

Von Veedels- und Winkelzügen

Darstellung und Berechnung prozentualer Anteile mit Hilfe von Flächensegmenten

Düsseldorf, den 9. April 2014 Wenn jemand von einem kostümierten Kölner spricht, meint er dann unter den Kölnern einen, der kostümiert ist, oder erkundigt er sich gerade bei einer Gruppe von Kostümierten, ob da jemand aus Köln dabei sei? Oder denkt er gar zufällig an die vielen Narren, die jedes Jahr zum Rosenmontag den Karnevalszug säumen und überlegt sich, wie groß wohl die Wahrscheinlichkeit sei, unter all diesen Menschen jemanden zu treffen, der Kölner und außerdem kostümiert ist?

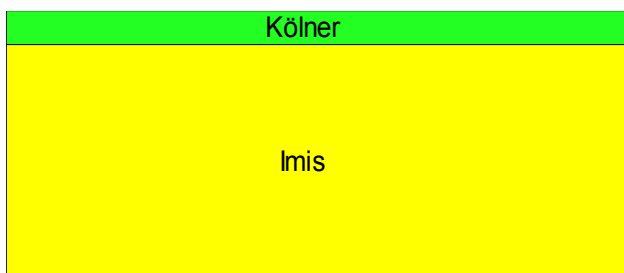


Abbildung 1: Es gibt 7mal so viele **Imis** wie **Kölner**

Die hiermit verbundenen Berechnungen prozentualer Anteile – bei auf den ersten Blick eher diffus anmutenden Verteilungen bestimmter Ausprägungen hinsichtlich verschiedener Merkmale – gehören zu den erwarteten Kompetenzen und inhaltlichen Schwerpunkten des Mathematikunterrichtes bis zum Ende der Einführungsphase.

Der Kernlehrplan für die Sekundarstufe II verweist in diesem Zusammenhang innerhalb der Stochastik auf die so genannten *Bedingten Wahrscheinlichkeiten* inklusive der Verwendung von *Vier-Felder-Tafeln*, um alle hier vorliegenden Beziehungen exakt quantifizieren zu können.

Im Folgenden stellen wir einen Lösungsansatz vor, der die in diesem Kontext beteiligten Grundgesamtheiten, sprich die 100%-Anteile

bzw. die so genannten Grundwerte¹, mit Hilfe von Flächensegmenten modelliert, die im Falle von zwei unterschiedlichen Merkmalen einerseits in einen horizontal orientierten Kontext eingebunden sind, andererseits aber auch vertikal notierten Strukturierungen genügen müssen. Damit gelingt es, die bei der Berechnung verwendeten Vier-Felder-Tafeln als unmittelbare Abstraktionen der involvierten Flächensegmente zu interpretieren bzw. zu verstehen. Betrachten wir hierzu ein erstes Beispiel:

In Köln findet man am Rosenmontag unter den Karnevalisten bzw. Narren entlang des Zugweges 7mal so viele Touristen wie Einheimische. 80% der Einheimischen tragen ein Kostüm. Darunter versteht man hier im Allgemeinen etwas mehr als eine Pappnase. Dagegen ist nur jeder vierte Tourist – in Köln auch „Imi“ genannt – in diesem Sinne kostümiert. Kaum haben wir uns in den Trubel auf der Domplatte gestürzt, winkt uns bereits ein Lappen-Clown entgegen.

Die erste Frage lautet:

*Mit welcher Wahrscheinlichkeit ist er ein **Kölner** bzw. ein **Imi**?*



Abbildung 2: Im Veedelszoch in Köln am Karnevalssonntag

¹ Im Sinne des 100%-Anteils innerhalb der Prozentrechnung.

Zeile ↓	Spalte	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	Kölner	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	o	o	o	o	Kostüm
2		m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
3		m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
4		m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
5	Imis	m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	Kostüm
6		m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
7		m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
8		m	m	m	m	m	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	

Abbildung 3: Der rote Winkel (m) verdeutlicht sämtliche Kostümierten unter Kölnern und Imis

Um die Antwort nachvollziehbar zu formulieren, betrachten wir ein geeignet eingefärbtes Rechteck, vgl. Abbildung 1, wo zunächst nur die Anteile der **Kölner**, respektive der **Imis** übersichtlich und horizontal orientiert darstellt sind.

Anschließend ergänzen wir Abbildung 1 hinsichtlich des vertikal orientierten Merkmals Kostümierung durch die Ausprägungen **mit** bzw. **ohne** Kostüm, wobei wir den jeweils vorliegenden Anteil zusätzlich durch die Größe der Segmente illustrieren (vgl. Abbildung 3).

Offenbar können wir – jetzt, wo sämtliche Anteile bekannt sind – die oben notierte Frage endlich beantworten. Ausschließlich unter den **Kostümierten** suchen wir nach einem **Kölner** bzw. nach einem **Imi**. Die graphische Verdeutlichung der relevanten Grundgesamtheit geschieht dabei weder durch einen vertikalen, noch durch einen horizontalen Streifen. Stattdessen ist eine Kombination beider Darstellungsformen in Gestalt eines **roten Winkels** (100%) angesagt².

Wir finden schließlich³:

$$P(\text{Kölner}|\text{mit Kostüm} \approx 100\%) = \frac{16}{16+35} = \frac{16}{51} = 31,37\%$$

bzw.

$$P(\text{Imi}|\text{mit Kostüm} \approx 100\%) = \frac{35}{16+35} = \frac{35}{51} = 68,63\%$$

² Die Gesamtheit aller Kostümierten wird in Abbildung 3 durch einen roten Winkel ($\approx 100\%$) symbolisiert.

Entsprechend erhalten wir bezogen auf den blauen Winkel (100%):

$$P(\text{Kölner}|\text{ohne Kostüm} \approx 100\%) = \frac{4}{4+105} = \frac{4}{109} = 3,67\%$$

$$P(\text{Imi}|\text{ohne Kostüm} \approx 100\%) = \frac{105}{4+105} = \frac{105}{109} = 96,33\%$$

Wir haben die betreffende Wahrscheinlichkeit zweifarbig als 31,37% formatiert, was sich von der zweifarbig Darstellung für die Wahrscheinlichkeit einen Kostümierten unter den Kölnern (100%) zu treffen, das heißt:

$P(\text{Kostümierten}|\text{Kölnern} \approx 100\%) = 80\%$, farblich offenbar nicht unterscheidet.

		mit Kostüm	ohne Kostüm	Summe
	Kölner	4/5		1
7/8	Imis	1/4		1
1	Summe			

Abbildung 4: Eingangsdaten für die Vier-Felder-Tafel

Die Ausdifferenzierung aller vorliegenden Situationen allein mithilfe von zwei Farben stößt

³ Die Schreibweise „ $P(\text{Kölner}|\text{Kostümiert})$ “ lesen wir als „Wahrscheinlichkeit für einen Kölner unter der Bedingung, wir haben einen Kostümierten (100%) vor uns“. Kurz: Das Zeichen „|“ lesen wir als „unter der Bedingung“, womit wir natürlich explizit die betreffende Grundgesamtheit bzw. den relevanten Grundwert ($\approx 100\%$) ansprechen. De facto reden wir jedoch meistens nur über den Anteil unter, etwa in der Form $P(\text{Anteil der Kölner unter den Kostümierten})$. Daneben verwenden wir Schreibweisen wie $P(\text{Kölner}|\text{Kostümiert} \approx 100\%)$ bzw. $P_{\text{Kostümiert}}(\text{Kölner})$.

mitunter an ihre Grenzen. Denn wir notieren sowohl die Wahrscheinlichkeiten

$$P(\text{Kölner}|\text{mit Kostüm}) = P_{\text{mit Kostüm}}(\text{Kölner}),$$

als auch

$$P(\text{mit Kostüm}|\text{Kölner}) = P_{\text{Kölner}}(\text{mit Kostüm})$$

und die Variante $P(\text{Kölner} \cap \text{Kostümierten})$ jeweils in derselben Fassung als $p\%$ ⁴.

Damit sind Rückschlüsse von der farblichen Kennzeichnung auf die vorliegende Situation nicht immer eindeutig möglich, weil unter Umständen mehr als zwei Alternativen vorliegen. Die Frage nach dem Anteil der **Kölner** unter sämtlichen Narren (100%) hingegen lässt natürlich nur die Notation $p\%$ zu.

In einem nächsten Schritt kann man Abbildung 4 leicht weiter ergänzen. Aus

$$P(\text{Imis}|\text{Narren} \approx 100\%) = \frac{7}{8} = 87,5\%.$$

ergibt sich:

$$P(\text{Kölner}|\text{Narren} \approx 100\%) = 1 - \frac{7}{8} = \frac{1}{8} = 12,5\%$$

Innerhalb der **Kölner** und **Imis** argumentieren wir nun vertikal im Sinne von:

$$P(\text{mit Kostüm}|\text{Kölner} \approx 100\%) = \frac{4}{5} = 80\%$$

führt zu:

$$P(\text{ohne Kostüm}|\text{Kölner} \approx 100\%) = 1 - \frac{4}{5} = \frac{1}{5} = 20\%$$

Analog findet sich mit Hilfe von:

$$P(\text{mit Kostüm}|\text{Imi} \approx 100\%) = \frac{1}{4} = 25\%$$

der Wert

$$P(\text{ohne Kostüm}|\text{Imi} \approx 100\%) = 1 - \frac{1}{4} = \frac{3}{4} = 75\%.$$

Wir relativieren somit die waagrecht notierten (**grünen** und **gelben**) Informationen (Merk-

mal *Wohnort*) in Hinblick auf ihre senkrecht verbuchten (**roten** und **blauen**) Ausdifferenzierungen (Merkmal *Kostümierung*).

In der 2. Fassung (Abbildung 5) wechseln wir die Perspektive und schauen zuerst auf die vertikalen Unterscheidungen. Zunächst betrachten wir den **roten Winkel der Kostümierten**. Dabei überlegen wir, wie groß der Anteil der **Kölner** bzw. der **Imis** unter sämtlichen Narren (100%) ist, die jeweils **kostümiert** sind.

Hier ergibt sich z.B. aus dem Anteil der **Kölner** unter den Karnevalisten (100%), hier $\frac{1}{8}$, und dem Anteil der **Kostümierten** unter diesen **Kölnern** (100%) hier $\frac{4}{5}$, der Anteil an Personen unter den Narren (100%), die gleichzeitig **Kölner** und **kostümiert** sind. Bei dieser Betrachtungsweise müssen wir zweimal auswählen. Unter allen Narren $\approx 100\%$ fokussieren wir zunächst die **Kölner** und anschließend unter diesen **Kölnern** $\approx 100\%$ die **Kostümierten**. Die Anteile der Anteile berechnen wir durch ihr Produkt.

		mit Kostüm	ohne Kostüm	Summe
1/8	Kölner	4/5	1/5	1
		1/8 * 4/5	1/8 * 1/5	
7/8	Imis	7/8 * 1/4	7/8 * 3/4	
		1/4	3/4	1
1	Summe			

Abbildung 5: 2. Fassung der Vier-Felder-Tafel

		mit Kostüm	ohne Kostüm	Summe
1/8	Kölner	4/5	1/5	1
		16/160	4/160	
7/8	Imis	35/160	105/160	
		1/4	3/4	1
1	Summe	51/160	109/160	

Abbildung 6: Komplette Fassung der Vier-Felder-Tafel

Das heißt⁵:

5 Etwas präziser wäre die Notation:

$$P_{\text{Narren}}(\text{Kölner mit Kostüm}) =$$

$$P_{\text{Narren}}(\text{Kölner} \cap \text{Kostümierten}) =$$

$$P(\text{Kölner}|\text{Narren}) \cdot P(\text{Kostümierten}|\text{Kölnern})$$

4 Korrekter wäre die Notation $100 \cdot p\%$, weil sich damit z.B. aus $p=0,25125$ der Wert $25,125\%$ ergäbe.

$$P(\text{Kölner mit Kostüm} | \text{Narren} \approx 100\%) = \frac{1}{8} \cdot \frac{4}{5} = \frac{16}{160} = 10\%$$

bzw.

$$P(\text{Imis mit Kostüm} | \text{Narren} \approx 100\%) = \frac{7}{8} \cdot \frac{1}{4} = \frac{35}{160} = 21,875\%$$

Die mittleren, hier farbig hinterlegten Felder in Abbildung 6 beschreiben demnach nicht nur den tabellarischen Schnitt des waagrecht notierten Merkmals *Wohnort* mit dem senkrecht angeordneten Merkmal *Kostümierung*, sondern bilden auch mengentheoretische Schnitte im Sinne von $\text{Kölner} \cap \text{Kostümierte}$.

Die folgenden beiden dezimal notierten Vier-Felder-Tafeln (vgl. Abbildung 7 und Abbildung 8) demonstrieren darüber hinaus die Kommutativität des \cap -Operators, indem sie unter anderem den Anteil $(\text{Kostümierte} \cap \text{Kölner})$ versus $(\text{Kölner} \cap \text{Kostümierte})$ transparent machen. Dabei wurden natürlich die waagerechten und senkrechten Differenzierungen vertauscht.

In exakten Brüchen notiert, erhalten wir selbstverständlich ebenfalls die Gleichheit dieser beiden – auf unterschiedlichen Wegen berechneten – Wahrscheinlichkeiten:

	mit Kostüm	ohne Kostüm	
0,125	0,8	0,2	1
Kölner	0,1	0,025	
Imis	0,21875	0,65625	
0,875	0,25	0,75	1
1	0,31875	0,68125	1

Abbildung 7: Erste dezimale Fassung

Unter allen Narren, die den Zug säumen gilt:

$$P(\text{unter allen Narren einen Kölner zu treffen, der kostümiert ist}) = \frac{20}{160} \cdot \frac{16}{20} = \frac{16}{160} = 10\%$$

bzw.

$$P(\text{unter allen Narren einen Kostümierten zu treffen, der}$$

$$\text{Kölner ist}) = \frac{51}{160} \cdot \frac{16}{51} = \frac{16}{160} = 10\%$$

	Kölner	Imis	
0,31875	0,31372549	0,68627451	1
mit Kostüm	0,1	0,21875	
ohne Kostüm	0,025	0,65625	
0,68125	0,03669725	0,963302752	1
1	0,125	0,875	1

Abbildung 8: Zweite dezimale Fassung

Nachdem wir somit sämtliche Beziehungen zwischen zwei voneinander abhängigen Merkmalen anhand der zugehörigen Winkelfelder und den daraus resultierenden Vier-Felder-Tafeln angesprochen haben, lenken wir unsere Aufmerksamkeit abschließend auf das Problem von so genannten a-priori- bzw. a-posteriori-Wahrscheinlichkeiten, die beim wiederholten Testen auf ein bestimmtes Merkmal bzw. seine Ausprägungen hin eine entscheidende Rolle spielen.

Wir betrachten ein medizinisches Testverfahren, das bzgl. einer sehr selten auftretenden Infektionskrankheit mit einer Sicherheit von 99 Prozent korrekte Testergebnisse für bereits Infizierte liefert. Aufgrund von ausgereiften Labortechniken liegt die Fehlerquote bei Gesunden (nicht Infizierten) bei circa 2 Prozent. Andererseits kann man momentan davon ausgehen, dass unter 10 Millionen Menschen ca. 2500 Infizierte vorkommen.

Neben dem Gesundheitszustand mit den Ausprägungen *krank (infiziert)* und *gesund (nicht infiziert)* betrachten wir das Merkmal Testergebnis mit den Ausprägungen *positiv* und *negativ*.

Wir haben somit die in Abbildung 9 beschriebenen Eingangsdaten vorliegen.

		positiver Test	negativer Test
0,0002500	infiziert	0,9900000	
	gesund	0,0200000	

Abbildung 9: Testen auf eine Infektion, Eingangsdaten

Wie weiter oben ausführlich beschrieben, gewinnen wir hieraus:

		positiver Test	negativer Test	Summe
0,00.025	infiziert	0,99	0,01	1
		0,00.024.75	0,00.000.25	
0,99.975	gesund	0,01.999.5	0,97.975.5	
		0,02	0,98	1
1	Summe	0,02.024.25	0,97.975.75	1

Abbildung 10: komplettierte Vier-Felder-Tafel zum Test

Und weiter:

$$\frac{0,00.024.75}{0,02.024.25} = 0,01.222.675.065 = 1,223\%$$

als Wahrscheinlichkeit, unter den als **positiv Getesteten** (100%) einen tatsächlich **Infizierten** anzutreffen

bzw.

$$\frac{0,01.999.5}{0,02.024.25} = 0,98.777.324.94 = 98,78\%$$

als Wahrscheinlichkeit, unter den als **positiv Getesteten** (100%) einen **Gesunden** anzutreffen.

Dies ist natürlich ein ernüchterndes Resultat, zeigt es doch, dass unsere Sicherheit; unter den als positiv Getesteten einen **Infizierten** durch den Test auch als einen solchen zu erkennen, auf bescheidene 1,223% angestiegen ist. Vor unserem Test durften wir – in dem zu diesem Zeitpunkt noch wesentlich größeren Testfeld – nur mit einer Wahrscheinlichkeit von 0,025% davon ausgehen, tatsächlich einen **Infizierten** vor uns zu haben.

P(infiziert positiver Test)=	0,0122268
P(gesund positiver Test)=	0,9877732
Testfeld	10000000
davon positiv & infiziert	2475
davon positiv & gesund	199950
Summe	202425

Abbildung 11: Ergebnisse des ersten Tests

Nichtsdestotrotz liegt die Fehlerquote für **Gesunde**, die irrtümlich als **Infizierte** klassifiziert wurden, nach dem ersten Test nun bei 98,78%, was selbstverständlich in dieser Größenordnung weiterhin als wenig akzeptabel erscheint.

Bei einem Testfeld von 10 Millionen Personen (100%) ergäben sich demnach 202.425 als positiv getestete Personen, wovon allerdings nur 2475 tatsächlich infiziert sind, während der Rest von 199.950 Personen in Wirklichkeit gesund ist. Als Ausweg aus diesem Dilemma bleibt nur die Möglichkeit offen, alle als infiziert klassifizierten Personen, das sind immerhin 202.425, erneut zu testen.

Der zweite Test basiert somit auf folgenden Eingangsdaten:

2. Test: Eingangsdaten		positiver Test	negativer Test
0,0122268	infiziert	0,9900000	
	gesund	0,0200000	

Abbildung 12: Eingangsdaten für den dritten Test

Die restlichen Daten lassen sich – wie beim ersten Test dargestellt – entsprechend ergänzen und führen damit zur komplettierten Vier-Felder-Tafel bzgl. des zweiten Tests (vgl. Abbildung 13).

2. Test		positiver Test	negativer Test
0,0122268	infiziert	0,9900000	0,0100000
		0,0121045	0,0001223
0,9877732	gesund	0,0197555	0,9800000
		0,0200000	0,9800000
		0,0318599	0,9801223

Abbildung 13: komplettierte Vier-Felder-Tafel zum 2. Test

Wir erhalten hieraus:

$$\frac{0,01.210.45}{0,03.185.99} = 0,37.992.79 = 37,99\%$$

als Wahrscheinlichkeit, unter den nun als **positiv Getesteten** (100%) einen tatsächlich **Infizierten** anzutreffen

bzw.

$$\frac{0,01.210.45}{0,03.185.99} = 0,62.007.21 = \mathbf{62,01\%}$$

als Wahrscheinlichkeit, unter den als **positiv Getesteten** (100%) einen **Gesunden** anzutreffen.

P(infiziert positiver Test)=	0,3799279
P(gesund positiver Test)=	0,6200721
Testfeld	202425
davon positiv & infiziert	2450
davon positiv & gesund	3999
Summe	6449

Abbildung 15: Ergebnisse des zweiten Tests

Bei dem reduzierten Testfeld von 202.425 Personen (100%) ergäben sich demnach **6449** als positiv getestete Personen, wovon jetzt **2450** tatsächlich infiziert sind, während der Rest von **3999** Personen gesund ist. Um unsere Sicherheiten weiter zu verbessern, sollten wir in einen dritten Test investieren, indem wir die verbleibenden als infiziert klassifizierten Personen, das sind inzwischen nur noch **6449**, erneut testen.

3. Test: Eingangsdaten		positiver Test	negativer Test
0,3799279	infiziert	0,9900000	
	gesund	0,0200000	

Abbildung 16: Eingangsdaten für den dritten Test

3. Test		positiver Test	negativer Test
0,3799279	infiziert	0,9900000	0,0100000
		0,3761286	0,0037993
0,6200721	gesund	0,0124014	0,9800000
		0,0200000	0,9800000
		0,3885301	0,9837993

Abbildung 17: komplettierte Vier-Felder-Tafel zum 3. Test

P(infiziert positiver Test) =	0,9680811
P(gesund positiver Test) =	0,0319189
Testfeld	6449
davon positiv & infiziert	2426
davon positiv & gesund	80
Summe	2506

Abbildung 18: Ergebnisse des dritten Tests

Bei dem im dritten Durchgang auf 6449 Personen (100%) reduzierten Testfeld ergäben sich **2506** als positiv getestete Personen, wovon jetzt **2426** tatsächlich infiziert sind, während der Rest von **80** Personen gesund ist. Mit der Sicherheit von knapp **3%** für positive, falsche Ergebnisse geben wir uns zufrieden. Nach dem dritten Test ist es also gelungen, **2506** Personen als infiziert einzustufen. Darunter befinden sich nur noch **80** falsche, positive Befunde.

Bei allen drei Testergebnissen liegt die Anzahl der korrekt positiv Getesteten durchgängig bei ca. **2500** Personen und zwar unabhängig vom Umfang des Testfeldes, was ja im ersten Durchgang immerhin ein Volumen von 10 Millionen Personen aufwies. Dagegen ließ sich die Anzahl der fälschlicherweise positiv Getesteten nach jedem Test erheblich eingrenzen. Hier reduzierte sich das Volumen fast dramatisch von **199.950** über **3999** auf letztendlich **80** Personen.

rm