

Zu Abschnitt 3.4.2: Statistische Tests auf Gruppenebene

Statistische Tests für Häufigkeiten: Realnormvergleich

Beispiel 4: Gibt es in Berlin mehr HIV-Positive als im Bundesdurchschnitt?

Die Inzidenzrate für HIV-Neuinfektionen lag in Deutschland im Jahre 2005 bei 3,02 pro 100.000 (= 0,03‰; vgl. Robert-Koch-Institut, 2006). Diesen Wert betrachten wir im Folgenden als Populationskennwert. In Berlin wurden im Jahre 2005 insgesamt 370 HIV-Neuinfektionen gemeldet. Die Inzidenzrate liegt also hier mit ca. 1,1‰ (370/3.400.000) deutlich höher. Weicht nun die Inzidenzrate für Berlin signifikant von der Inzidenzrate für das gesamte Bundesgebiet (das wir hier als Population definieren) ab?

Diese Frage kann mit Hilfe eines Binomialtests beantwortet werden (s. Eid et al., 2013; Abschn. 10.4). Hierfür benötigen wir *keine Originaldaten*; der Test kann sehr leicht „von Hand“ mit Hilfe der entsprechenden Formel berechnet werden. Da es sich hier um eine sehr große Stichprobe handelt und der exakte Binomialtest in einem solchen Fall fast perfekt durch einen Gauß-Test approximiert wird, führen wir hier der Einfachheit halber den Gauß-Test (z-Test) durch. Der Wert der Prüfgröße berechnet sich wie folgt:

$$z = \frac{h - n \cdot \pi}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}$$

mit

- h = Absolute Häufigkeit, die gegen einen fixen Wert getestet werden soll (hier also die Anzahl der HIV-Neuinfektionen in Berlin im Jahre 2005, $h = 370$),
- n = Stichprobengröße (hier: Anzahl der in Berlin lebenden Personen; $n = 3400000$) und
- π = Wahrscheinlichkeit des Auftretens des Merkmals in der Population (hier: HIV-Inzidenz im Bundesdurchschnitt; $\pi = 0,0302\text{‰} = 0,0000302$).

Der Binomialtest gibt am Ende also einen z -Wert aus, mit dessen Hilfe sich die Wahrscheinlichkeit des empirischen Ergebnisses unter allen möglichen Ergebnissen (gegeben die Nullhypothese, dass die Inzidenz in der Subpopulation nicht von der Populationsrate abweicht) ermittelt werden kann. In unserem Datenbeispiel lässt sich dieser z -Wert wie folgt berechnen:

$$z = \frac{h - n \cdot \pi}{\sqrt{n \cdot \pi \cdot (1 - \pi)}} = \frac{370 - 3400000 \cdot 0,0000302}{\sqrt{3400000 \cdot 0,0000302 \cdot (1 - 0,0000302)}} = \frac{267,32}{10,13} \approx 26,38$$

Die Wahrscheinlichkeit für diesen (oder jeden größeren) z -Wert ist außerordentlich klein: Der entsprechende p -Wert würde selbst bei 30 Nachkommastellen nur Nullen aufweisen. Das bedeutet: In Berlin ist die HIV-Inzidenz in der Tat signifikant höher als im Bundesgebiet.