4 Der Qualitätstest unter der rechnerischen Lupe

In diesem Kapitel wollen wir die Wahrscheinlichkeiten α und β für die Fehler 1. und 2. Art bei unserem Qualitätstest *berechnen.* Nach dem Studium dieses Kapitels werden Sie in der Lage sein, α und β für einen vergleichbaren neuen Test *selbständig* zu bestimmen. Darüber hinaus lernen Sie, mit zwei neuen Begriffen, nämlich dem *Signifikanzniveau* und der *Macht* des Tests, sachverständig umzugehen.

4.1 Der Fehler 1. Art - das Signifikanzniveau

Wiederholen wir: Bei unserem Qualitätstest machen wir den Fehler 1. Art, wenn mehr als 13 schwarze Kugeln in die Stichprobe geraten, obwohl die Urne nicht mehr als 4000 schwarze Kugeln enthält. Wie gross ist die Wahrscheinlichkeit α für diesen Fehlentscheid?

Die Wahrscheinlichkeit hängt, wie erwähnt, von der tatsächlichen Anzahl schwarzer Kugeln in der Urne ab; sie ist am grössten, wenn gerade 4000 schwarze Kugeln in der Urne sind und dieser Maximalwert heisst *Irrtumswahrscheinlichkeit* α.

Berechnung von α

Wie Sie gleich feststellen werden, ist es gar nicht so schwierig, α zu berechnen. Gehen Sie einfach die folgenden Umformungen aufmerksam durch:

$$\begin{aligned}\alpha &= P(\text{"mehr als 13 schwarze Kugeln in der Stichprobe"})\\ &= P(\text{"14 oder 15 oder ... 199 oder 200 schwarze Kugeln"})\\ &= 1-P(\text{"0 oder 1 oder ... 13 schwarze Kugeln"})\\ &= 1-\left[P(\text{"0 schwarze"})+P(\text{"1 schwarze"})+\ \dots\ +P(\text{"13 schwarze"})\right]\end{aligned}$$

Für diese Wahrscheinlichkeiten haben wir im letzten Kapitel eine Formel hergeleitet. Damit der Term überblickbar bleibt, schreiben wir ihn mit dem Summenzeichen:

$$\alpha = 1-\sum_{s=0}^{13} P(\text{"s schwarze Kugeln"}) = 1-\sum_{s=0}^{13}\frac{\binom{S}{s}\binom{N-S}{n-s}}{\binom{N}{n}}$$

Wir können die Formel fast eins zu eins in den Taschenrechner eingeben (Fig. 4.1).

Nach einigen Minuten Rechenzeit erhalten wir „grosso modo“ das gleiche Resultat wie mit der Computersimulation:

$$\alpha = 0.0310 = 3.10\ \%.$$

Nochmals die Interpretation für die Realität: Wenn der Einkäufer diesen Test *häufig* durchführt, so wird er durchschnittlich 3 von 100 Sendungen *irrtümlicherweise ablehnen.*

100000→nn

4000→ss

200→n

$$1 - \sum_{s=0}^{13} \frac{\mathrm{nCr(ss,s) \cdot nCr(nn\text{-}ss,n\text{-}s)}}{\mathrm{nCr(nn,n)}}$$

0.0310

Fig. 4.1: Die Irrtumswahrscheinlichkeit α auf einem Taschenrechner

Das Signifikanzniveau

Wenn der Statistiker mehr als 13 defekte Knallkörper in der Stichprobe findet, sagt er, das Testergebnis sei *signifikant* (bezeichnend) für eine Sendung mit unzulässiger Qualität. Es ist in der Statistik üblich, von *Signifikanz* zu sprechen, wenn die Irrtumswahrscheinlichkeit α höchstens 5% ist. Man sagt dann auch, man habe auf dem *5%-Signifikanzniveau* entschieden.
Diese Konvention gilt eventuell nicht, wenn es um sehr wichtige Entscheidungen geht, beispielsweise um Menschenleben. Dann versucht man oft auf dem 1%-Niveau zu entscheiden.
Merken wir uns:

Bei einem statistischen Test sprechen wir (üblicherweise) von *Signifikanz*, wenn die Irrtumswahrscheinlichkeit $\alpha \leq 5\%$ ist.

Aufgabe 4.1

Wir ändern den besprochenen Qualitätstest wie folgt ab: Die Sendung wird abgelehnt, wenn mehr als ein Dutzend defekter Knallkörper in die Stichprobe geraten. Die neue Irrtumswahrscheinlichkeit sei α‘.

a

Entscheiden Sie, ohne zu rechnen: Ist α‘ grösser oder kleiner als α?

b

Berechnen Sie α‘.

c

Ermitteln Sie α‘ aufgrund der Computersimulation in Kap. 2.2, Seite 2.6.

d

Sind auch beim neuen Entscheidungsschema signifikante Entscheide möglich?

4.2 Der Fehler 2. Art - die Macht des Tests

Einen Fehler 2. Art begehen wir bekanntlich bei unserem Qualitätstest, wenn wir nicht merken, dass die Sendung nicht in Ordnung ist und sie also irrtümlicherweise akzeptieren. Wie gross ist die Wahrscheinlichkeit β für diesen Irrtum?

Im Urnenmodell lautet die Frage: Wie gross ist die Wahrscheinlichkeit β, dass in die Stichprobe 13 oder weniger schwarze Kugeln geraten, obwohl die Grundgesamtheit mehr als 4000 schwarze Kugeln enthält.

Diese Frage lässt sich *nicht generell* beantworten: Es kommt darauf an, wie viele schwarze Kugeln die Urne tatsächlich enthält. Wir können beispielsweise die Wahrscheinlichkeit berechnen, dass unser Test nicht merkt, dass die Urne 8000, statt nur 4000, schwarze Kugeln enthält. Das wollen wir im folgenden tun.

Berechnung von β

$$\begin{aligned}\beta &= P(\text{"höchstens 13 schwarze in der Stichprobe"})\\ &= P(\text{"0 oder 1 oder 2 oder .... oder 13 schwarze"})\\ &= P(\text{"0 schwarze"}) + P(\text{"1 schwarze"}) + ... + P(\text{"13 schwarze"})\end{aligned}$$

Schreiben wir die Summe mit Hilfe des Summenzeichens, ersetzen wir die Wahrscheinlichkeiten mit der hergeleiteten Formel und werten wir schliesslich den Term auf dem Taschenrechner aus, so erhalten wir:

$$\beta = \sum_{s=0}^{13} \frac{\binom{S}{s}\binom{N-S}{n-s}}{\binom{N}{n}} = 0.264 = 26.4\%$$

Dabei haben wir folgende Werte eingegeben:

- N = 100‘000
- S = 8000
- n = 200

Das Resultat liegt in der Nähe des Wertes von 25%, den wir mit der Computersimulation erhalten haben.

Die Wahrscheinlichkeit ist, wie schon früher erwähnt, unerwartet gross: In durchschnittlich rund einem Viertel der Tests würden wir *nicht* merken, dass die Qualität doppelt so schlecht ist, wie vereinbart. Wenn der Einkäufer sich dieses Risiko nicht leisten kann oder will, muss er seinen Test verbessern.

Die Macht des Tests

Wenn β die Wahrscheinlichkeit ist, dass wir *nicht* entdecken, dass die Sendung 8000 defekte Knallkörper enthält, so ist eben $1–\beta$ die Wahrscheinlichkeit, diesen Tatbestand aufzudecken:

$$1–\beta = 1–26.4\% = 73.6\%.$$

Man bezeichnet deshalb $1-\beta$ als die *Macht des Tests bei einem effektiven Ausschussanteil von 8% (8000 unter 100'000).*

Merken wir uns:

Unter der *Macht* eines Tests verstehen wir die Wahrscheinlichkeit $1–\beta$, mit welcher ein Fehler 2. Art verhindert wird. Sie hängt vom tatsächlichen Ausschussanteil ab.

Aufgabe 4.2

Machen Sie mindestens einen Vorschlag, wie man β beim ursprünglichen Qualitätstest verkleinern, bzw. wie man die Macht des Tests vergrössern könnte.