Begonnen am	Donnerstag, 13. Juli 2023, 11:00
Status	Beendet
Beendet am	Donnerstag, 13. Juli 2023, 12:00
Verbrauchte Zeit	1 Stunde
Punkte	14,00/19,00
Bewertung	**73,68** von 100,00

Frage **1**

Richtig

Erreichte Punkte 1,00 von 1,00

Ein renommiertes Unternehmen sucht ei Kandidati (m/w/d) für eine (hoch dotierte) Führungsposition. Ein Managementberatungsunternehmung führt ein Assessmentcenter durch, welches pro Kandidati eine positive bzw. negative Empfehlung ergibt. Aus früheren Erfahrungen heraus wissen die Berater, dass die tatsächlich geeigneten Kandidaten (Ereignis E wie *eligible*) mit 68% eine positive Empfehlung für die Stelle ausgesprochen bekommen (Ereignis R wie *recommendation*). Weiterhin bekommen von den *nicht* geeigneten Kandidaten 69% eine negative Empfehlung. Insgesamt wissen die Berater, dass 8% der Bewerber/innen tatsächlich geeignet sind.

Geben Sie den Wert folgender Kenngröße aus der entsprechenden Kontingenztabelle an: $E \cap \overline{R}$!

Hinweise:

- $\overline{R} = R^C = \neg R$ (logische Verneinung).
- $\cap$ meint das logische "Und".
- Geben Sie Wahrscheinlichkeiten nicht als Prozentzahlen, sondern als Anteile an.
- Runden Sie auf zwei Dezimalstellen.
- Achten Sie darauf, das richtige Dezimaltrennzeichen Ihres Systems zu verwenden.

Antwort: 0,026 ✔

Einige Wahrscheinlichkeiten lassen sich direkt aus dem Text errechnen:

$$P(E \cap R) = P(R|E) \cdot P(E) = 0.68 \cdot 0.08 = 0.0544 = 5.44\%$$

$$P(\overline{E} \cap \overline{R}) = P(\overline{R}|\overline{E}) \cdot P(\overline{E}) = 0.69 \cdot 0.92 = 0.6348 = 63.48\%$$

Die restlichen gemeinsamen Wahrscheinlichkeiten lassen sich durch Addieren und Subtrahieren in der Kontingenztabelle errechnen:

	R	$\overline{R}$	**Summe**
E	**5.44**	*2.56*	**8.00**
$\overline{E}$	*28.52*	**63.48**	*92.00*
Summe	*33.96*	*66.04*	**100.00**

Lösung: Der gesuchte Wert lautet: 0.03.

Die richtige Antwort ist: 0,03

Frage **2**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `msleep` zu analysieren. Er ist unter dieser [Quelle](https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/msleep.csv) erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/msleep.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Zur Lösung dieser Aufgabe ist folgendes lineares Modell zu berechnen:

AV: awake.

UVs: sleep_rem, bodywt.

Im Folgenden wird *ein* Prädiktor aus der Menge der UV näher betrachtet. Der gewählte Prädiktor ist die erste oben genannte UV: `sleep_rem` (uv1).

Aufgabe: **Geben Sie den Anteil der Stichproben der Posteriori-Verteilung an (für den gewählten Prädiktor), die außerhalb des ROPE liegen!**

Beachten Sie die Bearbeitungshinweise.

Antwort: 1 ✔

MUSTERLÖSUNG

Pakete starten:

Daten importieren:

```
d <- read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/msleep.csv')
```

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1 + uv2,
                data = d,
                refresh = 0,
                # chains = config$chains,  # ignorieren Sie diese Zeile
                seed = 42)
```

So kann man sich den gesuchten Wert ausgeben lassen:

```
rope(mod)
```

```
## # Proportion of samples inside the ROPE [-0.45, 0.45]:
##
## Parameter   | inside ROPE
## -------------------------
## (Intercept) |      0.00 %
## uv1         |      0.00 %
## uv2         |    100.00 %
```

Für die Lösung ist aber der *Anteil* gefragt, der *außerhalb* des ROPE liegt. Daher ist noch $1-x$ zu rechnen.

Lösung: `1.00`.

Aufgaben-ID: predict-dyn2 , Toleranzbreite: 0.100

Die richtige Antwort ist: 1

Frage **3**

Falsch

Erreichte Punkte 0,00 von 1,00

Im Folgenden ist der Datensatz `education` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/robustbase/education.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Zur Lösung dieser Aufgabe ist folgendes lineares Modell zu berechnen:

AV: Y.

UVs: X3, X1.

Der gewählte Prädiktor ist die erste oben genannte UV: `X3` (uv1).

Aufgabe: **Wie hoch ist die Wahrscheinlichkeit, dass der Effekt der gewählten UV größer ist als 0.7-SD-Einheiten der AV?**

Beachten Sie die Bearbeitungshinweise.

Antwort: 1 ✖

MUSTERLÖSUNG

Der Datensatz heißt:

```
## [1] "education"
```

Quelle: https://vincentarelbundock.github.io/Rdatasets/csv/robustbase/education.csv.

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1 + uv2,
                data = d,
                refresh = 0,
                chains = 4,  # ignorieren Sie diese Zeile
                seed = 42)
```

Dann berechnen wir die SD der AV:

```
sd_av <- sd(d$av)
sd_av
```

```
## [1] 61.3401
```

Es ist komfortabel, sich den Grenzwert für den gesuchten Effekt (hier `effect_size` genannt) vor Augen zu führen:

```
grenzwert_effekt <- effect_size * sd_av
grenzwert_effekt
```

```
## [1] 42.9381
```

Als nächstes zählen wir die Anzahl der Stichproben der Post-Verteilung, die die gewünschte Bedingung erfüllen (die ein Treffer sind, hier mit `is_hit` bezeichnet):

```
mod |>
  as_tibble() |>
  count(is_hit = uv1 > grenzwert_effekt) |>
  mutate(prop = n/sum(n))
```

```
## # A tibble: 2 × 3
##   is_hit     n  prop
##   <fct>  <int> <dbl>
## 1 TRUE       0     0
## 2 FALSE   4000     1
```

Zur Erinnerung: Wir suchen den Anteil der *Treffer*, also `is_hit == TRUE`.

Lösung: Der laut Modell vorhergesagte Werte beträgt 0.00.

Aufgaben-ID: post-effekt-1sd, Toleranzbreite: 0.050

Die richtige Antwort ist: 0

Frage **4**

Falsch

Erreichte Punkte 0,00 von 1,00

Die kausale Abhängigkeit zwischen drei Variablen A, B, C sieht wie folgt aus:

- B hängt ab von A.
- C hängt ab von A.
- C hängt ab von B.

Wie lautet die R Formel, um den totalen kausalen Effekt von A auf C zu bestimmen?

Wählen Sie eine Antwort:

- a. `A ~ C`
- b. `C ~ A`
- c. `A ~ C + B` ✖ Falsch
- d. Der totale kausale Effekt kann nicht bestimmt werden.

a. Richtig
b. Falsch
c. Falsch
d. Falsch

Die richtige Antwort ist: `C ~ A`

Frage **5**

Falsch

Erreichte Punkte 0,00 von 1,00

Im Folgenden ist der Datensatz `CASchools` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/AER/CASchools.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Zur Lösung dieser Aufgabe ist folgendes lineares Modell zu berechnen:

AV: math.

UVs: lunch, district.

Aufgabe: **Was ist der Mittelwert des adjustierten Priors für den Intercept des Modells?**

Beachten Sie die Bearbeitungshinweise.

Antwort: 46,89 ✖

MUSTERLÖSUNG

Datenquelle: https://vincentarelbundock.github.io/Rdatasets/csv/AER/CASchools.csv.

Pakete starten:

```
library(tidyverse)
library(easystats)
library(rstanarm)
```

Daten importieren:

```
d <- read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/AER/CASchools.csv')
```

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1 + uv2,
                data = d,
                refresh = 0,
                chains = config$chains,  # ignorieren Sie diese Zeile
                seed = 42
                )
```

Dann lässt man sich die Prior-Informationen ausgeben:

```
prior_summary(mod)
```

```
## Priors for model 'mod'
## ------
## Intercept (after predictors centered)
##   Specified prior:
##     ~ normal(location = 653, scale = 2.5)
##   Adjusted prior:
##     ~ normal(location = 653, scale = 47)
##
## Coefficients
##   Specified prior:
##     ~ normal(location = [0,0], scale = [2.5,2.5])
##   Adjusted prior:
##     ~ normal(location = [0,0], scale = [1.729,0.014])
##
## Auxiliary (sigma)
##   Specified prior:
##     ~ exponential(rate = 1)
##   Adjusted prior:
##     ~ exponential(rate = 0.053)
## ------
## See help('prior_summary.stanreg') for more details
```

Und kann dort den gesuchten Wert ablesen.

Lösung: Der Wert beträgt:

```
## [1] "653.34"
```

Aufgaben-ID: prior1, Toleranzbreite: 65.330

Die richtige Antwort ist: 653,34

Frage **6**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `mtcars` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Zur Lösung dieser Aufgabe ist folgendes lineares Modell zu berechnen:

AV: mpg.

UVs: vs, hp.

Aufgabe: **Geben Sie die erklärte Varianz (R-Quadrat) des Modells an (den Punktschätzer)!**

Beachten Sie die Bearbeitungshinweise.

Antwort: 0,61 ✔

MUSTERLÖSUNG

Der Datensatz heißt:

```
## [1] "mtcars"
```

Quelle: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv.

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1 + uv2,
                data = d,
                refresh = 0,
                chains = 4,  # ignorieren Sie diese Zeile
                seed = 42)
```

So kann man sich den gesuchten Wert ausgeben lassen:

```
r2_bayes(mod)
```

```
## # Bayesian R2 with Compatibility Interval
##
##   Conditional R2: 0.603 (95% CI [0.414, 0.769])
```

Und dann liest man einfach den gesuchten Wert von der Ausgabe ab.

Lösung: `0.60`.

Aufgaben-ID: bayes-r2-dyn, Toleranzbreite: 0.075

Die richtige Antwort ist: 0,6

Frage **7**

Falsch

Erreichte Punkte 0,00 von 1,00

Im Folgenden ist der Datensatz `mtcars` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Aufgabe: **Berechnen Sie folgende Wahrscheinlichkeit: $Pr(Y \cap X)$!**

X und Y sind Ereignisse, die wie folgt definiert sind:

1. $Y := \text{av} > Md(av)$
2. $X := \text{uv} > Md(\text{uv})$

Dabei ist mit `Md` der Median gemeint.

`av` ist mpg und `uv` ist wt.

Beachten Sie die Bearbeitungshinweise.

Antwort: 0,24

MUSTERLÖSUNG

Datenquelle: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv.

Pakete starten:

```
library(tidyverse)
library(easystats)
#library(rstanarm)
```

Daten importieren:

```
d <- read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv')
```

X und Y berechnen:

```
d2 <-
  d %>%
  select(av, uv1) %>%
  mutate(Y = av > median(av, na.rm = TRUE),
         X = uv1 > median(uv1, na.rm = TRUE))
```

Bedingte Wahrscheinlichkeit, $Pr(Y|X)$ berechnen:

```
pr_y_geg_x_tab <-
  d2 %>%
  filter(X = TRUE) %>%
  count(Y) %>%
  mutate(Pr_Y_geg_X = n / sum(n))

pr_y_geg_x_tab
```

```
##       Y  n Pr_Y_geg_X
## 1 FALSE 17    0.53125
## 2  TRUE 15    0.46875
```

Die Lösung lässt sich ablesen bei `Y == TRUE` in der Spalte `Pr_Y_geg_X`.

Gemeinsame Wahrscheinlichkeit, $Pr(Y \cap X)$, berechnen:

```
pr_y_und_x_tab <-
  d2 %>%
  mutate(Y_and_X = X == TRUE & Y == TRUE) %>%
  mutate(Y_and_X = factor(Y_and_X, levels = c("TRUE", "FALSE"))) %>%
  count(Y_and_X, .drop = FALSE) %>%
  mutate(Y_and_X_prob = n / sum(n))

pr_y_und_x_tab
```

```
##   Y_and_X  n Y_and_X_prob
## 1    TRUE  0            0
## 2   FALSE 32            1
```

Lösung:

```
## [1] "0.00"
```

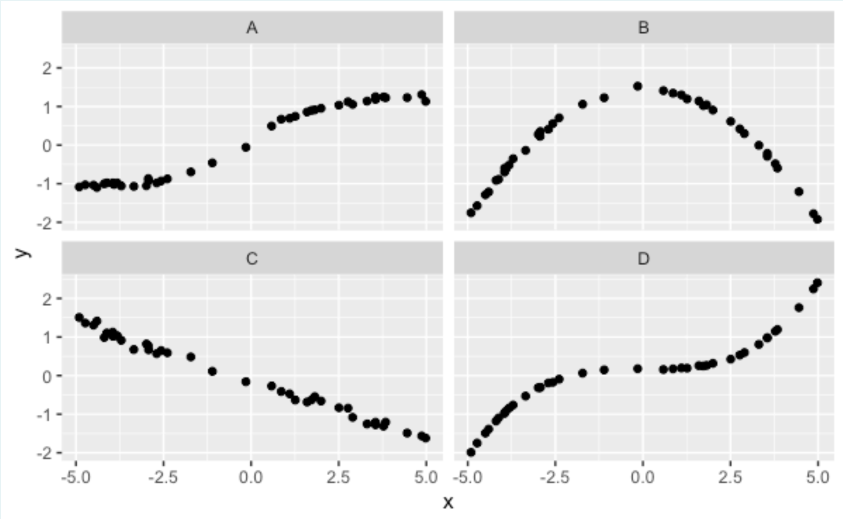
Aufgaben-ID: Bed-gem-Wskt, Toleranzbreite: 0.025

Die richtige Antwort ist: 0

Frage **8**

Richtig

Erreichte Punkte 1,00 von 1,00



Bei welcher der Abbildungen ist eine lineare Regression (am ehesten) angemessen?

Wählen Sie eine Antwort:

- a. A
- b. D
- c. C ✔ Richtig
- d. B

a. Falsch
b. Falsch
c. Richtig
d. Falsch

Die richtige Antwort ist: C

Frage **9**

Richtig

Erreichte Punkte 1,00 von 1,00

Gegeben sei der DAG `g` (s. u.). Der DAG verfügt über $n = 6$ Variablen, die als Knoten im Graph dargestellt sind und mit $x_1, x_2, \ldots x_n$ bezeichnet sind.

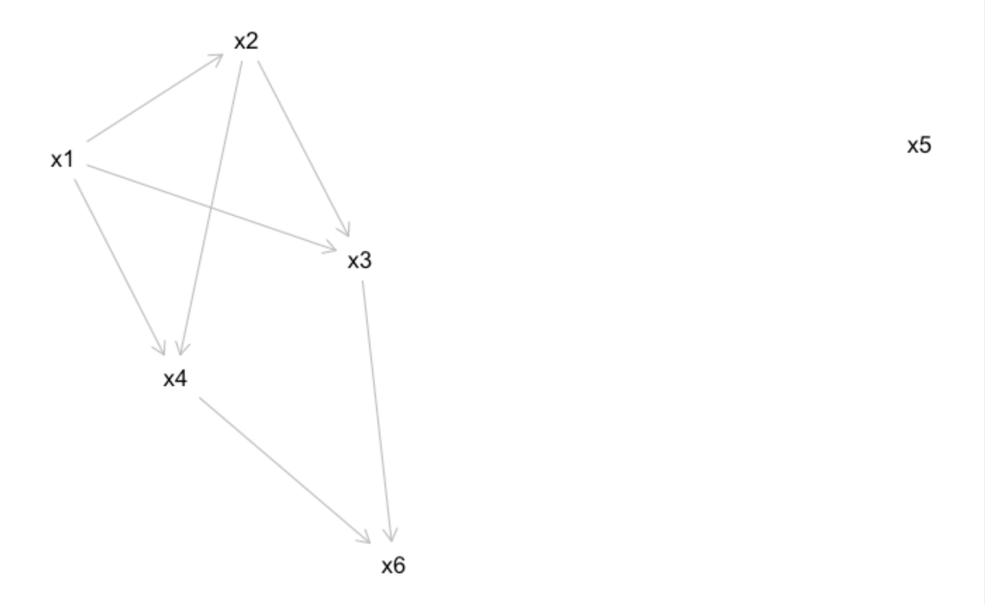
Welche minimale Variablenmenge muss kontrolliert werden, um den totalen kausalen Effekt von der UV zur AV zu identifizieren?

UV: x3.

AV: x5.

Hinweise:

- Mengen sind mittels geschweifter Klammern gekennzeichnet, z.B. `{x1, x2}` meint die Menge mit den zwei Elementen `x1` und `x2`.
- Die leere Menge `{ }` bedeutet, dass keine Variable kontrolliert werden muss, um den Effekt zu identifizieren.
- Alle Variablen werden als gemessen vorausgesetzt.
- Es ist möglich, dass es keine Lösung gibt, dass es also keine Adjustierungsmenge gibt, um den totalen kausalen Effekt zu identifizieren. Wenn dies der Fall sein sollte, wählen Sie "/".
- Es ist möglich, dass einzelne Variablen keine Kanten besitzen, also keine Verbindung zu anderen Variablen (Knoten) haben. Diese Variablen sind dann kausal unabhängig von den übrigen Variablen.



Wählen Sie eine Antwort:

- a. { x2 }
- b. { x4, x6 }
- c. { } ✔ Richtig
- d. /
- e. { x2, x5 }

Folgende minimale Variablenmenge muss kontrolliert werden, um den kausalen Effekt der UV auf die AV zu identifizieren: `{ }`.

Es wird also nach dem *minimal adjustment set* gefragt. Zu beachten ist, dass es weitere, aber größere Mengen geben kann, die den kausalen Effekt identifizieren, aber gefragt ist nach der minimalen Menge. Außerdem ist es möglich, dass es weitere minimale Adjustierungsmengen gibt, die aber nicht in der Auswahlliste vorkommen.

Erläuterung:

UV und AV sind nicht (durch mindestens einen offenen Pfad) verbunden ("d-connected").

Folgende Unabhängigkeitsbeziehungen (`_||_`) werden durch den DAG definiert:

```
## x1 _||_ x5
## x1 _||_ x6 | x3, x4
## x2 _||_ x5
## x2 _||_ x6 | x3, x4
## x3 _||_ x4 | x1, x2
## x3 _||_ x5
## x4 _||_ x5
## x5 _||_ x6
```

UV und AV werden durch folgende Pfade verbunden:

```
## $paths
## list()
##
## $open
## list()
```

Offenen Pfade sind der/die Pfad/e, die in der obigen Ausgabe unter `$open` mit `TRUE` gekennzeichnet sind.

Der DAG ist wie folgt definiert:

```
dag {
x1
x2
x3
x4
x5
x6
x1 -> x2
x1 -> x3
x1 -> x4
x2 -> x3
x2 -> x4
x3 -> x6
x4 -> x6
}
```

Aufgaben-ID: randomdag1

---

a. Falsch
b. Richtig
c. Falsch
d. Falsch
e. Falsch

Die richtige Antwort ist: { }

Frage **10**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `Affairs` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/AER/Affairs.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Aufgabe: **Berechnen Sie die Schiefe von education!**

Beachten Sie die Bearbeitungshinweise.

Antwort: -0,25

MUSTERLÖSUNG

Datenquelle: https://vincentarelbundock.github.io/Rdatasets/csv/AER/Affairs.csv.

Pakete starten:

```
library(tidyverse)
library(easystats)
#library(rstanarm)
```

Daten importieren:

`d <- read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/AER/Affairs.csv')`

Die Schiefe lässt sich z.B. so berechnen:

```
d %>%
  describe_distribution()
```

```
## Variable |  Mean |   SD |  IQR |         Range | Skewness | Kurtosis |   n | n_Missing
## ----------------------------------------------------------------------------------------
## av       |  1.46 | 3.30 | 0.50 | [0.00, 12.00] |     2.35 |     4.26 | 601 |         0
## uv1      | 16.17 | 2.40 | 4.00 | [9.00, 20.00] |    -0.25 |    -0.30 | 601 |         0
## uv2      |  3.93 | 1.10 | 2.00 |  [1.00, 5.00] |    -0.84 |    -0.20 | 601 |         0
```

Dann liest man den gesuchten Wert aus der Tabelle in der Spalte `Skewness` ab.

Lösung:

```
## [1] "-0.25"
```

Aufgaben-ID: skew-dny1, Toleranzbreite: -0.025027

Die richtige Antwort ist: -0,25

Frage **11**

Richtig

Erreichte Punkte 1,00 von 1,00

Eine Studentin muss oder will (?) eine Statistik-Klausur schreiben. Die Klausur besteht ausschließlich aus 18 Richtig-Falsch-Aufgaben, Aufgaben also, die mit entweder *Ja* oder *Nein* zu beantworten sind (per Ankreuzen). Nach (mehr oder weniger) reiflicher Überlegung entschließt sie sich, die Klausur nur durch Münzwurf zu beantworten.

Die Münze, die die Studentin benutzt, hat eine Wahrscheinlichkeit für einen "Treffer" (richtige Antwort angekreuzt) von $p = 0.5$.

Wie groß ist die Wahrscheinlichkeit für *genau* $k = 9$ Treffer in der Klausur?

Beachten Sie die Bearbeitungshinweise.

Antwort: 0,19

```
sol <- dbinom(x = k_treffer,  # Anzahl Treffer
              size = anz_aufgaben,  # Anzahl Aufgaben in der Klausur
              prob = p_treffer)  # Wahrscheinlichkeit für einen Treffer
```

Antwort: Der gesuchte Werte beträgt: 0.19.

Aufgaben-ID: Klausur-raten, Toleranzbreite: 0.025

Die richtige Antwort ist: 0,19

Frage **12**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `gtcars` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/gt/gtcars.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Aufgabe: **Berechnen Sie die Korrelation von msrp und mpg_h!**

Beachten Sie die Bearbeitungshinweise.

Antwort: -0,54

MUSTERLÖSUNG

Datenquelle: https://vincentarelbundock.github.io/Rdatasets/csv/gt/gtcars.csv.

Pakete starten:

```
library(tidyverse)
library(easystats)
#library(rstanarm)
```

Daten importieren:

`d <- read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/gt/gtcars.csv')`

Die Korrelation lässt sich z.B. so berechnen:

```
d %>%
  correlation()
```

```
## # Correlation Matrix (pearson-method)
##
## Parameter1 | Parameter2 |     r |         95% CI | t(44) |         p
## --------------------------------------------------------------------
## av         |        uv1 | -0.54 | [-0.72, -0.30] | -4.29 | < .001***
## av         |        uv2 |  0.45 | [ 0.19,  0.66] |  3.39 | 0.001**
## uv1        |        uv2 | -0.63 | [-0.78, -0.42] | -5.40 | < .001***
##
## p-value adjustment method: Holm (1979)
## Observations: 46
```

Dann liest man den gesuchten Wert aus der Tabelle ab.

Lösung:

```
## [1] "-0.54"
```

Aufgaben-ID: corr-dyn1, Toleranzbreite: 0.025

Die richtige Antwort ist: -0,54

Frage **13**

Richtig

Erreichte Punkte 1,00 von 1,00

Sie werfen eine Münze $k = 4$ Mal; die Trefferchance betrage $p = 0.8$. Die Münzwürfe seien unabhängig voneinander. Dabei werden $h = 0, 1, \ldots, k$ Treffer (*h*its) erzielt.

Aufgabe: Wie hoch ist die Wahrscheinlichkeit für (genau) $h \neq k$ Treffer?

Beachten Sie die Bearbeitungshinweise.

Antwort: 0,59

Trefferchance bei jedem Wurf:

```
## [1] 0.8
```

Anzahl Würfe/Münzen:

```
## [1] 4
```

Aufgrund der Multiplikationsregel der Wahrscheinlichkeitsrechnung sind die Wahrscheinlichkeiten der h Ereignisse (Treffer) zu multiplizieren, da unabhängig. Da es nur eine Möglichkeit gibt, bei k Würfen k Treffer zu erzielen, gibt es nur einen "Pfad" für k Treffer.

Das gesuchte Ereignis ist das logische Gegenteil, das Komplementärereignis also.

```
sol <- 1 - p^k
sol
```

```
## [1] 0.5904
```

Aufgaben-ID: k-coins-nicht-k-hits, ex_tol: 0.025

Die richtige Antwort ist: 0,59

Frage **14**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `iris` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/iris.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Zur Lösung dieser Aufgabe ist folgendes lineares Modell zu berechnen:

AV: Sepal.Length.

UVs: Petal.Width (uv1).

Im Modell sind alle Variablen sind z-standardisiert, also z-transformieren wir:

```
library(easystats)
d_z <- standardize(d)
```

```
mod <- stan_glm(av ~ uv1,
                data = d_z,
                refresh = 0,
                chains = 4,  # ignorieren Sie diese Zeile
                seed = 42)
```

Hier sind die Parameter des Modells:

```
parameters(mod)
```

```
## Parameter   |   Median |        95% CI |     pd |  Rhat |     ESS |                    Prior
## -------------------------------------------------------------------------------------------
## (Intercept) | 8.18e-04 | [-0.10, 0.09] | 50.60% | 1.000 | 3730.00 | Normal (-4.48e-16 +- 2.50)
## uv1         |     0.82 | [ 0.72, 0.91] |   100% | 1.000 | 3602.00 |      Normal (0.00 +- 2.50)
```

Betrachten Sie dazu folgende Beziehung (Gleichung bzw. Ungleichung):

$$Pr(\text{Sepal.Length}_i = 0 | \text{Petal.Width}_i = 0, \alpha, \beta, \sigma) \quad \square \quad Pr(\text{Sepal.Length}_i = 0 | \text{Petal.Width}_i = 1, \alpha, \beta, \sigma)$$

Die in der obigen Beziehung angegebenen Parameter beziehen sich auf dargestellte Regressionsmodell, `mod`.

Aufgabe: Geben Sie als Antwort *genau eines* der folgenden Zeichen ein:

- `<`
- `>`
- `=`
- `?`

Hinweise:

- Das Fragezeichen `?` als Antwortoption bedeutet, dass keine Antwort möglich ist (auf Basis der Angabe).
- Verwenden Sie Standardwerte der R-Funktionen, soweit nicht anders angegeben.
- Verwenden Sie Methoden der Bayes-Statistik für inferenzstatistische Analysen.
- Geben Sie keine Prozentzahlen an, sondern Anteile (also nicht "50%", sondern "0.5" etc.)
- Findet sich in einer Auswahlliste möglicher Antworten nicht die exakte Lösung, wählen Sie die am besten passende.
- Beziehen Sie sich im Zweifel auf den Stoff, so wie im Unterricht behandelt.
- Z-Skalieren Sie alle Variablen (d.h. Prädiktoren und AV).
- Geben Sie *nur ein Zeichen, wie oben angegeben an*, andere Eingaben werden als *falsch* gewertet.
- Gehen Sie davon aus, dass der Likelihood normalverteilt ist.

Antwort: >

Das Modell kann so berechnet werden:

```
mod <- stan_glm(av ~ uv1,
                data = d_z,
                refresh = 0,
                chains = 4,  # ignorieren Sie diese Zeile
                seed = 42)
```

Die Aufgabe fragt, wie wahrscheinlich es ist, dass das Modell $y = 0$ (also den Mittelwert der AV) vorhersagt, wenn $x = 0$ bzw. wenn $x = 1$ ist.

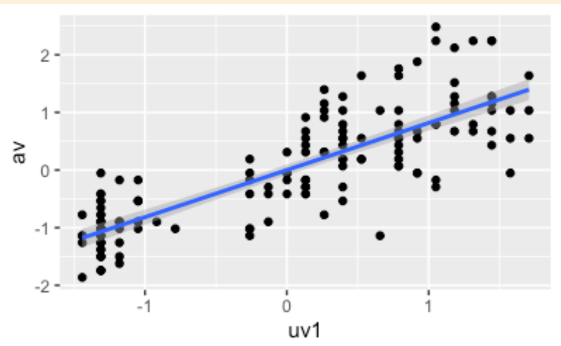
Daher lassen wir uns die Vorhersagen zunächst ausgeben.

Je nachdem, wie Ihr Prädiktor heißt, könnten Sie sich die Vorhersagen z.B. so ausgeben lassen: `predict(mod, newdata = tibble(Petal.Width = c(0,1))`.

```
##           1           2
## 0.000218648 0.818420556
```

Eine Visualisierung ist hilfreich:

```
ggplot(d_z, aes(x = uv1, y = av)) +
  geom_point() +
  geom_smooth(method = "lm")
```



Die Regressionsgerade zeigt (mehr oder weniger) deutlich, bei welchem Wert, $x = 0$ oder $x = 1$ die vorhergesagten Werte näher an Null liegen. Wenn der Punktschätzer einer Vorhersage näher an Null ist, so ist es wahrscheinlicher, dass der Wert Null ist (im Verhältnis zu einer Vorhersage, deren Punktschätzer weiter weg von Null liegt).

Also: Je näher der Wert von y an Null ist, desto größer seine Wahrscheinlichkeit, Null zu sein.

Man betrachtet also die beiden Vorhersagen (x=0 vs. x=1) und sucht die Zahl, die näher an Null ist. Für diesen Wert ist die Wahrscheinlichkeit höher, dass $y = 0$.

In diesem Fall gilt also:

$$\begin{aligned} Pr(\text{Sepal.Length}_i = 0 | \text{Petal.Width}_i = 0, \alpha, \beta, \sigma) &> Pr(\text{Sepal.Length}_i = 0 | \text{Petal.Width}_i = 1, \alpha, \beta, \sigma) \\ Pr(\text{Sepal.Length}_i = 0 | \text{Sepal.Width}_i = 0, \alpha, \beta, \sigma) &> Pr(\text{Sepal.Length}_i = 0 | \text{Sepal.Width}_i = 1, \alpha, \beta, \sigma) \end{aligned}$$

Die richtige Lösung lautet also: **>**.

Aufgaben-ID: Bed-Post-Wskt2

Die richtige Antwort ist: >

Frage **15**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `msleep` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/msleep.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Berechnen Sie das folgende lineare Modell:

AV: awake.

UVs: sleep_rem.

Aufgabe: **Was ist der Anteil der Posterior-Verteilung des Regressionskoeffizienten der Variable `sleep_rem`, der außerhalb des ROPE liegt? Gehen Sie von z-standardisierten Variablen aus.**

Hinweise:

Beachten Sie die Bearbeitungshinweise.

Antwort: 1

MUSTERLÖSUNG

Der Datensatz heißt:

```
d_name
```

```
## [1] "msleep"
```

Quelle: https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/msleep.csv.

So sehen die ersten Zeilen des Datensatzes aus:

```
head(d)  # Rohdaten
```

```
##     av uv1     uv2
## 1 11.9  NA      NA
## 2  7.0 1.8 0.01550
## 3  9.6 2.4      NA
## 4  9.1 2.3 0.00029
## 5 20.0 0.7 0.42300
## 6  9.6 2.2      NA
```

Prädiktor(en) und AV z-transformieren:

```
d <- standardize(d)
head(d)  # z-standardisiert
```

```
##          av        uv1       uv2
## 1 -0.374537         NA        NA
## 2 -1.475145 -0.0580841 -0.272509
## 3 -0.891149  0.4040630        NA
## 4 -1.003456  0.3270385 -0.288086
## 5  1.444835 -0.9053536  0.144835
## 6 -0.891149  0.2500140        NA
```

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1,
                data = d,
                refresh = 0,
                chains = 4,  # ignorieren Sie diese Zeile
                seed = 42)
```

Gefragt war nach diesem Prädiktor `sleep_rem`:

So kann man sich das ROPE berechnen lassen:

```
library(easystats)
rope(mod)
```

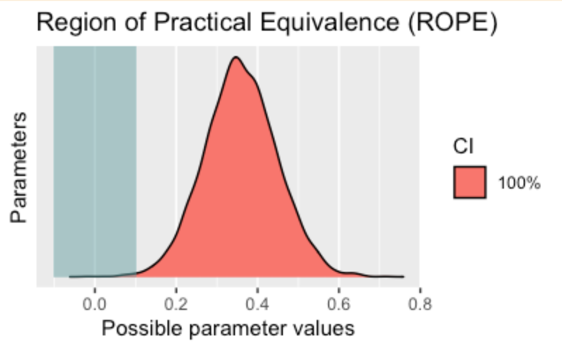
```
## # Proportion of samples inside the ROPE [-0.10, 0.10]:
##
## Parameter   | inside ROPE
## -------------------------
## (Intercept) |     77.89 %
## uv1         |      0.00 %
```

Dann sucht man sich die richtige Zeile und liest den Prozentwert `p` ab. Dann rechnet man `1 - p` und reicht das Ergebnis (als Anteil, nicht als Prozentwert) als Lösung ein.

Antwort: 1.00.

Oft ist ein Diagramm zum Verstehen hilfreich, das kann man so erhalten:

```
plot(rope(mod))
```



Wie man sieht liegt also dieser Teil außerhalb des ROPE: 1.

Aufgaben-ID: Post2-rope, Toleranzbreite: 0.100

Die richtige Antwort ist: 1

Frage **16**

Richtig

Erreichte Punkte 1,00 von 1,00

Gegeben sei der DAG `g` (s. u.). Der DAG verfügt über $n = 6$ Variablen, die als Knoten im Graph dargestellt sind und mit $x_1, x_2, \ldots x_n$ bezeichnet sind.

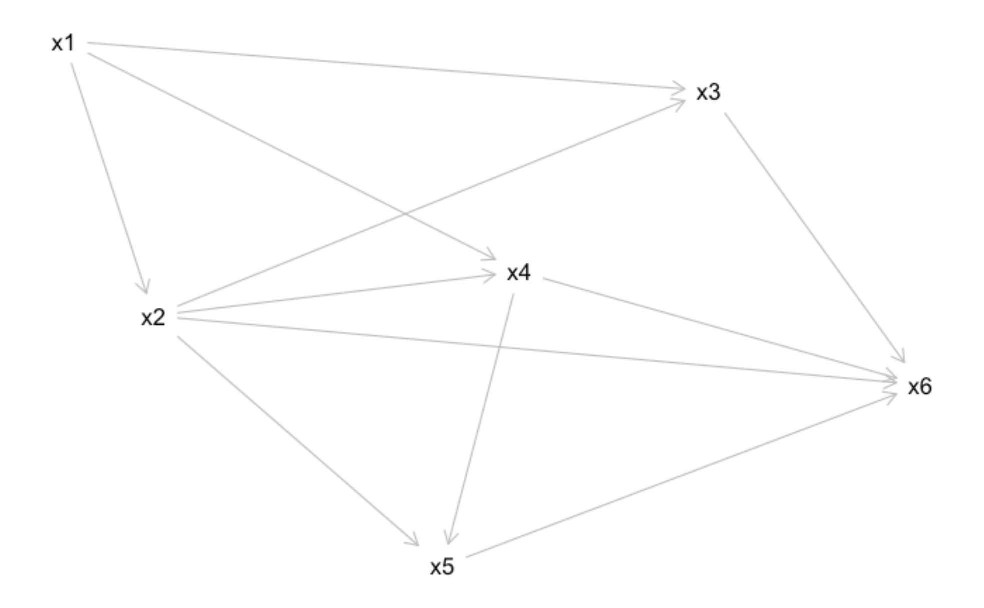
Welche minimale Variablenmenge muss kontrolliert werden, um den totalen kausalen Effekt von der UV zur AV zu identifizieren?

UV: x5.

AV: x6.

Hinweise:

- Mengen sind mittels geschweifter Klammern gekennzeichnet, z.B. `{x1, x2}` meint die Menge mit den zwei Elementen `x1` und `x2`.
- Die leere Menge `{ }` bedeutet, dass keine Variable kontrolliert werden muss, um den Effekt zu identifizieren.
- Alle Variablen werden als gemessen vorausgesetzt.
- Es ist möglich, dass es keine Lösung gibt, dass es also keine Adjustierungsmenge gibt, um den totalen kausalen Effekt zu identifizieren. Wenn dies der Fall sein sollte, wählen Sie "/".
- Es ist möglich, dass einzelne Variablen keine Kanten besitzen, also keine Verbindung zu anderen Variablen (Knoten) haben. Diese Variablen sind dann kausal unabhängig von den übrigen Variablen.



Wählen Sie eine Antwort:

a. { x6 }

b. { x3, x6 }

c. { x1, x6 }

d. { x5, x6 }

e. { x2 , x4 } Richtig

Folgende minimale Variablenmenge muss kontrolliert werden, um den kausalen Effekt der UV auf die AV zu identifizieren: `{ x2 , x4 }`.

Es wird also nach dem *minimal adjustment set* gefragt. Zu beachten ist, dass es weitere, aber größere Mengen geben kann, die den kausalen Effekt identifizieren, aber gefragt ist nach der minimalen Menge. Außerdem ist es möglich, dass es weitere minimale Adjustierungsmengen gibt, die aber nicht in der Auswahlliste vorkommen.

Erläuterung:

UV und AV sind (durch mindestens einen offenen Pfad) verbunden ("d-connected").

Folgende Unabhängigkeitsbeziehungen (`_||_`) werden durch den DAG definiert:

```
## x1 _||_ x5 | x2, x4
## x1 _||_ x6 | x2, x3, x4
## x3 _||_ x4 | x1, x2
## x3 _||_ x5 | x2, x4
## x3 _||_ x5 | x1, x2
```

UV und AV werden durch folgende Pfade verbunden:

```
## $paths
##  [1] "x5 -> x6"                         "x5 <- x2 -> x3 -> x6"
##  [3] "x5 <- x2 -> x3 <- x1 -> x4 -> x6" "x5 <- x2 -> x4 -> x6"
##  [5] "x5 <- x2 -> x4 <- x1 -> x3 -> x6" "x5 <- x2 -> x6"
##  [7] "x5 <- x2 <- x1 -> x3 -> x6"       "x5 <- x2 <- x1 -> x4 -> x6"
##  [9] "x5 <- x4 -> x6"                   "x5 <- x4 <- x1 -> x2 -> x3 -> x6"
## [11] "x5 <- x4 <- x1 -> x2 -> x6"       "x5 <- x4 <- x1 -> x3 -> x6"
## [13] "x5 <- x4 <- x1 -> x3 <- x2 -> x6" "x5 <- x4 <- x2 -> x3 -> x6"
## [15] "x5 <- x4 <- x2 -> x6"             "x5 <- x4 <- x2 <- x1 -> x3 -> x6"
##
## $open
##  [1]  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [12]  TRUE FALSE  TRUE  TRUE  TRUE
```

Offenen Pfade sind der/die Pfad/e, die in der obigen Ausgabe unter `$open` mit `TRUE` gekennzeichnet sind.

Der DAG ist wie folgt definiert:

```
dag {
x1
x2
x3
x4
x5
x6
x1 -> x2
x1 -> x3
x1 -> x4
x2 -> x3
x2 -> x4
x2 -> x5
x2 -> x6
x3 -> x6
x4 -> x5
x4 -> x6
x5 -> x6
}
```

Aufgaben-ID: randomdag1

---

a. Falsch
b. Richtig
c. Falsch
d. Falsch
e. Falsch

Die richtige Antwort ist: { x2 , x4 }

Frage **17**

Falsch

Erreichte Punkte 0,00 von 1,00

Im Folgenden ist der Datensatz `iris` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/iris.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Berechnen Sie das folgende lineare Modell:

AV: Sepal.Length.

UVs: Sepal.Width, Petal.Width.

Denken Sie daran, auch beim Erstellen der PPV die Zufallszahlen zu fixieren. Sie können dazu den Parameter `seed` auf 42 setzen.

Aufgabe: **Geben Sie die Breite des 50%-PI der PPV an für eine Beobachtung, bei der alle Prädiktoren den Wert -1 aufweisen! Gehen Sie für alle Variablen von z-Werten aus.**

Hinweise:

Beachten Sie die Bearbeitungshinweise.

Antwort: 0,62 ✖

MUSTERLÖSUNG

Der Datensatz heißt:

```
d_name
```

```
## [1] "iris"
```

Quelle: https://vincentarelbundock.github.io/Rdatasets/csv/datasets/iris.csv.

Prädiktor(en) und AV z-transformieren:

```
head(d)  # Rohdaten
```

```
##    av uv1 uv2
## 1 5.1 3.5 0.2
## 2 4.9 3.0 0.2
## 3 4.7 3.2 0.2
## 4 4.6 3.1 0.2
## 5 5.0 3.6 0.2
## 6 5.4 3.9 0.4
```

```
d <- standardize(d)
head(d)  # z-standardisiert
```

```
##          av        uv1      uv2
## 1 -0.897674  1.0156020 -1.31105
## 2 -1.139200 -0.1315388 -1.31105
## 3 -1.380727  0.3273175 -1.31105
## 4 -1.501490  0.0978893 -1.31105
## 5 -1.018437  1.2450302 -1.31105
## 6 -0.535384  1.9333146 -1.04867
```

Die Prädiktoren (UVs) des Modells lauten: Sepal.Width, Petal.Width.

Laut Angabe haben die Prädiktoren den Wert `-1`.

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1 + uv2,
                data = d,
                refresh = 0,
                chains = 4,  # ignorieren Sie diese Zeile
                seed = 42)
```

Als nächstes berechnen Sie die PPV, z.B. so `ppv <- estimate_prediction(mod, newdata = tibble(Sepal.Width = -1, Petal.Width = -1, seed = 42))`.

```
estimate_prediction(mod, tibble(uv1 = preds_k, uv2 = preds_k), ci = .5, seed = 42)
```

```
## Model-based Prediction
##
## uv1   |   uv2 | Predicted |   SE |         50% CI
## -------------------------------------------------
## -1.00 | -1.00 |     -1.11 | 0.56 | [-1.48, -0.74]
##
## Variable predicted: av
```


```
## Model-based Prediction
##
## uv1   |   uv2 | Predicted |   SE |         50% CI
## -------------------------------------------------
## -1.00 | -1.00 |     -1.11 | 0.56 | [-1.48, -0.74]
##
## Variable predicted: av
```


Dann berechnet man die Breite, in dem man die Differenz der oberen und unteren Grenze des Intervalls bildet.

```
sol_raw <- ppv$CI_high - ppv$CI_low
sol <- exams::fmt(sol_raw, 2)
```

Antwort: Die Lösung lautet: 0.74.

Aufgaben-ID: PPV3a-num, Toleranzbreite: 0.070

Die richtige Antwort ist: 0,74

Frage **18**

Richtig

Erreichte Punkte 1,00 von 1,00

Im Folgenden ist der Datensatz `gtcars` zu analysieren. Er ist unter dieser Quelle erhältlich: https://vincentarelbundock.github.io/Rdatasets/csv/gt/gtcars.csv.

Eine Möglichkeit, den Datensatz zu beziehen, ist diese Sammlung an Datensätzen. Suchen Sie dort nach dem Namen des Datensatzes. Importieren Sie dann die Daten in R.

Hilfe zum Datensatz ist auf dieser Webseite abrufbar.

Zur Lösung dieser Aufgabe ist folgendes lineares Modell zu berechnen:

AV: msrp.

UVs: hp_rpm, trq.

Aufgabe: **Was ist der Wert des Punktschätzers für eine Beobachtung, bei der alle Prädiktoren den jeweiligen Mittelwert aufweisen?**

Beachten Sie die Bearbeitungshinweise.

Antwort: 193973,5 ✔

MUSTERLÖSUNG

Datenquelle: https://vincentarelbundock.github.io/Rdatasets/csv/gt/gtcars.csv.

Pakete starten:

```
library(tidyverse)
library(easystats)
library(rstanarm)
```

Daten importieren:

`d <- read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/gt/gtcars.csv')`

Zunächst berechnet man das Modell, etwa so:

```
mod <- stan_glm(av ~ uv1 + uv2,
                data = d,
                refresh = 0,
                chains = config$chains,  # ignorieren Sie diese Zeile
                seed = 42
                )
```

Dann sagen wir eine Beobachtung mit den gewünschten Werten vorher. In Pseudo-Syntax:

```
predict(mod, newdata = data.frame(uv1 = <hier MW von UV 1 eintragen>, uv2 = <hier MW von UV 1 eintragen>))
```

Für die Lösung (`sol`) berechnen wir noch die Mittelwerte jeder UV:

```
sol <- predict(mod, newdata = data.frame(uv1 = mean(d$uv1, na.rm = TRUE), uv2 = mean(d$uv2, na.rm = TRUE)))
```

Lösung: Der laut Modell vorhergesagte Werte beträgt:

```
## [1] "193973.55"
```

Aufgaben-ID: predict-dyn2, Toleranzbreite: 19397.350

Die richtige Antwort ist: 193973,55

Frage **19**

Richtig

Erreichte Punkte 1,00 von 1,00

Betrachten Sie die folgenden Modelldefinitionen. Nur *eine* davon ist korrekt bzw. sinnvoll. Wählen Sie diese aus!

Hinweise:

- Es ist *ein* Prädiktor im Modell.
- Der Achsenabschnitt ist Null.
- Die Prädiktoren sind z-standardisiert.
- Gehen Sie davon aus, dass Daten und notwendige R-Pakete und sonstige technischen Voraussetzungen erfüllt sind.
- $N()$ meint die Normalverteilung und $E()$ die Exponentialverteilung.

Modelldefinition A:

$y_i \sim N(\mu_i, \sigma)$
$\mu_i = 0 + \beta_1 x_i$
$\beta_1 \sim N(3.1, -0.6)$
$\sigma \sim E(-0.1)$

Modelldefinition B:

$y \sim N(\mu, \sigma)$
$\mu = 0 + \beta_1 x$
$\beta_1 \sim N(3.1, 0.6)$
$\sigma \sim E(0.1)$

Modelldefinition C:

$y_i \sim N(\mu_i, \sigma)$
$\mu_i = 0 + \beta_1 x_i$
$\beta_1 \sim N(3.1, 0.6)$
$\sigma \sim E(0.1)$

Modelldefinition D:

$y_i = N(\mu_i, \sigma)$
$\mu_i = 0 + \beta_1 x_i$
$\beta_1 = N(3.1, 0.6)$
$\sigma = E(0.1)$

Modelldefinition E:

$y_i \sim N(\mu_i, \sigma)$
$\mu_i = 1 + \beta_1 x_i$
$\beta_1 \sim N(3.1, 1.6)$
$\sigma \sim E(1.1)$

Modelldefinition F:

$\mu_i \sim N(\mu_i, \sigma)$
$\sigma = 0 + \beta_1 x_i$
$\beta_1 \sim N(3.1, 0.6)$
$y_i \sim E(0.1)$

Wählen Sie eine Antwort:

- a. B
- b. A
- c. C — Richtig. korrekt
- d. E
- e. D
- f. F

Aufgaben-ID: Modelldef2

a. Falsch. Streuungen dürfen kein negatives Vorzeichen haben
b. Falsch. Die Indizes i fehlen (bei mu)
c. Richtig. korrekt
d. Falsch. Die Tilde-Zeichen sind fälschlich durch Gleichheitszeichen ersetzt
e. Falsch. Der Achsenabschnitt ist per Angabe Null
f. Falsch. Die linken Seiten der Gleichungen passen nicht zu den rechten Seiten

Die richtige Antwort ist: C

◄ Beachten Sie die Bearbeitungshinweise für die Prüfung

Direkt zu:

Notenschwellen ►