

Wahrscheinlichkeiten auf Gleichheit prüfen

Um mithilfe einer Datenerhebung zu prüfen, ob z. B. zwei Wirkstoffe gleich wahrscheinlich heilen („+“), vergleicht man die zugehörigen relativen Heilungshäufigkeiten. Je weiter sie auseinanderliegen, desto eher bezweifelt man die Gleichheit der Wahrscheinlichkeiten.

Aus einer Datenerhebung gehen nebenstehende Werte hervor. Man erhält für das Placebo

$$h_1(+) = \frac{75}{200} = 0,375 \text{ und für das Medikament } h_2(+) = \frac{189}{400} = 0,4725.$$

Der Abstand der beiden relativen Heilungshäufigkeiten beträgt $\Delta = |h_1 - h_2| = 0,0975$.

Ist dieser Abstand groß genug, um an der Gleichheit der Wahrscheinlichkeiten zu zweifeln?

Wenn sich die zugehörigen Konfidenzintervalle ($\beta = 95,4\%$) mit den Radien

$$r_1 = 2\sqrt{\frac{0,375 \cdot 0,625}{200}} = 0,06846 \text{ und } r_2 = 2\sqrt{\frac{0,4725 \cdot 0,5275}{400}} = 0,04992 \text{ nicht überlappen würden, dann}$$

gäbe es keine gemeinsame Heilungswahrscheinlichkeit, die zum Placebo und zum Medikament passt. Der Unterschied der Heilungswahrscheinlichkeiten wäre *statistisch abgesichert*.

Die Konfidenzintervalle überlappen sich aber und in der Nähe der gemeinsamen relativen

Heilungshäufigkeit $h(+) = \frac{264}{600} = 0,44$ gibt es viele Heilungswahrscheinlichkeiten, die sowohl zum Placebo als auch zum Medikament passen. Der Nachweis, dass die Wirkstoffe mit unterschiedlichen Wahrscheinlichkeiten heilen, wäre somit gescheitert.

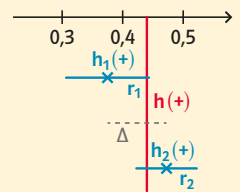
Die Forderung, dass der Abstand Δ zwischen h_1 und h_2 größer als die Summe der Konfidenzintervallradien $r_1 + r_2 = 0,11838$ sein soll, erweist sich als zu einschneidend. Simulationen zeigen: Für den Nachweis, dass sich die Wahrscheinlichkeiten statistisch signifikant voneinander unterscheiden, muss nur $\Delta > \sqrt{r_1^2 + r_2^2} = 0,084727$ gelten, und das ist hier der Fall.

Info

Um zu prüfen, ob zwei Bernoulli-Ketten mit den Längen n_1 und n_2 eine gleiche Trefferwahrscheinlichkeit zugrunde liegt, verfährt man wie folgt:

1. Man bestimmt die relativen Häufigkeiten $h_1 = \frac{x_1}{n_1}$ und $h_2 = \frac{x_2}{n_2}$ sowie $\Delta = |h_1 - h_2|$.
2. Man berechnet die Radien der Konfidenzintervalle $r_1 = 2\sqrt{\frac{h_1(1-h_1)}{n_1}}$ und $r_2 = 2\sqrt{\frac{h_2(1-h_2)}{n_2}}$.
3. Wenn $\Delta > \sqrt{r_1^2 + r_2^2}$ gilt, bezweifelt man die Gleichheit der Trefferwahrscheinlichkeiten.

	-	+	Summe
Placebo	125	75	200
Medikament	211	189	400
Summe	336	264	600



Eine Simulation wird in Aufgabe 2 durchgeführt.

In der Statistik heißt dieses Prüfverfahren **χ^2 -Vierfeldertest**. Statistikprogramme verwenden die Näherung

$$\begin{aligned} & \sqrt{r_1^2 + r_2^2} \\ &= 2\sqrt{\frac{h_1(1-h_1)}{n_1} + \frac{h_2(1-h_2)}{n_2}} \\ &\approx 2\sqrt{h(1-h) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \end{aligned}$$

mit $h = \frac{x_1 + x_2}{n_1 + n_2}$.

Wenn Sie die je 20 Würfe in zwei 10er-Serien teilen, können Sie zusätzlich prüfen, ob sich die Trefferwahrscheinlichkeit während des Versuchs ändert.

146 **interaktive Anwendung**
Mit Konfidenzintervallen die Gleichheit von Wahrscheinlichkeiten prüfen.

1 Experimentieren

- a) Jede Person Ihres Kurses wirft abwechselnd 20 rote Gegenstände mit der rechten und 20 lila-farbene mit der linken Hand von einer festen Position aus in einen Eimer. Notieren Sie in einer Vierfeldertafel, wie häufig Sie mit welcher Hand treffen bzw. nicht treffen.
- b) Fassen Sie die Daten des gesamten Kurses zusammen.
- c) Bestimmen Sie die relative Trefferhäufigkeit h_1 für die linken Hände, h_2 für die rechten Hände und berechnen Sie mit der Näherungsformel die zu h_1 und h_2 gehörenden Konfidenzintervalle zum Konfidenzniveau $\beta = 95,4\%$ sowie deren Radien r_1 und r_2 .
- d) Prüfen Sie, ob in Ihrem Kurs die Trefferwahrscheinlichkeiten der linken und der rechten Hand als gleich angesehen werden können.

2 Simulieren

- a) Simulieren Sie 100-mal zwei Bernoulli-Ketten mit den Längen $n_1 = 200$ und $n_2 = 400$ und der gemeinsamen Trefferwahrscheinlichkeit $p = 0,45$.
- b) Berechnen Sie die Differenz Δ sowie die Radien r_1 und r_2 der beiden Konfidenzintervalle.
- c) Bestätigen Sie, dass in ca. 95,4% aller Fälle $\Delta \leq \sqrt{r_1^2 + r_2^2}$ gilt.
- d) Variieren Sie die Parameter n_1 , n_2 , und p und beschreiben Sie Ihre Beobachtungen.
- e) Überprüfen Sie, zu welchen Ergebnissen die Simulation führt, wenn man den beiden Bernoulli-Ketten unterschiedliche Trefferwahrscheinlichkeiten zuordnet.

