1. Mogeln I

Hast Du bei der Klassenarbeit gemogelt?

Es wird ein Stapel Spielkarten eingesetzt. Die Schüler/innen werden aufgefordert, falls sie eine Kreuz-Karte ziehen, ehrlich die Frage zu beantworten "Hast du bei der Klassenarbeit gemogelt?". Falls sie irgendeine andere Karte ziehen, beantworten sie eine harmlose Frage, die unvermeidlich mit "Ja" zu beantworten ist, wie z. B. "Gehst Du zur Schule?".

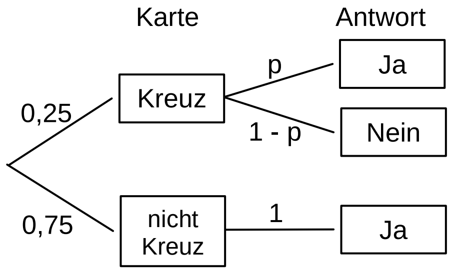
a1) Der normale Fall: Man kennt den Anteil p der Mogler nicht. Aber bei der Befragung melden sich 80 % Ja-Sager. Wie hoch ist der Mogler-Anteil zu schätzen?
a2) Wie wird bei diesem Verfahren der ehrliche Ja-Antworter anonymisiert?

b1) Kann es die Umfrage-Ergebnisse 90 %, 60 %, 70 %, 85 % geben?
b2) Leite die allgemeine Formel zur Berechnung von P(Mogel) aus dem Ja-Anteil der Umfrage her. Prüfe sie anhand der Beispiele oben.
b3) Welche Umfrage-Ergebnisse sind überhaupt möglich? Zeige das am Baumdiagramm und an der Formel.
b4) Wie ändert sich die Bandbreite möglicher Umfrage-Ergebnisse, wenn "Kreuz-Karte" durch "schwarze Karte" ersetzt wird?

c1) Bei der Berechnung hat sich ein Mogler-p von 60 % ergeben. Mit welcher Wahrscheinlichkeit kann man aus einem Ja auf ein ehrliches Mogler-Ja schließen?
c2) Wenn 80 % der Befragten mogeln, mit welcher Wahrscheinlichkeit kann man sie dann aufgrund eines "Ja" identifizieren?

d1) Wie ändert sich die Identifizierungswahrscheinlichkeit, wenn in dem Verfahren zu c1) "Kreuz-Karte" durch "schwarze Karte" ersetzt wird?
d2) Die ehrliche Antwort wird verlangt bei "Kein Ass", bei Ass ein "Ja". …
d3) Auf die Spitze getrieben: Die ehrliche Antwort wird verlangt bei "kein Kreuz-Ass". …
d4) Diskutiere den Gegensatz der beiden Forderungen: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungswahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1), d1), d2), d3) zusammen.

a1)



$0{,}25\,p + 0{,}75 = 0{,}8 \Leftrightarrow p = 0{,}2$

Es gibt unter den Befragten 20 % Mogler.

a2) Es ist nicht zu unterschieden, ob ein "Ja" eine ehrliche Antwort auf die Frage oder ein automatisches "Ja" ist. Insofern ist der ehrliche Ja-Anwortende nicht zu erkennen.

b1) 60 % und 70 % sind als Umfrage-Ergebnisse nicht möglich, da der untere Pfad schon ein Ergebnis von 75 % liefert. 85 % und 90 % wären möglich.

b2) $u = 0{,}25\,p + 0{,}75$

$p = 4 \cdot (u - 0{,}75)$

$ = 4u - 3$

Beispiele von oben:

$u = 80\,\% \Rightarrow p = 0{,}2$

$u = 90\,\% \Rightarrow p = 0{,}6$

$u = 85\,\% \Rightarrow p = 0{,}4$

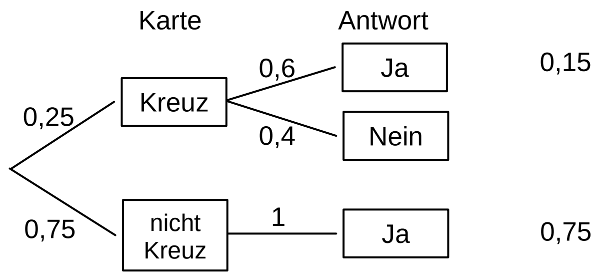
$u = 60\,\% \Rightarrow p = -0{,}6 \rightarrow$ unmöglich

$u = 70\,\% \Rightarrow p = -0{,}2 \rightarrow$ unmöglich

b3)
- ★ Formel: $p = 0 \Rightarrow u = 0{,}75$; $p = 1 \Rightarrow u = 1$
- ★ Baumdiagramm: Wegen des unteren Astes gibt es immer 75 % Ja-Antworten oder mehr. Durch den oberen Ast kommen maximal 25 % dazu.
- ★ Die Umfragewerte liegen zwischen 75 % und 100 %.

b4) Da die Hälfte der Spielkarten schwarz ist, steht an den ersten beiden Ästen im Baumdiagramm jetzt jeweils 0,5 statt 0,25|0,75.
- ★ Formel: $0{,}5 \cdot p + 0{,}5 \cdot 1 = u$
  $p = 0 \Rightarrow u = 0{,}5$
  $p = 1 \Rightarrow u = 1$
- ★ Baumdiagramm: Wegen des unteren Pfades gibt es immer mindestens 50 % Ja-Stimmen. Durch den oberen Pfad zum Ja kommen maximal 50 % hinzu.
- ★ Die Umfragewerte liegen breiter als in b3): zwischen 50 % und 100 %.

c1)



Befragt man auf diese Weise 1000 Personen, so werden 150 ehrlich "Ja" antworten, 750 wegen der gezogenen Nicht-Kreuz-Karte. Der Anteil der ehrlichen Ja-

Sager an der Gesamtzahl beträgt $\frac{150}{150+750} \approx 16{,}7\ \%$.
Nur mit geringer Wahrscheinlichkeit lässt sich aus einer Ja-Antwort auf ein ehrliches Ja rückschließen. Der Anonymisierungsgrad dieses Verfahrens ist sehr hoch.
Zu Bayes: Statt mit den absoluten Zahlen kann man auch sofort mit den berechneten Wahrscheinlichkeiten für "Ja" die berechneten Wahrscheinlichkeiten für "Ja" die bedingte Wahrscheinlichkeit berechnen:

$P(\text{Mogler}|\text{Ja}) = \frac{0{,}15}{0{,}15+0{,}75} \approx 16{,}7\ \%$

Allgemein: $P = \frac{\text{"günstige Wahrscheinichkeit" für Mogler und Ja}}{\text{"mögliche Wahrscheinlichkeiten für Ja"}}$

c2) Die Astwahrscheinlichkeiten nach der Kreuzkarte in c1) lauten jetzt 0,8/0,2 statt 0,6/0,4. Als Pfadwahrscheinlichkeit ergibt sich rechts oben 0,2 statt 0,15.

$P(\text{Mogler}|\text{Ja}) = \frac{0{,}2}{0{,}2+0{,}75} \approx 21{,}1\ \%$

Die Identifizierungswahrscheinlichkeit liegt mit 21,1 % niedrig.

d1) Pfadwahrscheinlichkeit: P(schwarze Karte und Mogler-Ja) = 0,5 · 0,6 = 0,3
Pfadwahrscheinlichkeit: P(rote Karte und Ja) = 0,5 · 1 = 0,5

$P(\text{Mogler}|\text{Ja}) = \frac{0{,}3}{0{,}3+0{,}5} \approx 37{,}5\ \%$

Der Anonymisierungsrad ist immer noch gut, wenn auch geringer als in c1).

d2) P(kein Ass und Mogler-Ja) = $\frac{28}{32} \cdot 0{,}6 = 52{,}5\ \%$

P(Ass und Ja) = $\frac{4}{32} \cdot 1 = 12{,}5\ \%$

$P(\text{Mogler}|\text{Ja}) = \frac{52{,}5\ \%}{52{,}5\ \% + 12{,}5\ \%} \approx 80{,}8\ \%$

Der Anonymisierungsgrad ist deutlich schlechter als in c2) und c3), da sehr viele Ja-Antworten aus dem "ehrlichen Pfad" stammen.

d3) P(kein Kreuz und Mogler-Ja) = $\frac{31}{32} \cdot 0{,}6 \approx 58{,}1\ \%$

P(Kreuz-Ass und Ja) = $\frac{1}{32} \cdot 1 \approx 3{,}1\ \%$

$P(\text{Mogler}|\text{Ja}) = \frac{58{,}1\%}{58{,}1\% + 3{,}1\%} \approx 94{,}9\ \%$

Die Anonymisierung ist kaum noch gewährleistet. Mit hoher Sicherheit lässt sich aus einem "Ja" auf einen Mogler rückschließen. Die Befragten werden vermutlich nicht mehr ehrlich antworten.

d4)

Verfahren (zu p = 60 %)	c1	d1	d2	d3
Anteil ehrlich Antwortender	25 %	50 %	87,5 %	96,9 %
Identifizierungswahrscheinlichkeit	16,7 %	37,5 %	80,8 %	94,9 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

2. Mogeln II

Hast Du bei der Klassenarbeit gemogelt?

Aus einem Stapel gewöhnlicher Spielkarten zieht jede/r Schüler/in verborgen eine zufällige Karte und steckt sie gleich wieder zurück. Nach jedem Ziehen wird kurz neu gemischt. Die Spielfarbe Kreuz bedeutet Mogler, die anderen Farben stehen für Nicht-Mogler. Jeder Schüler und jede Schülerin ist dann aufgefordert, die folgende Frage mit "Ja" oder "Nein" zu beantworten: "Passt die Karte zu der Antwort, die du ehrlicherweise geben würdest?"

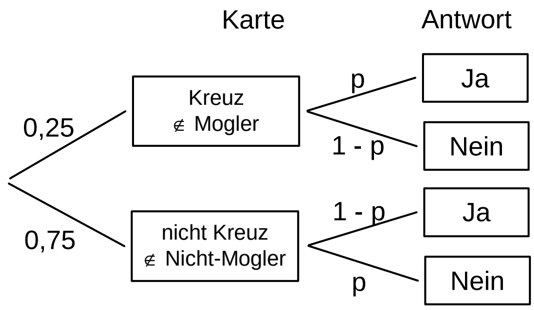
a1) Wenn es 40 % Ja-Stimmen bei einer Befragung gibt, mit welchem Prozentsatz an Moglern ist dann zu rechnen?
a2) Wie wird bei diesem Verfahren sicher gestellt, dass man von einem "Ja" nicht auf "Mogler" schließen kann?

b1) Notiere eine allgemeine Formel zur Berechnung des p-Wertes für Mogeln aus dem Ja-Umfrage-Ergebnis u. Prüfe deine Formel mit den Daten aus a1).
b2) Welche Umfrage-Ergebnisse kann es maximal, welche minimal geben?
b3) Was ergibt sich in a1), wenn unbeobachtet eine Münze geworfen wird und Zahl für Mogler, Wappen für Nicht-Mogler steht?
b4) Durch zweimaligen Münzwurf kann man das oben beschriebene Verfahren doch durchführen. Wie muss eine Regel lauten?

c1) Aus einer Befragung hat sich ein Mogler-p von 70 % ergeben. Mit welcher Wahrscheinlichkeit kann man aus einer Ja-Antwort auf ein ehrliches Mogler-Ja rückschließen?
c2) Welche Identifizierungswahrscheinlichkeit ergibt sich bei 30 % Moglern?

d1) Wie ändern sich Identifizierungswahrscheinlichkeiten, wenn in dem Verfahren bei Nicht-Kreuz-Karte ehrlich geantwortet werden soll?
d2) Wie ändern sich Identifizierungswahrscheinlichkeiten, wenn in dem Verfahren bei Nicht Kreuz-Ass ehrlich geantwortet werden soll?
d3) Diskutiere den Gegensatz der beiden Forderungen an: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungs-Wahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1), d1), d2) zusammen.

a1)



$0{,}25 \cdot p + 0{,}75\,(1 - p) = 0{,}4 \Leftrightarrow p = 0{,}7$

70 % der Befragten sind Mogler.

a2) Das Ja kann Mogler (bei Kreuz-Karte) und Nicht-Mogler (bei anderer Karte) bedeuten. Solange die gezogene Karte unbekannt ist, bleibt die Zuordnung unbekannt.

b1)
$$\begin{aligned} u &= 0{,}25\,p + 0{,}75\,(1 - p) \\ &= 0{,}75 - 0{,}5\,p \\ p &= 2 \cdot (0{,}75 - u) \\ &= 1{,}5 - 2 \cdot u \end{aligned}$$

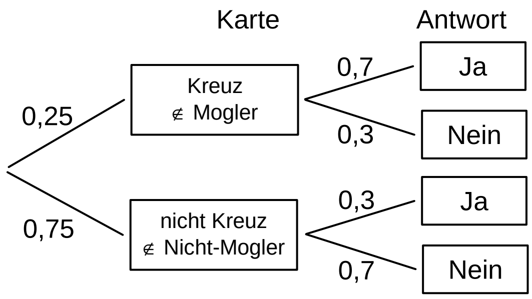
Beispiel von oben:
$u = 0{,}4 \Rightarrow p = 1{,}5 - 2 \cdot 0{,}4 = 0{,}7$

b2)
- ★ Formel:
  $p = 0 \Rightarrow u = 0{,}75$
  $p = 1 \Rightarrow u = 0{,}25$
- ★ Baumdiagramm:
  Ist p = 0, so ergibt sich aus dem oberen "Ja-Pfad" kein Beitrag, aus dem unteren 75 %.
  Ist p = 1, liefert der untere "Ja-Pfad" nichts, der obere 25 %.
- ★ Die Umfragewerte-Ergebnisse liegen zwischen 25 % und 75 %.

b3) In dem Baumdiagramm ist an den ersten beiden Ästen jeweils 0,5 einzutragen statt 0,25|0,75.
$0{,}5 \cdot p + 0{,}5 \cdot (1 - p) = 0{,}4$
ist nicht lösbar, da p auf der linken Gleichungsseite verschwindet. Das Verfahren ist untauglich zur p-Bestimmung.

b4) Wirft man 2-mal eine Münze, so bedeutet WW Mogler, die anderen Fälle Nicht-Mogler; der Rest wie im Kasten.

c1)



17,5 %

22,5 %

$P(\text{Mogler}|\text{Ja}) = \frac{17{,}5\,\%}{17{,}5\,\% + 22{,}5\,\%} = 43{,}75\,\%$

Die Wahrscheinlichkeit, dass sich hinter einer Ja-Antwort ein Mogler verbirgt, beträgt 43,75 %. Die Anonymisierung ist gewährleistet, da sie unter 50 % liegt.

c2) In dem Baumdiagramm zu c1) sind die Astwahrscheinlichkeiten zu tauschen:
0,7/0,3 wird ersetzt durch 0,3/0,7 und unten umgekehrt.
Der obere Ja-Pfad liefert dann 0,25 · 0,3 = 0,075, der untere 0,75 · 0,7 = 0,525.

$P(\text{Mogler}|\text{Ja}) = \frac{7{,}5\,\%}{7{,}5\,\% + 52{,}5\,\%} = 12{,}5\,\%$

Nur in einem Achtel der Fälle steckt hinter einem "Ja" ein Mogler. Damit sind Mogler bei diesem Verfahren und diesem p-Wert sehr gut "versteckt".

d1) Nicht-Kreuzkarte
Die Wahrscheinlichkeiten an den beiden Eingangsästen sind zu tauschen.

$P(\text{Mogler}|\text{Ja}) = \frac{0{,}75 \cdot 0{,}7}{0{,}75 \cdot 0{,}7 + 0{,}25 \cdot 0{,}3} = 87{,}5\,\%$

d2) Nicht Kreuz-Ass

Die Eingangswahrscheinlichkeiten lauten $\frac{31}{32} / \frac{1}{32}$ statt 0,25/0,75.

$P(\text{Mogler}|\text{Ja}) = \frac{31/32 \cdot 0{,}7}{31/32 \cdot 0{,}7 + 1/32 \cdot 0{,}3} \approx 98{,}6\,\%$

Die Identifizierungswahrscheinlichkeitnimmt in d1) auf 87,5 % und d2) auf 98,6 % zu. Die Anonymisierung ist nicht mehr gewährleistet.

d3)

Verfahren (zu p = 70 %)	c1	d1	d2
Anteil ehrlich Antwortender	25 %	50 %	96,9 %
Identifizierungswahrscheinlichkeit	43,75 %	87,5 %	98,6 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

3. Rauchen

Rauchst du durchschnittlich mehr als 4 Zigaretten pro Tag?

Wirf einen Würfel, bevor du antwortest. Beachte dabei, dass die Augenzahl von anderen nicht gesehen werden kann.
Ist die Augenzahl 1, 2, 3 oder 4, so beantworte die Frage wahrheitsgemäß mit "Ja" oder mit "Nein".
Ist die Augenzahl 5, dann antworte unabhängig vom Frageinhalt mit "Ja".
Ist die Augenzahl 6, dann antworte unabhängig vom Frageinhalt mit "Nein".

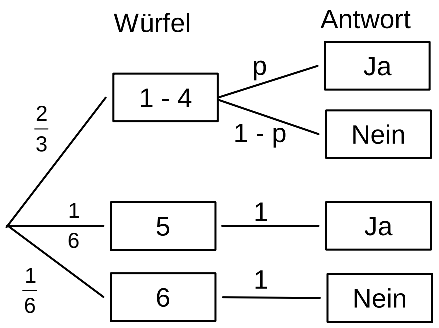
a1) Bei einer Befragung antworten 35 % der Personen mit Ja. Welches p für Rauchen lässt sich schätzen?
a2) Wie wird in dem Frageverfahren abgesichert, dass niemand als Raucher zu identifizieren ist?

b1) Wie hängt der p-Wert für Rauchen vom Umfragewert u für "Ja" ab? Teste das Umfrageergebniss für "Ja" aus a1) und u = 90 %.
b2) Welche Werte für u sind überhaupt möglich? Argumentiere mit der Formel und mit dem Baumdiagramm.

c1) Bei einer Befragung ließ sich ein Raucher-p von 65 % errechnen. Mit welcher Wahrscheinlichkeit lässt sich aus einem "Ja" ein ehrliches Raucher-Ja identifizieren.
c2) Welche Identifizierungswahrscheinlichkeit ergibt sich, wenn 50 % rauchen?

d1) Wie ändern sich die Identifizierungswahrscheinlichkeiten, wenn in dem Verfahren die drei Antwortfälle auf die Würfelzahlen 1/2, 3/4, 5/6 verteilt werden?
d2) Wie ändern sich die Identifizierungswahrscheinlichkeiten, wenn in dem Verfahren die ehrliche Antwort auf 1 - 5, die Ja-Antwort auf 6 gegeben wird (automatisches Nein entfällt)?
d3) Diskutiere den Gegensatz der beiden Forderungen: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungswahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1), d1), d2 zusammen.

a1)



$\frac{2}{3} \cdot p + \frac{1}{6} \cdot 1 = 0{,}35 \Rightarrow p = 0{,}275$

Es ist mit 27,5 % Rauchern zu rechnen.

a2) Eine Ja-Antwort kann von einem Raucher stammen oder seinen Grund in der gewürfelten 5 haben. Das ist bei verdecktem Würfelwurf nicht auseinander zu halten.

b1) $\frac{2}{3}\ p + \frac{1}{6} = u \Leftrightarrow \frac{2}{3}\ p = u - \frac{1}{6} \Leftrightarrow p = 1{,}5\ u - 0{,}25$

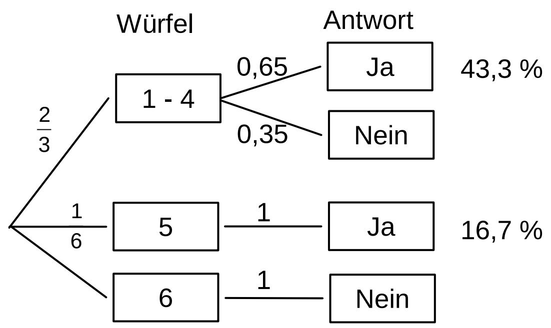
Beispiele von oben:

$u = 35\ \% \Rightarrow p = 0{,}275$

$u = 90\ \% \Rightarrow p = 1{,}10$. Das ist unsinnig.

b2)
- ★ Formel: $p = 0 \Rightarrow u = \frac{1}{6}$; $p = 1 \Rightarrow u = \frac{5}{6}$
- ★ Baumdiagramm:
  Die gewürfelte 5 sorgt für mindestens $\frac{1}{6}$ Ja-Antworten. Durch 1 - 4 kommen maximal $\frac{2}{3}$ dazu und es ergibt sich $\frac{5}{6}$.
- ★ Der Anteil der Ja-Antworten in der Befragung kann zwischen 16,7 % und 83,3 % liegen.

c1)



$$P(\text{Raucher}|\text{Ja}) = \frac{43{,}3\ \%}{43{,}3\ \% + 16{,}7\ \%} \approx 72{,}2\ \%$$

Mit einer Wahrscheinlichkeit von 72,2 % ist von einem "Ja" auf einen Raucher zu schließen. Der Anonymisierungsgrad ist nicht hoch.

c2) Im Baumdiagramm zu c1) ist 0,65/0,35 durch 0,5/0,5 zu ersetzen. Die obere Pfadwahrscheinlichkeit beträgt dann 33,3 %.

P(Raucher|Ja) = $\frac{33{,}3\,\%}{33{,}3\,\% + 16{,}7\,\%} \approx 66{,}7\,\%$

Mit einer Wahrscheinlichkeit von $\frac{2}{3}$ kann man aus einem Ja auf Rauchen schließen. Der Anonymisierungsgrad ist zwar besser als in c1), liegt aber noch deutlich über 50 %.

d1) Die 3 Eingangswahrscheinlichkeiten ändern sich zu $\frac{1}{3} / \frac{1}{3} / \frac{1}{3}$.

P(Raucher|Ja) = $\frac{\frac{1}{3} \cdot 0{,}65}{\frac{1}{3} \cdot 0{,}65 + \frac{1}{6} \cdot 1} \approx 56{,}5\,\%$

d2) Der untere Baumdiagramm-Pfad entfällt, die anderen haben die Eingangswahrscheinlichkeiten $\frac{5}{6} / \frac{1}{6}$.

P(Raucher|Ja) = $\frac{\frac{5}{6} \cdot 0{,}65}{\frac{5}{6} \cdot 0{,}65 + \frac{1}{6} \cdot 1} \approx 76{,}5\,\%$

Die Identifizierungswahrscheinlichkeit ist im ersten Fall mit 56,5 % in Ordnung (fast "fifty-fifty"), im zweiten Fall mit 76,5 % zu hoch.

d3)

Verfahren (zu p = 65 %)	c1	d1	d2
Anteil ehrlich Antwortender	66,7 %	33,3 %	83,3 %
Identifizierungswahrscheinlichkeit	72,2 %	56,5 5	76,5 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

4. Steuererklärung

Eine Befragung von Erwachsenen:
Haben Sie schon einmal in Ihrer Steuererklärung eine nicht wahrheitsgemäße Angabe gemacht, um mehr Geld zurückzubekommen?
Ehe Sie antworten, werfen Sie verdeckt einen Würfel. *Ist die Augenzahl 1 oder 2, dann antworten Sie ganz unabhängig von der Wahrheit mit "Ja" – egal, ob das stimmt.* *Ist die Augenzahl 3 oder 4, dann antworten Sie unabhängig von der Wahrheit standhaft mit "Nein" – auch egal, ob es stimmt.* *Ist die Augenzahl 5 oder 6, dann antworten Sie ehrlich.*

a1) Das p für Steuerschwindler kennt man nicht, deshalb macht man die Befragung. Wenn 60 % "Ja" antworten, welcher Anteil p von Schwindlern ist dann zu schätzen?
a2) Das Verfahren ist offen, aber anonym. Wieso?

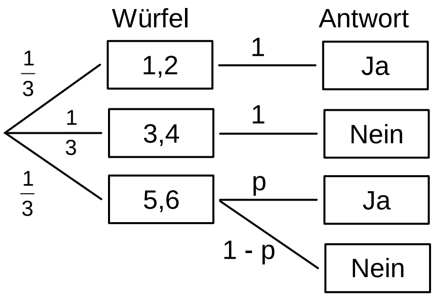
b1) Der Anteil der Ja-Antworten beträgt 80 %. …
b2) Welche Umfragewerte können überhaupt vorkommen?

c1) Wenn eine bestimmte Person mit Ja antwortet, lässt sich nicht mit Sicherheit sagen, ob sie schwindelt. Aber mit welcher Wahrscheinlichkeit hat sie dann ein ehrliches Schwindler-Ja gesagt im Fall a1) (p = 80 %)?
c2) Sind unter 2000 Befragten 800 Steuerbetrüger, mit welcher Wahrscheinlichkeit sind sie dann durch ein "Ja" zu identifizieren?

d1) Verfahrensänderungen bewirken andere Identifizierungswahrscheinlichkeiten (zu p = 0,8 wie in a1/c1):
Die Würfelergebnisse werden anders verteilt: 1; 2; 3 - 6
d2) Der mittlere Ast entfällt. Es bleibt 1; 2 - 6.
d3) Diskutiere den Gegensatz der beiden Forderungen: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungswahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1), d1), d2) zusammen.

e1) Wie groß ist die in c1) gefragte Wahrscheinlichkeit für eine Schwindlerquote p?
e2) Welchen Wert nimmt die Wahrscheinlichkeit in e1) maximal an? Erläutere das auch am Verfahren.
e3) Untersuche entsprechend die Abhängigkeit der Identifizierungswahrscheinlichkeit vom gewählten Verfahren – siehe c1), d1), d2). Beurteile die Verfahren.

a1)



$P(\text{Ja}) = \frac{1}{3} + \frac{1}{3} \cdot p = 0{,}6$

$\frac{1}{3}\,p = 0{,}6 - \frac{1}{3}$

$p = 1{,}8 - 1$

$p = 0{,}8$

Man kann den Anteil der Steuer-schwindler auf rund 80 % schätzen.

a2) Das "Ja" kann wegen der Würfel 1/2 kommen oder ehrlich. Es lässt sich nicht rück-verfolgen.

b1) $\frac{1}{3} + \frac{1}{3}\,p = 0{,}8 \Rightarrow p = 1{,}4$. Das ist unsinnig.

b2) ★ Argumentation mit der Formel: u $\notin$ Ja-Stimmenanteil im Umfrage-Ergebnis

$\frac{1}{3} + \frac{1}{3}\,p = u$

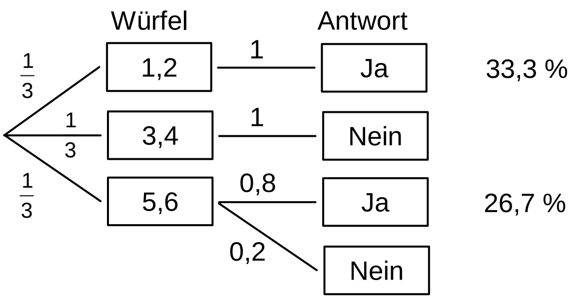
Ist p = 0 (es mogelt niemand), so gilt $u = \frac{1}{3}$. Mogeln alle (p = 1), so ist $u = \frac{2}{3}$.

Es können nur Ja-Antworten zwischen $\frac{1}{3}$ und $\frac{2}{3}$ vorkommen, 80 % ist nicht möglich. (VORSICHT: u = 0,8 ist nicht möglich, p = 0,8 sehr wohl – siehe a1.)

★ Argumentation mit dem Baumdiagramm:

$\frac{1}{3}$ Ja-Antworten gibt es immer von dem Würfel-1,2-Pfad. Hinzu kommt maximal (p = 1) ein weiteres Drittel über den Würfel-5,6-Pfad. Insgesamt liegt u zwischen $\frac{1}{3}$ und $\frac{2}{3}$.

c1)



$P(\text{Betrüger}|\text{Ja}) = \frac{26{,}7\,\%}{26{,}7\,\% + 33{,}3\,\%} = 44{,}5\,\%$

Ein Ja weist mit einer Wahrscheinlichkeit von 44,5 % auf einen Betrüger hin. Der Anonymisierungsgrad ist in Ordnung (unter fifty-fifty).

c2) Im Baumdiagramm zu c1) ist 0,8/0,2 durch 0,4/0,6 zu ersetzen. Der zweite Ja-Pfad führt auf $\frac{1}{3} \cdot 0{,}4 \approx 13{,}3\,\%$.

$P(\text{Betrüger}|\text{Ja}) = \frac{13{,}33\,\%}{13{,}33\,\% + 33{,}33\,\%} \approx 28{,}6\,\%$

Die Wahrscheinlichkeit der Identifikation auf Grund eines "Ja" ist klein. Die Anonymität ist gesichert.

d1) Die Eingangswahrscheinlichkeiten betragen jetzt $\frac{1}{6}/\frac{1}{6}/\frac{4}{6}$.

$$P(\text{Betrüger}|\text{Ja}) = \frac{\frac{4}{6}\cdot 0{,}8}{\frac{4}{6}\cdot 0{,}8+\frac{1}{6}\cdot 1} \approx 76{,}2\ \%$$

d2) Die beiden Eingangswahrscheinlichkeiten lauten hier $\frac{1}{6}/\frac{5}{6}$.

$$P(\text{Betrüger}|\text{Ja}) = \frac{\frac{5}{6}\cdot 0{,}8}{\frac{5}{6}\cdot 0{,}8+\frac{1}{6}\cdot 1} \approx 80{,}0\ \%$$

Der Identifizierungsgrad nimmt deutlich auf 76,2 % bzw. 80,0 % zu. Die Anonymität ist nicht mehr gut gewährleistet.

d3)

Verfahren (zu p = 80 %)	c1	d1	d2
Anteil ehrlich Antwortender	33,3 %	50 %	83,3 %
Identifizierungswahrscheinlichkeit	44,5 %	76,2 %	80,0 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

e1) $P(\text{Betrüger}|\text{Ja}) = \frac{\frac{1}{3}\cdot p}{\frac{1}{3}\cdot p+\frac{1}{3}} = \frac{p}{p+1}$

e2) Ist p = 1, so ist ein Ja-Antworter zu 50 % auch ein Mogler. Das ergibt sich aus dem Verfahren: $\frac{1}{3}$ der Befragten sagt Ja wegen der 1 oder 2 beim Würfeln, $\frac{1}{3}$ wegen Betrugs, $\frac{1}{3}$ sagt Nein. Die Hälfte der Ja-Sager sind die in der Befragung ehrlichen Betrüger.
Das Verfahren gewährleistet immer die Anonymität, da die Identifikationswahrscheinlichkeit maximal 50 % beträgt.

e3)

Verfahren	c1	d1	d2
Anteil ehrlich Antwortender	33,3 %	50 %	83,3 %
Identifizierungswahrscheinlichkeit zur Schwindlerquote p	$\frac{p}{p+1}$	$\frac{4p}{4p+1}$ 1)	$\frac{5p}{5p+1}$
maximale Identifizierungswahrscheinlichkeit	50 %	80 %	83,3 %

1) $p(\text{Betrüger}|\text{Ja}) = \frac{\frac{4}{6}\cdot p}{\frac{4}{6}\cdot p+\frac{1}{6}\cdot 1} = \frac{4p}{4p+1}$ (s. o. d1)

Verfahren c1) sichert eine akzeptabel niedrige Identifizierungswahrscheinlichkeit von maximal 50 %. Aber nur ein Drittel der Befragten nehmen "richtig" an der Befragung teil.

Verfahren d1) erhöht die "Teilnehmenden"-Zahl auf 50 %, ist aber nur bis p = 25 % zu gebrauchen, da bis zu dem p-Wert die Identifizierungswahrscheinlichkeit von 50 % nicht überschritten wird: $\frac{4 \cdot 0,25}{4 \cdot 0,25 + 1}$ = 50 %.

Bei Verfahren d2) nehmen 83,3 % "richtig" an der Befragung teil. Aber p-Werte oberhalb von 20 % bewirken eine Identifizierungswahrscheinlichkeit über 50 %, denn $\frac{5 \cdot 0,2}{5 \cdot 0,2 + 1}$ = 50 %.

Liegt eine begründete Vermutung vor über den p-Wert vor der Befragung, so kann das Verfahren gewählt werden, das einen möglichst hohen Anteil ehrlich Antwortender hat und die Identifizierungswahrscheinlichkeit unter 50 % hält.

5. Gammelfleisch

Eine Befragung von Fleischvermarktern:

Beteiligen Sie sich an Verarbeitung und Verkauf von Gammelfleisch?

Werfen Sie außerhalb der Sicht- und Hörweite des Interviewers eine Münze in die Luft. Wenn sie Wappen zeigt, so beantworten Sie bitte die Frage: "Vermarkten Sie Gammelfleisch?" Falls sie aber Zahl zeigt, so werfen Sie die Münze ein zweites Mal und beantworten dann die Frage: "Zeigte die Münze beim zweiten Mal Wappen?"

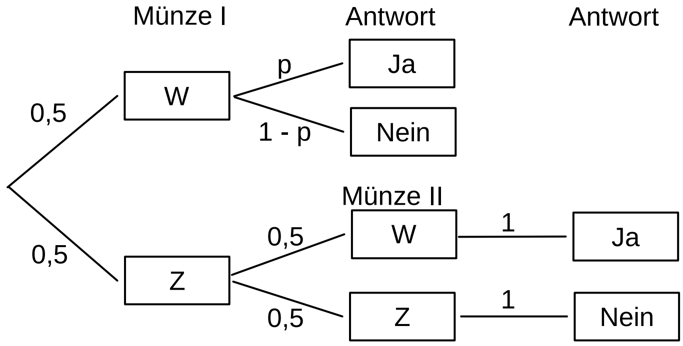
a1) Nimm an, dass 35 % der Antworten "Ja" lauten. Schätze dann den Anteil der Gammelfleisch-Vermarkter.
a2) Ehrlichkeit kann man erwarten, wenn die Antwort anonym ist. Ist sie es mit dem Verfahren?

b1) Welche p ergibt sich bei einem "Ja"-Anteil von 20 %?
b2) Erläutere den Bereich, in dem es überhaupt "Ja"-Anteile u (wegen Umfrage-Ergebnis) geben kann – an der Formel und am Baumdiagramm.
b3) Man könnte doch bei Zahl im ersten Münzwurf einfach die Gegenfrage beantworten lassen: "Beteiligen Sie sich nicht an der Gammelfleisch-Vermarktung?"

c1) Nimm an, der Anteil der Gammelfleisch-Vermarkter sei 50 %. Wie groß ist dann die Wahrscheinlichkeit, dass ein Mensch ein ehrliches Gammelfleisch-Ja geantwortet hat, wenn seine Antwort "Ja" lautete?
c2) Wenn von 80 % der Befragten Gammelfleisch vermarktet wird, mit welcher Wahrscheinlichkeit kann man die Personen dann aus einem "Ja" identifizieren?

d1) Wie ändert sich die Wahrscheinlichkeit in c1), wenn beim 3-fachen Münzwurf bei WWW immer "Ja" geantwortet wird, sonst immer ehrlich?
d2) Wie hoch ist die Identifizierungswahrscheinlichkeit, wenn die Eingangsbedingungen von d1) vertauscht werden.
d4) Diskutiere den Gegensatz der beiden Forderungen: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungswahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1), d1), d2 zusammen.

a1)



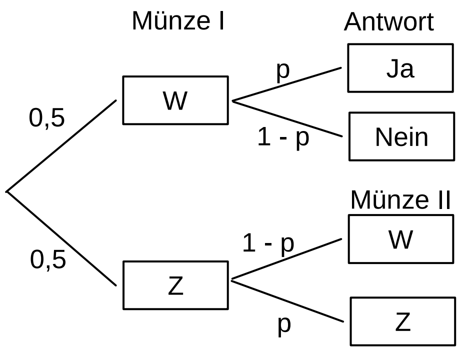
$0{,}5 \cdot p + 0{,}25 = 0{,}35$
$p = 0{,}2$
Der Anteil der Gammelfleisch-Vermarkter ist auf 20 % zu schätzen.

a2) Eine Möglichkeit für ein ehrliches "Ja" stammt aus dem 1. Wappenwurf, ein zweites notwendiges "Ja" aus zweitem Wappenwurf. Aus welchem Grund "Ja" geantwortet wird, ist nicht zurückzuverfolgen.

b1) $0{,}5 \cdot p + 0{,}25 = 0{,}2 \Rightarrow p = -10\ \%$
Das Ergebnis ist unsinnig.

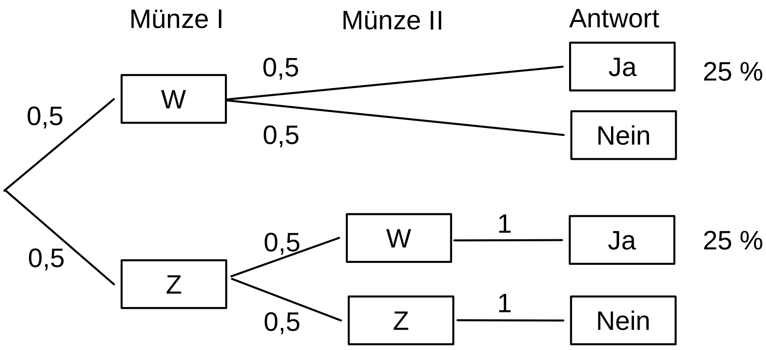
b2)
- ★ Formel: $0{,}5 \cdot p + 0{,}25 = u$
  $p = 0 \Rightarrow u = 0{,}25$; $p = 1 \Rightarrow u = 0{,}75$
- ★ Baumdiagramm:
  Z, W liefert immer 25 % Ja-Antworten
  Ist $p = 1$, kommen maximal $0{,}5 \cdot 1 = 50\ \%$ hinzu.
- ★ Die Umfrage liefert einen Ja-Anteil zwischen 25 % und 75 %.

b3)



$0{,}5\, p + 0{,}5\, (1 - p) = 0{,}35$
ist nicht lösbar, da p auf der linken Gleichungsseite verschwindet. Das Verfahren taugt nicht.

c1)



P(Gammelfleisch-Vermarkter|Ja) = $\frac{25\,\%}{25\,\% + 25\,\%} \approx 50\,\%$

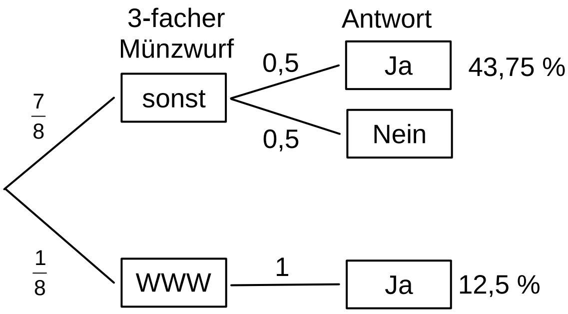
Die Chance, dass hinter einem "Ja" ein Gammelfleisch-Vermarkter steckt, ist fifty-fifty. Die Anonymität ist durch das Verfahren gewahrt.

c2) Im Baumdiagramm zu c1) sind die Ast-Wahrscheinlichkeiten 0,5/0,5 durch 0,8/0,2 zu ersetzen. Als obere Pfadwahrscheinlichkeit ergibt sich dadurch 0,5 · 0,8 = 0,4.

P(Gammelfleisch-Vermarkter|Ja) = = $\frac{40\,\%}{40\,\% + 25\,\%} \approx 61{,}5\,\%$

Mit 61,5 % ist die Identifizierungswahrscheinlichkeit so hoch, dass nicht mehr mit Ehrlichkeit bei allen Befragten zu rechnen ist.

d1)



P(Gammelfleisch-Vermarkter|Ja) = $\frac{43{,}75\,\%}{43{,}75\,\% + 12{,}5\,\%} \approx 77{,}8\,\%$

Die Identifizierungswahrscheinlichkeit ist hoch. Vermutlich werden Befragte lügen, da der Anonymisierungsgrad zu gering ist.

d2) Die Eingangswahrscheinlichkeiten $\frac{7}{8}$ und $\frac{1}{8}$ werden getauscht.

P(Gammelfleisch-Vermarkter|Ja) = $\frac{\frac{1}{8} \cdot 0{,}5}{\frac{1}{8} \cdot 0{,}5 + \frac{7}{8} \cdot 1} \approx 6{,}7\,\%$

Da nur mit sehr geringer Wahrscheinlichkeit ein tatsächlicher Gammelfleisch-Vermarkter aus einer Ja-Antwort identifiziert wird, ist der Anonymisierungsgrad sehr hoch.

d3)

Verfahren (zu p = 50 %)	c1	d1	d2
Anteil ehrlich Antwortender	50 %	87,5 %	12,5 %
Identifizierungswahrscheinlichkeit	50 %	77,8 %	6,7 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

6. Ladendiebe

Hast du schon einmal einen Ladendiebstahl begangen?

Der Kriminologe legt den Befragten einen Stapel mit 10 Kärtchen vor. Auf vier von ihnen steht die Frage 1: "Hast du schon einmal einen Ladendiebstahl begangen?", auf den restlichen die Frage 2: "Ist es richtig, dass du noch keinen Ladendiebstahl begangen hast?" Der Befragte zieht nun eine Karte und beantwortet die dort gestellte Frage ehrlich, ohne dass der Kriminologe weiß, welche er beantwortet hat. Nach Zurückstecken der Karte und Mischen des Stapels zieht der Nächste.

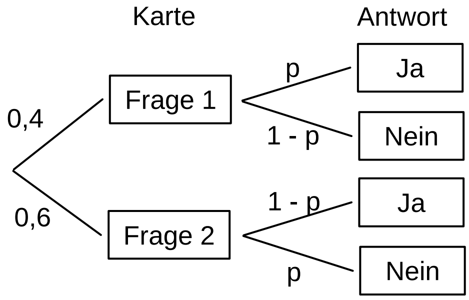
a1) In 47 % aller Fälle erhält der Kriminologe die Antwort "Ja". Wie groß ist der Anteil der Ladendiebe unter den Befragten?
a2) Erläutere: Auch wenn ehrlich geantwortet wird, outet sich der Ladendieb bei dem Verfahren nicht.

b1) 70 % der Befragten antworten "Ja". Welches p für Ladendiebe lässt sich dann schätzen?
b2) In welchen Grenzen ist das Umfrage-Ergebnis u zu erwarten? Begründe mit dem Frageverfahren bzw. dem Baumdiagramm.
b3) Zeige die möglichen Werte für das Umfrage-Ergebnis u auch an der Gleichung, die u mit dem p für Ladendiebe verbindet.
b4) Lässt sich der enge Bereich, in dem Umfrage-Ergebnisse u liegen können, vergrößern? Probiere andere Kärtchen-Aufteilungen.
b5) Könnte man es nicht einfacher machen und nur unbeobachtet eine Münze werfen? Bei Wappen wäre Frage 1, bei Zahl Frage 2 zu beantworten.

c1) Jemand hat in a1) "Ja" geantwortet. Mit welcher Wahrscheinlichkeit verbirgt sich dann hinter dem "Ja" ein ehrliches Ladendieb-Ja? Beurteile die Identifizierungswahrscheinlichkeit bzw. den Anaonymisierungsgrad.
c2) Wie ist die Identifizierungswahrscheinlichkeit zu beurteilen bei einem 35 %-Anteil der Ladendiebe?

d1) Bestimme und beurteile die Identifizierungswahrscheinlichkeit für den Fall, dass auf 9 Karten die Frage 1 steht,
d2) Bestimme und beurteile die Identifizierungswahrscheinlichkeit für den Fall, dass auf einer Karte die Frage 1 steht.
d3) Diskutiere den Gegensatz der beiden Forderungen: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungs-Wahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1), d1), d2) zusammen.

a1) p sei die Wahrscheinlichkeit, auf Frage 1 mit Ja zu antworten. Dann wird Frage 2 mit der Wahrscheinlichkeit 1 - p mit Ja beantwortet.



$P(Ja) = 0{,}4 \cdot p + 0{,}6 \cdot (1 - p) = 0{,}47$
$\Leftrightarrow 0{,}4\,p + 0{,}6 - 0{,}6\,p = 0{,}47$
$\Leftrightarrow 0{,}6 - 0{,}2\,p = 0{,}47$
$\Leftrightarrow p = 0{,}65$

Man muss mit 65 % Ladendieben rechnen!

a2) Es ist nicht auszumachen, ob ein "Ja" als Antwort auf Frage 1 oder Frage 2 gegeben wurde. Insofern outet sich niemand durch ein Ja als Ladendieb.

b1) $0{,}6 - 0{,}2\,p = 0{,}7 \Leftrightarrow p = -50\ \%$. Das ist unsinnig.

b2) Für p = 0 ergibt sich nur über den Frage 2-Ja-Pfad u = 0,6. Dagegen führt p = 1 nur über den Frage 1-Ja-Pfad auf u = 0,4. Das Umfrage-Ergebnis ist zwischen 40 % und 60 % zu erwarten.

b3)
$$\begin{aligned} 0{,}4\,p - 0{,}6\,(1 - p) &= u \\ 0{,}6 - 0{,}2\,p &= u \qquad \text{(I)} \\ -0{,}2\,p &= u - 0{,}6 \\ p &= 3 - 5\,u \qquad \text{(II)} \end{aligned}$$

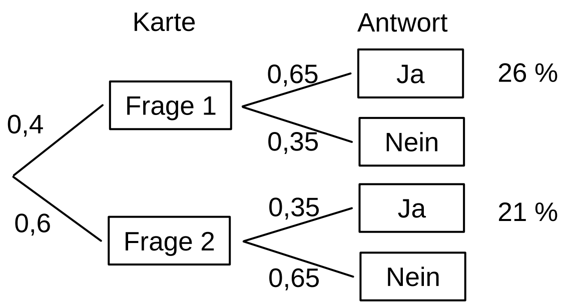
An Gleichung I sind für p = 0 und p = 1 die Grenzen für u abzulesen. Mit Gleichung II lässt sich zu jedem u das zugehörige p berechnen.

b4) Enthält nur eine Karte die Frage 1, die restlichen 9 Karten aber Frage 2, so ergibt sich für "Ja": $0{,}1\,p + 0{,}9\,(1 - p) = u$ bzw. $u = 0{,}9 - 0{,}8\,p$. Damit liegen Umfragewerte zwischen 10 % (für p = 1) und 90 % (für p = 0). Dieselbe Bandbreite ergibt sich für u, wenn die Frage 1/Frage 2-Kartenzahlen getauscht werden.

Dies ist auch schon die größte Bandbreite.
Bei zwei Karten mit Frage 1 erhält man 80 %/20 %, bei 3 Karten 70 %/30 % usw.

b5) Zu "Ja" gehört dann die Gleichung: $0{,}5\,p + 0{,}5\,(1 - p) = u$ bzw. $u = 0{,}5$. Das Umfrage-Ergebnis endet immer bei 50 % und man kann nicht auf p rückschließen. Die Eingangswahrscheinlichkeiten zu Frage 1/Frage 2 dürfen nicht gleich sein, da sich dann die p-Terme aus der Gleichung aufheben. p wäre nicht mehr bestimmbar.

c1)



P(Ladendieb|Ja) = = $\frac{26\,\%}{26\,\% + 21\,\%} \approx 55{,}3\,\%$

Die Identifikationswahrscheinlichkeit liegt mit 55,3 % nur knapp über einer fifty-fifty-Wahrscheinlichkeit und wird vermutlich noch akzeptiert.

c2) Im Baumdiagramm zu c1) ist oben 0,65/0,35 durch 0,35/0,65 zu ersetzen und unten umgekehrt. Als Wahrscheinlichkeit für den oberen Ja-Pfad ergibt sich $0{,}4 \cdot 0{,}35 = 0{,}14$, für den unteren $0{,}6 \cdot 0{,}65 = 0{,}39$.

P(Ladendieb|Ja) = $\frac{14\,\%}{14\,\% + 39\,\%} \approx 26{,}4\,\%$

Mit gut $\frac{1}{4}$ liegt die Identifizierungswahrscheinlichkeit deutlich unter 50 % und ist akzeptabel.

d1) Die Eingangswahrscheinlichkeiten im Baumdiagramm oben lauten 0,9/0,1 statt 0,4/0,6.

P(Ladendieb|Ja) = $\frac{0{,}9 \cdot 0{,}65}{0{,}9 \cdot 0{,}65 + 0{,}1 \cdot 0{,}35} \approx 94{,}4\,\%$

d2) Die ersten beiden Äste im Baumdiagramm erhalten die Wahrscheinlichkeiten 0,1/0,9 statt 0,4/0,6.

P(Ladendieb|Ja) = $\frac{0{,}1 \cdot 0{,}65}{0{,}1 \cdot 0{,}65 + 0{,}9 \cdot 0{,}35} \approx 17{,}1\,\%$

Im Fall d1) liegt die Identifikationswahrscheinlichkeit mit 94,4 % nicht akzeptabel hoch, im Fall d2) ist die Anonymität dagegen sehr gut gewahrt.

d3)

Verfahren (zu p = 65 %)	c1	d1	d2
Anteil ehrlich Antwortender	40 %	90 %	10 %
Identifizierungswahrscheinlichkeit	55,3 %	94,4 %	17,1 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

7. Sexualpraktik I

Nach einem Vortrag über bestimmte Sexualpraktiken,
die gefährlich sind, wird das Publikum befragt:

Haben Sie die geschilderte Sexualpraktik ausgeführt?

Der Interviewer bittet alle, eine Münze aus der Geldbörse zu nehmen und diese verdeckt zu werfen. Zeigt sie Kopf, soll die Person die Frage ehrlich mit "Ja" oder "Nein" beantworten. Erscheint Zahl, soll die betreffende Person in jedem Fall mit "Ja" antworten.

a1) Erläutere, wieso die Anonymität der Befragten gesichert ist, obwohl sie öffentlich antworten?
a2) Welche Wahrscheinlichkeit für das Ausüben der Sexualpraktik lässt sich mit diesem Verfahren schätzen, wenn 620 von 1000 Antworten Ja lauten?

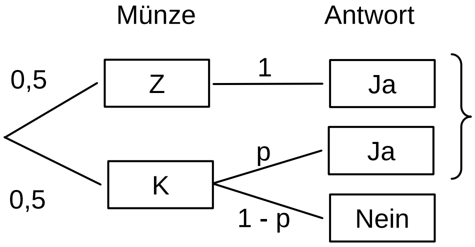
b1) Was ergibt sich als Wahrscheinlichkeit, wenn der Umfragewert 40 % lautet?
b2) Welche Werte kommen als Umfrageergebnis u für die Antwort ja überhaupt in Frage?

c1) Aus einer Ja-Antwort kann man nicht auf einen Sexualpraktiker schließen, das sichert das anonymisierte Verfahren. Aber die Wahrscheinlichkeit, dass sich hinter einem Ja ein ehrliches Sexualpraktik-Ja verbirgt, ist berechenbar. Gehe von 60 % Sexualpraktikern aus. Beurteile die Identifizierungs-Wahrscheinlichkeit.
c2) Ist das Identifizierungsrisiko auch bei 80 % Sexualpraktikern unter den Befragten noch akzeptabel?

d1) Lässt man die Münze 2-mal werfen und bei ZZ "Ja" antworten, sonst ehrlich, so ergibt sich ein anderer Anonymisierungsgrad in c1).
d2) Beim 3-fachen Münzwurf und bei ZZZ "Ja", sonst ehrliches Antworten, …
d3) Notiere jeweils, wie viele Personen von 1000 Befragten bei c1), d1), d2) tatsächlich an der Befragung teilnehmen (ehrlich antworten) und wie hoch jeweils der Anonymisierungsgrad ist.
d4) Diskutiere die gegenläufigen Entwicklungen in ihren Vor- und Nachteilen.

a1) Ein Ja ist (bei verdecktem Münzwurf) nicht eindeutig zuzuordnen, da es zwei Gründe für ein Ja gibt.

a2) Steht p für den Prozentsatz der Personen, die die Sexualpraktik ausüben, so ergibt sich:



$\frac{620}{1000} = 0{,}62$

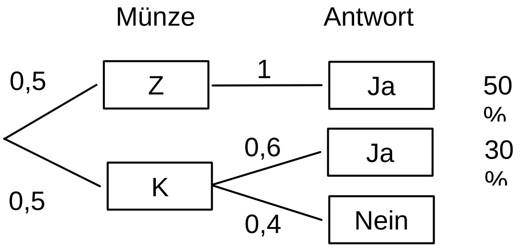
$$0{,}5 + 0{,}5 \cdot p = 0{,}62$$
$$0{,}5 \cdot p = 0{,}12$$
$$p = 0{,}24$$

Aus dem Umfrage-Ergebnis 62 % lässt sich schließen, dass 24 % der Befragten die entsprechende Sexualpraktik ausüben.

b1) $0{,}5 + 0{,}5\,p = 0{,}4$
$p = -20\ \%$ unsinnig

b2) $0{,}5 + 0{,}5\,p = u$
Für p = 0 gilt u = 0,5; für p = 1 ergibt sich u = 1.
Die Umfragewerte liegen zwischen 50 % und 100 %.
Argumentation mit dem Befragungsverfahren bzw. dem Baumdiagramm:
50 % antworten bei Zahl immer mit Ja. Hinzu kommen maximal noch einmal 50 %, falls p = 1.

c1)



P(Sexualpraktiker|Ja) = $\frac{30\ \%}{30\ \% + 50\ \%} = 37{,}5\ \%$.

Da die Identifikationswahrscheinlichkeit deutlich unter 50 % liegt, ist die akzeptabel.

c2) In dem Baumdiagramm zu c1) sind die Astwahrscheinlichkeiten 0,6/0,4 durch 0,8/0,2 zu ersetzen. Die Wahrscheinlichkeit für den zweiten Ja-Pfad errechnet sich dann zu $0{,}5 \cdot 0{,}8 = 0{,}4$.

P(Sexualpraktiker|Ja) = $\frac{40\ \%}{40\ \% + 50\ \%} = 44{,}4\ \%$.

Auch bei hohem Anteil der Sexualpraktiker ist die Identifizierungswahrscheinlichkeit mit 44,4 % noch akzeptabel, da sie die fifty-fifty-Grenze unterschreitet.

d1) Die Eingangswahrscheinlichkeiten im Baumdiagramm lauten jetzt 0,25/0,75 statt 05,/05.

P(Sexualpraktiker|Ja) = $\frac{0{,}75 \cdot 0{,}6}{0{,}75 \cdot 0{,}6 + 0{,}25 \cdot 1} \approx 64{,}3\ \%$

Die Identifikationswahrscheinlichkeit liegt mit 64,3 % deutlich höher als in c1) und führt vielleicht schon zu einigen Lügen.

d2) Eingangswahrscheinlichkeiten lauten jetzt 0,125/0,875 statt 0,5/0,5.

P(Sexualpraktiker|Ja) = $\frac{0{,}875 \cdot 0{,}6}{0{,}875 \cdot 0{,}6 + 0{,}125 \cdot 1} \approx 80{,}8\ \%$

Die Anonymität ist nicht mehr gewährleistet.

d3)

Verfahren (zu p = 60 %)	c1	d1	d2
Anteil ehrlich Antwortender (n = 1000)	500	750	875
Identifizierungswahrscheinlichkeit	37,5 %	64,3 %	80,8 %

d4) Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

8. Sexualpraktik II

Haben Sie die geschilderte Sexualpraktik ausgeführt?

Ein Interviewer bittet die befragten Personen, verdeckt einen Würfel zu werfen. Kommt 1 bis 4, so sollen sie die Frage ehrlich beantworten, ansonsten lügen.

a1) Erläutere, wieso die Anonymität der Befragten gesichert ist, obwohl sie öffentlich antworten?
a2) Welche Wahrscheinlichkeit für das Ausüben der Sexualpraktik lässt sich mit diesem Verfahren schätzen, wenn 620 von 1000 Antworten "Ja" lauten?

b1) Kann man die Ausübungsrate mit diesem Verfahren auch ermitteln, wenn über Ehrlichkeit/Lüge durch Wappen/Zahl beim Münzwurf entschieden wird?
b2) Gibt es eine Möglichkeit mit Münzwürfen dieses Verfahren trotzdem durchzuführen? Notiere ein passendes Baumdiagramm. Erläutere, wie die Ehrlichkeit/Lüge-Wahl erfolgen muss.
b3) Beim Verfahren im Kasten ergeben sich 70 % Ja-Antworten …
b4) Welche Ja-Anteile sind überhaupt möglich? Begründe anhand des Baumdiagramms und mit Hilfe einer Berechnungsformel für p bzw. u ($\notin$ Umfrageergebnis für "Ja").

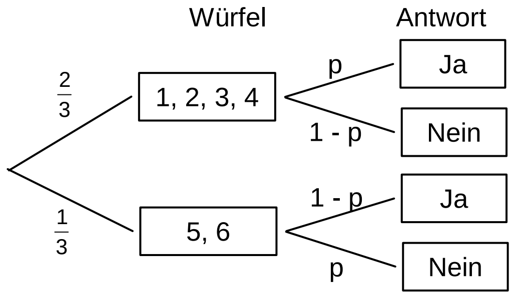
c1) Mit welcher Wahrscheinlichkeit kann man aus der Situation a2) auf ein ehrliches "Ja" zurückschließen?
c2) Wenn unter den Befragten rund 33 % die Sexualpraktik ausüben, wie groß ist dann die Wahrscheinlichkeit auf Grund eines "Ja" identifiziert zu werden?

d1) Welche Identifikationswahrscheinlichkeiten ergeben sich, wenn nur bei einer gewürfelten 1 ehrlich geantwortet werden soll?
d2) Welche Identifikationswahrscheinlichkeiten ergeben sich, wenn bei 1 - 5 ehrlich geantwortet werden soll?
d3) Diskutiere den Gegensatz der beiden Forderungen an: Eine möglichst hohe Zahl tatsächlich Beteiligter unter den Befragten und ein möglichst hoher Anonymisierungsgrad bzw. möglichst kleine Identifizierungs-Wahrscheinlichkeit.
Stelle dazu die entsprechenden Daten aus c1, d1, d2 zusammen.

e1) Wenn der Anteil der Sexualpraktiker im Verfahren a2/c1 95 % beträgt, mit welcher Wahrscheinlichkeit kann man dann aus einem "Ja" auf den Sexualpraktiker zurückschließen?
e2) Wie ist die Situation bei p = 5 %?
e3) Beurteile den Einfluss der p-Größe auf den Anonymisierungsgrad; siehe c1), e1), e2).
e4) Untersuche entsprechend die Abhängigkeit der Identifizierungswahrscheinlichkeit von der p-Größe in den verschiedenen Verfahren – siehe c1), d1), d2). Beurteile die Verfahren.

a1) Der Antwort Ja ist nicht zu entnehmen, ob sie eine ehrliche Antwort auf die gestellte Frage ist oder eine verlangte Lüge ist. Deshalb stellt sie den Antwortenden nicht bloß.

a2)

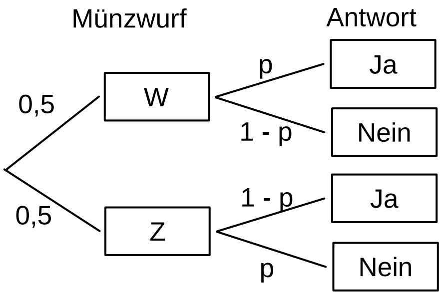


p ∉ Anteil der Personen, die die Sexualpraktik ausüben

$$\frac{2}{3}\,p + \frac{1}{3}\,(1 - p) = 0{,}62$$
$$\frac{1}{3}\,p + \frac{1}{3} = 0{,}62$$
$$p = 0{,}86$$

86 % der Befragten üben die Sexualpraktik aus.

b1)

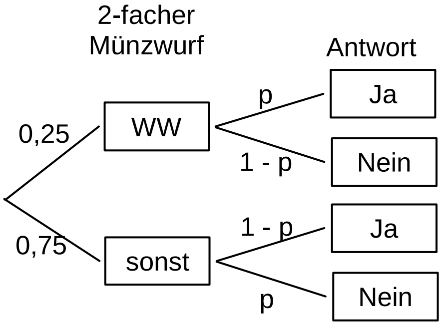


p wie in b); u ∉ Umfrage-Ergebnis

$$0{,}5 \cdot p + 0{,}5 \cdot (1 - p) = u$$
$$0{,}5 = u$$

So ist p nicht zu ermitteln, denn es ergibt sich als Ja-Anteil immer 50 %, unabhängig vom p-Wert. Das Verfahren ist unbrauchbar.

b2) Die Wahrscheinlichkeit für die Entscheidung Ehrlichkeit/Lüge darf nicht gleich groß (50 %) sein. Wirf z. B. eine Münze 2-mal. Kommt Wappen-Wappen, so antworte ehrlich, sonst lüge.
Da es WW,WZ, ZW, ZZ als Ausfälle gibt, folgt:



p, u wie in c)

$$0{,}25 \cdot p + 0{,}75 \cdot (1 - p) = u$$
$$-0{,}5\,p = u - 0{,}75$$
$$p = 1{,}5 - 2\,u$$

b3)
$$\frac{2}{3}\,p + \frac{1}{3}\,(1 - p) = 0{,}7$$
$$p = 1{,}1$$

Das ist unsinnig.

b4) ★ Formel: $\frac{2}{3}\,p + \frac{1}{3}\,(1 - p) = u$

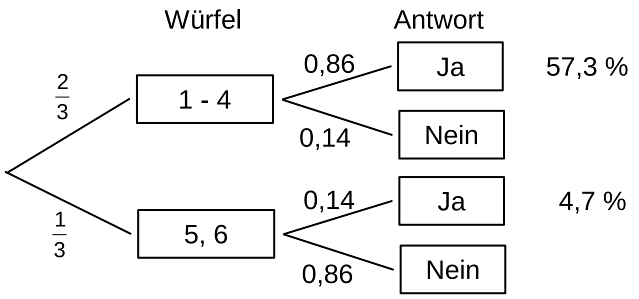
$$\frac{1}{3}\,p + \frac{1}{3} = u$$

$p = 0 \Rightarrow u = \frac{1}{3}$; $p = 1 \Rightarrow u = \frac{2}{3}$

★ Baumdiagramm:
Mit p = 1 liefert der oberste Pfad im Baumdiagramm mit einer Wahrscheinlichkeit von $\frac{2}{3}$ "Ja", der zweite Pfad trägt nichts bei. Zu p = 0 geht der oberste Pfad leer aus, der zweite Ja-Pfad liefert $\frac{1}{3}$.

★ Umfrage-Ergebnisse liegen bei diesem Verfahren zwischen $\frac{1}{3}$ und $\frac{2}{3}$.

c1)



P(Sexualpraktiker|Ja) = $\dfrac{57{,}3\,\%}{57{,}3\,\% + 4{,}7\,\%} \approx 92{,}4\,\%$

Mit zu hoher Wahrscheinlichkeit (92,4 %) ist ein Ja-Sager als Sexualpraktiker zu identifizieren. Es ist zu erwarten, dass Befragte lügen.

c2) Im Baumdiagramm zu c1) sind die Astwahrscheinlichkeiten oben 0,86/0,14 durch 0,33/0,67 zu ersetzen, in den unteren Ästen in entsprechend anderer Reihenfolge.
Als Pfadwahrscheinlichkeit zum oberen "Ja" errechnet man $\frac{2}{3} \cdot 0{,}33 = 22\,\%$, zum unteren "Ja" $\frac{1}{3} \cdot 0{,}66 = 22\,\%$.

P(Sexualpraktiker|Ja) = $\dfrac{22\,\%}{22\,\% + 22\,\%} = 50\,\%$

Die Identifizierungswahrscheinlichkeit von 50 % wird vermutlich von den Befragten noch akzeptiert.

d1) In dem Baumdiagramm lauten die Eingangswahrscheinlichkeiten $\frac{1}{6}/\frac{5}{6}$ statt $\frac{2}{3}/\frac{1}{3}$.

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{1}{6} \cdot 0{,}86}{\frac{1}{6} \cdot 0{,}86 + \frac{5}{6} \cdot 0{,}14} \approx 55{,}1\,\%$

Der Anonymisierungsgrad ist noch okay, da die Identifikationswahrscheinlichkeit nur knapp über der fifty-fifty-Marke liegt.

d2) Im Baumdiagramm werden die Eingangswahrscheinlichkeiten getauscht, also $\frac{5}{6}/\frac{1}{6}$ statt $\frac{2}{3}/\frac{1}{3}$.

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{5}{6}\cdot 0{,}86}{\frac{5}{6}\cdot 0{,}86+\frac{1}{6}\cdot 0{,}14}$ ≈ 96,8 %

Die Identifikationswahrscheinlichkeit ist nicht mehr akzeptabel. Mit sehr hoher Wahrscheinlichkeit ist von einem "Ja" auf einen Sexualpraktiker zu schließen.

d3)

Verfahren (zu p = 86 %)	c1	d1	d2
Anteil ehrlich Antwortender	66,7 %	16,7 %	83,3 %
Identifizierungswahrscheinlichkeit	92,4 %	55,1 %	96,8 %

Je mehr befragte Personen ehrlich antworten, desto höher ist die Wahrscheinlichkeit, sie aus einer Ja-Antwort zu identifizieren bzw. desto geringer ist der Anonymisierungsgrad des Verfahrens.

e1) Die Wahrscheinlichkeiten in der zweiten Stufe lauten jetzt 0,95/0,05 statt 0,86/0,14.

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{2}{3}\cdot 0{,}95}{\frac{2}{3}\cdot 0{,}95+\frac{1}{3}\cdot 0{,}05}$ ≈ 97,4 %

Fast sicher ist ein Ja-Sager Sexualpraktiker. Die Anonymität ist nicht gewahrt.

e2) Die Wahrscheinlichkeiten in der zweiten Stufe des Baumdiagramms lauten jetzt 0,05/0,95 statt 0,86/0,14.

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{2}{3}\cdot 0{,}05}{\frac{2}{3}\cdot 0{,}05+\frac{1}{3}\cdot 0{,}95}$ ≈ 9,5 %

Der Anonymisierungsgrad des Verfahrens ist ausgezeichnet.

e3)

Aufgabenstellung (zum Verfahren im Kasten)	c1	e1	e2
Anteil p der Sexualpraktiker	86 %	95 %	5 %
Identifizierungswahrscheinlichkeit	92,4 %	97,4 %	9,5 %

Je höher der gesuchte Anteil p desto weniger ist die Anonymität der Befragten gewahrt.

e4)

Verfahren	c1	d1	d2
Anteil ehrlich Antwortender	66,7 %	16,7 %	83,3 %
Identifizierungswahrscheinlichkeit zur Sexualpraktiker-Quote p	$\frac{2p}{p+1}$ 1)	$\frac{p}{5-4p}$ 2)	$\frac{5p}{4p+1}$ 3)
maximale Identifizierungswahrscheinlichkeit	100 %	100 %	100 %

1) Siehe c1) bzw. e1).

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{2}{3}\cdot p}{\frac{2}{3}\cdot p+\frac{1}{3}\cdot(1-p)}=\dfrac{2p}{2p+1-p}=\dfrac{2p}{p+1}$

2) Siehe e1).

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{1}{6}\cdot p}{\frac{1}{6}\cdot p+\frac{5}{6}\cdot(1-p)}=\dfrac{p}{p+5-5p}=\dfrac{p}{5-4p}$

3) Siehe d2).

P(Sexualpraktiker|Ja) = = $\dfrac{\frac{5}{6}\cdot p}{\frac{5}{6}\cdot p+\frac{1}{6}\cdot(1-p)}=\dfrac{5p}{5p+1-p}=\dfrac{5p}{4p+1}$

Alle Verfahren haben den Nachteil, dass sie sich bei hohem p-Wert mit der Identifizierungswahrscheinlichkeit der 100 % nähern.
Das ist bei Verfahren, die mit Frage und Gegenfrage (wie bei Nr. 2, 5b3, 6, 8) arbeiten, immer so. Bei hohem p-Wert spielt das Gegenfrage-Ja nur noch eine verschwindend kleine Rolle. Fast alle Befragten "laufen" über den Ehrlich-Pfad und sind damit identifizierbar.
Erwartet man einen hohen gesuchten Anteil, so sind solche reinen Frage/Gegenfrage-Verfahren unbrauchbar.

- ★ Verfahren d1) wäre bis zu p = 83,3 % noch akzeptabel, da dann die Identifizierungswahrscheinlichkeit noch unter 50 % liegt, denn $\frac{0{,}833}{5-4\cdot 0{,}833} \approx 49{,}9\ \%$. Allerdings werden nur $\frac{1}{6}$ der Befragten "echt" befragt.
- ★ Verfahren c1) erfasst mit $\frac{2}{3}$ mehr tatsächlich Befragte. Der p-Wert sollte allerdings 33,3 % nicht überschreiten, da sonst die Identifizierungswahrscheinlichkeit 50 % überschreiten würde, denn $\frac{2\cdot 0{,}333}{0{,}333+1} \approx 50\ \%$.
- ★ Verfahren d2) beteiligt zwar $\frac{5}{6}$ der Befragten ernsthaft an der Auswertung, erlaubt aber nur p-Werte bis 16,7 %, denn $\frac{5\cdot 0{,}167}{4\cdot 0{,}167+1} \approx 50\ \%$.
- ★ Gibt es eine begründete Vermutung über den p-Wert vor der Befragung, so kann das Verfahren gewählt werden, das einen möglichst hohen Anteil ehrlich Antwortender hat und die Identifizierungswahrscheinlichkeit unter 50 % hält. Bei hohen p-Werten taugt das Verfahren überhaupt nicht (s. o.)