

China im Fieberwahn – Der Satz von BAYES

Seit November 2002 versetzt das *Schwere Akute Respiratorische Syndrom (SARS)* China in Angst und Schrecken. In der Region Guangdong, in welcher das Coronavirus erstmalig aufgetaucht ist, sind bereits nach 4 Monaten 200 der etwa 100 000 Einwohner infiziert.

Innerhalb kürzester Zeit wurde ein Massentestverfahren entwickelt, welches bei 80 % der Kranken „positiv“ ausfällt, d.h. es zeigt an, dass die Person an dieser Krankheit erkrankt ist. Leider liefert das Testverfahren aber auch bei 10 % der gesunden Menschen ein „positives“ Ergebnis.

Eine Person aus der Guangdong-Region wird mit diesem Testverfahren getestet; das Ergebnis ist „positiv“. Muss sich diese Person daraufhin ernstlich Sorgen um ihre Gesundheit machen?

Problemfelder

- 1) Erstelle mit Hilfe der Tabellenkalkulation je eine geeignete Vierfeldertafel mit den absoluten bzw. den relativen Häufigkeiten und berechne daraus die bedingte Wahrscheinlichkeit dafür, dass die Person Virenträger ist.
- 2) Erstelle zudem ein geeignetes Baumdiagramm (mithilfe der relativen Häufigkeiten). Berechne damit die Wahrscheinlichkeit, dass die Krankheit tatsächlich die Ursache für das „positive“ Testergebnis ist, d.h. dass die getestete Person wirklich krank ist.
- 3) Ein alternatives Testverfahren gibt bei nur 75 % der Kranken ein „positives“ Ergebnis, fällt aber nur bei 5 % der Gesunden „positiv“ aus. Welches Verfahren hältst du für zuverlässiger?
- 4) Schreibe einen Zeitungsartikel, in welchem du deine Erkenntnisse beschreibst und erklärst. (Es muss kein ‚Fachchinesisch‘ sein!)
- 5) Besorge dir aus Internet Daten der SARS-Ausbreitung in den Jahren 2002/03. Entwickle und löse anhand dieser Daten eigene Problemstellungen! (Variiere dazu die Vierfeldertafeln und das Baumdiagramm.)

Analyse

- Der Problemkomplex muss zunächst als ein zweistufiges Zufallsexperiment erkannt werden (1. Stufe: „krank“ vs. „gesund“; 2. Stufe „Testergebnis positiv“ vs. „Testergebnis negativ“).
- Es ist sinnvoll, die Vierfeldertafel zunächst mit absoluten Häufigkeiten zu gestalten, das Baumdiagramm direkt mit den angegebenen relativen Häufigkeiten.
- Die Vierfeldertafel der absoluten Häufigkeiten könnte z. B. so aussehen:

	Krank	Gesund	Summe
Test +	160		
Test -			
Summe	200		100 000

	Krank	Gesund	Summe
Test +	160	998	1158
Test -	40	98802	98842
Summe	200	99800	100 000

- Anknüpfungspunkte sind bei dieser Aufgabenstellung die bekannten Pfadrechenregeln. Außerdem können an rechnerisch einfachen Beispielen Aspekte der bedingten Wahrscheinlichkeit (Satz von BAYES) untersucht werden:

$$P(\text{krank} | \text{Test} +) = \frac{\text{Anzahl der positiv getesteten Kranken}}{\text{Anzahl der positiv Getesteten}}$$

Dies gilt sowohl für die Berechnung über die relativen Häufigkeiten im Baumdiagramm als auch für die absoluten und relativen Häufigkeiten in den Vierfeldertafeln.

Rechenblatt in CellSheet™ (TI-83)

BAY3	A	B	C
1	TEST	KRANK	GESUN
2	POS	160	998
3	NEG	40	98802
4			
5	SUMME	200	99800
6	KRANKLEBNPOS		
B5: =.002*100000			

Bild 1

BAY3	C	D	E
1	GESUND		SUMME
2	998		1158
3	98802		98842
4			100000
5	99800	100000	1
6			
E5: =If(D5=E4,1,0) (Menu)			

Bild 2

BAY3	A	B	C
2	POS	160	998
3	NEG	40	98802
4			
5	SUMME	200	99800
6	KRANKLEBNPOS		
7	.13817		
A7: =B2/E2			

Bild 3

BAY2	A	B	C
2		KRANK	
3	NEG	.002	POS
4	.2		.8
5	KR:NEG		KR:POS
6	4E-4		.0016
7			
A6: =B3*A4			

Bild 4

BAY2	D	E	F
2		GESUND	
3	POS	.998	POS
4	.01		.99
5	GE:POS		GE:NEG
6	.00998		.98802
7			
F6: =E3*F4			

Bild 5

BAY2	C	D	E
4	.8	.01	
5	KR:POS	GE:POS	
6	.0016	.00998	
7			
8	KRANKLEBNPOS	ITIV	
9	.13817		
C9: =C6/(C6+D6)			

Bild 6

Hinweise

- Die Umsetzung des Problems mit Hilfe einer Tabellenkalkulation mag zunächst fragwürdig erscheinen, die Zusammenhänge zwischen Baumdiagramm, Vierfeldertafeln und den einzelnen Angaben werden aber durch die Programmierung von den Schülern besser durchschaut.
- Außerdem können dieselbe Tabelle und das Baumdiagramm für vergleichbare Probleme genutzt werden. Die Schüler erkennen, dass unabhängig von der jeweiligen inhaltlichen Formulierung jedes Mal dieselbe Rechenmethode Anwendung findet. Der Vorteil der Tabellenkalkulationsnutzung liegt darin, dass man die Parameter in 3) leicht variieren und als Grundlage für die Diskussion über den Stichprobenumfang nutzen kann.

Auf jeden Fall sollten vor der Programmierung in CellSheet™ die Vierfeldertafeln und das Baumdiagramm bereits von Hand erstellt werden. Erst bei der Vertiefung und der damit einhergehenden Variation der Parameter ist der Rechner sinnvoll einsetzbar.

- In der Umsetzung der Vierfeldertafel der absoluten Häufigkeiten (Bild 1 bis 3) können durch eine *If*-Anweisung die Summen kontrolliert werden. Leider ist jedoch eine Berechnung innerhalb der Anweisung nur mit dem Voyage möglich, da mit dem TI-83 keine Zellbezüge verwendet werden können. Dazu ist die *cellIf*-Anweisung zu verwenden, welche in der Funktionsliste *F6* zu finden ist.
- Bild 4 bis 6 zeigen eine mögliche Darstellung des Baumdiagramms in Tabellenform.