Diabetes-Test

Blutspenden werden auf Diabetes untersucht, das mit 8 % in der Bevölkerung verbreitet ist. Dabei werden an Diabetes Erkrankte mit einer Wahrscheinlichkeit von 95 % erkannt, während 2 % als Diabetiker eingestuft werden, obwohl sie es nicht sind.

1. Baumdiagramm

a) Stellen Sie ein Baumdiagramm zu dem Sachverhalt auf.
b) Mit welcher Wahrscheinlichkeit lautet das Testergebnis "Diabetiker"?
c) Mit welcher Wahrscheinlichkeit ist ein als Diabetiker eingestufter Spender tatsächlich nicht erkrankt?

Tipps zu Aufgabe 1c:

Tipp I: Überlegen Sie, welche Anzahlen sich erwartbar ergeben, wenn 100 000 Personen getestet werden.

Mit Hilfe der Zahlen können Sie die Wahrscheinlichkeit zu b berechnen, wie üblich:

$$\frac{\text{Zahl der "günstigen" Fälle}}{\text{Zahl der "möglichen" Fälle}}.$$

Tipp II: Statt mit den absoluten Zahlen können Sie auch direkt mit den Wahrscheinlichkeiten rechnen, indem sie alle absoluten Zahlen wieder durch 100 000 dividieren.

Tipp III: Leiten Sie eine immer direkt anwendbare Formel her zur Berechnung der sogenannten "bedingten Wahrscheinlichkeit".

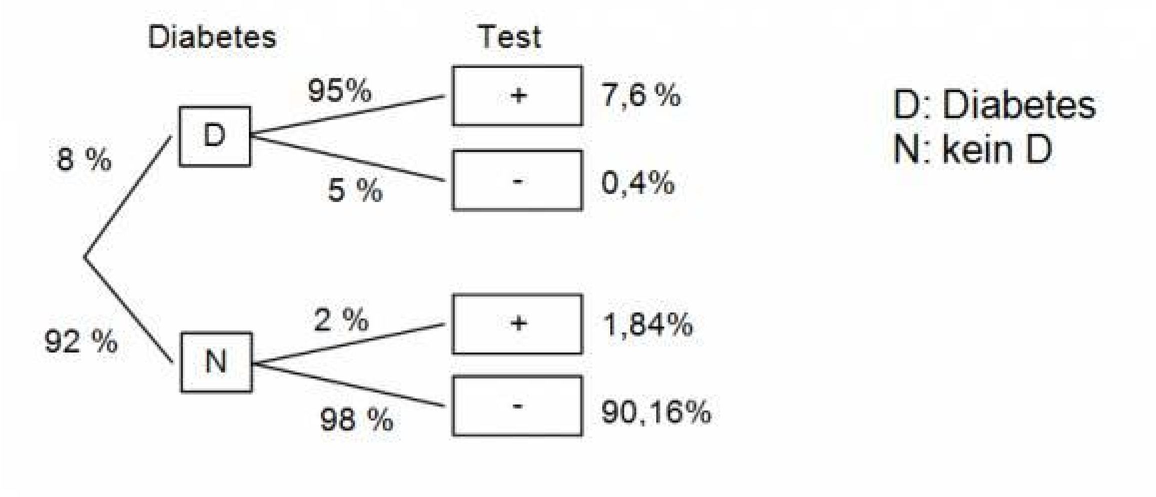
Information: Bedingte Wahrscheinlichkeiten schreibt man mit einem senkrechten Strich, wobei hinter dem Strich die Bedingung steht. Beispiel zur Fragestellung c P(Nichtdiabetiker|Test positiv); kurz P(N|T+). Gesprochen: "P von N unter der Bedingung T+". Die Klammer wird mit "von", der senkrechte Strich mit "unter der Bedingung" benannt.

2. Vierfeldertafel

a) Stellen Sie eine Vierfeldertafel zu dem Sachverhalt auf.
b) Suchen Sie in der Vierfeldertafel die Wahrscheinlichkeiten, die Sie zur Beantwortung der Fragen 1b und c benötigen.
c) Notieren Sie eine allgemeine Formel für die Berechnung einer bedingten Wahrscheinlichkeit mithilfe einer Vierfeldertafel.
d) Vergleichen Sie die Berechnungswege über das Baumdiagramm und die Vierfeldertafel. Benennen Sie Vor- und Nachteile für beide Vorgehensweisen.

1. Lösung mit einem Baumdiagramm und Einführungsempfehlung

a)



Hier zunächst die Lösungen in formaler Schreibweise:

b) Das Testergebnis "Diabetiker" bedeutet, der Test ist positiv ausgefallen. Im einen Fall trifft das Testergebnis zu, im anderen ist es falsch. Aber nach dem Test weiß niemand, welcher Fall zutrifft. Einen Test positiv erhalten 7,6 % bzw. 1,84 % der Getesteten.
P(T+) = 7,6 % + 1,84 % = 9,44 %
Der Test diagnostiziert 9,44 % als Diabetiker (obwohl nur 8 % Diabetiker sind!).

c) $P(N|T+) = \frac{1{,}84\,\%}{1{,}84\,\% + 7{,}6\,\%} = 19{,}5\,\%$

Von den Getesteten, deren Ergebnis "Diabetes" lautet, sind rund 20 % entgegen dem Testurteil nicht an Diabetes erkrankt.
Dieses Ergebnis überrascht die Schülerinnen und Schüler in der Regel, da der Test zunächst sehr sicher aussieht.

Zu den Tipps zu c)

Um auf die Lösung zu kommen, sollten Sie die Schülerinnen und Schüler zunächst selbst probieren und erkunden lassen.
Das geht mit den Tipps zum Umgang mit den absoluten Zahlen i. d. R. ganz gut.
Mit Tipp I ergeben sich rechts die absoluten Zahlen: 7600, 400, 1840 und 90 160. Ein positives Testergebnis erhalten 7600 + 1840 = 9440 Getestete. Davon sind 1840 aber nicht erkrankt. Mit den absoluten Zahlen können die Schülerinnen und Schüler die Lösung finden, da es sich um die bekannte Wahrscheinlichkeitsberechnung mit der Laplace-Regel handelt.

$$P(N|T+) = \frac{\text{Zahl der "günstigen" Fälle}}{\text{Zahl der "möglichen" Fälle}} = \frac{1840}{7600 + 1840} = \frac{1840}{9440} \approx 19{,}5\,\%$$

Die Division ergibt einen Prozentsatz von knapp 20 % wie oben.

Tipp II: Dividiert man in dem Bruch Zähler und (beide Summanden im) Nenner durch die "investierten" 100 000, so ergibt sich unmittelbar die oben notierte Bayes-Berechnung mit den Prozentsätzen.

$$P(N|T+) = \frac{\frac{1840}{100\,000}}{\frac{1840}{100\,000} + \frac{7600}{100\,000}} = \frac{1{,}84\,\%}{1{,}84\,\% + 7{,}6\,\%} \approx 19{,}5\,\%$$

Vielen Schülern und Schülerinnen ist die Division von Prozentsätzen nicht ganz geheuer. Die können immer auch über den Umweg mit den absoluten Zahlen zum richtigen Ergebnis kommen.

Tipp III:

$$P(N|T+) = \frac{P(N \text{ und } T+)}{P(T+)}$$

Im Nenner steht immer die Summe der beiden Pfadwahrscheinlichkeiten, die zur Bedingung gehören; hier die Wahrscheinlichkeiten der beiden Pfade, die zu T+ führen.
Im Zähler steht immer eine der beiden Wahrscheinlichkeiten aus dem Nenner, nämlich der untersuchte Fall, hier die Wahrscheinlichkeit zusätzlich kein Diabetes zu haben.
Diese vereinfachte Formel gilt immer für zweistufige Baumdiagramme mit jeweils nur zwei Ausfällen (die sich also in einer Vierfeldertafel darstellen lassen).

a)

	Test positiv T+	Test negativ T-	Summe
Diabetes D	7,6 %	0,4 %	8 %
kein Diabetes N	1,84 %	90,16 %	92 %
Summe	9,44 %	90,56 %	100 %

Ist schon ein Baumdiagramm vorhanden, so ergeben sich die vier Innenfelder der Vierfeldertafel direkt aus den vier Pfadwahrscheinlichkeiten im Baumdia-gramm (deshalb auch der Name). Die "Vierfeldertafel" müsste in der oben üblichen Form eigentlich "Vierfeldertafel mit Summenrand" heißen! Da das ein sehr umständlicher Name ist, bleibt es bei der verkürzten Variante, gemeint ist aber immer die lange. Der Rand ist einfach als Summe der Spalten bzw. Zeilen zu ergänzen. Rechts unten muss sich immer (bis auf kleine Rundungsabweichungen) 100 % als Summe der Summenspalte bzw. -zeile ergeben, denn in beiden Fällen sind alle Fälle, also 100 %, des Zufallsversuchs erfasst.

b) P(T+) = 7,6 % + 1,84 % = 9,44 % ist hier direkt als Summenwahrscheinlichkeit unten ablesbar. In P(N|T+) = $\frac{1{,}84\,\%}{1{,}84\,\% + 7{,}6\,\%} = \frac{1{,}84\,\%}{9{,}44\,\%}$ ≈ 19,5 % sind Zähler (1,84 %) und Nenner (9,44 %) direkt in der Vierfeldertafel abzulesen.

c) Allgemeine Regel für die Berechnung bedingter Wahrscheinlichkeiten mit Hilfe einer Vierfeldertafel:

P(N|T+) = $\frac{P(N \text{ und } T+)}{P(T+)}$

Im Nenner steht immer die Summe der beiden Wahrscheinlichkeiten, die zur Bedingung gehören; hier die Summe zu T+ (9,44 %).
Im Zähler steht immer eine der beiden Wahrscheinlichkeiten aus der Nennersumme, nämlich der untersuchte Fall; hier die Wahrscheinlichkeit zusätzlich kein Diabetes zu haben (1,84 %).
Diese vereinfachte Formel gilt immer für Vierfeldertafeln.

d) Wenn die Vierfeldertafel erstellt ist, dann ist die Berechnung der bedingten Wahrscheinlichkeit einfach: Man benutzt nur die Zahlen der Bedingungszeile oder -spalte; hier die Spalte zu T+. Daraus wird der Zähler (zu kein Diabetes N) entnommen und der Nenner (die Summe zu T+).
Allerdings ist die Erstellung einer Vierfeldertafel komplizierter, wenn Daten wie hier gegeben sind, denn sie passen direkt zur Aufstellung eines Baumdia-gramms.
Hier ist die Erstellung des Baumdiagramms mit den gegebenen Daten leicht. Die Berechnung der bedingten Wahrscheinlichkeit muss mit den Pfadwahrscheinlichkeiten erfolgen. Dazu muss die Summe im Nenner zuerst noch gebildet werden.

Nachtrag zur Erstellung der Vierfeldertafel
aus den anfangs gegebenen Daten

- Direkt gegeben ist die Zeilensumme zu D mit 8 %, damit auch direkt die andere zu N mit 92 %.
- Von den 8 % Diabetesfällen werden vom Test 95 % positiv getestet. Das Produkt $8\ \% \cdot 95\ \% = 0{,}08 \cdot 0{,}95 = 0{,}076 = 7{,}6\ \%$ kann in dem Feld zur Zeile D und zur Spalte T+ notiert werden.
- Von den 92 % Nicht-Diabetesfällen erhalten 2 % irrtümlich einen positiven Test. Das Produkt $92\ \% \cdot 2\ \% = 0{,}92 \cdot 0{,}02 = 0{,}0184 = 1{,}84\ \%$ kann in dem Feld zur Zeile N und zur Spalte T+ eingetragen werden.
- Die restlichen zwei Innenfelder ergeben sich durch Differenzbildung: $8\ \% - 7{,}6\ \% = 0{,}4\ \%$ und $92\ \% - 1{,}84\ \% = 90{,}16\ \%$.
- Die beiden Summenfelder unten lauten dann $7{,}6\ \% + 1{,}84\ \% = 9{,}44\ \%$ und $0{,}4\ \% + 90{,}16\ \% = 90{,}56\ \%$.
- Die Summenzeile und die Summenspalte ergänzen sich korrekt zu 100 %.

ZUM ARBEITSBLATT

Mit dem Arbeitsblatt geht es um ein Stück Lehrerfortbildung: Wie kann die bedingte Wahrscheinlichkeit eingeführt werden – und zwar berechnet mit Hilfe eines Baumdiagramms oder einer Vierfeldertafel? Zudem wird gezeigt, wie man Vierfeldertafeln aus Textdaten erstellt oder leichter, wenn schon ein Baumdiagramm vorhanden ist.
Der Mathematikinhalt ist in der Einführungsphase in NRW seit diesem Schuljahr verpflichtend, in anderen Ländern früher oder später auch.
Eingeführt wird hier am Thema Testunsicherheit medizinischer Prüfungen, eine relevante Problematik, die jede und jeden betreffen kann und über die man informiert sein sollte, ehe man sich auf so einen Test einlässt, was ich – so informiert – trotzdem rate!