

5 Statistische Auswertungsmethode

Für die Auswertung der Ergebnisse des ABX-Tests ist zunächst die Wahl einer statistischen Methode notwendig, um anschließend mathematisch fundierte und interpretierbare Aussagen treffen zu können. Von der ABX-Company wird hier das statistische Modell der Binomialverteilung als sehr nützlich empfohlen.

5.1 Die Binomialverteilung

Bei einem Zufallsexperiment mit lediglich zwei möglichen Ausgängen gilt, dass das Ereignis A mit der Wahrscheinlichkeit p ($0 < p < 1$) und das zu A komplementäre Ereignis B mit der Wahrscheinlichkeit $q = 1 - p$ eintritt. Wird dieses Zufallsexperiment mehrmals hintereinander ausgeführt, so spricht man vom sogenannten Bernoulli-Prozeß. Voraussetzung hierfür ist, dass die aufeinanderfolgenden Experimente sich nicht gegenseitig beeinflussen, d.h., dass der Ausgang der i -ten Durchführung unabhängig vom Ausgang der j -ten Durchführung ist.

Die Wahrscheinlichkeiten p für das Ereignis A und q für das Ereignis B gelten auch für jede Wiederholung des Experiments. Das jeweilige Ereignis tritt also grundsätzlich mit der gleichen Wahrscheinlichkeit ein.

Finden bei einem Bernoulli-Prozeß n Versuche statt, so kann das Ereignis A unter den n beobachteten Ergebnissen 0- bis n -mal auftreten. Ist nun X die Anzahl der Ergebnisse A unter den n beobachteten Ergebnissen, dann liefert die Binomialverteilung die Wahrscheinlichkeit für das Ereignis ($X = x$).

Sei $BBA \dots AB$ eine Folge der Länge n , in der x -mal der Ausgang A und $(n-x)$ -mal der Ausgang B auftritt, so lässt sich aufgrund der Unabhängigkeit der einzelnen Versuche die Wahrscheinlichkeit $P(BBA \dots AB)$ wie folgt berechnen:

$$P(BBA \dots AB) = P(B)P(B)P(A) \dots P(A)P(B) = p^x(1-p)^{n-x}$$

Zwar sind bei diesem Versuch die Anzahl der Ereignisse A interessant, nicht aber deren konkrete Platzierung innerhalb der Abfolge, d.h., im wievielten Versuch ein Ereignis A

5 Statistische Auswertungsmethode

aufgetreten ist. Unter den n möglichen Positionen sind genau x Positionen mit dem Ereignis A zu besetzen, was auf $\binom{n}{x}$ verschiedene Möglichkeiten erfolgen kann. Hieraus ergibt sich die gesuchte Wahrscheinlichkeit $P(X=x)$ der Binomialverteilung:

$$P(X=x) = p_x = \binom{n}{x} p^x q^{n-x} \quad \text{mit } x = 0, 1, \dots, n$$

$$= \binom{n}{x} p^x (1-p)^{n-x} \quad \text{und} \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

p : Wahrscheinlichkeit für das Eintreten des Ereignisses A (beim Einzelversuch) mit $(0 < p < 1)$

q : Wahrscheinlichkeit für das Eintreten des zum Ereignis A komplementären Ereignisses B (beim Einzelversuch) mit $q = 1 - p$

n : Anzahl der durchgeführten Experimente/ Einzelversuche

x : Anzahl der Versuche, bei denen das gewünschte Ereignis eintritt

Bei solch einem Auswertungsverfahren soll geprüft werden, ob eine zuvor aufgestellte Hypothese H aufgrund der Ergebnisse des Tests beibehalten werden kann oder ob man sie ablehnt und sich damit für die sogenannte Gegenhypothese G entscheidet.

Bei Entscheidungen, die auf der Grundlage von Tests getroffen werden, können grundsätzlich jedoch auch Fehlentscheidungen bezüglich der Interpretation der Ergebnisse auftreten:

- Entscheidung für G , obwohl H richtig ist – dieser Fehler wird als Fehler erster Art bezeichnet
- Entscheidung für H , obwohl G richtig ist – dieser Fehler wird als Fehler zweiter Art bezeichnet

5 Statistische Auswertungsmethode

Die Möglichkeiten, sich richtig oder falsch zu entscheiden, lassen sich aus untenstehender Tabelle ersehen:

	H trifft zu (Realität)	G trifft zu (Realität)
Entscheidung für H	richtige Entscheidung	Fehler zweiter Art
Entscheidung für G	Fehler erster Art	richtige Entscheidung

Es lässt sich zeigen, dass die Wahrscheinlichkeit für einen Fehler erster Art umso kleiner ist, je größer die Wahrscheinlichkeit für einen Fehler zweiter Art ist. Man kann also nicht bei einer vorgegebenen Anzahl von Einzelversuchen die Wahrscheinlichkeiten für beide Fehlermöglichkeiten klein halten! In der Praxis geht man daher so vor, dass man die Wahrscheinlichkeit für den Fehler erster Art vorgibt, um zumindest eine Fehlerart unter Kontrolle halten zu können.

Die Wahrscheinlichkeit für den Fehler erster Art bezeichnet man auch als Signifikanzniveau oder Irrtumswahrscheinlichkeit, da durch ihn – durch Konvention – bestimmt wird, ab wann ein Ergebnis eine signifikante Abweichung zu dem unter der Hypothese erwarteten Ergebnis zeigt. Das Signifikanzniveau gibt somit an, ab welcher Abweichung man nicht mehr bereit ist, diese als zufällig zu bezeichnen. Die kritische Wahrscheinlichkeit $P(K)$ wird in den meisten naturwissenschaftlichen Gebieten mit 5 % angegeben, d.h., $P(K) \leq 0,05$.⁷⁴

Da lediglich die Entscheidung für G aus statistischer Sicht „ziemlich sicher“ ist⁷⁵, versucht man, den Fehler zweiter Art zu vermeiden. Dies ist nur gewährleistet, wenn man sich bei der Auswertung der Ergebnisse für die Gegenhypothese entscheidet. Somit sollte die eigentliche Hypothese als Gegenhypothese ausgesprochen werden. Man nimmt also an, dass die beobachtete zufällige Größe dem Modell der Hypothese H entspricht – liegt dann der beobachtete Wert im Bereich der kritischen Wahrscheinlichkeit $P(K)$, sieht man dies als statistischen Beweis dafür an, dass die Hypothese H ungültig, die Gegenhypothese G dagegen gültig ist.

⁷⁴ Stahel 1995: 175.

⁷⁵ Autorenkollektiv k.A.: 421.