

Statistische Inferenz | 01 | Fisher-Exakt-Test

Prof. Dr. Roland Schäfer | Germanistische Linguistik FSU Jena

3. November 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 und ein falscher „Tee zuerst“-Tipp von insgesamt 4 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.

2 Vorgriff auf Kollostruktionsanalysen mit Fisher-Test

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

	<i>kaufen</i>	<i>umarmen</i>
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?
3. Wie groß war die gesamte Stichprobe?
4. Was ist das Verhältnis zwischen Passiv und Aktiv bei *kaufen* und bei *umarmen*?
5. Wie würden Sie den p-Wert berechnen? Es reicht der Lösungsweg, wenn Ihr Taschenrechner bei den großen Zahlen aussteigt.
6. Der p-Wert ist ungefähr $1,03 \cdot 10^{-133}$. Überführen Sie diese Zahl aus der Exponentialschreibweise in herkömmliche Fließkommanotation, falls Sie das noch können.
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*.

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

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.