

Statistische Inferenz | 01 | Fisher-Exakt-Test

Musterlösung

Prof. Dr. Roland Schäfer | Germanistische Linguistik FSU Jena

17. Oktober 2024

1 Fisher-Exakt-Test und Stichprobengröße

1. Rekapitulieren Sie die Berechnung des klassischen *Tea Tasting Lady*-Experiments für 6 richtige Tassen bei 8 Tassen insgesamt (also 3 richtige „Tee zuerst“-Tipps und ein falscher Tipp von vier möglichen richtigen „Tee zuerst“-Tipps) auf den Folien.
2. Berechnen Sie die Wahrscheinlichkeiten bzw. den p-Wert für dasselbe Verhältnis von richtigen Tassen, aber bei einer zehn Mal größeren Stichprobe, also 60 Tassen korrekt vorhergesagt bei 80 Tassen insgesamt.
3. Interpretieren Sie das Ergebnis.
4. Stellen Sie die Anfangswerte dieser Berechnung als Vier-Felder-Tabelle dar.

Lösung zu 2

$$\frac{\binom{40}{30}\binom{40}{10}}{\binom{80}{40}} = \frac{\frac{40!}{30!(40-30)!} \frac{40!}{10!(40-10)!}}{\frac{80!}{40!(80-40)!}} = \frac{\frac{40!}{30!10!} \frac{40!}{10!30!}}{\frac{80!}{40!40!}} = \frac{\left(\frac{40!}{30!10!}\right)^2}{\frac{80!}{40!^2}} \approx \frac{\left(\frac{8,16 \cdot 10^{47}}{9,63 \cdot 10^{38}}\right)^2}{\frac{7,16 \cdot 10^{118}}{6,66 \cdot 10^{95}}} \approx \frac{7,19 \cdot 10^{17}}{1,08 \cdot 10^{23}} \approx 6,68 \cdot 10^{-6} \approx 0,00000668$$

Lösung zu 3

Das ist vor dem Experiment die Wahrscheinlichkeit gewesen, durch Raten genau 30 Richtige zu erhalten. Zusätzliche Überlegung: Ist das wirklich das, was uns interessiert? Eigentlich interessiert uns für unsere Schlussfolgerung doch eher, wie wahrscheinlich es war, mindestens ein so gutes Ergebnis zu erzielen. Das ist das, was der tatsächliche Fisher-Test typischerweise berechnet, und das ergibt in diesem Fall $p \approx 7,44 \cdot 10^{-6}$. Man kann sich das herleiten als:

$$\frac{\binom{40}{30}\binom{40}{10}}{\binom{80}{40}} + \frac{\binom{40}{31}\binom{40}{9}}{\binom{80}{40}} + \frac{\binom{40}{32}\binom{40}{8}}{\binom{80}{40}} + \dots + \frac{\binom{40}{40}\binom{40}{0}}{\binom{80}{40}}$$

Lösung zu 4

30	10
10	30

Summe der Werte 4. In der fiktiven Korpusstudie haben wir zwar für beide Verben 500 Belege gezogen, und die Spalten summieren sich daher jeweils zu 500, aber wie viele Passive und Aktive wir jeweils finden würden, konnten wir vor der Durchführung der Studie nicht wissen. Die erste Zeile summiert sich zu 150, die zweite zu 850, es hätte aber auch ganz anders kommen können. Daher ist der Fisher-Test eigentlich nicht geeignet für solche Studien. Überlegen Sie, warum. Das ist aber wirklich extrem fortgeschritten. Kaum jemand in der Linguistik weiß das überhaupt, ganz zu schweigen davon, zu wissen, wie es mathematisch zu begründen ist.