Zusammenhänge nominalskalierter Daten – Einheitsquadrate mit *ProVis*

VICTORIA DÖLLER, WIEN; STEFAN GÖTZ, WIEN

Der neue Lehrplan Mathematik für die Sekundarstufe I sieht unter anderem „Darstellen, Ergänzen und Interpretieren von absoluten und relativen Häufigkeiten in Kreuztabellen, insbesondere in Vierfeldertafeln“ vor. Sogenannte Einheitsquadrate sind zentrale Werkzeuge zur Untersuchung von Fragestellungen in diesem Zusammenhang im Unterricht. Aufgrund des Konstruktionsaufwands ist ihr Einsatz per Hand aufwändig, das kann aber durch eine geeignete Software behoben werden. Im Beitrag werden neben fachlichen Hinweisen didaktische Anforderungen an eine solche Software erörtert und die darauf basierende Software *ProVis* – Probability Visualized – wird anhand von Beispielen vorgestellt. Mit ihr können auch Baumdiagramme erzeugt werden, wie sie zur Veranschaulichung von (zweistufigen) Zufallsexperimenten jetzt auch in der Sekundarstufe I gebraucht werden.

1. Einleitung

1.1 Motivation

Im neuen Lehrplan Mathematik für die Sekundarstufe I ist in der achten Schulstufe „Darstellen, Ergänzen und Interpretieren von absoluten und relativen Häufigkeiten in Kreuztabellen, insbesondere in Vierfeldertafeln“ vorgesehen (RIS 2024). In Vierfeldertafeln werden üblicherweise Häufigkeiten nominalskalierter Daten von zwei Merkmalen mit jeweils zwei Ausprägungen dargestellt. Ein typisches Beispiel ist Geschlecht (Ausprägungen männlich – weiblich) und Rauchverhalten (Rauchen – Nichtrauchen). Es gibt dann in der Regel rauchende Männer, nichtrauchende Männer, rauchende Frauen und nichtrauchende Frauen, deren Anzahlen in einer Stichprobe in eine Vierfeldertafel eingetragen werden. Was eine:n meist interessiert an einem solchen Datensatz ist die Frage, ob das eine Merkmal vom anderen abhängig ist oder nicht:

> „**Männer und Frauen** gleichen sich in ihrem Rauchverhalten im Verlauf der Jahrzehnte einander an. Frauen rauchen allerdings nach wie vor etwas seltener […] und im Durchschnitt weniger Zigaretten pro Tag als Männer. Laut ATHIS 2019 rauchen 18 Prozent der Frauen und 24 Prozent der Männer täglich. Täglich rauchende Österreicher rauchen im Durchschnitt 17 Zigaretten, täglich rauchende Österreicherinnen rund 13 Zigaretten pro Tag.“

(Schmutterer 2022, S. 11, Hervorhebung im Original; ATHIS steht für „Austrian Health Information Survey“, die Österreichische Gesundheitsbefragung). Um den Grad der Abhängigkeit zu quantifizieren, ist der Begriff der „bedingten relativen Häufigkeit“ unvermeidlich. Seine Einführung im Mathematikunterricht der Sekundarstufe I kann als Vorbereitung für den Begriff der bedingten Wahrscheinlichkeit in der Sekundarstufe II dienen. Mit dem Satz der totalen Wahrscheinlichkeit erschließt sich dann das BAYES’sche Theorem (zumindest in seiner einfachen Form) auch für relative Häufigkeiten (Abschnitt 3.).

Der Grad der Abhängigkeit kann aber auch graphisch dargestellt werden, dazu dienen sogenannte „Einheitsquadrate“, die aus den Vierfeldertafeln gewonnen werden können. Ihre Konstruktion ist händisch ein wenig mühsam, das in diesem Artikel vorgestellte Programm *ProVis* erledigt diese Aufgabe auf Knopfdruck.

Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), Heft 56, 2024, S. 21 - 41.

1.2 Ein Beispiel

Teilzeit als Beschäftigungsausmaß wird in der Öffentlichkeit in Österreich derzeit heftig diskutiert. Im Standard online vom 6.2.2024 findet sich dazu (Dang 2024):

> „Ob Viertagewoche, Teilzeit oder Überstunden: Die Arbeitszeit ist zu einem Debattenthema geworden. Der Trend in Richtung weniger Wochenstunden zeigt sich auch in den Ergebnissen einer neuen Auswertung der Jobplattform Karriere.at. Mit 76 Prozent entfielen demnach im Vorjahr zwar die meisten Inserate auf Vollzeitstellen – diese werden aber jedes Jahr weniger. Zum Vergleich: 2021 waren noch 82 Prozent aller Positionen in Vollzeit gesucht, 2022 waren es schon nurmehr 78 Prozent. Analysiert wurden rund 500.000 auf der Plattform ausgeschriebene Stellen.“

Tabelle 1 zeigt dazu eine Vierfeldertafel mit Daten der Statistik Austria (https://www.statistik.at/fileadmin/pages/263/11_Teilzeitarbeit_Teilzeitquote_2022.ods, Tabelle 5), die die Anzahl der Erwerbstätigen in Tausend nach Geschlecht und Beschäftigungsausmaß im Bereich Erziehung und Unterricht 2022 in Österreich wiedergibt.

Tab. 1: Anzahl (in Tausend, gerundet) der voll- und teilzeitbeschäftigten Personen im Bereich Erziehung und Unterricht im Jahr 2022 in Österreich

	männlich (M)	weiblich (W)
Vollzeit (V)	60	99
Teilzeit (T)	23	98

Wir leiten daraus folgende Fragen ab:

Arbeiten mehr Frauen als Männer in Vollzeit in diesem Bereich? – In absoluten Zahlen lautet die Antwort natürlich „Ja!“.

Heißt das (auch), dass Frauen eher in Vollzeit arbeiten?

Um das Verhältnis Geschlecht – Beschäftigungsausmaß in dieser Branche 2022 näher zu untersuchen – und zwar mit elementaren Mitteln der Sekundarstufe I! –, wollen wir im nächsten Schritt die Daten in Tabelle 1 graphisch darstellen.

Es sei zuvor noch angemerkt, dass das Arbeiten mit realen Daten unbedingt in den Statistikunterricht (der Sekundarstufen) integriert gehört (Götz & Döller 2020). Das Internet ermöglicht den einfachen Zugriff auf eine Fülle von zu analysierenden realen aktuellen Daten.

2. Darstellung und Auswertung nominalskalierter zweidimensionaler Daten mit zwei Ausprägungen

Die Daten (in Tausend, gerundet) von Tabelle 1 in Abschnitt 1.2 werden in Abbildung 1 mit einem mit dem Programm *ProVis* erstellten (wie alle anderen Einheitsquadrate auch in diesem Artikel) sogenannten Einheitsquadrat dargestellt (Eichler & Vogel 2013, S. 81).

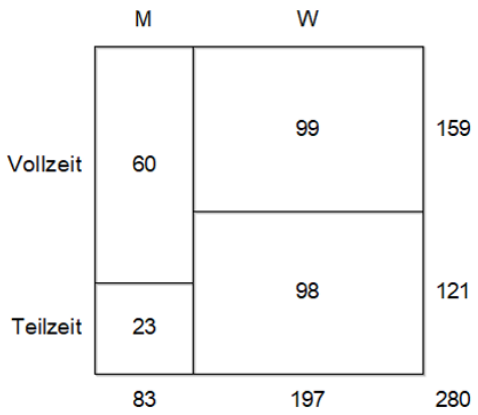


Abb. 1: Ein Einheitsquadrat mit den Daten von Tabelle 1.

Die absoluten Häufigkeiten aus Tabelle 1 werden als Flächeninhalte von (Teil-)Rechtecken dargestellt. Die Spaltensummen geben über die Anzahlen das Geschlecht betreffend Auskunft (Abbildung 1 unten), die Zeilensummen beziehen sich auf das Beschäftigungsausmaß (Abbildung 1 rechts). Die Gesamtanzahl der Beteiligten findet sich rechts unten in Abbildung 1. Für die Konstruktion eines solchen Einheitsquadrats erhebt sich die Frage: Wie werden die Seitenverhältnisse der (Teil-)Rechtecke bestimmt?

Die inhaltliche Beantwortung der zweiten Frage in Abschnitt 1.2 kann auf das Verhältnis der Anzahlen von teilzeit- zu vollzeitarbeitenden Frauen abzielen. Dieses Verhältnis wird mit dem analogen bei den Männern verglichen. Dabei ist der Anteil an der Grundgesamtheit jeweils ausschlaggebend. Die Grundgesamtheit wird zur Einheit, auf die die anderen Werte bezogen werden. So ergibt sich aus dem Kontext heraus der Übergang zu relativen Häufigkeiten.

Formal repräsentiert das Quadrat in Abbildung 1 die Grundgesamtheit als Einheit. Das Einheitsquadrat sehen wir nun als Quadrat mit Seitenlänge 1. Die Flächeninhalte der Teilrechtecke des Quadrats entsprechen den relativen UND-Häufigkeiten (konjunktive Häufigkeiten). Die senkrechte Unterteilung des Quadrats richtet sich nach den relativen Anteilen des sogenannten dominierenden Merkmals, in Abbildung 1 ist das das Geschlecht. Für die waagrechten Unterteilungen des Quadrats in Abbildung 1 muss die gesuchte Höhe des jeweiligen Teilrechtecks multipliziert mit dem relativen Anteil der Männer bzw. Frauen die entsprechende relative konjunktive Häufigkeit ergeben. Das bedeutet zum Beispiel für die Höhe des linken oberen Teilrechtecks des Quadrats in Abbildung 1: Höhe $\cdot\, h(M) = h(M \cap V)$ bzw.

$$\text{Höhe} = \frac{h(M \cap V)}{h(M)} =: h(V|M),$$

wobei *h* für die (absolute oder) relative Häufigkeit steht (Eichler & Vogel 2013, S. 82). Damit legen wir eine neue Grundgesamtheit fest, hier: Männer in der linken bzw. Frauen in der rechten Spalte des Quadrats in Abbildung 1 als Ausprägungen des dominierenden Merkmals „Geschlecht“. Diese Festlegung motiviert die Interpretation der Höhen der Teilrechtecke im Quadrat von Abbildung 1 als *relative bedingte Häufigkeiten* $h(.|.)$. Sie kann in der Sekundarstufe II als Konzept der bedingten Wahrscheinlichkeit weitergeführt werden. Ganz wichtig sind mögliche Verbalisierungen dieses Konzepts wie zum Beispiel „wie viele … von denen, die …“ oder „… unter jenen, die …“.

Wir können nun zwei unterschiedliche Vergleiche zur Beantwortung der ersten Frage in Abschnitt 1.2 anstellen, der Kontext leitet uns jeweils. Erstens ist der Anteil der Vollzeitarbeitenden unter den Männern $h(V \mid M) = \frac{60}{83} = 0.723$ (gerundet). Der entsprechende Anteil bei den Frauen beträgt $h(V|W) = \frac{99}{197} = 0.503$ (gerundet), ist also deutlich geringer. Beziehen wir die Häufigkeiten von Männern bzw. Frauen auf die ursprüngliche Grundgesamtheit, so landen wir zweitens beim gängigeren

Vergleich: der Anteil der vollzeitarbeitenden Männer in diesem Bereich ist $h(V \cap M) = \frac{60}{280} = 0.214$, jener der Frauen ist $h(V \cap W) = \frac{99}{280} = 0.354$ (beides gerundet). Jetzt lautet der Befund mehr Frauen als Männer arbeiten Vollzeit in diesem Bereich. – Was stimmt denn nun? Eine Analyse des Zustandekommens der verschiedenen Anteile klärt die aufgetretenen Differenzen (wie schon festgestellt arbeiten absolut deutlich mehr Frauen als Männer im Bereich Erziehung und Unterricht), ein fairer Vergleich wird wohl auf die bedingten relativen Häufigkeiten hinauslaufen.

In Abbildung 2 (die dort angeführten Ergebnisse sind ebenfalls mit *ProVis* berechnet worden) ist das Einheitsquadrat von Abbildung 1 mit relativen Häufigkeiten zu sehen. Wie oben schon festgestellt, treten bedingte relative Häufigkeiten als Seitenlängen der Teilrechtecke des Quadrats auf, konjunktive relative Häufigkeiten als Flächeninhalte derselben.

Für die Beantwortung der zweiten Frage in Abschnitt 1.2 ändern wir diese unter dem Eindruck des eben Analysierten ab: Heißt das (auch), dass Frauen eher in Teilzeit arbeiten? – Wieder gibt es zwei unterschiedliche Antworten: Wenn wir einerseits $h(V|M) = 0.723$ mit $h(V|W) = 0.503$ von eben miteinander vergleichen, dann lautet die Antwort „Ja!“. Denn die Gegenwahrscheinlichkeiten $h(T|M) = 1 - h(V|M) = 0.277$ bzw. $h(T|W) = 0.497$ sprechen eine eindeutige Sprache. Andererseits unterscheidet sich $h(V|W)$ von $h(T|W)$, wie wir eben gesehen haben, kaum. Daher ist unsere Antwort „Nein!“.

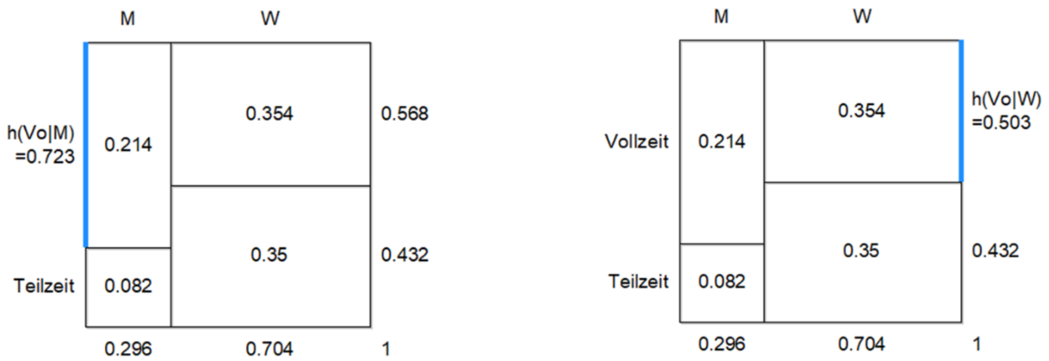


Abb. 2: Konjunktive und bedingte relative Häufigkeiten im Vergleich („Vo“ steht für „Vollzeit“).

Diese unterschiedlichen Herangehensweisen zeigen, wie vielschichtig das Anwenden von Mathematik in realen Kontexten sein kann:

> „Erscheinungen der Welt um uns, die uns alle angehen oder angehen sollten, aus Natur, Gesellschaft und Kultur, in einer spezifischen Art wahrzunehmen und zu verstehen, […]“

lautet die erste von drei Grunderfahrungen, die nach Heinrich WINTER der Mathematikunterricht ermöglichen sollte (Winter 1995, S. 37). Wir haben gesehen, dass sowohl das Analysieren von als auch das Entscheiden in realen Situationen nicht unbedingt eindeutige Lösungen mit sich bringen muss. Vor allem in Hinblick auf das Entscheiden ist das eine wichtige Einsicht. Entscheidungen können von den getroffenen Voraussetzungen oder gewählten Analysemethoden abhängen, die oft bei Bekanntgabe einer Entscheidung nicht transparent gemacht werden. Ein berühmtes Beispiel wird im Zusammenhang mit dem Paradoxon von SIMPSON immer wieder genannt: die Zulassungszahlen von Frauen und Männern an der University of California, Berkeley (Bickel et al. 1975). Je nachdem, ob man die Aufnahmeraten bei den einzelnen Fakultäten (Departments) miteinander verglich oder jene der gesamten Universität, kam es zu unterschiedlichen Ergebnissen und auch Schlüssen daraus.

3. Lernziele

Wir wollen nun die im vorigen Abschnitt beschriebenen Konzepte systematisieren, um relevante Lernziele zu definieren. Zunächst geht es um die Interpretation von Seitenlängen und Flächeninhalten im Einheitsquadrat. In Abbildung 3 stellt der markierte Bereich, die „linke Spalte“, die neue, eingeschränkte Grundgesamtheit dar. Sein Flächeninhalt gibt den Anteil einer Ausprägung des dominierenden Merkmales (hier: das Geschlecht, siehe Abschnitt 2.) wieder. Ebenfalls aus Abbildung 3 lesen wir heraus, dass der Summe der Flächeninhalte der Teilrechtecke einer „Zeile“ die (relative) Häufigkeit einer Ausprägung des nicht dominierenden Merkmals (hier: Beschäftigungsausmaß, siehe Abschnitt 2.) entspricht (Eichler & Vogel 2013, S. 81 f.).

Aus Abbildung 3 erkennen wir auch, woraus sich die die relative Häufigkeit einer Ausprägung des nicht dominierenden Merkmals formal zusammensetzt. Für die Ausprägung „Teilzeit“ gilt zum Beispiel
$$h(T) = h(T \cap M) + h(T \cap W) = h(T|M) \cdot h(M) + h(T|W) \cdot h(W).$$

Unschwer ist die Struktur des *Satzes von der totalen Wahrscheinlichkeit* zu erkennen.

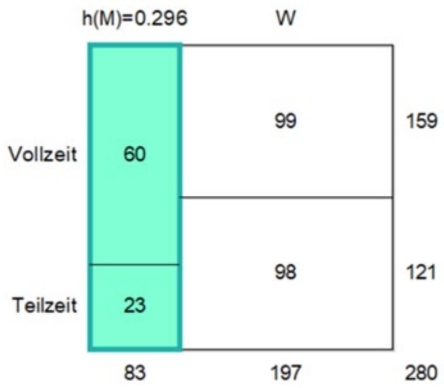


Abb. 3: Die „Spalte“ eines Einheitsquadrats.

Setzt man nun zu guter Letzt den Flächeninhalt eines Teilrechteckes ins Verhältnis zur Gesamtfläche der zugehörigen „Zeile“, so landet man beim *Satz von* BAYES für relative Häufigkeiten (Eichler & Vogel 2013, S. 82):

$$h(M|T) = \frac{h(M \cap T)}{h(T)} = \frac{h(T|M) \cdot h(M)}{h(T)}.$$

Die Einheitsquadrate von Abbildung 4 illustrieren das eben Gesagte.

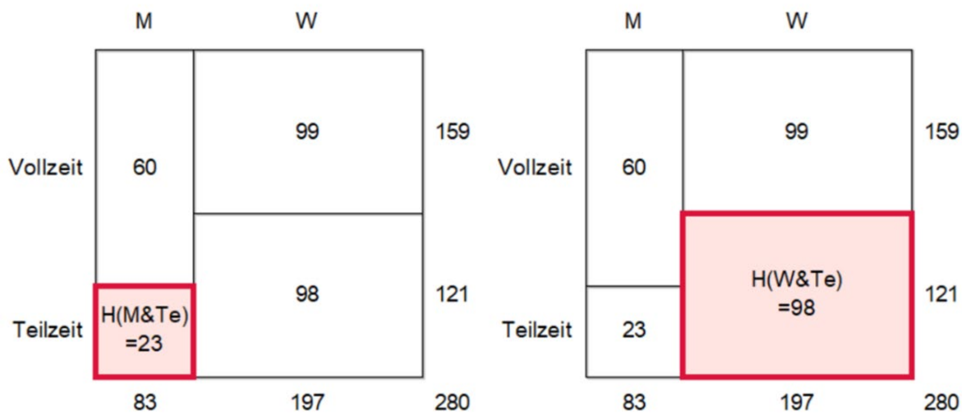


Abb. 4: Dividend und Divisor im Satz von BAYES („Te“ steht für „Teilzeit“).

Natürlich ist nicht gesagt, dass das Geschlecht das dominierende Merkmal in der Vierfeldertafel (Tabelle 1) sein muss. Auch dem Beschäftigungsausmaß (Abschnitt 2.) kann man diese Rolle zuschreiben. Wir nennen das daraus resultierende Einheitsquadrat das *transponierte Einheitsquadrat* (Döller & Götz 2021). Abbildung 5 zeigt das graphische Resultat dieser Neudefinition mit absoluten Häufigkeiten.

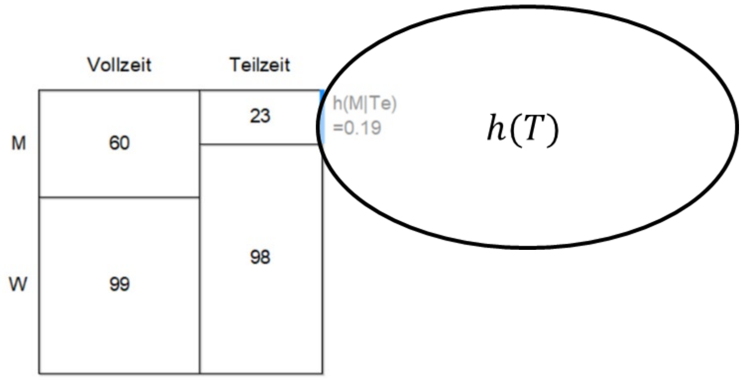


Abb. 5: Einheitsquadrat mit Beschäftigungsausmaß als dominierendes Merkmal („Te“ steht für „Teilzeit“).

ProVis liefert (Abbildung 5) $h(M|T) = 0.19 < h(W|T) = 0.81$ (gerundet), das bedeutet, dass der Anteil der Frauen unter den Teilzeitarbeitenden wesentlich größer ist als der Anteil der Männer. Wir sehen eine weitere Antwortmöglichkeit auf die zweite Frage in Abschnitt 1.2.

Kommen wir noch einmal auf Abbildung 2 zurück. Uns interessiert der Unterschied zwischen $h(V|M)$ und $h(V|W)$, wobei wir schon $h(V|M) > h(V|W)$ festgestellt haben. Graphisch äußert sich dieser Unterschied als Stufe im Einheitsquadrat (Eichler & Vogel 2013, S. 82 ff.): Abbildung 6.

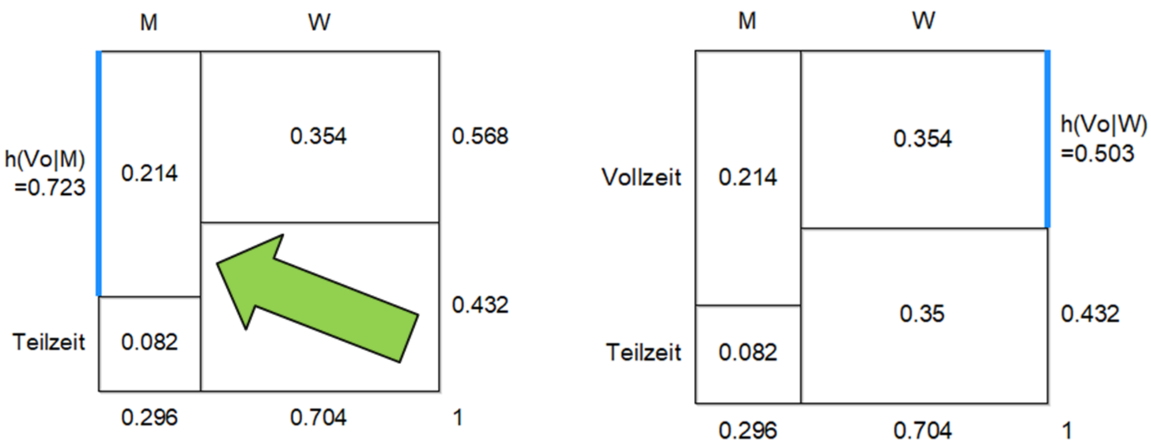


Abb. 6: Der graphische Unterschied von $h(V|M)$ und $h(V|W)$ im Einheitsquadrat („Vo“ steht für „Vollzeit“).

Dieser Unterschied, graphisch die Höhe der Stufe, wird das *Assoziationsmaß* $\mathcal{A}$ genannt. Es ist ein elementares Maß für die (Un-)Abhängigkeit zweier Merkmale, hier: Geschlecht und Beschäftigungsausmaß im Bereich Erziehung und Unterricht 2022 in Österreich. Das Programm *ProVis* berechnet es auf Knopfdruck: Abbildung 7. Konkret ist also $\mathcal{A}_E = h(V|M) - h(V|W) = 0.22$ (gerundet) in dem in Rede stehenden Beispiel.

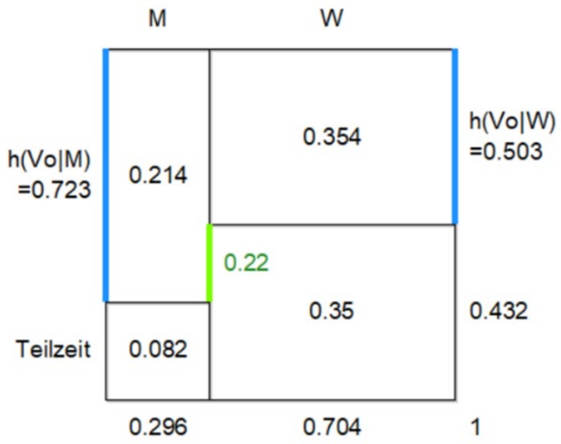


Abb. 7: Das Assoziationsmaß im Einheitsquadrat („Vo“ steht für „Vollzeit“).

Je höher die Stufe ist, desto größer ist die Abhängigkeit zwischen den beiden Merkmalen (Eichler & Vogel 2013, S. 84). Sie kann betragsmäßig höchstens 1 werden.

Abbildung 8 zeigt zusammenfassend die (relativen) Häufigkeiten, die im Einheitsquadrat dargestellt sind (Döller & Götz 2022, p. 489). Die Unterscheidung zwischen konjunktiven und bedingten Häufigkeiten wird graphisch durch verschiedene Darstellungsformen – Rechtecksfläche bzw. Rechtecksseite – unterstützt. Die Zuordnung, Interpretation und Anwendung dieser Häufigkeiten sind mögliche Lernziele zum in Abschnitt 1.1 zitierten Lehrplanabschnitt.

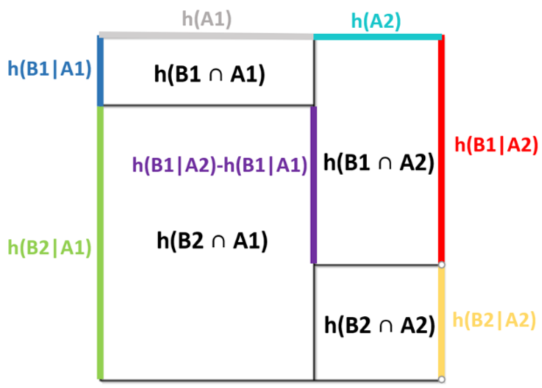


Abb. 8: Die Häufigkeiten im Einheitsquadrat.

4. Inhalte für Fortgeschrittene

Für das in Abbildung 5 zu sehende transponierte Einheitsquadrat $\hat{E}$ kann natürlich auch das Assoziationsmaß (Abschnitt 3.) (mit *ProVis*) berechnet werden: Abbildung 9. Es unterscheidet sich im Allgemeinen von jenem des ursprünglichen Einheitsquadrates E (Abbildung 7).

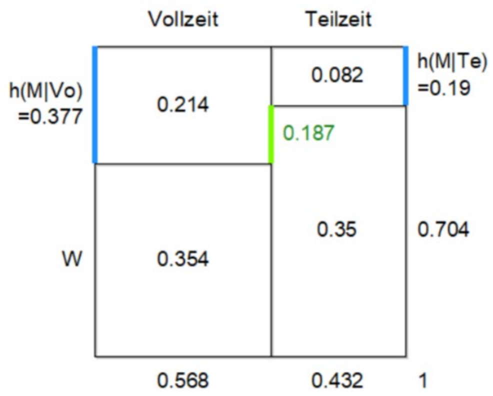


Abb. 9: Das transponierte Einheitsquadrat $\hat{E}$ mit dem Assoziationsmaß.

Es gilt

$$\mathcal{A}_E \cdot h(M) \cdot h(W) = \mathcal{A}_{\hat{E}} \cdot h(V) \cdot h(T)$$

für das Assoziationsmaß $\mathcal{A}_E$ des ursprünglichen Einheitsquadrats und das Assoziationsmaß $\mathcal{A}_{\hat{E}}$ des transponierten Einheitsquadrats (Döller 2020, S. 21). Es sind hier also nur Randhäufigkeiten im Spiel. Ein Beweis gelingt durch Einsetzen und Nachrechnen. Die zugehörige konkrete Rechnung lautet

$$0.22 \cdot 0.296 \cdot 0704 = 0.187 \cdot 0.568 \cdot 0.432$$

mit den relativen Häufigkeiten aus Abbildung 7 und Abbildung 9.

In der Statistik gibt es unterschiedliche Methoden, Abhängigkeiten von Merkmalen zu quantifizieren. Ein sehr bekanntes Verfahren ist der Chi-Quadrat-Unabhängigkeitstest (für die Vierfeldertafel, z. B. https://de.wikipedia.org/wiki/Chi-Quadrat-Test).

Die zugehörige Testgröße X^2, deren Verteilung man wenigstens asymptotisch kennt, lautet

$$X^2 = \frac{n \cdot (a \cdot d - b \cdot c)^2}{(a+c) \cdot (b+d) \cdot (a+b) \cdot (c+d)}.$$

Die Variablen a, b, c, d und n sind in Tabelle 2 erklärt, die zusätzlich zu Tabelle 1 noch die Randsummen aufweist. Die Testgröße X^2 ist näherungsweise chi-quadrat-verteilt mit einem Freiheitsgrad (ebd.).

Tab. 2: Anzahl (in Tausend, gerundet) der voll- und teilzeitbeschäftigten Personen im Bereich Erziehung und Unterricht im Jahr 2022 in Österreich inklusive Randsummen

	männlich (M)	weiblich (W)	Summe
Vollzeit (V)	60 (*a*)	99 (*b*)	159
Teilzeit (T)	23 (*c*)	98 (*d*)	121
Summe	83	197	280 (*n*)

Für die Realisierung von X^2 (gerundet) erhalten wir mit den Daten (Randsummen) von Tabelle 2

$$x^2 = \frac{280 \cdot (60 \cdot 98 - 99 \cdot 23)^2}{83 \cdot 197 \cdot 159 \cdot 121} = 11.55 > \mathcal{X}^2_{1;0.999} = 10.83.$$

Dabei bezeichnet $\mathcal{X}^2_{1;0.999}$ das 0.999-Quantil der Chi-Quadrat-Verteilung mit einem Freiheitsgrad (ebd.). Das heißt, die Abhängigkeit von Geschlecht und Beschäftigungsausmaß war 2022 im Bereich Erziehung und Unterricht in Österreich höchstsignifikant!

Es gilt nach Döller & Götz (2021): $n \cdot \mathcal{A}_E \cdot \mathcal{A}_{\hat{E}} = X^2$.

Noch ein (Un-)Abhängigkeitsmaß, und zwar wiederum ein elementares, ist die *odds ratio* q. Mit den in Abbildung 10 erklärten Variablenbezeichnungen lautet die Definition der odds ratio

$$q := \frac{\alpha}{\beta} : \frac{\gamma}{\delta} = \frac{\alpha}{1-\alpha} : \frac{\gamma}{1-\gamma} = \frac{\alpha}{1-\alpha} \cdot \frac{1-\gamma}{\gamma}.$$

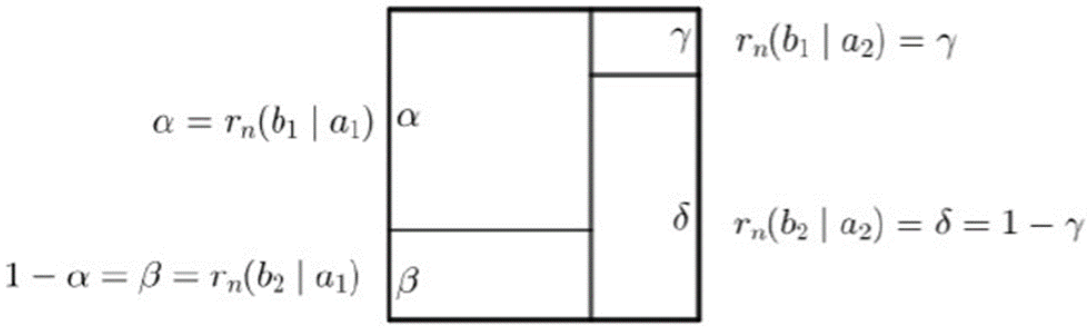


Abb. 10: Zur Definition der odds ratio *q*.

Mit den Werten von Abbildung 7 ergibt sich konkret $q = \frac{0.723}{1-0.723} \cdot \frac{1-0.503}{0.503} = 2.58$ (gerundet) für das in Rede stehende Beispiel. Auch hier gilt: Je größer q ist, desto größer ist die Abhängigkeit der beiden Merkmale voneinander (Eichler & Vogel 2011, S. 55).

Lustigerweise gilt allgemein $q_E = q_{\hat{E}}$ (zeigen durch Nachrechnen)!

5. Zwei Aufgaben

Zum Kolloquium der Lehrveranstaltung „Schulmathematik Stochastik“ aus dem Sommersemester 2023, gehalten vom Zweitautor, wurde u. a. gegeben:

> „Für den Lehrlingsgesundheitsbericht aus 2023 (Health Behaviour in School-aged Children (HBSC): vgl. https://www.sozialministerium.at/Themen/Gesundheit/Kinder--und-Jugendgesundheit.html) wurden insgesamt 3062 Lehrlinge (männlich – weiblich) in Österreich zu ihrem aktuellen Sportverhalten („mehrSport“ bedeutet mindestens einmal pro Woche Sport) befragt. Das gegebene Einheitsquadrat schlüsselt die erhaltenen Daten graphisch auf: Abbildung 11.

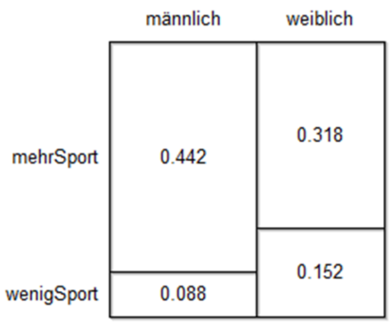


Abb. 11: Einheitsquadrat zu einer Prüfungsaufgabe.

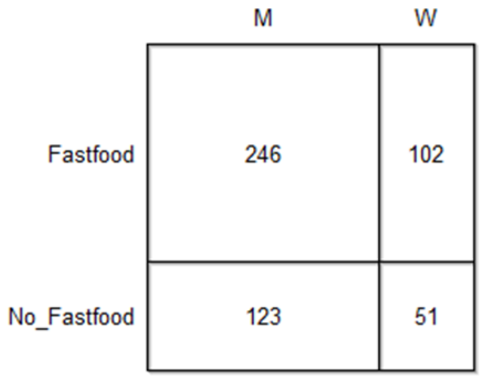
1. Bestimmen Sie den relativen Anteil der mehr Sport Betreibenden unter den weiblichen Lehrlingen in dieser Stichprobe!

2. Wie groß ist der relative Anteil der Lehrlinge, die wenig Sport betreiben, in dieser Stichprobe?
3. Berechnen Sie den relativen Anteil der männlichen Lehrlinge unter jenen, die mehr Sport betreiben, in dieser Stichprobe!
4. Erklären Sie, wie der Satz von Bayes mit Hilfe dieser Aufgabe vorbereitet werden kann!“

Die Angabe einer anderen Aufgabe mit einem analogen Kontext und fingierten Daten lautete:

> „Die Schüler:innen einer HTL wurden über ihren Konsum von Fastfood befragt. Unter $n = 522$ Befragten gaben 153 an, weiblich (*W*) zu sein. Der Rest nannte männlich (*M*) als Geschlecht. Von den männlichen Schülern sagten zwei Drittel aus, regelmäßig Fastfood (*F*) zu konsumieren. Andererseits teilten 51 der Mädchen mit, unregelmäßig bzw. selten (¬*F*) Fastfood zu sich zu nehmen.“

Das verlangte Erstellen eines zugehörigen Einheitsquadrates ergibt Abbildung 12.



:

Abb. 12: Ergebnis der HTL-Umfrage.

Auch ohne Rechnen sieht (!) man $\mathcal{A} = 0$ und $q = 1$. Übersetzt heißt das $h(F \,|\, M) - h(F \,|\, W) = 0$ bzw. $h(F \,|\, M) = h(F|W)$. Das bedeutet, dass das Essverhalten an dieser Schule unter den Befragten unabhängig vom Geschlecht ist. Im Einheitsquadrat von Abbildung 12 ist keine Stufe (Abschnitt 3.) vorhanden. Unabhängigkeit von Merkmalen wird also durch ein (betragsmäßig) kleines $\mathcal{A}$ bzw. q gekennzeichnet. Umgekehrt gilt also, je größer (betragsmäßig) $\mathcal{A}$ bzw. q sind, desto stärker ist die Abhängigkeit der untersuchten Merkmale (Abschnitt 3.).

6. Nochmals für Fortgeschrittene

Fachlich kommen wir nun zum Knackpunkt der Geschichte, nämlich zur Aufzählung der *unterschiedlichen Charakterisierungen der Unabhängigkeit von Merkmalen*. Zunächst einmal gilt für das Assoziationsmaß und die odd ratio (Abschnitt 3. und Abschnitt 4.) $\mathcal{A} = 0$ genau dann, wenn $q = 1$ ist. *Beweis* mit den Bezeichnungen in Abbildung 10:

i. $\mathcal{A} = 0 \Rightarrow \alpha = \gamma \Rightarrow 1 - \alpha = 1 - \gamma \Rightarrow \frac{\alpha}{1-\alpha} = \frac{\gamma}{1-\gamma} \Rightarrow q = 1$ und

ii. $q = 1 \Rightarrow \frac{\alpha}{1-\alpha} : \frac{\gamma}{1-\gamma} = 1 = \frac{\alpha}{1-\alpha} \cdot \frac{1-\gamma}{\gamma} \Rightarrow \frac{1-\alpha}{\alpha} = \frac{1-\gamma}{\gamma} \Rightarrow \frac{1}{\alpha} - 1 = \frac{1}{\gamma} - 1 \Rightarrow \alpha = \gamma \Rightarrow \mathcal{A} = 0.$ □

Mit den Bezeichnungen in Abbildung 8 ist weiters

- $\mathcal{A} = 0 \Leftrightarrow h(B1|A1) = h(B1)$ und
- $\mathcal{A} = 0 \Leftrightarrow h(B1 \cap A1) = h(B1) \cdot h(A1)$.

Beweis durch Nachrechnen. Die Charakteristik $h(B1|A1) = h(B1)$ zeigt, dass es auf die Bedingung $A1$ im Falle der Unabhängigkeit der Merkmale A und B nicht ankommt. Die Charakteristik $h(B1 \cap A1) = h(B1) \cdot h(A1)$ zielt auf die Symmetrie der Unabhängigkeit zweier Merkmale A und B ab.

7. Andere graphische Darstellungen dichotomer nominalskalierter Merkmale

Der Lehrplan Mathematik für die Sekundarstufe I verlangt in der sechsten Schulstufe das „Arbeiten mit relativen Anteilen und relativen Häufigkeiten in zweistufigen, allenfalls dreistufigen Situationen, insbesondere mit Hilfe von Baumdiagrammen“ (RIS 2024). Aus der Vierfeldertafel in Tabelle 1 bzw. je einem Einheitsquadrat (Abbildung 7 bzw. Abbildung 9) kann man mit dem Programm *ProVis* die zugehörigen Baumdiagramme auf Knopfdruck erstellen: Abbildung 13 (vgl. Binder et al. 2023). Die Ausgangssumme 280 (Tausend) und die Randsummen sind in der erweiterten Vierfeldertafel in Tabelle 2 eingetragen. Die bedingten relativen Häufigkeiten entnehmen wir (teilweise) den Abbildungen 7 und 9.

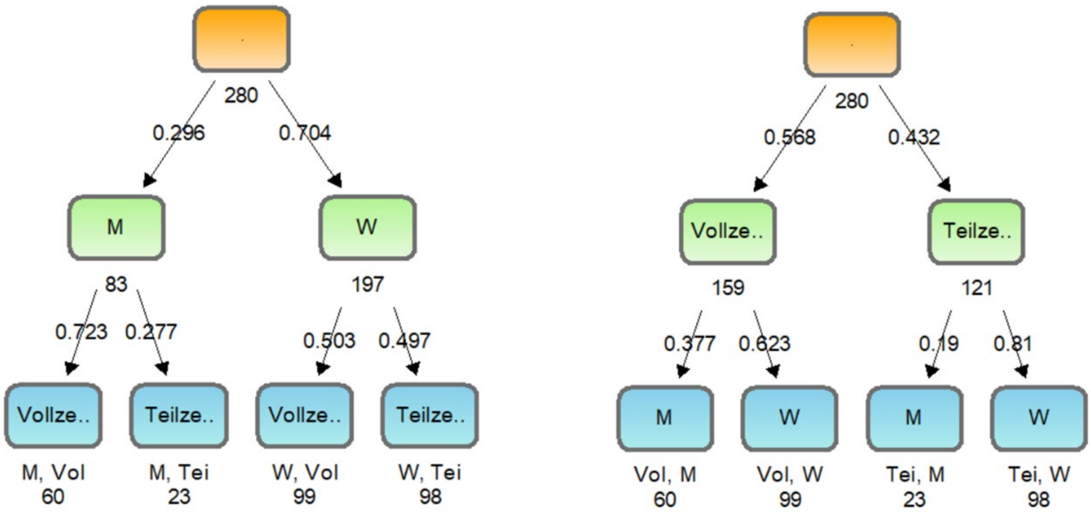


Abb. 13: Zwei Baumdiagramme mit relativen und absoluten Häufigkeiten (in Tausend).

Die beiden Baumdiagramme in Abbildung 13 entsprechen also dem ursprünglichen Einheitsquadrat in Abbildung 7 bzw. dem transponierten in Abbildung 9. Mit anderen Worten: in Abbildung 13 links ist das Geschlecht das dominierende Merkmal, in Abbildung 13 rechts das Beschäftigungsausmaß. Wichtig ist zu beachten, dass die Baumdiagramme keine Wahrscheinlichkeiten enthalten, sondern relative (und absolute) Häufigkeiten. Es handelt sich also um keinen zweistufigen Zufallsversuch, wie die vertraute Darstellung suggerieren könnte.

> „Während in den in der Schule üblicherweise verwendeten Vierfeldertafeln *Schnittwahrscheinlichkeiten* (aber keine bedingten Wahrscheinlichkeiten) fokussiert werden, stehen bei den schulisch typischerweise eingesetzten **Baumdiagrammen** umgekehrt *bedingte Wahrscheinlichkeiten* (**aber keine Schnittwahrscheinlichkeiten**) im Fokus.“ (Binder et al. 2023, S. 477; Hervorhebung fett S. G., kursiv im Original).

Im Gegensatz dazu zeigt Abbildung 14, dass man mit *ProVis* sehr wohl auch Baumdiagramme mit relativen „Schnitthäufigkeiten“ erstellen lassen kann. Sie werden gemäß den Pfadregeln in der letzten Zeile angezeigt.

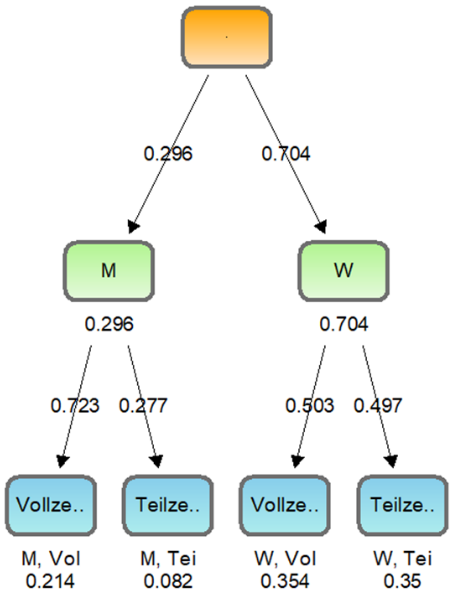


Abb. 14: Ein Baumdiagramm mit konjunktiven relativen Häufigkeiten.

Zwei Zitate von Eichler & Vogel (2013, S. 209 f.) fassen Vor- und Nachteile der beiden Darstellungsarten für dichotome nominalskalierte Daten zusammen:

> „[Das Baumdiagramm] ist […] eine Hilfestellung zur gedanklichen Strukturierung eines zufälligen Vorgangs. Es ist nicht selbst Lösung eines Problems, sondern eine Heuristik, um eine Lösung zu finden. Die gleiche Funktion hat auch das Einheitsquadrat […]. Gegenüber dem Baum hat es den Nachteil, dass es weniger universal ist. So ist es sinnvoll nur in Situationen einsetzbar, in denen die stochastische Abhängigkeit von Ereignissen untersucht werden soll."

versus

> „Bei der Untersuchung von stochastischen Abhängigkeiten hat das Einheitsquadrat gegenüber dem Baum den Vorteil, gleichzeitig eine numerische (bzw. formale) und eine geometrische Repräsentation des Problems zu bieten. Neben den Termen oder Wahrscheinlichkeiten können die Flächeninhalte und deren Verhältnis untereinander in die Analyse einbezogen werden. Schließlich bietet – im Hinblick auf die Flächenverhältnisse – das Einheitsquadrat insbesondere dann einen Vorteil, wenn es interaktive Parameteränderungen erlaubt. […]"

In Bea & Scholz (1995, S. 307) finden wir empirisch-didaktische Vergleiche zwischen Baumdiagramm und Einheitsquadrat und folgende Ergebnisse:

- das Einheitsquadrat übertrifft das Baumdiagramm in der Darstellung und der Veranschaulichung von Rechenregeln und Zusammenhängen,
- seine Konstruktion ist aber schwieriger: hier hilft *ProVis*.

Tabelle 3 zeigt Details (ebd.). Der Punkt „Qualitative Zusammenhänge (Satz von der totalen Wahrscheinlichkeit, Satz von Bayes)" ist in Abschnitt 3. ausführlich mit Hilfe von Einheitsquadraten besprochen worden und soll die empirisch gewonnene Bewertung „+++" für das Einheitsquadrat erklären und unterstreichen.

Tab. 3: Empirisch-didaktischer Vergleich Baumdiagramm-Einheitsquadrat im Detail (Bea & Scholz 1995, S. 307)

Kriterium	Wahrscheinlichkeitsbaum	Einheitsquadrat
Aufstellung des Modells	++	+
Repräsentation des Ereignisraumes (numerisch, grafisch)	+	++

Problemlösung (Rechenregeln ablesbar)	+	+++
Qualitative Zusammenhänge (Satz von der totalen Wahrscheinlichkeit, Satz von Bayes)	+	+++
Berücksichtigung von Aspekten der menschlichen Informationsverarbeitung	+	++

Eine mathematikdidaktische und kognitionspsychologische Perspektive nimmt Böcherer-Linder (2017) bei ihren Erhebungen zu Einheitsquadrat und Baumdiagramm ein. Ein Auszug aus den erzielten Ergebnissen lautet (ebd., S. 52, Hervorhebungen im Original)

> „1. Das Einheitsquadrat ist günstiger als das Baumdiagramm für das Erkennen von *Teilmengenbeziehung*en. […]
>
> 2. Das Einheitsquadrat ist günstiger als das Baumdiagramm für die *Berechnung* der Formel von Bayes. […]
>
> 3. Das Einheitsquadrat ist günstiger als das Baumdiagramm für das *Verständnis* von Parameterabhängigkeiten bei bedingten Wahrscheinlichkeiten. […]“

Eine „Vereinigung“ der beiden Baumdiagramme in Abbildung 13 bietet der sogenannte *(Häufigkeits-) Doppelbaum* (Binder et al. 2018). Abbildung 15 zeigt ihn in waagrechter Ausrichtung mit absoluten Häufigkeiten (in Tausend), ebenfalls erstellt mit *ProVis.*

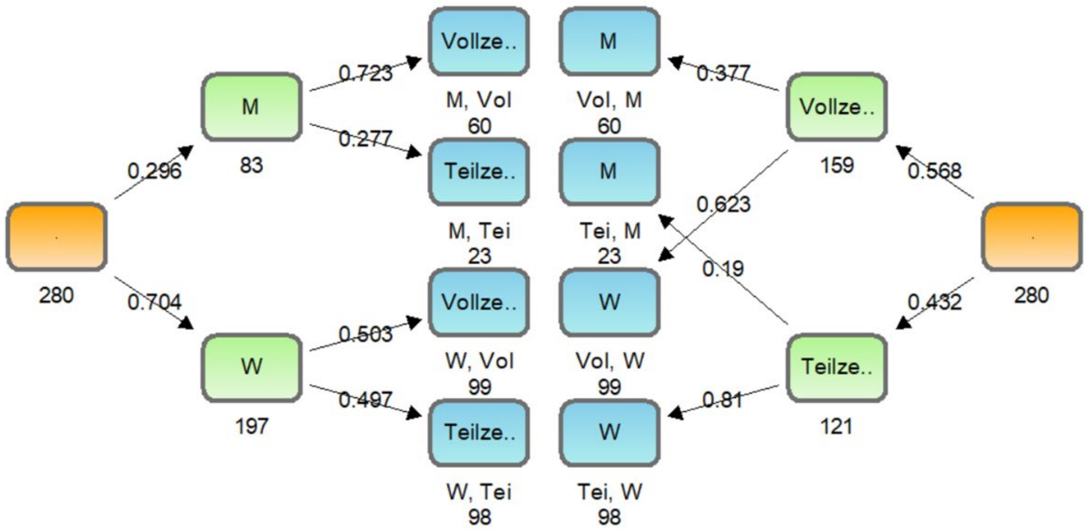


Abb. 15: Ein Häufigkeitsdoppelbaum.

In Binder et al. (2023, S. 477) finden wir dazu:

> „Aus didaktischer Sicht ist es bedauerlich, dass Häufigkeitsbäume schulisch nur eine untergeordnete Rolle spielen, da sich in empirischen Studien gerade Visualisierungen mit absoluten Häufigkeiten […] als hilfreich erwiesen haben.“

Mit Hilfe von *ProVis* kann dieses Desiderat zumindest gemildert werden. Das einmalige Kreuzen der Pfade in Abbildung 15 rechts ist nicht vermeidbar und erschwert die (händische) Konstruktion.

Ebenfalls in Binder et al. (2023) wird das sogenannte *Häufigkeitsnetz* auf seine didaktische Wirkung hin empirisch untersucht. Abbildung 16 zeigt das mit der Hand (!) erstellte Häufigkeitsnetz für das in Rede stehende Beispiel.

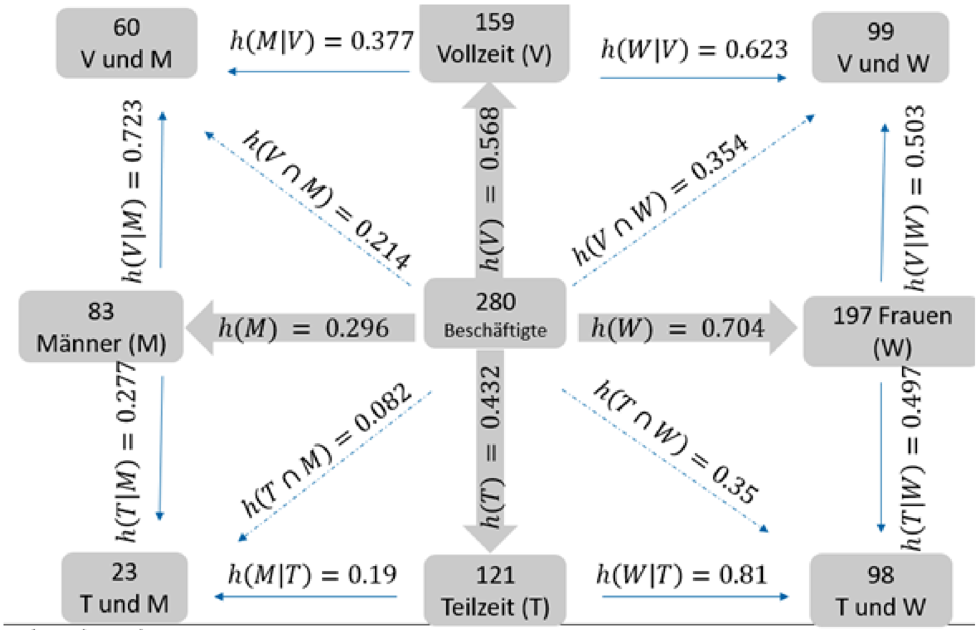


Abb. 16: Das Häufigkeitsnetz zur Vierfeldertafel in Tabelle 2.

Wie man Abbildung 16 entnehmen kann, sind im Häufigkeitsnetz alle absoluten Häufigkeiten der Vierfeldertafel in Tabelle 2, alle (bedingten) relativen Häufigkeiten aus den Baumdiagrammen (Abbildung 13) und alle konjunktiven relativen Häufigkeiten aus den Einheitsquadraten (Abbildung 7) vereint. Obwohl das Häufigkeitsnetz auf den ersten Blick sehr überladen wirkt, befinden Binder et al. (2023, S. 471, Hervorhebungen im Original):

> „Obwohl *Häufigkeitsdoppelbäume* und *Häufigkeitsnetze* den Schülerinnen und Schülern gänzlich unbekannt waren, unterstützten diese Visualisierungen die Schülerinnen und Schüler bei der Bearbeitung der Aufgaben am meisten.“

8. Didaktisches Resümee

Vier Indikationen für das *Spiralprinzip* bringt die Behandlung dichotomer nominalskalierter zweidimensionaler Daten, also Einträge einer Vierfeldertafel, mit sich:

1. Anders als im Artikel kann man vom wohlbekannten Baumdiagramm ausgehend auf das Einheitsquadrat als alternative Darstellungsform schließen. Die Betonung wandert dadurch von den bedingten relativen Häufigkeiten auf die konjunktiven, wenn man die fertige graphische Darstellung betrachtet (Produktorientierung). In der Entstehung dieser Darstellungen ist es genau umgekehrt: auf die bedingten relativen Häufigkeiten wird beim Einheitsquadrat fokussiert (Abschnitt 2.), auf die konjunktiven als Ergebnis einer Pfadregel (Produkt der relativen Häufigkeiten entlang eines Pfades ergibt die konjunktive relative Häufigkeit am Ende des Pfades) beim Baumdiagramm: Prozessorientierung. Der Doppelbaum nimmt die Idee des transponierten Einheitsquadrats für das Baumdiagramm auf. Das Häufigkeitsnetz schließlich vereint alle Eintragungen von Einheitsquadraten und Baumdiagrammen, für seine Tauglichkeit im Schulalltag gibt es erste Hinweise (Binder et al. 2023).
2. Von den elementaren absoluten Häufigkeiten werden standardmäßig relative Häufigkeiten abgeleitet (RIS 2024). Unser Vorschlag ist, auch relative *bedingte* Häufigkeiten im Zuge der Konstruktion bzw. Interpretation von Einheitsquadraten (Abschnitt 2. und Abschnitt 3.) zu thematisieren. Dies kann bei komplexeren stochastischen Situationen in der Sekundarstufe II (BAYES-Aufgaben) dazu führen, dass (bedingte) Wahrscheinlichkeiten mit künstlichen

absoluten Häufigkeiten illustriert werden, um die Behandlung eben dieser stochastischen Situationen zu erleichtern (Wassner et al. 2007; Krauss et al. 2020).

3. Vom elementar zu berechnenden und einfach zu interpretierenden Assoziationsmaß $\mathcal{A}$ (Abschnitt 3.) kann über die odds ratio q (Abschnitt 4.) bis zum Chi-Quadrat-Unabhängigkeitstest (Abschnitt 4.) der Bogen gespannt werden. (Dieser spielt in manchen Schulformen – an BHS – eine Rolle: z. B. https://ibc-vienna.at/index.php/de/spirit-of-ibc/mathe-online/150-kontingenzanalyse-chi-quadrat-test.) In jedem Fall wird so der wichtige Unabhängigkeitsbegriff bei Ereignissen oder später dann bei Zufallsvariablen in der Sekundarstufe II vorbereitet.
4. Der Vergleich von bedingten Häufigkeiten, wie er sich bei der Darstellung mit Hilfe von Einheitsquadraten und ihren Transponierten (Abschnitt 3.) ergibt, wird bei der Behandlung von BAYES-Aufgaben wieder aufgenommen (vgl. Punkt 2. in dieser Aufzählung und McDowell & Jacobs 2017).

Ebenfalls resultieren vier *Ziele* aus der in diesem Beitrag diskutierten Thematik:

1. Eine sichere Interpretation der jeweiligen Darstellung von (relativen) Häufigkeiten (Einheitsquadrat, Baumdiagramm) ist unabdingbar im Sinne einer statistical literacy. Sie werden als Seitenlängen oder Flächeninhalte wiedergegeben (Einheitsquadrat) oder einfach numerisch angegeben (Baumdiagramm: es kommt auf die Position des Wertes im Diagramm an).
2. Ein fundierter Wechsel zwischen den Darstellungen Einheitsquadrat und Baumdiagramm, unterstützt durch *ProVis*, eröffnet die Möglichkeit, die jeweils der Situation (Aufgabenstellung, vgl. Abschnitt 5.) angemessene Art der Häufigkeit(en) (konjunktiv, bedingt) zu betonen.
3. Die Schärfung des Konzepts Unabhängigkeit von Merkmalen durch unterschiedliche Charakterisierungen (Abschnitt 6.) ist eine tragfähige, weil noch vorstellbare Grundlage für darauffolgende abstrakte(re) inhaltliche Erweiterungen (siehe Punkt 3. in der vorigen Aufzählung).
4. Die Unterscheidung von relativen Häufigkeiten und Wahrscheinlichkeiten ist von eminenter Bedeutung bei der Einführung des frequentistischen Wahrscheinlichkeitsbegriffs (Malle & Malle 2003) in der Sekundarstufe I nach dem neuen Lehrplan (RIS 2024). Auch später bei der Behandlung von Textaufgaben aus dem Bereich Stochastik sollte immer wieder (nicht ständig!) darüber reflektiert werden, welcher Natur die numerischen Angaben sind und woher sie kommen. In diesem Artikel ist nur von (relativen) Häufigkeiten, nicht von Wahrscheinlichkeiten die Rede.

9. *ProVis* – Probability Visualized: Überblick und Einsatz

Die Software *ProVis – Probability Visualized* ist ein frei verfügbares Tool, das die einfache Verwendung von Einheitsquadraten und Baumdiagrammen für den Unterricht unterstützt. Für die Entwicklung wurden sieben Anforderungen formuliert, wobei im Kontext dieses Beitrags besonders die Anforderungen 4, 5 und 6 relevant sind (Döller 2020). Die Anforderungen sind im Folgenden knapp zusammengefasst:

1. **Anforderung 1 – Verschiedene Typen von Baumdiagrammen realisieren:** Wahrscheinlichkeitsbäume, Häufigkeitsbäume, vollständige, unvollständige und reduzierte Bäume

Anforderung 2 – Rechenarbeit automatisieren: automatische Berechnung der Häufigkeiten bzw. Wahrscheinlichkeiten in Bäumen mit Pfadregeln; automatische Berechnungen von bedingten Häufigkeiten und dem Assoziationsmaß im Einheitsquadrat

Anforderung 3 – Baumdiagramme nach dem Spiralprinzip unterstützen: sowohl niederschwelligen Einstieg in die Darstellungsformen ermöglichen als auch fortgeschrittene Funktionen anbieten

Anforderung 4 – Einheitsquadrate automatisch erstellen: Daten der Vierfeldertafel in einem Einheitsquadrat darstellen und Ableitung des transponierten Einheitsquadrats ermöglichen

Anforderung 5 – Baumdiagramm und Einheitsquadrat vernetzen: das entsprechende Baumdiagramm aus einem Einheitsquadrat ableiten; Vorteile beider Methoden verfügbar machen

Anforderung 6 – Grafische Darstellung dynamisieren, Visualisierung unterstützen: situationsadäquate Darstellung von Bäumen, visuelle Differenzierung von Strecken, Flächen und Assoziationsmaß in Einheitsquadraten bereitstellen

Anforderung 7 – Experimentieren anregen: „Was wäre, wenn …“-Fragen unterstützen durch Änderung der Flächen beim Variieren der Werte im Einheitsquadrat; neue Perspektiven eröffnen durch die Möglichkeit der Ableitung der einen Darstellungsform aus der anderen

Die vollständige Benutzeroberfläche von *ProVis* ist in Abbildung 17 dargestellt, rechts die Modellierungfläche, links daneben die (blau hinterlegte) schmale Leiste mit den Modellierungselementen, ganz links der Modell-Explorer und die Übersicht, oben die Menüeinträge für das Speichern, Exportieren und weitere stochastik-spezifische Funktionen.

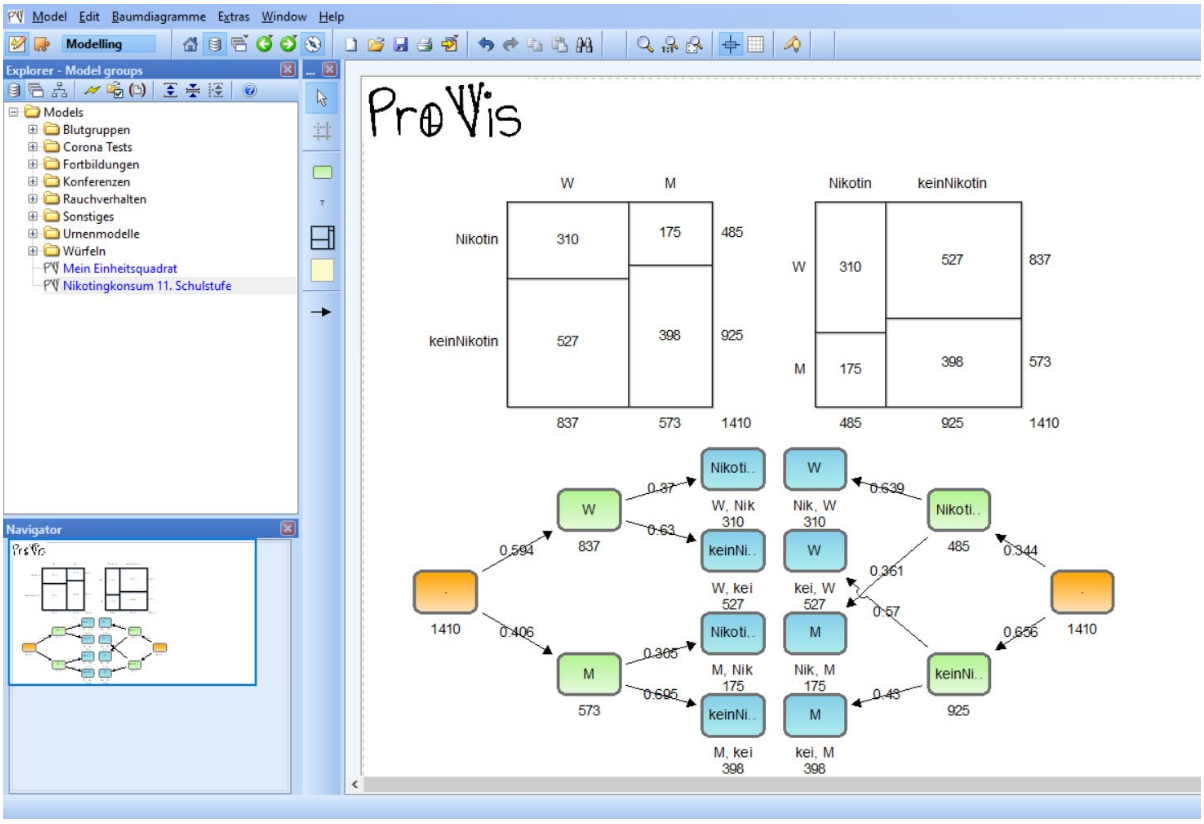


Abb. 17: Die Benutzeroberfläche von *ProVis*.

Die Funktionen und der mögliche Einsatz von *ProVis* sollen hier anhand eines Arbeitsauftrags mit realen Daten gezeigt werden:

Health Behaviour in School-aged Children (HBSC) Im HBSC-Bericht aus 2023 wurden österreichische Schüler:innen der 11. Schulstufe über ihren aktuellen Konsum von Nikotinprodukten (Zigaretten, E-Zigaretten, Wasserpfeife und andere Nikotinprodukte) befragt. In der folgenden Vierfeldertafel (Tabelle 4) sind die absoluten Häufigkeiten aktuell konsumierender und nicht-konsumierender Mädchen und Burschen der Stichprobe dargestellt.

Tab. 4: Daten des HBSC-Berichts 2021/22 (vgl. https://www.sozialministerium.at/Themen/Gesundheit/Kinder--und-Jugendgesundheit.html)

	Mädchen (W)	Burschen (M)	Summe
Nikotin-Konsum (K)	310	175	485
Kein Nikotin-Konsum (nK)	527	398	925
Summe	837	573	1410

1. Stellen Sie diese Daten in einem Einheitsquadrat dar! Wählen Sie als dominierendes Merkmal das Geschlecht!

Begründen Sie mithilfe des Assoziationsmaßes, ob ein Zusammenhang zwischen dem Geschlecht und dem Konsum von Nikotinprodukten bei den Schüler:innen besteht!

Geben Sie die bedingte relative Häufigkeit der Nikotin-Konsumentinnen unter den Mädchen an!

Leiten Sie aus dem Einheitsquadrat ein Baumdiagramm ab! Wählen Sie als dominierendes Merkmal wieder das Geschlecht!

Stellen Sie das transponierte Einheitsquadrat dar!

Geben Sie die bedingte relative Häufigkeit der Burschen an unter jenen, die kein Nikotin konsumieren!

Leiten Sie aus dem transponierten Einheitsquadrat das Baumdiagramm ab!

Eine Schritt-für-Schritt Anleitung zur Verwendung der Software für die Lösung dieses Arbeitsauftrags kann hier heruntergeladen werden: code.omilab.org/resources/modelling_tools/provis/-/blob/master/DOCUMENTS/ProVis_Hands-On.pdf.

Um mit Punkt 1. des Arbeitsauftrags zu beginnen, wird erst ein leeres Einheitsquadrat erstellt durch Auswählen des Einheitsquadrat-Symbols in der Leise mit den Modellierungselementen und einem Klick in die Modellierungsfläche, siehe Abbildung 17. Die Interaktion mit einem Element in der Modellierungsfläche erfolgt immer über dessen Notebook, das über Doppelklick auf das Element aufgerufen werden kann und aus mehreren Tabs besteht. In Abbildung 18 ist das Notebook eines Einheitsquadrats abgebildet, links der Tab „Merkmale", in dem die Namen der Merkmalsausprägungen eingegeben werden können, und rechts der Tab „Haeufigkeiten" zum Eintragen der (absoluten oder relativen) Häufigkeiten.

Sobald ein leeres Einheitsquadrat erstellt und die Daten wie oben eingegeben wurden, erscheint automatisch die Unterteilung in die vier Rechtecke, zu sehen in Abbildung 19. Damit ist Punkt 1. des Arbeitsauftrages erfüllt.

Um Punkt 2. und Punkt 3. zu erfüllen, werden die Konfigurationsmöglichkeiten im Tab „Grafische Darstellung" des Notebooks benötigt. Dort ist es möglich über Checkboxen den Wert des Assoziationsmaßes und die Zeilen- und Spaltensummen anzeigen zu lassen. Außerdem können alle relativen, bedingten und konjunktiven Häufigkeiten und die zugehörigen Flächen und Strecken farblich hervorgehoben und deren gerundete Werte angezeigt werden, siehe Drop-down Liste in Abbildung 20 links.

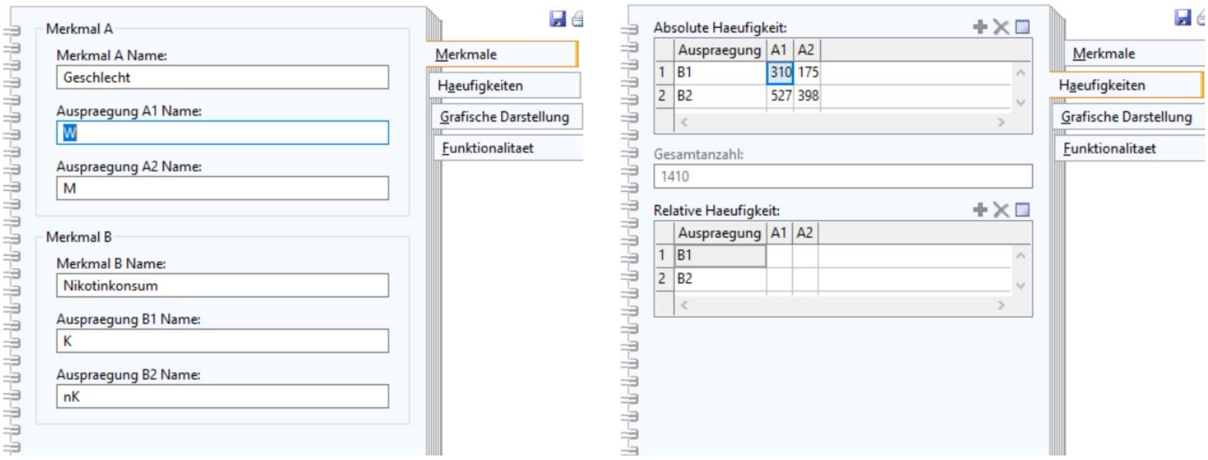


Abb. 18: Die zwei Tabs „Merkmale“ (links) und „Haeufigkeiten“ (rechts) des Notebooks eines Einheitsquadrats.

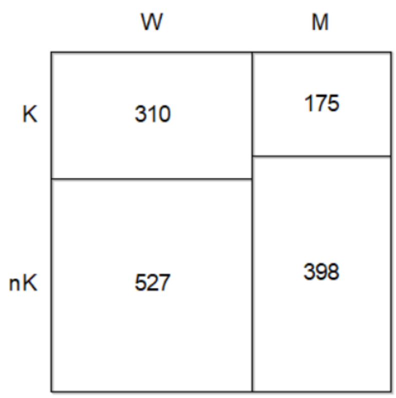


Abb. 19: Das Einheitsquadrat zu den Daten aus dem Arbeitsauftrag zum Nikotinkonsum von Schüler:innen der 11. Schulstufe.

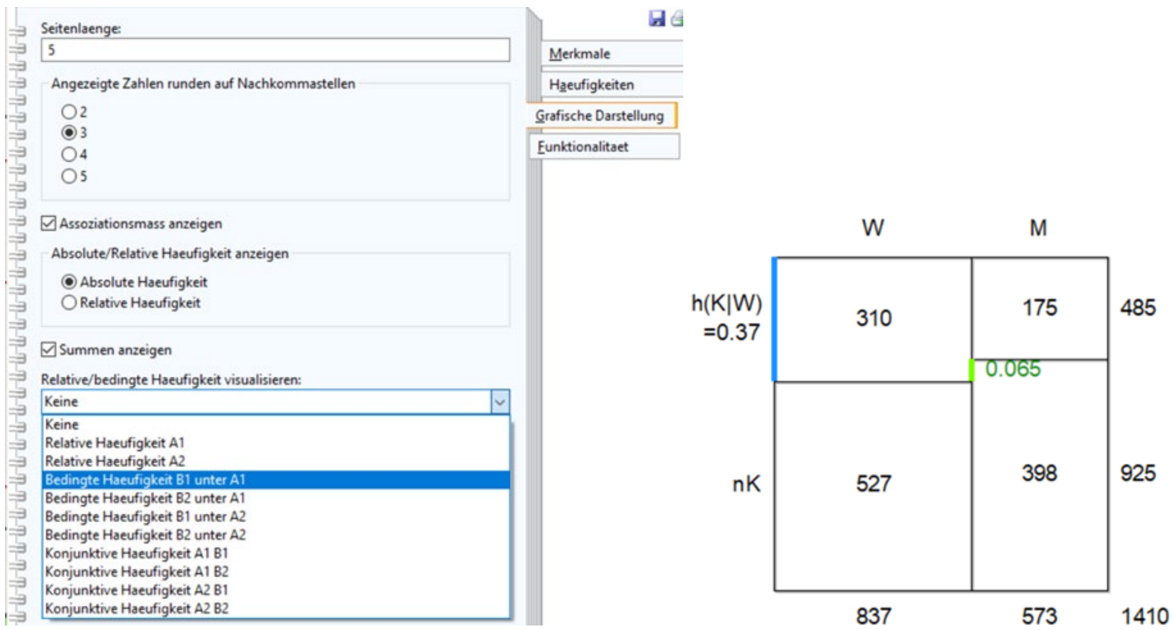


Abb. 20: Konfiguration der Anzeige und Hervorhebung ausgewählter Werte, Flächen und Strecken links und Einheitsquadrat mit farblichen Hervorhebungen und Werten rechts.

Nach Anwählen des Assoziationsmaßes, der Summenanzeige und Auswahl von „Bedingte Haeufigkeit B1 unter A1“ (in diesem Fall steht B1 für Nikotinkonsum K und A1 für Mädchen W) in der Drop-Down

Liste, erscheinen die gewünschten Größen farblich hervorgehoben und mit numerischen Werten wie in Abbildung 20 rechts. Der kleine Wert des Assoziationsmaßes von 0.065 (gerundet), wie man an der Abbildung 20 rechts ablesen kann, erlaubt eine Beantwortung von Punkt 2. des Arbeitsauftrags: Es lässt sich nicht auf einen Zusammenhang von Nikotinkonsum und Geschlecht schließen. Der relative Anteil der Nikotin-Konsumentinnen unter den Mädchen $h(K|W)$, wie in Punkt 3. gefragt, lässt sich ebenfalls ablesen: Er beträgt 37% (gerundet).

Um für die nächste Aufgabe das Baumdiagramm zu den Daten nicht von Neuem aus der Angabe heraus erstellen zu müssen, gibt es in *ProVis* die Möglichkeit, dieses per Knopfdruck automatisch abzuleiten; gleiches gilt für das transponierte Einheitsquadrat. Für beide Funktionen gibt es im Notebook im Tab „Funktionalitaet" je einen Button. Der Baum wird unterhalb des Einheitsquadrats erstellt, das transponierte Einheitsquadrat daneben. Für das transponierte Einheitsquadrat stehen die gleichen Funktionalitäten wie für das ursprüngliche Einheitsquadrat zur Verfügung, auch daraus lässt sich ein Baumdiagramm ableiten. Das Ergebnis ist in Abbildung 21 dargestellt. Damit sind die Punkte 4., 5. und 7. des Arbeitsauftrags jeweils mit einem Klick erledigt.

Das Ergebnis ist ein Zusammenspiel von vier Darstellungen (Abbildung 21) des gleichen Sachverhaltes, in jeder Darstellung werden andere Größen und Werte in den Vordergrund gerückt. Für deren Erstellung war nichts weiter nötig als die Namen und die vier Häufigkeiten der verschiedenen Merkmalsausprägungen (sie entsprechen den Eintragungen einer Vierfeldertafel) einzutragen und drei Buttons zu klicken. Die Antwort auf die Frage aus Punkt 6. nach dem Anteil der Burschen unter den Personen, die kein Nikotin konsumieren, lässt sich sowohl im transponierten Einheitsquadrat (durch explizites Auswählen der gesuchten Größe im Notebook: Abbildung 21 rechts oben) als auch im zugehörigen Baumdiagramm darunter (Übergang rechts unten in Abbildung 21) ablesen: Sie beträgt 43% (gerundet).

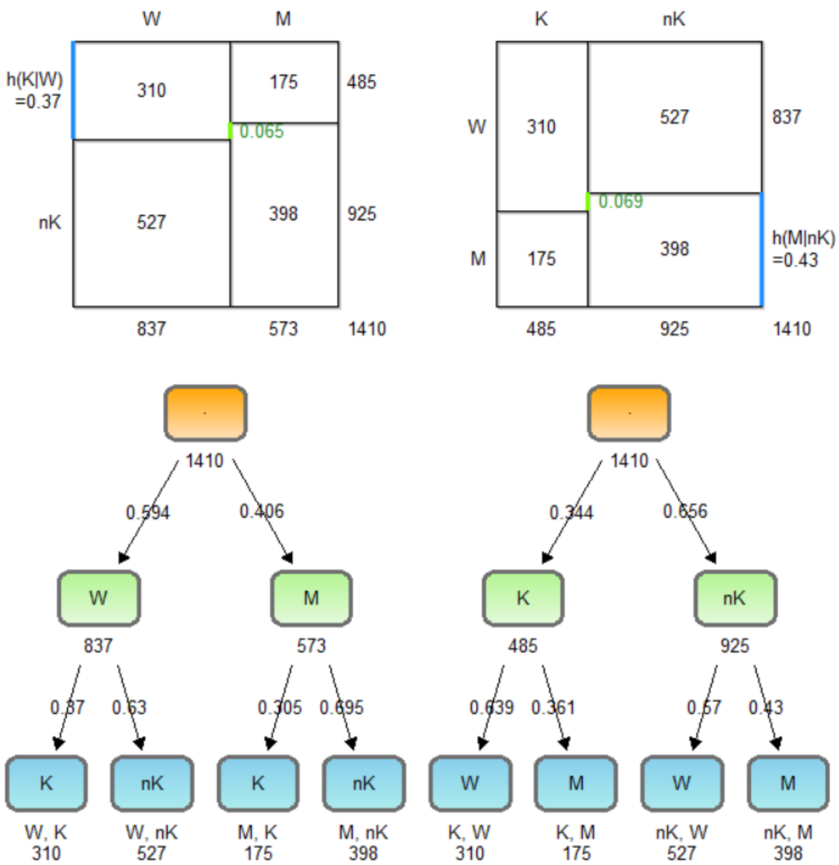


Abb. 21: Links oben das Einheitsquadrat zu den Daten des Arbeitsauftrags, darunter das entsprechende Baumdiagramm, rechts oben das transponierte Einheitsquadrat, darunter das Baumdiagramm zum transponierten Einheitsquadrat.

Durch ein wenig Umordnung der einzelnen Merkmalsgruppen (händisch oder teils automatisiert) können die beiden Bäume auch als Doppelbaum wie in Abbildung 15 angeordnet werden.

Literatur

Bea, W., Scholz, R. W. (1995): Graphische Modelle bedingter Wahrscheinlichkeiten im empirisch-didaktischen Vergleich. In: *Journal für Mathematik-Didaktik* 16(3/4), 299–327. https://doi.org/10.1007/BF03338820.

Bickel, P. J., Hammel, E. A., O'Connell, J. W. (1975): Sex Bias in Graduate Admissions: Data from Berkeley. *Science* 187(4175), 398–404. Online: https://en.wikipedia.org/wiki/Simpson%27s_paradox#cite_note-Bickel-16 (Zugriff: 26. 7. 2024).

Binder, K., Krauss, S., Wassner, C. (2018): Der Häufigkeitsdoppelbaum als didaktisch hilfreiches Werkzeug von der Unterstufe bis zum Abitur. In: *Stochastik in der Schule* 38(1), 2–11. Online: http://www3.math.uni-paderborn.de/~agbiehler/sis/sisonline/Jahrgang38-2018/Heft%201/Stochastik_1_2018_2_11.pdf (Zugriff: 29. 8. 2024).

Binder, K., Steib, N., Krauss, S. (2023): Von Baumdiagrammen über Doppelbäume zu Häufigkeitsnetzen – kognitive Überlastung oder didaktische Unterstützung? In: *Journal für Mathematik-Didaktik* 44(2), 471–503. https://doi.org/10.1007/s13138-022-00215-9.

Böcherer-Linder, K. (2017): *Visualisierung bedingter Wahrscheinlichkeiten. Eine Untersuchung aus kognitionspsychologischer, mathematikdidaktischer und schulpraktischer Perspektive.* Pädagogische Hochschule Freiburg: Dissertation. Online: https://www.researchgate.net/publication/327722172_Visualisierung_bedingter_Wahrscheinlichkeiten (Zugriff: 30. 7. 2024).

Dang, A. (2024): *Immer mehr Stellen in Teilzeit ausgeschrieben*. Online: https://www.derstandard.at/story/3000000206109/immer-mehr-stellen-in-teilzeit-ausgeschrieben (Zugriff: 23. 7. 2024).

Döller, V. (2020): *ProVis – Probability Visualized. Technologische Unterstützung für den Einsatz von Einheitsquadraten und Baumdiagrammen im Stochastikunterricht*. Universität Wien: Diplomarbeit. https://doi.org/10.25365/thesis.61461.

Döller, V., Götz, S. (2021): Baumdiagramme und Einheitsquadrate 4.0. In: *Stochastik in der Schule* 41(3), 9–19.

Döller, V., Götz, S. (2022): Tree Diagrams and Unit Squares 4.0: Digitizing Stochastic Classes with the Didactic Modeling Tool *ProVis*. In: Karagiannis, D.; Lee, M.; Hinkelmann, K.; Utz, W. (Eds.): *Domain-Specific Conceptual Modeling. Concepts, Methods and ADOxx Tools* (pp. 481–501). Cham: Springer. https://doi.org/10.1007/978-3-030-93547-4_21.

Eichler, A., Vogel, M. (2011): *Leitfaden Stochastik. Für Studierende und Ausübende des Lehramts*. Wiesbaden: Vieweg+Teubner | Springer. https://doi.org/10.1007/978-3-8348-9909-5.

Eichler, A., Vogel, M. (2013): *Leitidee Daten und Zufall. Von konkreten Beispielen zur Didaktik der Stochastik* (2., aktualisierte Auflage). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-00118-6.

Götz, S., Döller, V. (2020): Vorschläge für einen datenorientierten Stochastikunterricht in der Sekundarstufe. *R&E-SOURCE*. Online: https://journal.ph-noe.ac.at/index.php/resource/article/view/882 (Zugriff: 23. 7. 2024).

Krauss, S., Weber, P., Binder, K., Bruckmaier, G. (2020): Natürliche Häufigkeiten als numerische Darstellungsart von Anteilen und Unsicherheit – Forschungsdesiderate und einige Antworten. In: *Journal für Mathematik-Didaktik* 41(2), 485–521. https://doi.org/10.1007/s13138-019-00156-w.

Malle, G., Malle, S. (2003): Was soll man sich unter einer Wahrscheinlichkeit vorstellen? In: *mathematik lehren* 118, 52–56.

McDowell, M., Jacobs, P. (2017): Meta-analysis of the effect of natural frequencies on Bayesian reasoning. In: *Psychological Bulletin* 143(12), 1273–1312. https://doi.org/10.1037/bul0000126.

RIS (2024): *Bundesrecht konsolidiert: Lehrpläne – allgemeinbildende höhere Schulen Anl. 1, tagesaktuelle Fassung.* Online: https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568&Artikel=&Paragraf=&Anlage=1&Uebergangsrecht= (Zugriff: 23. 7. 2024).

Schmutterer, I. (2022): *Tabak- und Nikotinkonsum. Zahlen und Fakten 2022*. Wien: Gesundheit Österreich. Online: https://jasmin.goeg.at/id/eprint/2814 (Zugriff: 23. 7. 2024).

Wassner, C., Biehler, R., Martignon, L. (2007): Das Konzept der natürlichen Häufigkeiten im Stochastikunterricht. In: *Der Mathematikunterricht* 53(3), 33– 44.

Winter, H. (1995): Mathematikunterricht und Allgemeinbildung. In: *Mitteilungen der Gesellschaft für Didaktik der Mathematik* 61, 37–46. Online: https://ojs.didaktik-der-mathematik.de/index.php/mgdm/article/view/69 (Zugriff: 30. 7. 2024).

Verfasser:in

Victoria Döller
Universität Wien
Forschungsgruppe Knowledge Engineering
Währinger Straße 29
1090 Wien
victoria.doeller@univie.ac.at

Stefan Götz
Universität Wien
Fakultät für Mathematik
Oskar-Morgenstern-Platz 1
1090 Wien
stefan.goetz@univie.ac.at