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 2

$$\frac{\binom{40}{30}\binom{40}{10}}{\binom{80}{40}} = \frac{\frac{40!}{30!(40-30)!}\frac{40!}{10!(40-10)!}}{\frac{80!}{40!(80-40)!}} = \frac{\frac{40!}{30!10!}\frac{40!}{10!30!}}{\frac{80!}{40!40!}} = \frac{(\frac{40!}{30!10!})^2}{\frac{80!}{40!2}} \approx \frac{(\frac{816\cdot10^{47}}{9,63\cdot10^{38}})^2}{\frac{80!}{6.66\cdot10^{95}}} \approx \frac{7,19\cdot10^{17}}{1,08\cdot10^{23}} \approx 6,68\cdot10^{-6} \approx 0,00000668$$

Das ist vor dem Experiment die Wahrscheinlichkeit gewesen, durch Raten genau 30 Richtige zu erhalten. 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 berechnet, und das ergibt in diesem Fall $p \approx 7,44 \cdot 10^{-6}$. Man kann sich das herleiten als:

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

Lösung 4

Vorgriff auf Kollostruktionsanalysen mit Fisher-Test 2

Uns werden im weiteren Verlauf sogenannte Kollostruktionsanalysen begegnen. Diese sind ein Analyseverfahren für Korpusdaten und wurden ursprünglich (bei ihrer Erfindung vor 20 Jahren) mit Fisher-Exakt-Tests gerechnet. Ein möglicher Datensatz wäre folgender: 1

	kaufen	umarmen
im Passiv	120	30
im Aktiv	380	470

Konzeptionell sind die tabulierten Zahlen die Anzahlen der Vorkommnisse von (in diesem konkreten Fall) dem Verb kaufen im Aktiv und im Passiv und dem Verb umarmen – ebenfalls im Aktiv und im Passiv - in einer Stichprobe aus irgendeinem Korpus (= Textsammlung) des Deutschen. (Die Zahlen sind hier der Einfachheit halber ausgedacht.) Wir möchten mit so einer Untersuchung quantifizieren, wie hoch die Affinität der Verben im Vergleich zueinander zur Passivbildung ist.

- 1. Wie viele Vorkommnisse von kaufen und umarmen haben wir jeweils beobachtet?
- 2. Wie viele Sätze im Passiv und im Aktiv haben wir jeweils beobachtet? Im Passiv 150, im Aktiv 850.
- 3. Wie groß war die gesamte Stichprobe? Genau 1000.
- 4. Was ist das Verhältnis zwischen Passiv und Aktiv bei kaufen und bei umarmen? $kaufen: \frac{120}{380} \approx 0,32$ $umarmen: \frac{30}{470} \approx 0,06$ 5. Wie würden Sie den p-Wert berechnen? Es reicht der Lösungsweg, wenn Ihr Taschenrechner
- bei den großen Zahlen aussteigt.

$$\frac{\binom{500}{120}\binom{500}{380}}{\binom{1000}{500}} \approx 1,38 \cdot 10^{-63}$$

- 6. Der p-Wert ist ungefähr $1,38 \cdot 10^{-63}$. Überführen Sie diese Zahl aus der Exponentialschreibweise in herkömmliche Fließkommanotation, falls Sie das noch können. len zwischen dem Komma und der ersten Stelle der Basis.
- 7. Was sagt uns dieser p-Wert? Was ist das für eine Wahrscheinlichkeit? Das ist ja zunächst mal von der Idee her etwas völlig anderes als bei der *Tea Tasting Lady*. Das war vor der Korpusstudie die Wahrscheinlichkeit, genau 120 Passive und 380 Aktive von kaufen zu finden, wenn der wahre Anteil der Passive und Aktive bei kaufen dem von umarmen entspricht (also 30 Passive und 470 Aktive). Der empirisch interessantere p-Wert für "eine so starke oder stärkere Abweichung von der Verteilung bei umarmen" liegt bei $2,59 \cdot 10^{-16}$.
- 8. Es gibt zwischen den Designs der Tea Tasting Lady und der Kollostruktionsanalyse einen wesentlichen Unterschied bezüglich der Summen der Werte in den Spalten und den Zeilen der Tabelle. Finden Sie den? Das ist allerdings eine optionale Transferaufgabe auf sehr hohem Niveau.
 - Bei der Tea Tasting Lady sind die Summen der Werte in den Zeilen und Spalten der Tabelle durch das Design des Experiments festgelegt. Es wurde vereinbart, dass sie 8 Tassen bekommt, von denen in 4 die Milch zuerst eingeschenkt wurde. Außerdem sucht sie genau 4 Tassen aus. Egal, wie gut sie rät oder die Tassen erkennt, in jeder Zeile und Spalte der Tabelle ist die Summe der Werte 4. In der fiktiven Korpusstudie haben wir zwar für beide Verben 500 Belege gezogen, und die Spalten summieren sich jeweils zu 500, aber wie viele Passive und Aktive

¹Normalerweise nimmt man etwas andere Designs, aber das schauen wir uns dann in Ruhe an.

wir jeweils finden, wissen wir vorher nicht. Die erste Zeile summiert sich zu 150, die zweite zu 850, was wir vorher nicht wussten. 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.