modells? σ (Vorhersagefehler) quantifiziert die Streuung der Residuen $r_i = y_i - X_i \hat{\beta}$, mit $\hat{y}_i = X_i \hat{\beta}$.
- Anders gesagt: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots = X \hat{\beta}$.
- Anders gesagt gibt σ die "typische" Abweichung einer Beobachtung vom vorhergesagten Wert an.
- Es ist nützlich, σ in Bezug zu setzen zur Streuung der AV, $sd_y = s_y$:
- $R^2 = 1 - (\hat{\sigma}^2 / s_y^2)$.
- R^2 gibt damit den Anteil der vom Modell erklärten Varianz, V , an.
- Berechnet man das Modell mit der Methode der kleinsten Quadrate (nicht mit Bayes), dann ist der obige Ausdruck äquivalent zu:
- $R^2 = V_{i=1}^n \hat{y}_i / s_y^2$
- Die beiden obigen Ausdrücke nehmen \hat{y}_i als fix (sicher) an und vernachlässigen Ungewissheit; sie sind übergewiss aus Bayes-Sicht.

Bayes' R^2

- Besser ist es (aus Bayes-Sicht), die Ungewissheit der Vorhersagen bei der Berechnung der Modellgüte miteinzubeziehen:
- Bayes $R^2 = \frac{\text{erklärte Varianz}}{\text{Erklärte Varianz} + \text{Residualvarianz}} = \frac{V_{mod}}{V_{mod} + V_{res}}$
- V_{mod} ist die Varianz in der PPV mit $s = 1, \dots, S$ simulierten Stichproben, $V(\hat{y}_i)$ und V_{res} ist die Residualvarianz im Modell.
- Für jede Stichprobe s berechnet man die vorhergesagten Werte, \hat{y}_i^s , die Residualvarianz σ_s^2 und den Anteil der erklärten Varianz:
- Bayes $R_s^2 = \frac{V(\hat{y}_i^s)}{V(\hat{y}_i^s + \sigma_s^2)}$

(Gelman, Goodrich, Gabry, and Vehtari, 2019; Gelman, Hill, and Vehtari, 2021, Kap. 11.7)

Priori-Werte

- Unser Modell hat schwach informierte (weakly informative) Priors.
- Für Achsenabschnitt und die Regressionskoeffizienten werden Normalverteilungen angenommen mit Mittelwert entsprechend den Stichprobendaten und Streuung, die der 2.5-fachen der Streuung in der Stichprobe entspricht.
- Mehr Infos kann man sich so ausgeben lassen: `prior_summary(m10.6)`.
- Wo man über mehr inhaltliches Wissen verfügt, so wird man die Priors anpassen wollen, z.B.:

```
m10.6b <- stan_glm(price ~ cut, data = diamonds, refresh = 0,
                      prior = normal(location = c(100, 100, 100, 100),
                                    scale = c(100, 100, 100, 100)),
                      prior_intercept = normal(3000, 500))
coef(m10.6b)
```

```
##   (Intercept)      cutFair      cutGood
##   3746.62334    112.80267    105.05937
##   cutPremium  cutVery Good
##   237.13508     82.03354
```

```
sigma(m10.6b)
```

```
## [1] 3834.271
```

"Praktisch" kein Unterschied

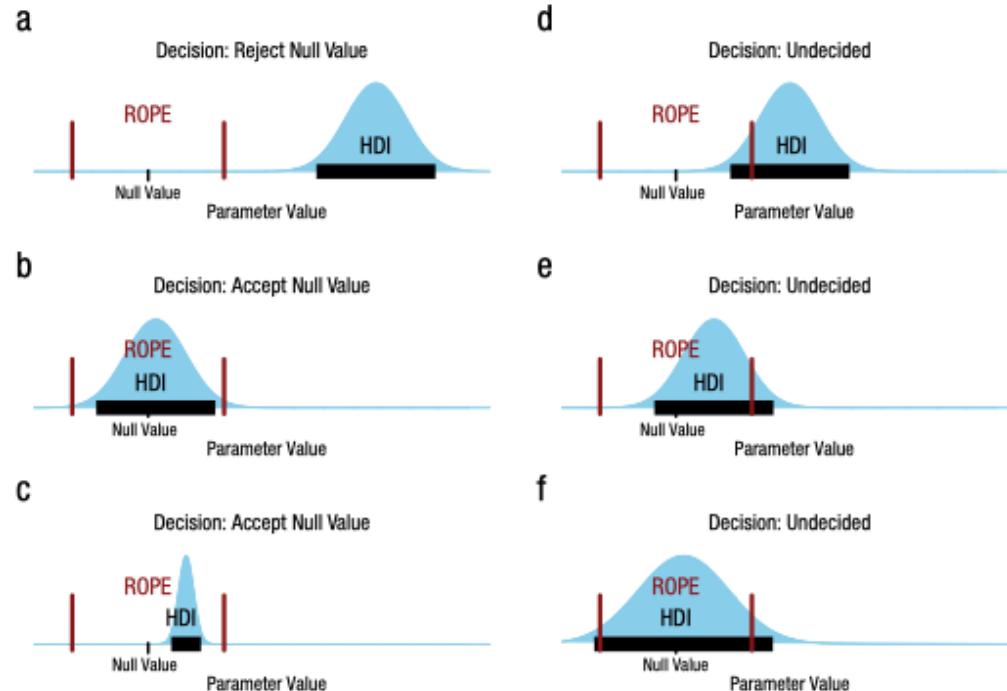
- Sagen wir, wenn sich zwei Preismittelwerte um höchstens $d = 100\text{€}$ unterscheiden, gilt dieser Unterschied für uns als "praktisch gleich", "praktisch kein Unterschied" bzw. vernachlässigbar.
- Nimmt man (praktisch) keinen Unterschied/Zusammenhang/Effekt an, spricht man von einer *Nullhypothese*: H_0 .
- Die Wahl von d ist *subjektiv* in dem Sinne als sie von inhaltlichen Überlegungen geleitet sein sollte.
- Diesen Bereich bezeichnen wir den *Indifferenzbereich* (Äquivalenzzone, Bereich eines vernachlässigbaren Unterschieds oder *Region of practical equivalence*, Rope).
- Jetzt prüfen wir, ob ein "Großteil" der Posteriori-Stichproben im Rope liegt.
- Unter "Großteil" wird häufig das *95%-HDI* verstanden.

Entscheidungsregel:

- Großteil liegt *innerhalb* von Rope → *Annahme* der Nullhypothese "praktisch kein Effekt", H_0
- Großteil liegt *außerhalb* von Rope → *Ablehnung* der Nullhypothese "praktisch kein Effekt", H_0
- Ansonsten → keine Entscheidung

(Kruschke, 2018)

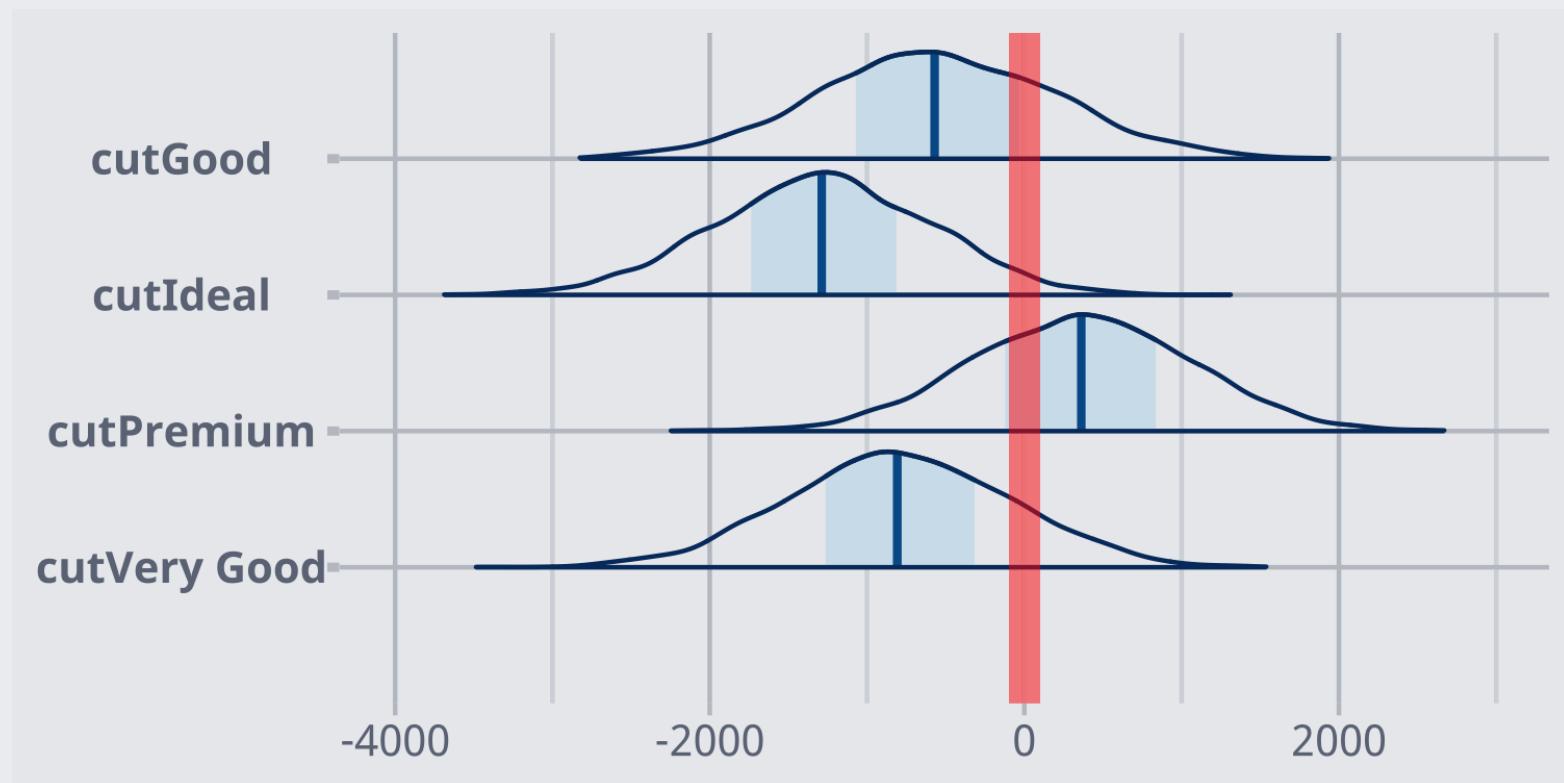
HDI-Rope-Entscheidungsregel visualisiert



(Kruschke, 2018), Abbildung 1, S. 272

Visualisierung unserer Rope-Werte, m10.6

- Ein Großteil der Posteriori-Masse von m10.6 liegt *nicht* innerhalb des Rope.
- Aber können wir umgekehrt sagen, dass ein Großteil außerhalb liegt? Das erkennt man optisch so gut.



Genaue Rope-Werte

```
library(bayestestR)
rope(m10.6, range = c(-100, 100))
```

Parameter	Anteil im ROPE	Entscheidung
(Intercept)	0.00 %	Nullhypothese verwerfen
cutGood	8.73 %	keine Entscheidung möglich
cutIdeal	1.47 %	Nullhypothese verwerfen
cutPremium	10.08 %	keine Entscheidung möglich
cutVery Good	6.42 %	keine Entscheidung möglich

Im Standard werden 95%-HDI berichtet, das kann man so ändern, wenn man möchte:

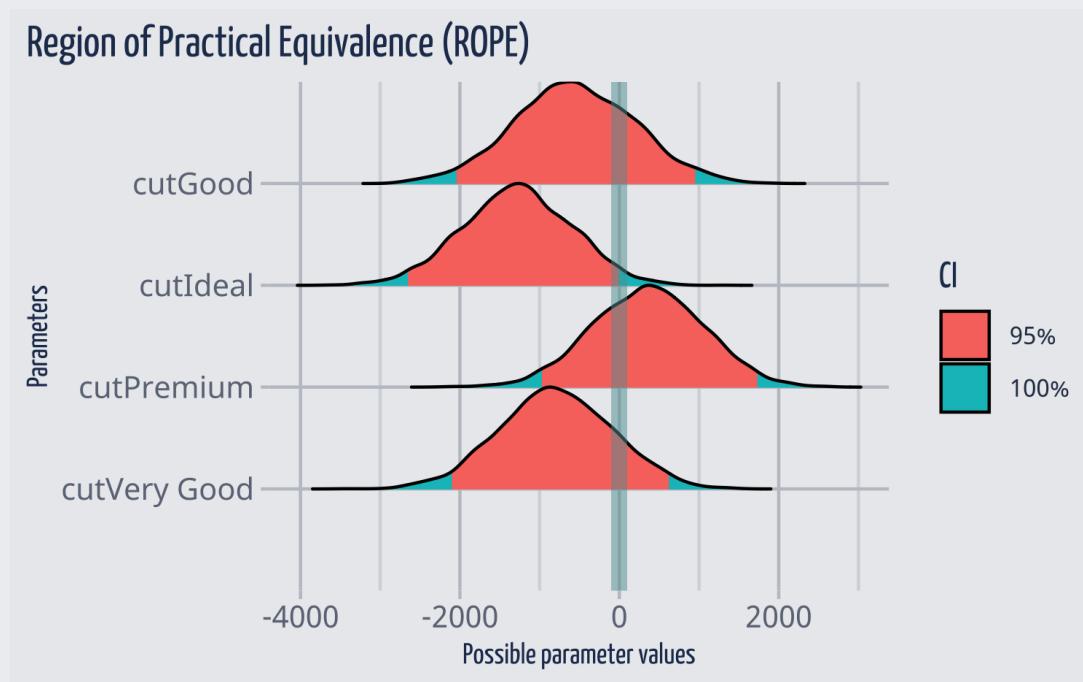
```
rope(m10.6, range, c(-100,100), ci = .89, ci_method = "ETI")
```

ETI (equal tails interval) steht für ein PI.

R-Befehl zur bequemen Visualisierung von Rope

Das Paket bayestestR bietet eine komfortable Visualisierungsfunktion:

```
plot(rope(m10.6, range = c(-100, 100)))
```



Beantwortung der Forschungsfrage

Nur das 95%-HDI für Schliffart "Ideal" schloss den Indifferenzbereich von $\pm 100\text{€}$ aus, die übrigen Mittelwertsdifferenzen nicht. Für die übrigen Differenzen ist keine klare inferenzstatistische Aussage hinsichtlich eines Indifferenzbereichs möglich: Es ist plausibel, laut dem Modell, dass es einen praktisch bedeutsamen Unterschied gibt, aber es ist auch plausibel, dass es keinen praktisch bedeutsamen Unterschied gibt.

Die 95%HDI für die Mittelwertsdifferenzen waren wie folgt: cutGood: [-2040, 953], cutIdeal: [-2645, -4], cutPremium: [-965, 1732], cutVeryGood: [-2102, 627]. Das Modell erklärte im Median ca. 3% der Varianz, also nur einen kleinen Teil.

Teil 4

Mehrere metrische UV

Forschungsfrage

| Stehen sowohl der IQ der Mutter als auch, unabhängig davon, das Alter der Mutter im Zusammenhang mit dem IQ des Kindes?

- Das ist wieder eine *deskriptive* Forschungsfrage. *Keine* Kausalwirkung (etwa "IQ der Mutter ist die Ursache zum IQ des Kindes") wird impliziert.
- Es geht rein darum, Zusammenhänge in den Daten - bzw. in der Population - aufzuzeigen.
- Viele Forschungsfragen gehen allerdings weiter und haben explizit Kausalwirkungen im Fokus. Für solche Fragen ist eine deskriptive Untersuchung nicht geeignet, sondern eine Kausalanalyse ist nötig.

Datenquelle als CSV-Datei oder alternativ:

```
library(rstanarm)
data("kidiq")
```

Was heißt, X hängt mit Y zusammen?

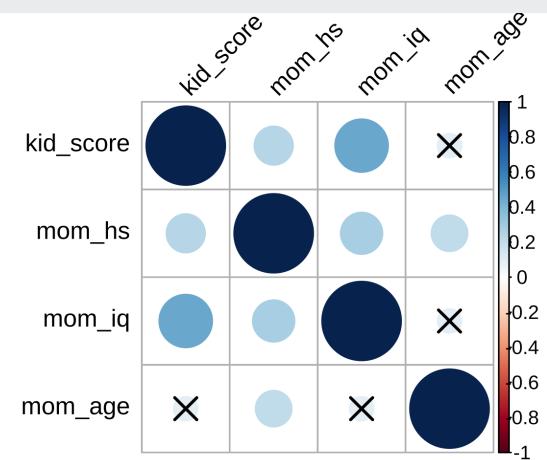
- Der Begriff "Zusammenhang" ist nicht exakt.
- Häufig wird er (für metrische Variablen) verstanden als
 - lineare Korrelation ρ bzw. r
 - lineare Regression β , bzw. b
- Der Regressionskoeffizient
 - misst die *Steigung* der Regressionsgerade
 - zeigt, wie groß der vorhergesagte Unterschied in Y, wenn man zwei Personen (Beobachtungseinheiten) vergleicht, die sich um eine Einheit in X unterscheiden
 - wird manchmal mit dem "Effekt von X auf Y" übersetzt. Vorsicht: "Effekt" klingt nach Kausalzusammenhang. Eine Regression ist keine hinreichende Begründung für einen Kausalzusammenhang.
- Der Korrelationskoeffizient
 - misst eine Art der Stärke des linearen Zusammenhangs
 - zeigt, wie klein die Vorhersagefehler der zugehörigen Regression im Schnitt sind.
 - **Korrelation ist nicht (automatisch) Kausation.**

Korrelationen zur Forschungsfrage

```
library(rstatix)
kidiq %>%
  cor_mat()
```

	kid_score	mom_hs	mom_iq	mom_age
kid_score	1.0	0.2	0.5	
mom_hs	0.2	1.0	0.3	
mom_iq	0.5	0.3	1.0	
mom_age	0.1	0.2	0.1	

```
kidiq %>%
  cor_mat() %>%
  cor_plot()
```



Univariate Regressionen

```
m10.7 <- stan_glm(kid_score ~ mom_iq, data = kidiq, refresh = 0)
m10.8 <- stan_glm(kid_score ~ mom_age, data = kidiq, refresh = 0)
```

```
coef(m10.7)
```

```
## (Intercept)      mom_iq
## 25.7278500    0.6105014
```

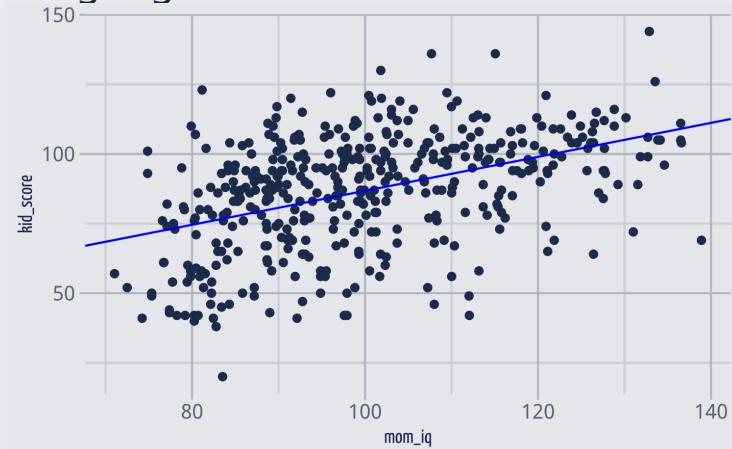
```
coef(m10.8)
```

```
## (Intercept)      mom_age
## 70.9402143    0.6997103
```

Visualisierung der univariaten Regressionen

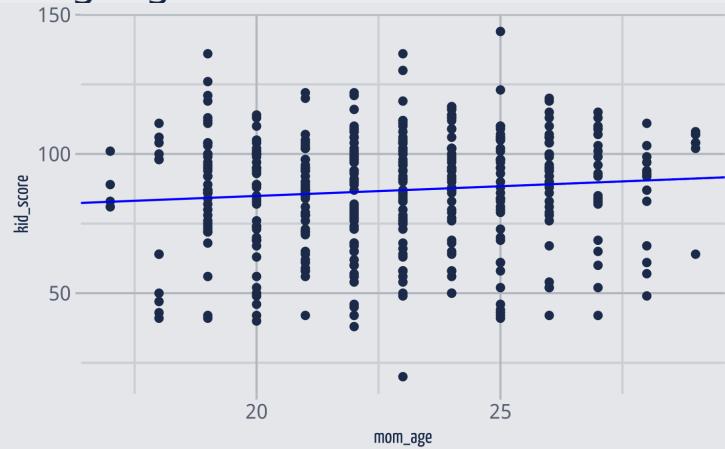
m10.7

Steigung: 0.6



m10.8

Steigung: 0.7



Multiples Modell (beide Prädiktoren)

```
m10.9 <- stan_glm(kid_score ~ mom_iq + mom_age,  
                     data = kidiq,  
                     refresh = 0)  
coef(m10.9)
```

```
## (Intercept)      mom_iq      mom_age  
## 17.5636666    0.6036358    0.3795824
```

💡 Die Regressionsgewichte unterscheiden sich zu den von den jeweiligen univariaten Regressionen.

- Bei einer multiplen Regression ist ein Regressionsgewicht jeweils *bereinigt* vom Zusammenhang mit dem (oder den) anderen Regressionsgewicht.
- Das bedeutet, man betrachtet den Zusammenhang eines Prädiktors mit der AV, wobei man gleichzeitig den anderen Prädiktor konstant hält.

```
coef(m10.8)
```

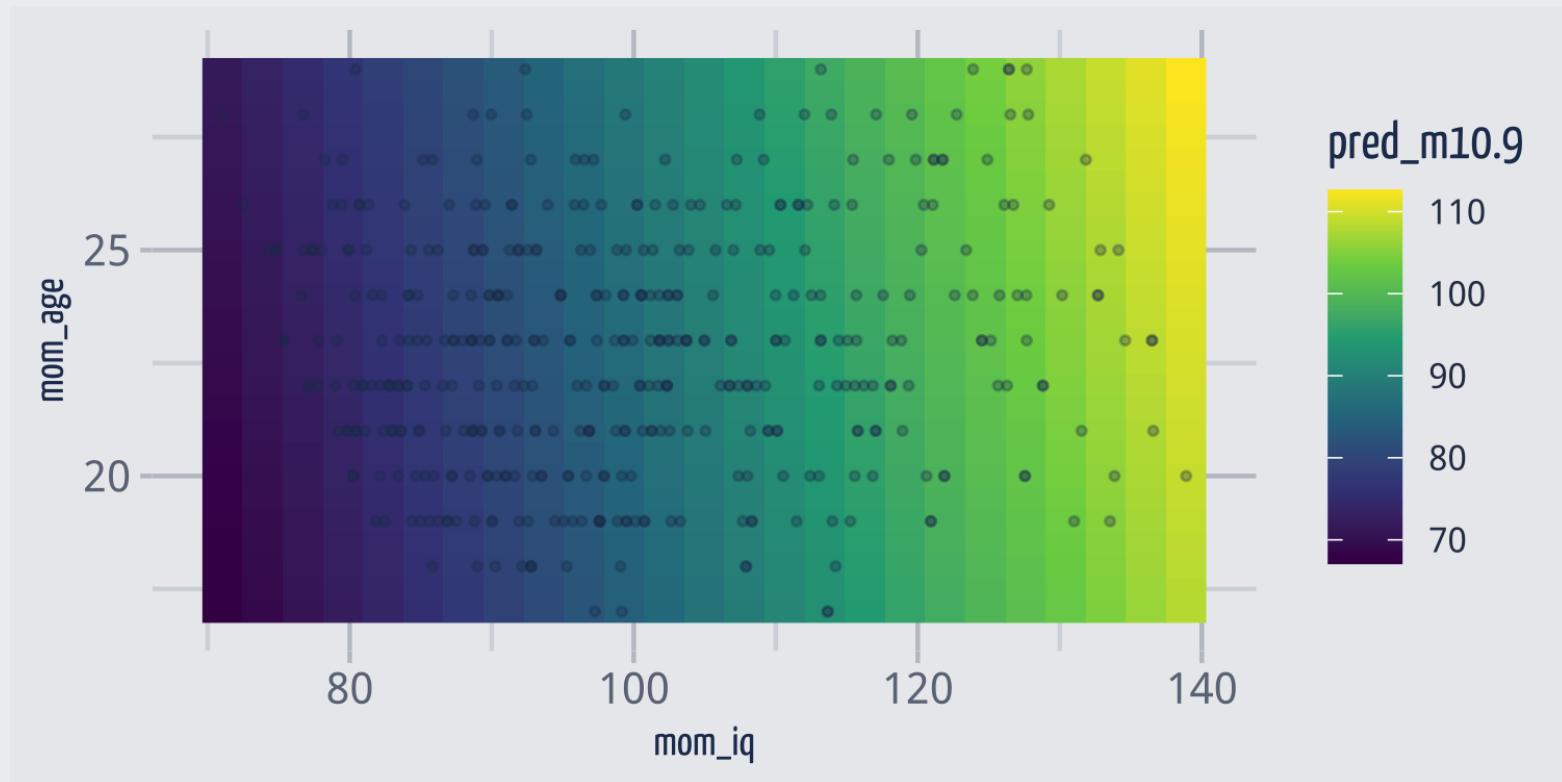
```
## (Intercept)      mom_age  
## 70.9402143    0.6997103
```

```
coef(m10.9)
```

3D-Visualisierung eines Modells mit zwei Prädiktoren 1

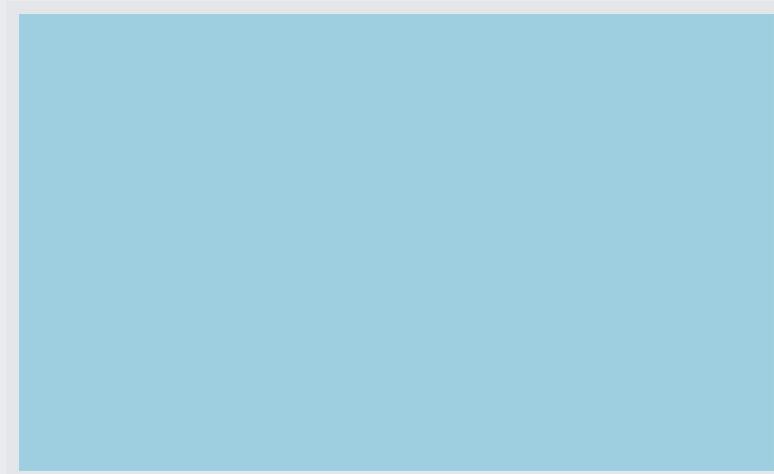


Visualisierung mit Farbe statt 3. Dimension



Auf der Achse von mom_iq erkennt man deutlich (anhand der Farbänderung) die Veränderung für die AV (kid_score). Auf der Achse für mom_age sieht man, dass sich die AV kaum ändert, wenn sich mom_age ändert.

Visualisierung in 10 Dimensionen



Leider macht mein Hirn hier nicht mit. Unsere Schwächen, eine große Zahl an Dimensionen zu visualisieren, ist der Grund, warum wir mathematische Modelle brauchen.

Daher kann man ein Modell verstehen als eine einfache Zusammenfassung eines (ggf. hochdimensionalen) Variablenraumes.

Relevanz der Prädiktoren

- Welcher Prädiktor ist nun "wichtiger" oder "stärker" in Bezug auf den Zusammenhang mit der AV, mom_iq oder mom_age?
- mom_iq hat den größeren Koeffizienten.
- mom_age hat weniger Streuung.
- Um die Relevanz der Prädiktoren vergleichen zu können, müsste man vielleicht die Veränderung von kid_score betrachten, wenn man von kleinsten zum größten Prädiktorwert geht.
- Allerdings sind Extremwerte meist instabil (da sie von einer einzigen Beobachtung bestimmt werden).
- Sinnvoller ist es daher, die Veränderung in der AV zu betrachten, wenn man den Prädiktor von "unterdurchschnittlich" auf "überdurchschnittlich" ändert.
- Das kann man mit *z-Standardisierung* erreichen.

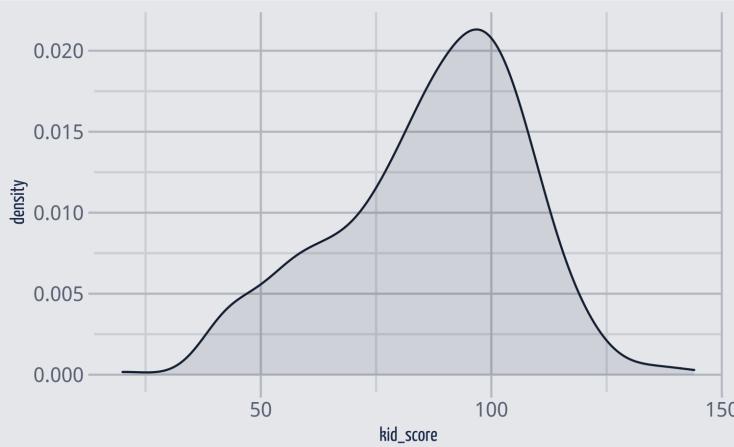
z-Standardisierung

- *z-Standardisierung* bedeutet, eine Variable so zu transformieren, dass sie über einen Mittelwert von 0 und eine SD von 1 verfügt:

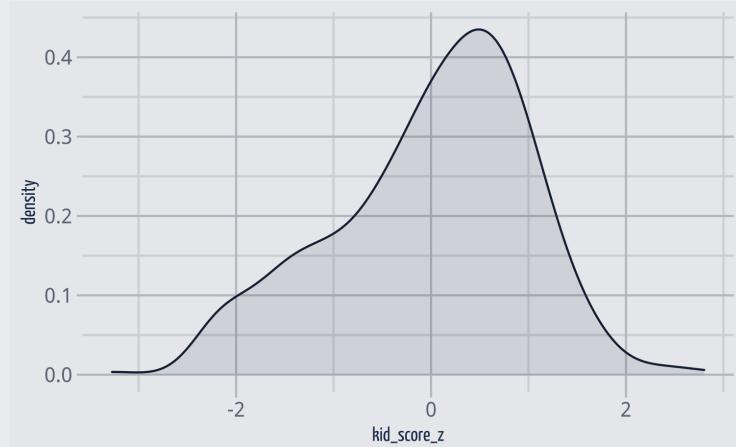
$$z = \frac{x - \bar{x}}{sd(x)}$$

```
kidiq <- kidiq %>%
  mutate(kid_score_z = ((kid_score - mean(kid_score)) / sd(kid_score)))
```

Rohwerte



Z-Werte



Statistiken zu den z-transformierten Variablen

variable	n	mean	sd
kid_score	434	86.797	20.411
kid_score_z	434	0.000	1.000
mom_age	434	22.786	2.701
mom_hs	434	0.786	0.411
mom_iq	434	100.000	15.000
pred_m10.9	434	86.576	9.205

So kann man auch die z-Transformation ("Skalierung") durchführen:

```
kidiq <- kidiq %>%
  mutate(mom_iq_z = scale(mom_iq),
        mom_age_z = scale(mom_age))
```

Modell mit standardisierten Prädiktoren, m10.10

```
m10.10 <- stan_glm(kid_score_z ~ mom_iq_z + mom_age_z, data = kidiq, rei  
coef(m10.10)
```

```
## (Intercept)      mom_iq_z      mom_age_z  
## -0.001246503  0.441916951  0.052042225
```

- Der Achsenabschnitt gibt den Mittelwert der AV (kid_score) an, da kid_score_z = 0 identisch ist zum Mittelwert von kid_score.
- Der Koeffizient für mom_iq_z gibt an, um wie viele SD-Einheiten sich kid_score (die AV) ändert, wenn sich mom_iq um eine SD-Einheit ändert.
- Der Koeffizient für mom_age_z gibt an, um wie viele SD-Einheiten sich kid_score (die AV) ändert, wenn sich mom_age um eine SD-Einheit ändert.

Jetzt sind die Prädiktoren in ihrer Relevanz (Zusammenhang mit der AV) vergleichbar:

- Man sieht, dass die Intelligenz der Mutter *deutlich wichtiger* ist als das Alter der Mutter (im Hinblick auf die Vorhersage bzw. den Zusammenhang mit der AV).

Was ist ein kleiner, was ein großer Effekt?

- Cohen (1988) definiert Effektstärken in Bezug auf Mittelwertsvergleiche anhand von $d = (\mu_1 - \mu_0)/\sigma$.
- Für kleine, mittlere und große Werte gab er folgende Richtwerte:
 - klein: $d \approx 0.2$
 - mittel: $d \approx 0.5$
 - groß: $d \approx 0.8$
- Auf dieser Basis schlägt Kruschke (2018) einen Rope von ± 0.1 vor.
- Fällt ein Intervall (mit vorab definierter Sicherheit, z.B. 95%) komplett in das Rope, so gilt der Effekt als "praktisch null".
- Richtlinien für Effektstärken sind nur Notlösungen, die durch Sachverstand ersetzt werden sollen, wo immer möglich.
- Man kann Effektstärken ineinander überführen, s. hier, z.B. von Korrelation (r) zu Cohens d oder R^2 .

Vernachlässigbarer Regressionseffekt

Kruschke (2018) schlägt vor, einen Regressionskoeffizienten unter folgenden Umständen als "praktisch Null" zu bezeichnen:

- Wenn eine Veränderung über "praktisch den ganzen Wertebereich" von x nur einen vernachlässigbaren Effekt auf y hat.
- Ein vernachlässigbarer Effekt ist dabei $\hat{y} = \pm 0.1sd_y$.
- Der "praktisch ganze Wertebereich" von x sei $\bar{x} \pm 2sd_x$.
- Resultiert der Vergleich von $\bar{x} - 2sd$ mit $\bar{x} + 2sd$ nur eine Veränderung in \hat{y} von $\bar{y} - 0.1sd_y$ auf $\bar{y} + 0.1sd_y$, so ist der Regressionskoeffizient praktisch Null, der Effekt also vernachlässigbar.
- Das impliziert Rope-Grenzen von $\beta_x = \pm 0.05$ für z-standardisierte Variablen.

Verwandtheit von Korrelation und Regression

- Sind X und Y *z-standardisiert*, so sind Korrelation und Regression identisch.

$$b = r \frac{sd_x}{sd_y}$$

```
m10.11 <-
  stan_glm(kid_score_z ~ mom_iq_z , data = kidiq, refresh = 0)
coef(m10.11)

##      (Intercept)      mom_iq_z
## -0.0008372507  0.4495398966
```

	kid_score	mom_iq	kid_score_z	mom_iq_z
kid_score	1.00	0.45	1.00	0.45
mom_iq	0.45	1.00	0.45	1.00
kid_score_z	1.00	0.45	1.00	0.45
mom_iq_z	0.45	1.00	0.45	1.00

Priori-Verteilung für m10.10

```
## Priors for model 'm10.10'  
## -----  
## Intercept (after predictors centered)  
## ~ normal(location = -2.8e-16, scale = 2.5)  
##  
## Coefficients  
##   Specified prior:  
##     ~ normal(location = [0,0], scale = [2.5,2.5])  
##   Adjusted prior:  
##     ~ normal(location = [0,0], scale = [2.50,2.50])  
##  
## Auxiliary (sigma)  
## ~ exponential(rate = 1)  
## -----  
## See help('prior_summary.stanreg') for more details
```

$$\text{kid_score} \sim \mathcal{N}(0, 2.5)$$

$$\mu_i = \alpha + \beta_1 \text{mom_iq}_i + \beta_2 \text{mom_age}_i$$

$$\alpha \sim \mathcal{N}(0, 2.5)$$

$$\beta_1 \sim \mathcal{N}(0, 2.5)$$

$$\beta_2 \sim \mathcal{N}(0, 2.5)$$

$$\sigma \sim \mathcal{E}(1)$$

Linearitätsannahme

Zentrale Annahme: Die AV ist eine *lineare* Funktion der einzelnen Prädiktoren:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots$$

- Hingegen ist es nicht erforderlich, dass die AV (y) normalverteilt ist.
- Zwar nimmt die Regression häufig normalverteilte Residuen an, aber diese Annahme ist nicht wichtig, wenn es nur darum geht, die Regressionskoeffizienten zu schätzen.

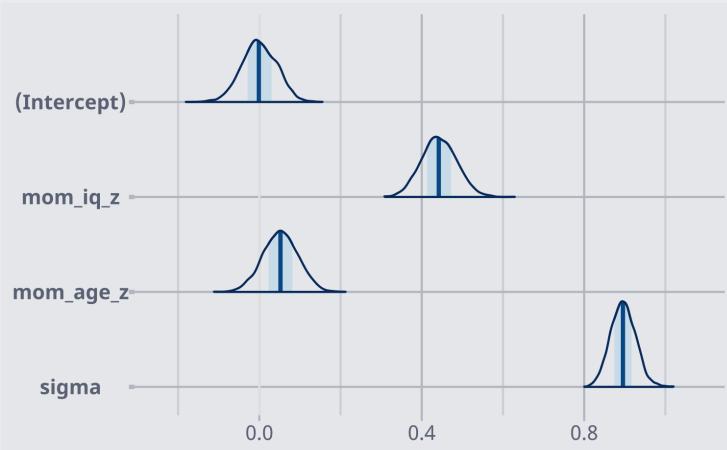
(Gelman, Hill, and Vehtari, 2021)

95%-PI

```
posterior_interval(m10.10) %>%  
  round(2)
```

```
##           5%   95%  
## (Intercept) -0.07  0.07  
## mom_iq_z     0.37  0.51  
## mom_age_z    -0.02  0.12  
## sigma        0.85  0.95
```

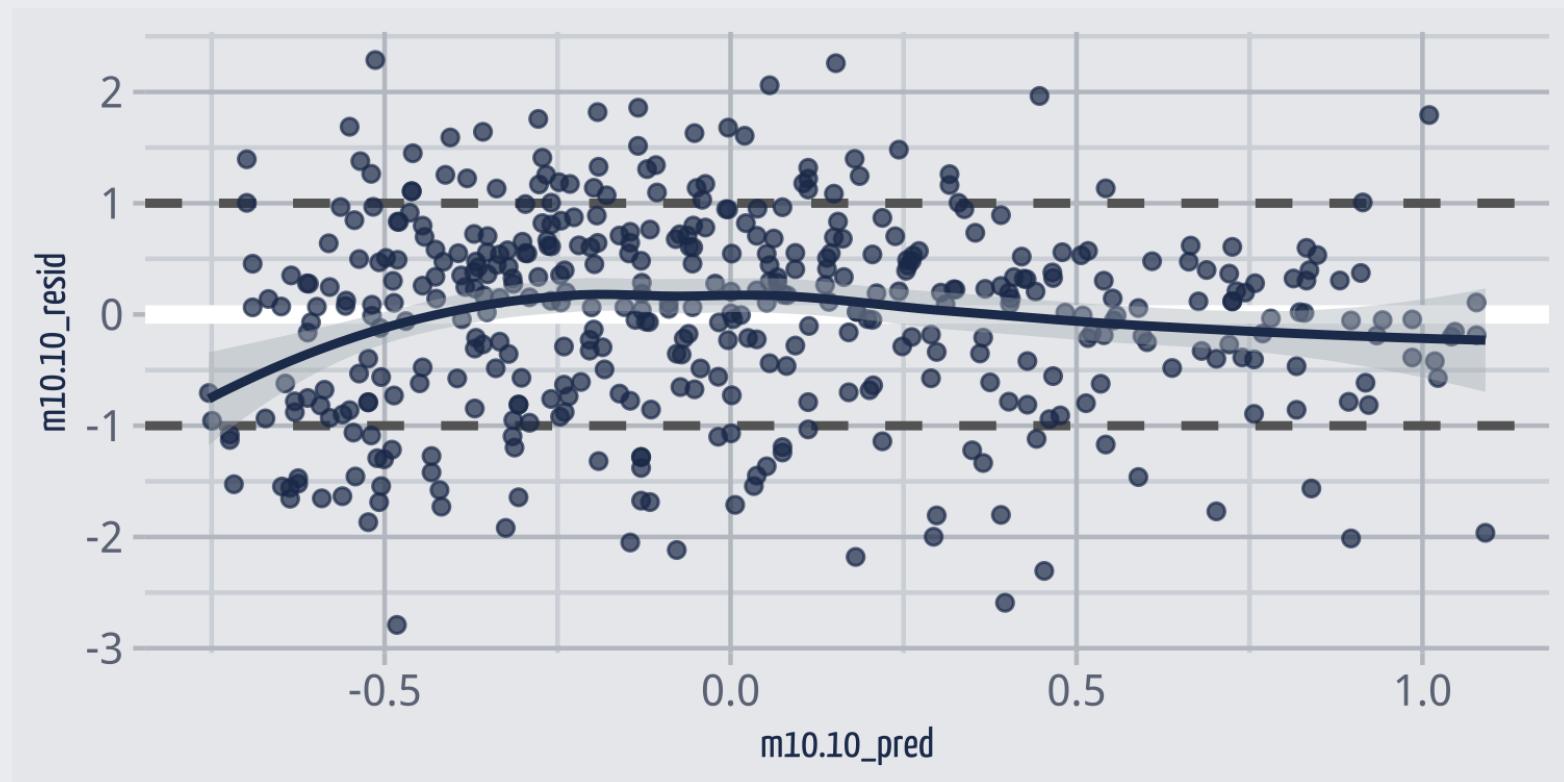
```
#plot(m10.10)  
library(bayesplot)  
mcmc_areas(m10.10)
```



Mit `hdi(m10.10)` (Paket `bayestestR`) bekommt man die HDI. Mit `plot(hdi(m10.10))` eine entsprechende Visualisierung. Allerdings ähneln sich PI und HDI im Falle von symmetrischer Posteriori-Verteilungen, so wie hier.

Prüfen der Linearitätsannahme

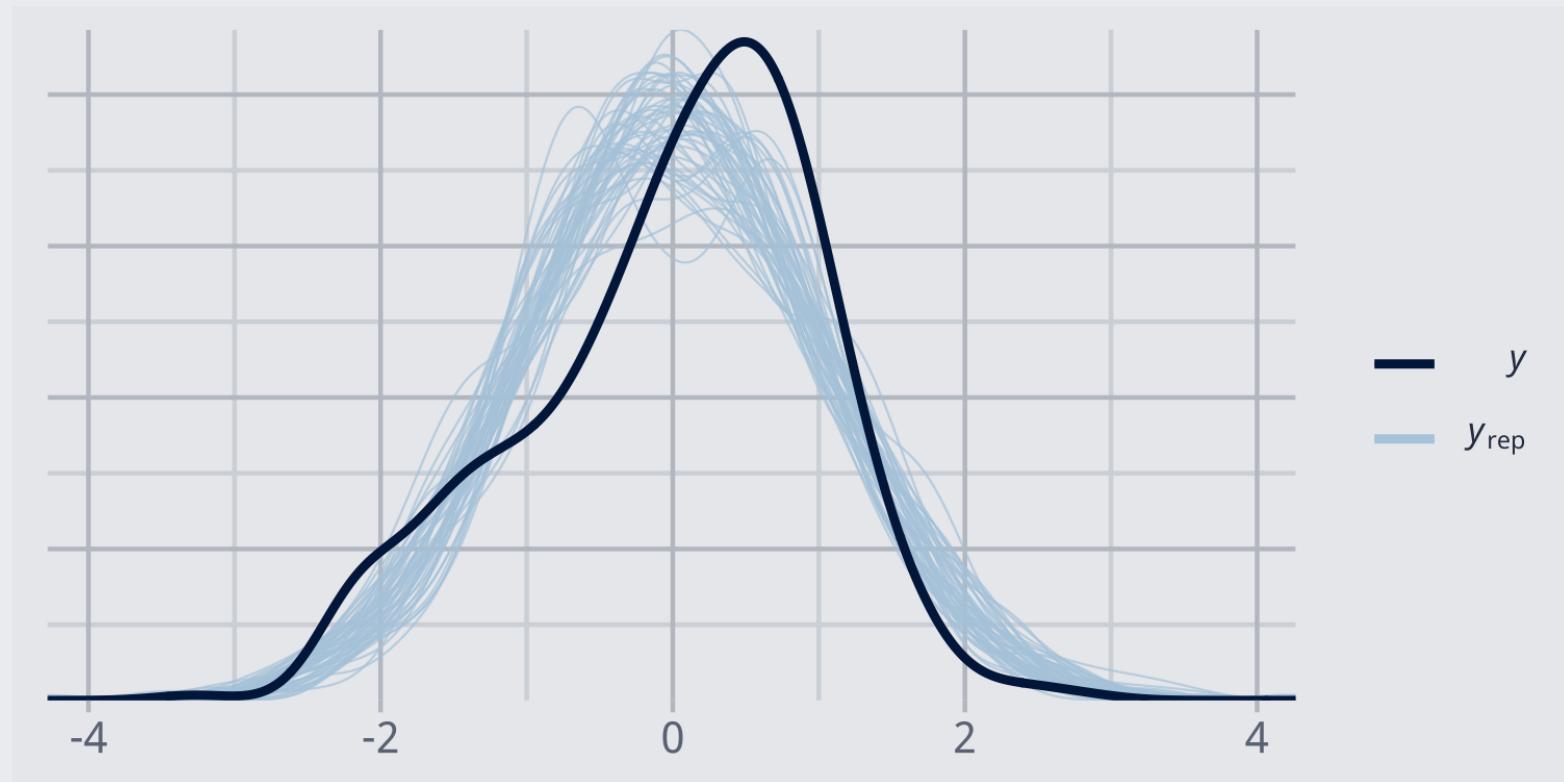
- Ist die Linearitätsannahme erfüllt, so sollte der Residualplot nur zufällige Streuung um $y = 0$ herum zeigen.



Hier erkennt man keine größeren Auffälligkeiten.

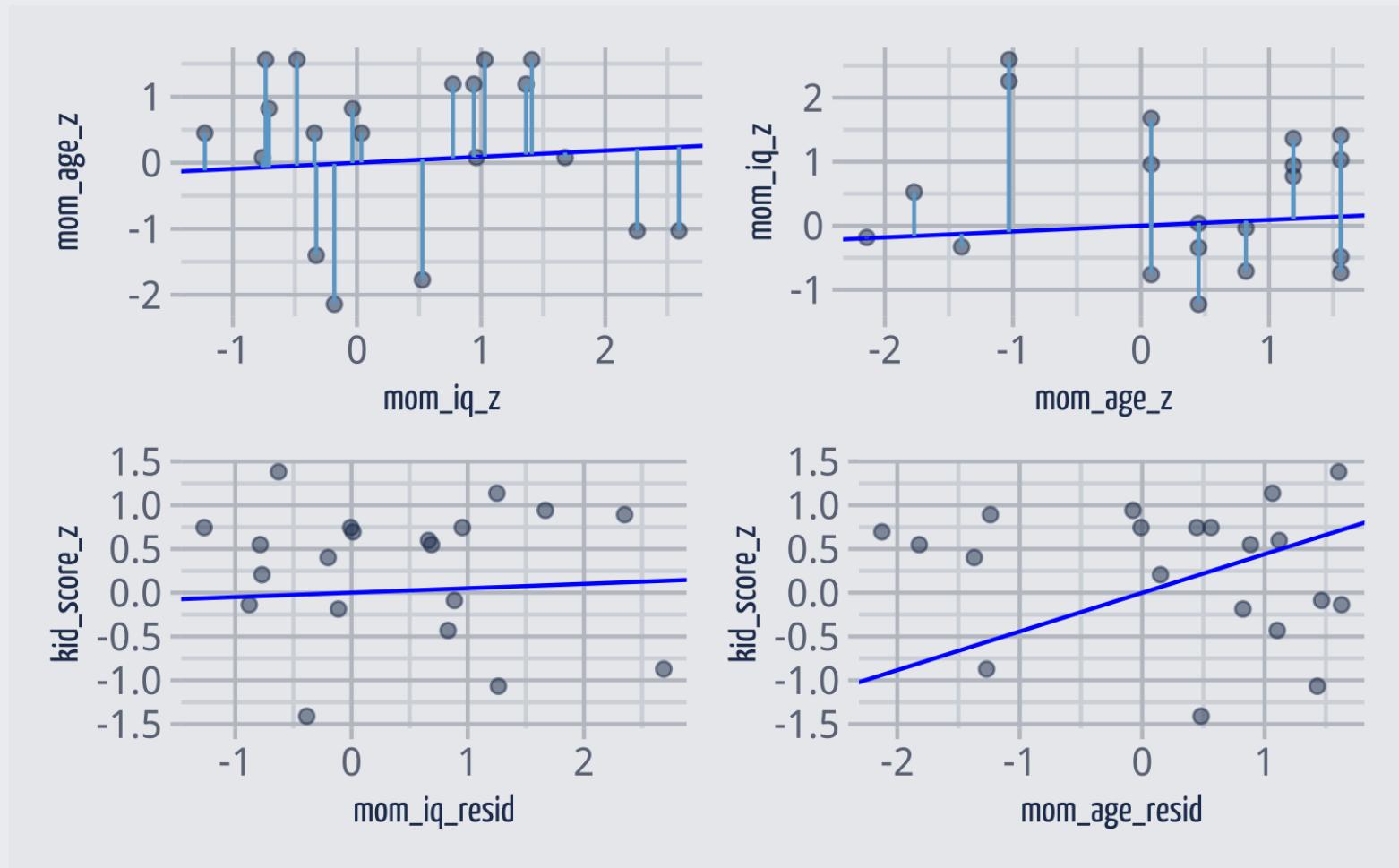
Modellprüfung mit der PPV

```
pp_check(m10.10)
```

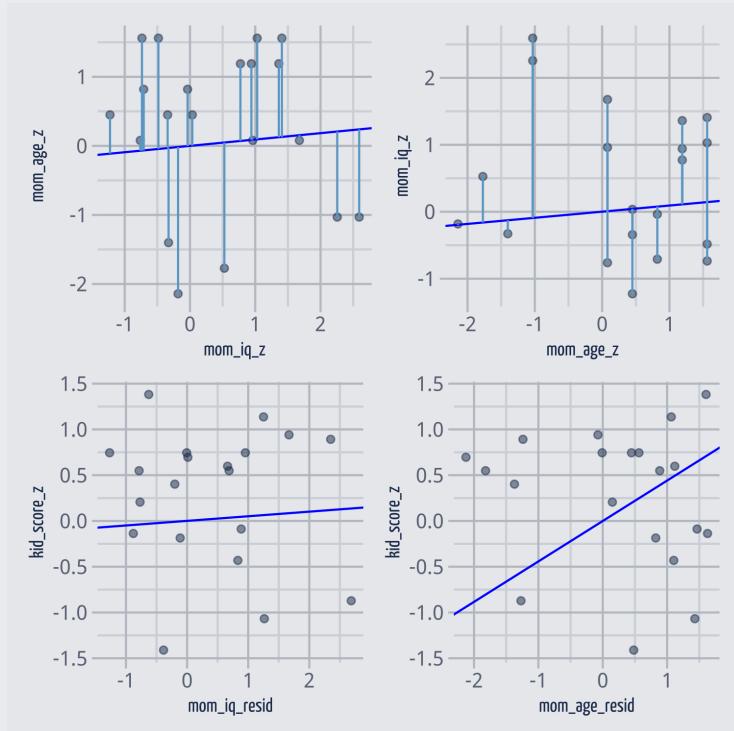


Unser Modell - bzw. die Stichproben unserer Posteriori-Verteilung, y_{rep} verfehlt den Mittelwert von y recht häufig.

Visualisierung der bereinigten Regressionskoeffizienten



Bereinigte Regressionskoeffizienten



Die vertikalen Balken zeigen die Residuen.

Zur Einfachheit nur $n = 20$.

- Obere Reihe: Regression eines Prädiktors auf den anderen anderen.
- Untere Reihe: Regression der Residuen der oberen Reihe auf die AV, `kid-score_z`.
- Unten links: Die Residuen von `mom_iq_c` sind kaum mit der AV assoziiert. Das heißt, nutzt man den Teil von `mom_age_z`, der nicht mit `mom_iq_z` zusammenhängt, um `kid_score` vorher zusagen, findet man nur einen kleinen Zusammenhang.
- Unten rechts: Die Residuen von `mom_age_c` sind stark mit der AV assoziiert. Das heißt, nutzt man den Teil von `mom_iq_z`, der nicht mit `mom_age_z` zusammenhängt, um `kid_score` vorher zusagen, findet man einen starken Zusammenhang.

Beantwortung der Forschungsfrage

Das Modell spricht sich klar für einen statistischen, linearen Effekt von Intelligenz der Mutter auf die Intelligenz des Kindes aus, wenn das Alter der Mutter statistisch kontrolliert wird (95%PI: [0.38, 0.51]). Hingegen zeigt das Modell, dass das Alter der Mutter statistisch eher keine Rolle spielt (95%PI: [-0.02, 0.12]). Alle Variablen wurden z-transformiert. Insgesamt erklärt das Modell im Median einen Anteil von 0.2 an der Varianz der Kinderintelligenz.

Hier wird von einem "statistischen Effekt" gesprochen, um klar zu machen, dass es sich lediglich um assoziative Zusammenhänge, und nicht um kausale Zusammenhänge, handelt.

Ausblick: Binäre AV

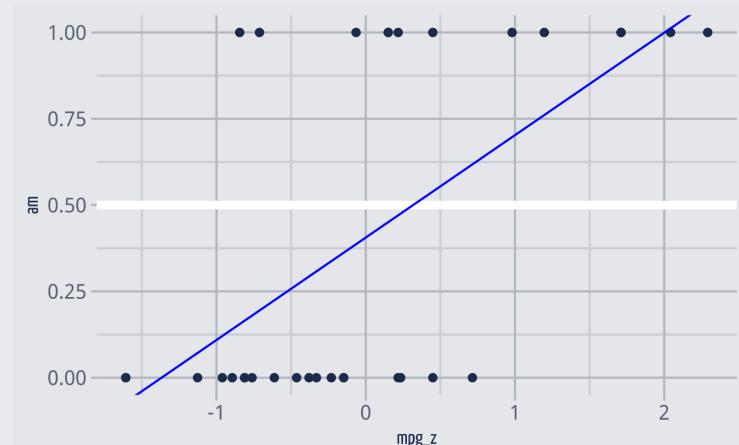
Forschungsfrage: Kann man anhand des Spritverbrauchs vorhersagen, ob ein Auto eine Automatik- bzw. ein manuelle Schaltung hat? Anders gesagt: Hängen Spritverbrauch und Getriebeart? (Datensatz mtcars)

```
data(mtcars)
mtcars <-
  mtcars %>%
  mutate(mpg_z = scale(mpg))
```

```
m11 <-
  stan_glm(am ~ mpg_z,
           data = mtcars,
           refresh = 0)
coef(m11)
```

```
## (Intercept)      mpg_z
##   0.4059633   0.2967602
```

Ab $\text{mpg_z} = 0.41$, 0.3 sagt das Modell $\text{am}=1$ (manuell) vorher. Ganz ok.



Für kleine Werte von mpg_z (<1.3) sagt unser Modell *negative* Werte für am voraus. Das macht keinen Sinn. Müssen wir mal bei Gelegenheit besser machen.

Wir waren fleißig



Quelle

Genug für heute 👍

Hinweise

Zu diesem Skript

- Dieses Skript bezieht sich auf folgende **Lehrbücher**:
 - Regression and other stories
- Dieses Skript wurde erstellt am 2021-12-01 10:12:18
- Lizenz: **MIT-Lizenz**
- Autor: Sebastian Sauer.
- Um diese HTML-Folien korrekt darzustellen, ist eine Internet-Verbindung nötig.
- Mit der Taste ? bekommt man eine Hilfe über Shortcuts.
- Wenn Sie die Endung .html in der URL mit .pdf ersetzen, bekommen Sie die PDF-Version der Datei.
- Alternativ können Sie im Browser Chrome die Folien als PDF drucken (klicken Sie auf den entsprechenden Menüpunkt).
- Den Quellcode der Skripte finden Sie **hier**.
- Eine PDF-Version kann erzeugt werden, indem man im Chrome-Browser die Webseite druckt (Drucken als PDF).



Homepage

Literatur

Diese R-Pakete wurden verwendet.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Dai, H., S. Saccardo, M. A. Han, L. Roh, N. Raja, et al. (2021). "Behavioural nudges increase COVID-19 vaccinations". In: *Nature* 597.7876, pp. 404-409. DOI: [10.1038/s41586-021-03843-2](https://doi.org/10.1038/s41586-021-03843-2).

Gelman, A., B. Goodrich, J. Gabry, and A. Vehtari (2019). "R-squared for Bayesian Regression Models". In: *The American Statistician* 73.3, pp. 307-309. DOI: [10.1080/00031305.2018.1549100](https://doi.org/10.1080/00031305.2018.1549100).

Gelman, A., J. Hill, and A. Vehtari (2021). *Regression and other stories*. Analytical methods for social research. Cambridge University Press.

Kruschke, J. K. (2018). "Rejecting or Accepting Parameter Values in Bayesian Estimation". In: *Advances in Methods and Practices in Psychological Science* 1.2, pp. 270-280. DOI: [10.1177/2515245918771304](https://doi.org/10.1177/2515245918771304).

Nasreen, S., H. Chung, S. He, K. A. Brown, J. B. Gubbay, et al. (2021). *Effectiveness of mRNA and ChAdOx1 COVID-19 vaccines against symptomatic SARS-CoV-2 infection and severe outcomes with variants of concern in Ontario*. , p. 2021.06.28.21259420. DOI: [10.1101/2021.06.28.21259420](https://doi.org/10.1101/2021.06.28.21259420).

Pormohammad, A., M. Zarei, S. Ghorbani, M. Mohammadi, M. H. Razizadeh, et al. (2021). "Efficacy and Safety of COVID-19 Vaccines: A Systematic Review and Meta-Analysis of Randomized Clinical Trials". In: *Vaccines* 9.5, p. 467. DOI: [10.3390/vaccines9050467](https://doi.org/10.3390/vaccines9050467).

Thompson, M. G., E. Stenehjem, S. Grannis, S. W. Ball, A. L. Naleway, et al. (2021). "Effectiveness of Covid-19 Vaccines in Ambulatory and Inpatient Care Settings". In: *New England Journal of Medicine* 385.15, pp. 1355-1371. DOI: [10.1056/NEJMoa2110362](https://doi.org/10.1056/NEJMoa2110362).

