Statistik

für Betriebswirtschaft, Internationales Management, Wirtschaftsinformatik und Informatik

Sommersemester 2017

HSA Statistik SS 2017 Sessionlist		
Datum	Statistik für BW/IM/I/Winf	Nr.
15.03.2017	Einführung Statistik	1
22.03.2017	Differentialrechnung, 2-dim Diff.Rechnung	2
29.03.2017	univ. deskr. Stat., Quantile, Plots	3
05.04.2017	Streuung, Konzentrationsmaße	4
12.04.2017	Kontingenztabellen, Mosaikplots, Korrelation	5
19.04.2017	Preisindizes, lineare Regression	6
26.04.2017	Wahrscheinlichkeitsbegriff	7
03.05.2017	Bedingte Wahrscheinlichkeit, Bayes	8
10.05.2017	diskrete Zufallsvariablen	9
17.05.2017	Stetige ZV, Gleichverteilung	10
24.05.2017	Pyramid	
31.05.2017	Normalverteilung, Verteilungsparameter	11
07.06.2017	Schätzfunktionen und Punktschätzer	12
14.06.2017	Konfidenzintervalle	13
21.06.2017	Wiederholung, Besprechung Probeklausur	14
28.06.2017	Prüfungswoche	15

Stefan Etschberger

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

- Zufallsvorgang: Geschehen mit ungewissem Ausgang, z.B. Münzwurf

kleines omega

- Elementarereignis ω: Ein möglicher Ausgang, z.B. „ Kopf “ Elementarereignisse schließen sich gegenseitig aus („ Kopf “ <u>oder</u> „ Zahl “)!
- Ergebnismenge Ω: Menge aller ω

- **Beispiel:** Werfen zweier Würfel:

ω_1

$$\Omega : \begin{Bmatrix} (1,1) & (1,2) & \cdots & (1,6) \\ (2,1) & (2,2) & \cdots & (2,6) \\ \vdots & \vdots & \ddots & \vdots \\ (6,1) & (6,2) & \cdots & (6,6) \end{Bmatrix}$$

$$\Rightarrow \Omega = \{(x_1, x_2) : x_1, x_2 \in \{1, \ldots, 6\}\}$$

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

Ereignisse und Wahrscheinlichkeiten

Elementarereignis → $\omega \in \Omega$

Ereignis → $\{\omega\} \subset \Omega$

- **Ereignis** A: Folgeerscheinung eines Elementarereignisses
- Formal:

$$A \subseteq \Omega$$

- Ereignisse schließen sich nicht gegenseitig aus!
- **Beispiel:** Werfen zweier Würfel:

Ereignis	verbal	formal
A	Augensumme $= 4$	$\{(1,3),\ (2,2),\ (3,1)\}$
B	Erste Zahl $= 2$	$\{(2,1),\ (2,2),\ \ldots,\ (2,6)\}$

- **Wahrscheinlichkeit** $P(A)$: Chance für das Eintreten von A
- **Laplace-Wahrscheinlichkeit:**

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}}$$

Laplace Wahrscheinlichkeit und Urnenmodell

- **Beispiel:** Werfen zweier Würfel:

$$\text{Augensumme} = 4: \quad A = \{(1{,}3),\ (2{,}2),\ (3{,}1)\}$$

$$|\Omega| = 36, \quad |A| = 3 \Rightarrow P(A) = \frac{3}{36} = \frac{1}{12} = 0{,}083$$

- Urnenmodell: Ziehe n Objekte aus einer Menge mit N Objekten
  Anzahl Möglichkeiten:

  mit Zurücklegen: N^n
  ohne Zurücklegen: $N \cdot (N-1) \cdots (N-(n-1)) = \frac{N!}{(N-n)!}$

- **Beispiel:**
  Wie groß ist die Wahrscheinlichkeit, aus einem gut gemischten 32-er Kartenblatt bei viermaligem Ziehen vier Asse zu bekommen?
  a) Ziehen mit Zurücklegen,
  b) Ziehen ohne Zurücklegen

Rechenregeln für Wahrscheinlichkeiten

$\bar{A} = \bar{A}_\Omega = \Omega \setminus A$

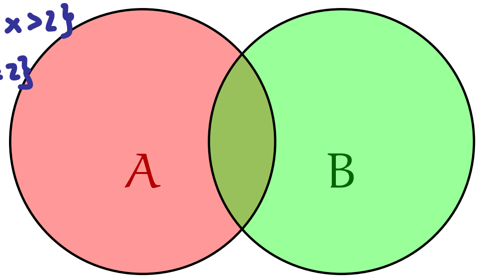
(Komplementärereignis)

z.B. $\Omega = \mathbb{R}$

$A = \{x \in \mathbb{R} : x > 2\}$

$\bar{A} = \{x \in \mathbb{R} : x \leq 2\}$

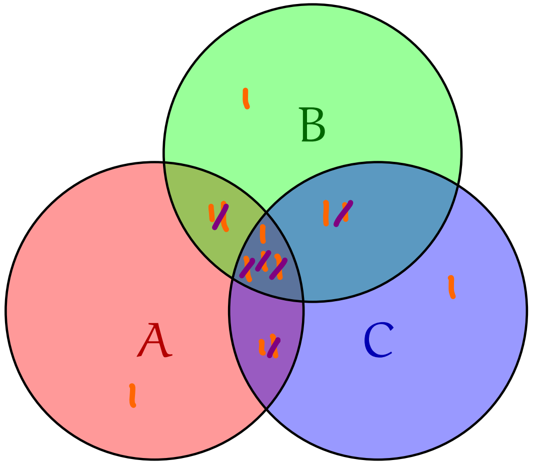
- Wichtige Rechenregeln:



1. $P(A) \leqq 1$
2. $P(\emptyset) = 0$

   $P(\{\}) =$
3. $A \subseteq B \Rightarrow P(A) \leqq P(B)$
4. $P(\bar{A}) = 1 - P(A)$
5. $P(A \cup B) =$

   $P(A) + P(B) - P(A \cap B)$



$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- **Beispiel:**

$$P(\text{„Augenzahl} \leqq 5\text{“}) = 1 - P(\text{„Augenzahl} = 6\text{“}) = 1 - \frac{1}{6} = \frac{5}{6}$$

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

1. Einführung

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

Kombinatorik

Zufall und Wahrscheinlichkeit

Zufallsvariablen und Verteilungen

Verteilungsparameter

5. Induktive Statistik

Quellen

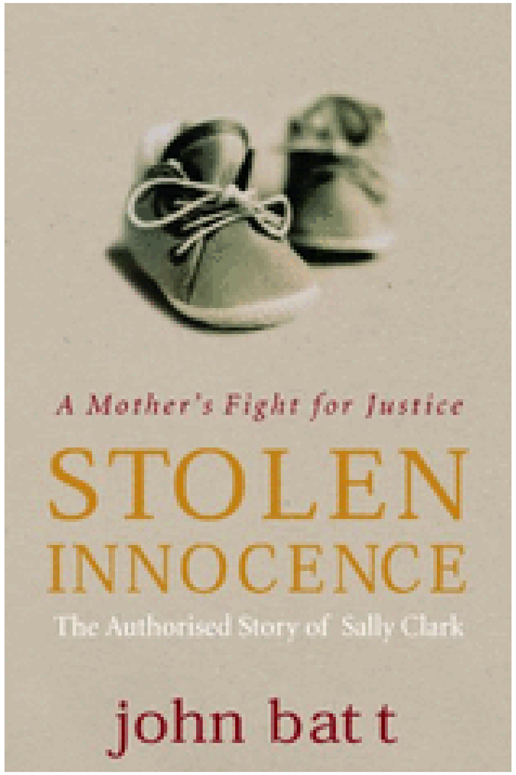
Tabellen

Der Fall Sally Clark

- Sally Clarks Söhne Christopher und Harry sterben 1996 und 1997 beide kurz nach der Geburt an plötzlichem Kindstod.
- Kinderarzt: „Wahrscheinlich Mord, da 2 maliger plötzlicher Kindstod sehr unwahrscheinlich!“ (ohne konkrete Hinweise)
- Gerichtliche Untersuchung
- Hauptargument der Anklage gestützt durch Gerichtsgutachter Sir Roy Meadow (renommierter Facharzt für Kinderheilkunde): Wahrscheinlichkeit für plötzlichen Kindstod ist 1:8500, d.h. Wahrscheinlichkeit für 2 maliges Auftreten in einer Familie

$$p = \left(\frac{1}{8500}\right)^2 \approx 1 : 72\,000\,000$$

- Urteil: Doppelmord; Strafe: 2 mal lebenslang; Inhaftierung von Sally Clark 1999



Der Fall Sally Clark

- Problem: Es gibt sehr viele Familien mit 2 Kindern
- Europa: ca. 80 Mio Familien mit Kindern, davon ca. 50% mit mindestens zwei Kindern, also ca. 40 Mio.

Gene find casts doubt on double 'cot death' murders

An expert said there was a one in 73 million chance Sally Clark's babies died naturally – and a jury agreed. Now new genetic research could help to clear her. **John Sweeney** and **Bill Law** investigate

- Wahrscheinlichkeit, dass in einer solchen Familie kein zweifacher plötzlicher Kindstod auftritt: $1 - \left(\frac{1}{8500}\right)^2$
  Annahmen:
  - Jede dieser Familien hat genau 2 Kinder; in Wirklichkeit: ca. 20% dieser Familien haben mindestens 3 Kinder
  - Zweiter plötzlicher Kindstod unabhängig von erstem (nicht untersucht)
- Wahrscheinlichkeit, dass in 40 Mio. Familien mindestens ein zweifacher plötzlicher Kindstod auftritt:

$$1 - \left(1 - \left(\frac{1}{8500}\right)^2\right)^{40\,000\,000} \approx 42{,}5\%$$

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

Beispiel Gegenereignis

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

Der Fall Sally Clark

Gene find casts doubt on double 'cot death' murders

An expert said there was a one in 73 million chance Sally Clark's babies died naturally – and a jury agreed. Now new genetic research could help to clear her. **John Sweeney** and **Bill Law** investigate

- Problem: Es gibt sehr viele Familien mit 2 Kindern
- Europa: ca. 80 Mio Familien mit Kindern, davon ca. 50% mit mindestens zwei Kindern, also ca. 40 Mio.
- Wahrscheinlichkeit, dass in einer solchen Familie kein zweifacher plötzlicher Kindstod auftritt: $1 - \left(\frac{1}{8500}\right)^2$
  Annahmen:
  - Jede dieser Familien hat genau 2 Kinder; in Wirklichkeit: ca. 20% dieser Familien haben mindestens 3 Kinder
  - Zweiter plötzlicher Kindstod unabhängig von erstem (nicht untersucht)
- Wahrscheinlichkeit, dass in 40 Mio. Familien mindestens ein zweifacher plötzlicher Kindstod auftritt:

$$1 - \left(1 - \left(\frac{1}{8500}\right)^2\right)^{40\,000\,000} \approx 42{,}5\%$$

- 2001: Royal Statistical Society interveniert
- 2003: Sally Clark wird nach Revision freigesprochen
- 2007 findet man sie tot in ihrer Wohnung auf - gestorben an einer akuten Alkoholvergiftung. Sie habe sich, so ihre Familie, von dem Justizirrtum nie erholt.

Bedingte Wahrscheinlichkeiten

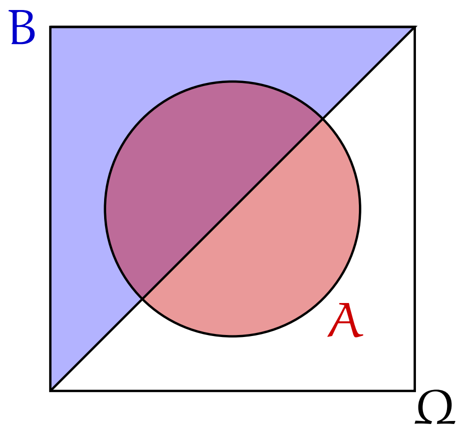
1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

- Wahrscheinlichkeit von A hängt von anderem Ereignis B ab.
  (B kann zeitlich vor A liegen, muss aber nicht!)
- **Beispiel:** Wahrscheinlichkeit für Statistiknote hängt von Mathenote ab.
- Formal:

Wahrscheinlichkeit von A
unter der Bedingung B

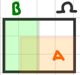
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Im Venndiagramm:

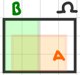


Bedingte Wahrscheinlichkeiten

$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

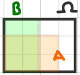


$P(A|B) = \frac{2/12}{6/12} = 1/3 < 1/2 = P(A)$



$P(A|B) = \frac{3/12}{6/12} = 1/2 = P(A)$

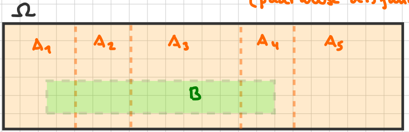
↳ hier: A, B unabhängig



$P(A|B) = \frac{4/12}{6/12} = 2/3 > 1/2 = P(A)$

Es gilt: $\bigcup_{i=1}^{n} A_i = \Omega$ und $A_i \cap A_j = \{\}$ $(i \neq j)$

(paarweise disjunkt)



- $P(B) = P(A_1\cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B)$
- $P(B|A_i) = \frac{P(B\cap A_i)}{P(A_i)} \Leftrightarrow P(B\cap A_i) = P(A_i)\cdot P(B|A_i)$

$$P(B) = P(A_1)\cdot P(B|A_1) + P(A_2)\cdot P(B|A_2) + \dots + P(A_n)\cdot P(B|A_n)$$

$$\Leftrightarrow \boxed{P(B) = \sum_{i=1}^{n} P(A_i)\cdot P(B|A_i)}$$

Satz von der totalen Wahrscheinlichkeit

Aufgabe 51

Der Bauer Bertram hat 3 Hühner (Erna, Lisa und Moni). Erna ist seine Lieblingshenne, denn sie liefert durchschnittlich 40 % aller pro Jahr gelegten Eier, während Lisa und Moni nur jeweils 30 % schaffen. Da die Eier ein Mindestgewicht haben müssen, gibt es einen gewissen Ausschuß (A). Bei Erna und Lisa beträgt er jeweils 3 % und bei Moni 5 %.

Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Ei

a) von Lisa stammt
b) zu klein ist
c) von Lisa stammt, wenn bekannt ist, dass es zu klein ist?

$L \mathrel{\hat=}$ Ei ist von Lisa , E, M, ...
$A \mathrel{\hat=}$ Ei ist zu klein

$P(A|E) = 0.03$, $P(A|L) = 0.03$, $P(A|M) = 0.05$
$P(E) = 0.4$, $P(L) = P(M) = 0.3$

a) 0.3

b) gesucht: $P(A)$

$$\begin{aligned} P(A) &= P(E)\cdot P(A|E) + P(L)\cdot P(A|L) + P(M)\cdot P(A|M) \\ &= 0.4\cdot 0.03 + 0.3\cdot 0.03 + 0.3\cdot 0.05 \\ &= 0.012 + 0.009 + 0.015 \\ &= 0.036 = 3.6\,\% \end{aligned}$$

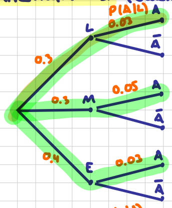
Alternative 1 (n-Felder-Tafel)

$P(A\cap E) = P(A|E)\cdot P(E)$

	E	L	M	
A	0.012	0.009	0.015	0.036
$\bar{A}$	0.388	0.291	0.285	0.964
	0.4	0.3	0.3	

c) $P(L|A) = \frac{P(L\cap A)}{P(A)} = \frac{0.009}{0.036} = 1/4 = 25\,\%$

Alternative 2 (Baumdiagramm

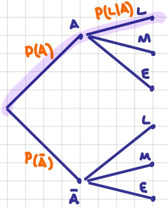


$P(L \cap A) = 0.3 \cdot 0.03 = 0.009$

$P(M \cap A) = 0.3 \cdot 0.05 = 0.015$

$P(E \cap A) = 0.4 \cdot 0.03 = 0.012$

$P(A) = 0.036$



$P(L \cap A)$

$P(L) \cdot P(A|L) = P(L \cap A) = P(A) \cdot P(L|A)$

$\Leftrightarrow \quad P(L|A) = \dfrac{P(L) \cdot P(A|L)}{P(A)}$

$\left(\text{z} \quad P(L|A) = \dfrac{0.3 \cdot 0.03}{0.036} = \dfrac{0.009}{0.036} = \frac{1}{4} = 25\% \right)$

allgemein

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B|A_i)}{P(B)}$$

Satz von Bayes

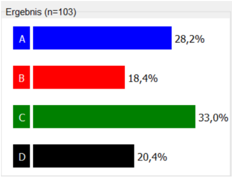
$$= \frac{P(A_i) \cdot P(B|A_i)}{\sum_{j=1}^{n} P(A_j) \cdot P(B|A_j)}$$

Beispiel: Das Ziegenproblem

Quizshow: Kandidat muss sich für eine Tür entscheiden (hinter 1 Tür: 1 Mio)

Nach Entsch. des Kandid.: Showmaster öffnet von den übrigen beiden Türen eine Ziegentür; danach: Kandidat könnte wechseln

Umfrage:
- A Wechseln ist schlecht
- B Wechseln ist gut
- C Macht keinen Unterschied
- D Ich kenne das Ergebnis



Experiment:

- Reihennummer von vorne abzählen

- Nr. 1,3,5,7,... -> Wechsler
  Nr. 2,4,6,8,... -> Bleiber

- Bilden Sie Zweiergruppen:
  Eine Person ist Showmaster,
  die andere Kandidat

- Führen Sie 10mal das Spiel durch:
  * Showmaster schreibt (verdeckt) eine Zahl s aus (1,2,3) auf
  * Kandidat wählt eine Zahl k
  * (bei den Wechslern:) SM nennt eine Zahl, die nicht s und nicht k ist; Kandidat wechselt
  * Falls s und k gleich: Treffer (aufschreiben) falls nicht: Ziege! (auch aufschreiben)

- Nach den 10 Spielen: Zählen Sie die Treffer

Ereignisse : A, B, C : Mio ist hinter Tür A, B, C

M_A, M_B, M_C : Moderator öffnet Tür A, B, C

Kandidat wählt Tür A:

$$P(M_B \mid A) = \tfrac{1}{2}$$
$$P(M_B \mid B) = 0$$
$$P(M_B \mid C) = 1$$

$$P(M_B) = P(M_B|A)\cdot P(A) + P(M_B|B)\cdot P(B) + P(M_B|C)\cdot P(C) = \tfrac{1}{2}\cdot\tfrac{1}{3} + 0\cdot\tfrac{1}{3} + 1\cdot\tfrac{1}{3} = \tfrac{1}{2}$$

$$P(A|M_B) = \frac{P(M_B|A)\cdot P(A)}{P(M_B)} = \frac{\tfrac{1}{2}\cdot\tfrac{1}{3}}{\tfrac{1}{2}} = \tfrac{1}{3}$$

↳ Kandidat bleibt bei Entscheidung

$$P(C|M_B) = \frac{P(M_B|C)\cdot P(C)}{P(M_B)} = \frac{1\cdot\tfrac{1}{3}}{\tfrac{1}{2}} = \tfrac{2}{3}$$

↳ Kandidat wechselt

Gewinn bei 10 Spielen			
Wechsler		Bleiber	
60,6%		39,3%	
Anz. Treffer	Häufigkeit	Anz. Treffer	Häufigkeit
0	0	0	1
1	0	1	1
2	2	2	6
3	2	3	3
4	3	4	8
5	5	5	2
6	7	6	4
7	5	7	2
8	3	8	1
9	4	9	0
10	1	10	0

Unabhängigkeit von Ereignissen

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
5. Induktive Statistik
Quellen
Tabellen

- A, B unabhängig: Eintreten von A liefert keine Information über $P(B)$.
- Formal:

$$P(A \mid B) = P(A)$$

- Bei Unabhängigkeit ist äquivalent dazu:

$$P(A \cap B) = P(A) \cdot P(B)$$

- Dann gilt:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$

- **Beispiel:** Werfen zweier Würfel:

$$\left.\begin{array}{rl} A: & \text{"'erster Würfel gleich 6"'} \\ B: & \text{"'zweiter Würfel gleich 6"'} \end{array}\right\} \Rightarrow P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

$$= P(A)$$