

4E2 Herleitung des p-Wertes

Um zu verstehen, was der p-Wert genau ist und wie er berechnet werden kann, müssen wir uns zunächst mit dem Begriff der *Nullhypothese* und der *Alternativhypothese* beschäftigen.

Die Nullhypothese ist die Annahme, dass es keinen Effekt (von x auf y , einer Maßnahme, ...) gibt. Die Alternativhypothese ist das Gegenteil, also die Annahme, dass es einen Effekt gibt.

Das bedeutet, unter der Nullhypothese nehmen wir an, dass in unserem Modell

$$y = a + bx$$

$b = 0$ ist. Wir können dann die Abweichung des geschätzten Wertes für b von 0 berechnen und nennen diese Statistik *t-Statistik*:

$$t = \frac{b - 0}{s/\sqrt{N}}$$

s ist hier die Standardabweichung von x und N die Anzahl der Datenpunkte.

Den p-Wert können wir nun beschreiben als die Wahrscheinlichkeit, dass der Betrag der t-Statistik (also die Abweichung des geschätzten Wertes für b von 0) mindestens so groß ist wie oben berechnet. Grafisch dargestellt sieht das wie folgt aus:

TODO

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.0
v purrr      1.0.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
x_vals <- seq(-4, 4, length = 100)

y_vals <- dnorm(x_vals)

data <- tibble(x_vals, y_vals)

data %>%
  ggplot(aes(x_vals, y_vals)) +
  geom_line(linewidth = 1, color = "black") +
  theme_classic() +
  theme(axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
        #axis.text.y = element_blank()
        ) +
  xlab("x") +
  ylab("y") +
  geom_vline(xintercept = 1.5, color = "red", linewidth = 1) +
  geom_vline(xintercept = -1.5, color = "red", linewidth = 1) +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 0.43)) +
  stat_function(fun = dnorm,
               xlim = c(1.5, 4),
               geom = "area",
               fill = "red",
               alpha = 0.3) +
  stat_function(fun = dnorm,
               xlim = c(-4, -1.5),
               geom = "area",
               fill = "red",
               alpha = 0.3) +
  annotate("text", x=-1.65, y=0.2, label="t", angle=90)
```

