Prof. Dr. U. Kamps
C. Empacher, D. Rauwolf

Einführung in die angewandte Stochastik

2. Globalübung - Lösungen

Aufgabe 8

Sei $r_{xy} \in \{-1,1\}$. Dann gilt mit $s_x, s_y > 0$:

$$1 = r_{xy}^2 \overset{\text{A } 7.27}{=} \frac{s_{xy}^2}{s_x^2\, s_y^2} \iff s_{xy}^2 = s_x^2\, s_y^2 \underset{(7.22)}{\overset{(3.29)}{\iff}}$$

$$\left(\sum_{i=1}^{n} \underbrace{(x_i - \overline{x})}_{=:\, a_i} \underbrace{(y_i - \overline{y})}_{=:\, b_i} \right)^2 = \left(\sum_{i=1}^{n} \underbrace{(x_i - \overline{x})^2}_{=\, a_i^2} \right) \left(\sum_{i=1}^{n} \underbrace{(y_i - \overline{y})^2}_{=\, b_i^2} \right)$$

Mit den Setzungen $a_i := x_i - \overline{x}$ und $b_i := y_i - \overline{y}$ für $i \in \{1, \dots, n\}$ existiert somit gemäß Hinweis ein $c \in \mathbb{R}$ mit

$$y_i - \overline{y} = c\,(x_i - \overline{x}) \quad \text{oder} \quad x_i - \overline{x} = c\,(y_i - \overline{y}) \tag{1}$$

für alle $i \in \{1, \dots, n\}$.

Hierbei ist in beiden Fällen $c \neq 0$, denn sonst würde folgen:

$$s_{xy} \overset{(7.22)}{=} \frac{1}{n} \sum_{i=1}^{n} \underbrace{(x_i - \overline{x})\,(y_i - \overline{y})}_{=\,0} = 0 \quad \text{und damit} \quad r_{xy} \overset{\text{A } 7.27}{=} \frac{s_{xy}}{s_x\, s_y} = 0$$

im Widerspruch zu $r_{xy} \in \{-1,1\}$.

Mit $c \neq 0$ ist (1) äquivalent zu

$$y_i = \overline{y} - c\,\overline{x} + c\,x_i = a + b\,x_i \ , \ i \in \{1, \dots, n\} \ , \tag{2}$$

mit $a := \overline{y} - c\,\overline{x}$ und $b := c$ im ersten Fall bzw. zu

$$y_i = \overline{y} - \frac{1}{c}\,\overline{x} + \frac{1}{c}\,x_i = a + b\,x_i \ , \ i \in \{1, \dots, n\} \ , \tag{3}$$

mit $a := \overline{y} - \frac{1}{c}\,\overline{x}$ und $b := \frac{1}{c}$ im zweiten Fall.

Aus (2),(3) folgt mit $\overline{y} = a + b\,\overline{x}$ (gemäß Def. von a und b):

$$\begin{aligned} s_{xy} &\underset{(2),(3)}{\overset{(7.22)}{=}} \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})\,(a + b\,x_i - (a + b\,\overline{x})) \\ &= b\,\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \overset{(3.29)}{=} b\,s_x^2 \end{aligned} \tag{4}$$

Damit erhält man in den einzelnen Fällen:

(i) $$r_{xy} = -1 \Longrightarrow \frac{s_{xy}}{s_x\, s_y} \overset{\text{A 7.27}}{=} r_{xy} < 0 \overset{s_x,s_y>0}{\Longrightarrow} s_{xy} < 0 \Longrightarrow b \overset{(4)}{=} \frac{s_{xy}}{s_x^2} < 0$$

(ii) $$r_{xy} = 1 \Longrightarrow \frac{s_{xy}}{s_x\, s_y} \overset{\text{A 7.27}}{=} r_{xy} > 0 \overset{s_x,s_y>0}{\Longrightarrow} s_{xy} > 0 \Longrightarrow b \overset{(4)}{=} \frac{s_{xy}}{s_x^2} > 0$$

Aufgabe 9

Zunächst Erstellung einer geeigneten Arbeitstabelle
(gleiche Tabelle zur Berechnung der Regressionsparameter wie zur Berechnung des Bravais-Pearson-Korrelationskoeffizienten)

Arbeitstabelle zur Berechnung der geschätzten Regressionsparameter (und des Bravais-Pearson-Korrelationskoeffizienten)

i	x_i	y_i	x_i^2	y_i^2	$x_i\, y_i$
1	7	16.5	49	272.25	115.5
2	9	22.0	81	484.00	198.0
3	6	15.0	36	225.00	90.0
4	11	24.5	121	600.25	269.5
5	8	19.0	64	361.00	152.0
6	7	17.0	49	289.00	119.0
Summe	48	114.0	400	2231.50	944.0

Mit den in der letzten Tabellenzeile angegebenen Summen erhält man:

$$\overline{x} = \frac{1}{6}\sum_{i=1}^{6} x_i \overset{\text{Tab.}}{=} \frac{1}{6}\cdot 48 = 8 \tag{1}$$

$$\overline{y} = \frac{1}{6}\sum_{i=1}^{6} y_i \overset{\text{Tab.}}{=} \frac{1}{6}\cdot 114 = 19 \tag{2}$$

$$\begin{aligned} s_x^2 &\underset{\text{vgl. A 6}}{\overset{\text{A 3.31}}{=}} \frac{1}{6}\sum_{i=1}^{6} x_i^2 - \overline{x}^2 \\ &\overset{\text{Tab.,(1)}}{=} \frac{1}{6}\cdot 400 - 8^2 = \frac{200}{3} - \frac{192}{3} = \frac{8}{3} \end{aligned} \tag{3}$$

$$\begin{aligned} s_{xy} &\overset{\text{A 7.25}}{=} \frac{1}{6}\sum_{i=1}^{6} x_i\, y_i - \overline{x}\,\overline{y} \\ &\overset{\text{Tab.,(1),(2)}}{=} \frac{1}{6}\cdot 944 - 8\cdot 19 = \frac{472}{3} - \frac{456}{3} = \frac{16}{3} \end{aligned} \tag{4}$$

Gemäß Vorl., A 8.4 folgt:

$$\hat{b} \overset{\text{Def.}}{=} \frac{s_{xy}}{s_x^2} \overset{(3),(4)}{=} \frac{16/3}{8/3} = 2 \tag{5}$$

$$\hat{a} \overset{\text{Def.}}{=} \overline{y} - \hat{b}\,\overline{x} \overset{(1),(2),(5)}{=} 19 - 2 \cdot 8 = 3$$

Die geschätzte Regressionsgerade ist damit gegeben durch:

$$y = \hat{f}(x) = \hat{a} + \hat{b}\,x = 3 + 2x \ , \ \ x \in \mathbb{R}$$

Für das zugehörige Bestimmtheitsmaß gilt:

$$B_{xy} \overset{\text{Def. A 8.13}}{=} 1 - \frac{\sum_{i=1}^{6} \big(y_i - (\hat{a} + \hat{b}\,x_i)\big)^2}{\sum_{i=1}^{6} (y_i - \overline{y})^2} \overset{\text{A 8.14}}{=} r_{xy}^2 \tag{6}$$

Hier Berechnung mittels r_{xy}^2, da Arbeitstabelle schon vorhanden und damit Berechnung des Bravais-Pearson-Korrelationskoeffizienten r_{xy} mit geringem Aufwand möglich. (Berechnung mittels Darstellung aus Def. A 8.13 erheblich aufwändiger!) Analog zu (3) erhält man:

$$s_y^2 \underset{\text{vgl. A 6}}{\overset{\text{A 3.31}}{=}} \frac{1}{6} \sum_{i=1}^{6} y_i^2 - \overline{y}^2 \tag{7}$$

$$\overset{\text{Tab.,(2)}}{=} \frac{1}{6} \cdot 2231.5 - 19^2 = \frac{4463}{12} - \frac{361 \cdot 12}{12} = \frac{131}{12}$$

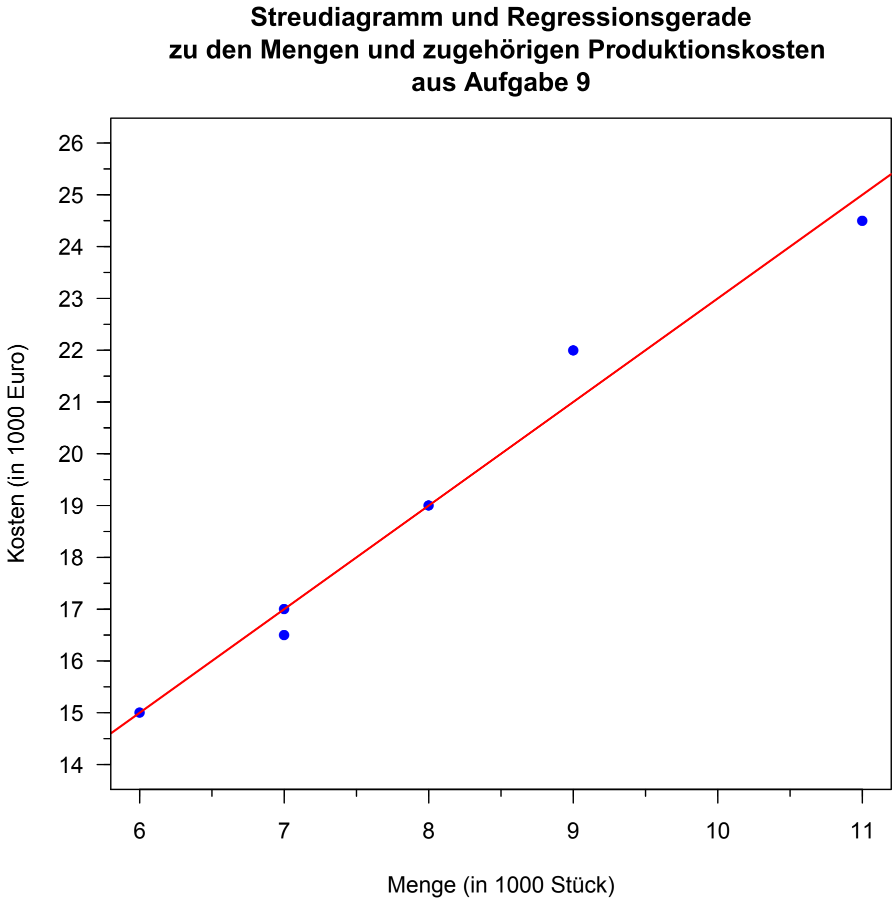
Es folgt:

$$B_{xy} \overset{(6)}{=} r_{xy}^2 \overset{\text{Def. A 7.27}}{=} \frac{s_{xy}^2}{s_x^2\, s_y^2} \overset{(3),(4),(7)}{=} \frac{\left(\frac{16}{3}\right)^2}{\frac{8}{3} \cdot \frac{131}{12}} = \frac{\frac{256}{9}}{\frac{2}{3} \cdot \frac{131}{3}} \tag{8}$$

$$= \frac{256}{262} \approx 0.977$$

Für den Bravais-Pearson-Korrelationskoeffizienten selbst erhält man hier:

$$r_{xy} \overset{\text{Def. A 7.27}}{=} \frac{s_{xy}}{s_x\, s_y} \overset{(3),(4),(7)}{=} \frac{\frac{16}{3}}{\sqrt{\frac{8}{3} \cdot \frac{131}{12}}} \underset{\text{vgl. (8)}}{\overset{\text{Kürzen!}}{=}} \frac{16}{\sqrt{262}} \approx \underbrace{0.988}_{>0} \tag{9}$$



Interpretationen/Bemerkungen zu Aufgabe 9:

(i) Der hohe Wert des Bestimmtheitsmaßes $B_{xy} \approx 0.977$ (nahe 1) zeigt ebenso wie die graphische Darstellung, dass die gegebenen Datenpunkte sehr gut durch die zugehörige Regressionsgerade approximiert werden.

(ii) Mit $r_{xy} \approx 0.988 > 0$ liegt eine hohe positive Korrelation und damit ein starker gleichsinniger (linearer) Zusammenhang zwischen den beiden betrachteten Merkmalen *Produzierte Menge* und *Zugehörige Produktionskosten* vor.

(iii) Hier wird das Bestimmtheitsmaß mit Hilfe des Bravais-Pearson-Korrelationskoeffizienten berechnet.

Falls umgekehrt das Bestimmtheitsmaß gegeben ist (z.B. als Ausgabe eines Statistik-Programms) und man hieraus den Bravais-Pearson-Korrelationskoeffizienten berechnen will, ist zu beachten, dass die quadratische Gleichung $B_{xy} = r_{xy}^2$ im Allgemeinen die beiden möglichen Lösungen $r_{xy} = -\sqrt{B_{xy}}$ und $r_{xy} = \sqrt{B_{xy}}$ besitzt.

Das Vorzeichen des Korrelationskoeffizienten kann man dann (z.B.) aus dem Vorzeichen der Steigung der Regressionsgeraden ablesen, denn mit $s_x > 0\,, s_y > 0$ gilt:

$$r_{xy} \overset{\text{Def.}}{=} \frac{s_{xy}}{s_x\, s_y} \begin{Bmatrix} > \\ < \end{Bmatrix} 0 \iff s_{xy} \begin{Bmatrix} > \\ < \end{Bmatrix} 0 \iff \hat{b} \overset{\text{Def.}}{=} \frac{s_{xy}}{s_x^2} \begin{Bmatrix} > \\ < \end{Bmatrix} 0\,.$$

(iv) Beim Einsetzen von vorweg berechneten Zwischenergebnissen besser Brüche anstelle gerundeter Dezimalzahlen verwenden, da sich Rundungs-Ungenauigkeiten insbesondere beim Potenzieren verstärken können.

Aufgabe 10

Es beschreiben

(i) B^c die Menge der Nichtraucher,

(ii) $A \cap B$ die Menge der männlichen Raucher,

(iii) $B \backslash A = B \cap A^c$ die Menge der weiblichen Raucher,

(iv) $(A \cup B)^c \overset{\text{de Morg.}}{=} A^c \cap B^c$ die Menge der weiblichen Nichtraucher.

(v) Zunächst gilt:

$$
\begin{aligned}
& (A \cup B^c) \cap (A^c \cup B) \\
\overset{\text{D-Ges.}}{=} \; & \big(A \cap (A^c \cup B)\big) \cup \big(B^c \cap (A^c \cup B)\big) \\
\overset{\text{D-Ges.}}{=} \; & \big(\underbrace{(A \cap A^c)}_{=\emptyset} \cup (A \cap B)\big) \cup \big((B^c \cap A^c) \cup \underbrace{(B^c \cap B)}_{=\emptyset}\big) \\
= \; & (A \cap B) \cup (A^c \cap B^c)
\end{aligned}
$$

Damit beschreibt $(A \cup B^c) \cap (A^c \cup B)$ die Menge der männlichen Raucher und weiblichen Nichtraucher (vgl. (ii) und (iv)).

Aufgabe 11

(a) Geeignete Grundmenge (Ergebnismenge):

$$\Omega = \big\{(i,j) \mid i,j \in \{1, \ldots ,6\}\big\}$$

mit folgender Interpretation:

Erste Komponente i gibt die Augenzahl des roten Würfels und zweite Komponente j die Augenzahl des schwarzen Würfels an.

Anzahl der Elemente von Ω: $\quad |\Omega| \overset{\text{B 1.11}}{=} 6^2 = 36\,.$

(i) Ereignis „Die Augensumme beträgt 8“ beschrieben durch

$$A = \big\{(i,j) \in \Omega \,\big|\, i + j = 8\big\} = \big\{(2{,}6), (3{,}5), (4{,}4), (5{,}3), (6{,}2)\big\}$$
$$\Longrightarrow \quad |A| = 5$$

(ii) Ereignis „Die Augensumme ist kleiner als 5“ beschrieben durch

$$B = \big\{(i,j) \in \Omega \,\big|\, i + j < 5\big\} = \big\{(1{,}1), (1{,}2), (1{,}3), (2{,}1), (2{,}2), (3{,}1)\big\}$$
$$\Longrightarrow \quad |B| = 6$$

(iii) Ereignis „Beide Augenzahlen betragen jeweils höchstens 3“ beschrieben durch

$$\begin{aligned} C &= \big\{(i,j) \in \Omega \,\big|\, i \leq 3\,, j \leq 3\big\} \\ &= \big\{(1{,}1), (1{,}2), (1{,}3), (2{,}1), (2{,}2), (2{,}3), (3{,}1), (3{,}2), (3{,}3)\big\} \end{aligned}$$
$$\Longrightarrow \quad |C| = 9$$

(iv) Ereignis „Die Augenzahl des roten Würfels ist geringer als die des schwarzen Würfels“ beschrieben durch

$$\begin{aligned} D &= \big\{(i,j) \in \Omega \,\big|\, i < j\big\} \\ &= \big\{(1{,}2), (1{,}3), (1{,}4), (1{,}5), (1{,}6), (2{,}3), (2{,}4), (2{,}5), \\ &\qquad (2{,}6), (3{,}4), (3{,}5), (3{,}6), (4{,}5), (4{,}6), (5{,}6)\big\} \end{aligned}$$
$$\Longrightarrow \quad |D| = 15$$

(b) Unverfälschte Würfel $\longrightarrow$ Beschreibung durch Laplace-Raum. d.h.:

$$P(E) = \frac{|E|}{|\Omega|} \overset{\text{(a)}}{=} \frac{|E|}{36}\,, \; E \subseteq \Omega\,.$$

Damit erhält man:

(i) $P(A) = \dfrac{|A|}{|\Omega|} \overset{\text{(a)}}{=} \dfrac{5}{36} \approx 0.139\,,$

(ii) $P(B) = \dfrac{|B|}{|\Omega|} \overset{\text{(a)}}{=} \dfrac{6}{36} = \dfrac{1}{6} \approx 0.167\,,$

(iii) $P(C) = \dfrac{|C|}{|\Omega|} \overset{\text{(a)}}{=} \dfrac{9}{36} = \dfrac{1}{4} = 0.25\,,$

(iv) $P(D) = \dfrac{|D|}{|\Omega|} \overset{\text{(a)}}{=} \dfrac{15}{36} = \dfrac{5}{12} \approx 0.417\,.$

Aufgabe 12

(a) Geeignete Grundmenge (Ergebnismenge):

$$\Omega = \big\{(\omega_1,\ldots,\omega_7)\,|\,\omega_1,\ldots,\omega_7 \in \{0,1\}\big\}$$

mit folgender Interpretation:

Für $i \in \{1,\ldots,7\}$ bedeutet $\omega_i = 0$ „Kopf" und $\omega_i = 1$ „Zahl" im i-ten Wurf.

Homogene Münze $\longrightarrow$ Beschreibung durch Laplace-Raum (Ω,P), d.h.:

$$P(E) = \frac{|E|}{|\Omega|}\ ,\ E \subseteq \Omega\ .$$

(b) Begründung für Festlegung von Ω in (a):

k-facher Münzwurf entspricht Urnenmodell B 1.11 (mit $n = 2$), da Auswahl *mit Wiederholung* und *Berücksichtigung der Reihenfolge*.

(Für Ereignis (iii) in (c) z.B. (1,0,1,0,1,0,1) und (1,1,1,1,0,0,0) zu unterscheiden.)

Damit Anzahl möglicher Versuchsausgänge gegeben durch

$$|\Omega| \underset{\text{B 1.11}}{\overset{\text{(a)}}{=}} 2^7 = 128\ .$$

(c) (i) Gesuchtes Ereignis beschrieben durch

$$A := \left\{(\omega_1,\ldots,\omega_7) \in \Omega \,\middle|\, \underbrace{\sum_{i=1}^{7}\omega_i = 6}_{\text{Genau eine Null} \,\hat{=}\, \text{Genau 6 Einsen}}\right\}$$

$$= \big\{(0,1,\ldots,1)\,,(1,0,1,\ldots,1)\,,\ldots,(1,\ldots,1,0)\big\}$$

$$\Longrightarrow \quad P(A) = \frac{|A|}{|\Omega|} = \frac{7}{128}$$

(ii) Gesuchtes Ereignis beschrieben durch

$$B := \left\{(\omega_1,\ldots,\omega_7) \in \Omega \,\middle|\, \underbrace{\sum_{i=1}^{7}\omega_i \geq 6}_{\text{Mind. 6 Einsen}}\right\} \overset{\text{(i)}}{=} A \cup \big\{(1,\ldots,1)\big\}$$

$$= \big\{(0,1,\ldots,1)\,,(1,0,1,\ldots,1)\,,(1,\ldots,1,0)\,,(1,\ldots,1)\big\}$$

$$\Longrightarrow \quad P(B) = \frac{|B|}{|\Omega|} = \frac{8}{128} = \frac{1}{16}$$

(iii) Es bezeichne C das gesuchte Ereignis.

Wichtiger Trick:

Übergang zum Komplementärereignis, da Beschreibung und Berechnung der zugehörigen Wahrscheinlichkeit erheblich einfacher ist.

Komplementärereignis C^c bedeutet:

Kein Mal zeigen aufeinanderfolgende Würfe das gleiche Symbol.

Damit folgt:

$$C^c = \big\{(1{,}0{,}1{,}0{,}1{,}0{,}1)\,,(0{,}1{,}0{,}1{,}0{,}1{,}0)\big\}$$

$$\Longrightarrow \quad P(C) \underset{\text{(B 4.1)}}{\overset{\text{Hinw.}}{=}} 1 - P(C^c) = 1 - \frac{|C^c|}{|\Omega|} = 1 - \frac{2}{128} = \frac{63}{64}$$

Aufgabe 13

(a) Sei $k \in \mathbb{N}$ (beliebig). In unserem Modell denken wir uns die 365 verschiedenen Tage des Jahres der Reihe nach durchnummeriert.

Die Verteilung der k Personen auf die 365 Tage des Jahres entspricht der k-fachen Ziehung aus einer Urne, in der sich 365 mit den Zahlen $1, \ldots, 365$ durchnummerierte Kugeln befinden. Hierbei gilt:

- Die Ziehung erfolgt *mit Zurücklegen*, da mehrere Personen am gleichen Tag Geburtstag haben können.
- Die einzelnen Personen werden unterschieden, d.h. die Ziehung erfolgt *unter Berücksichtigung der Reihenfolge.*

Gem. B 1.11 damit geeignete Ergebnismenge gegeben durch

$$\Omega_k = \big\{(\omega_1, \ldots, \omega_k) \,\big|\, \omega_1, \ldots, \omega_k \in \{1, \ldots, 365\}\big\} \ .$$

Zugehörige Interpretation:

Für $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, 365\}$ bedeutet $\omega_i = j$, dass die i-te Person am j-ten Tag des Jahres Geburtstag hat.

Es folgt (mit $n = 365$):

$$|\Omega_k| \overset{\text{B } 1.11}{=} 365^k \ .$$

Gesucht: Wahrscheinlichkeit des Ereignisses

E_k : „Mindestens zwei der k Personen haben am selben Tag Geburtstag“.

Gemäß Hinweis Betrachtung des zugehörigen Komplementärereignisses

E_k^c : „Alle k Personen haben an verschiedenen Tagen Geburtstag“.

Dann gilt:

$$E_k^c = \big\{(\omega_1, \ldots, \omega_k) \in \Omega_k \,|\, \omega_i \neq \omega_j \ \text{ für } \ i \neq j\big\} \ .$$

Dies ist die Menge aller (n,k)-Permutationen *ohne* Wiederholung mit $n = 365$ (vgl. B 1.13). Es folgt:

$$|E_k^c| = 365 \cdot 364 \cdot \ldots \cdot (365 - k + 1) = \frac{365!}{(365 - k)!} \ .$$

Vorbemerkung $\quad \longrightarrow \quad$ Laplace-Experiment

Mit P als (diskreter) Gleichverteilung auf Ω_k folgt dann:

$$P(E_k) \underset{\text{(B 4.1)}}{\overset{\text{Hinw.}}{=}} 1 - P(E_k^c) = 1 - \frac{|E_k^c|}{|\Omega_k|} = 1 - \frac{365!}{365^k\,(365-k)!} \qquad (*)$$

$$= \quad 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365-k+1)}{365^k} \, .$$

(b) Aus $(*)$ ergeben sich für $k \in \{10{,}22{,}23{,}40\}$ die folgenden Wahrscheinlichkeiten für das betreffende Ereignis E_k:

k	10	22	23	40
$P(E_k)$	0.117	0.476	0.507	0.891

Anhand der berechneten Wahrscheinlichkeiten erkennt man:

Bereits ab der Personenanzahl $k = 23$ ist die Wahrscheinlichkeit dafür, dass mindestens zwei der k Personen am selben Tag des Jahres Geburtstag haben, größer als 0.5 und damit größer als die Wahrscheinlichkeit dafür, dass alle k Personen an verschiedenen Tagen Geburtstag haben.