Statistik

für Betriebswirtschaft und internationales Management

Sommersemester 2015

HSA Statistik PLUS SS 2015 Sessionlist		
Datum	PLUS	Nr.
Dienstag, 17. März 2015	Einführung	1
Dienstag, 24. März 2015	univ. deskr. Stat., Konzentration	2
Dienstag, 31. März 2015	Konzentration	3
Dienstag, 7. April 2015	Ostern	
Dienstag, 14. April 2015	Korrelation, Preisindizes, Regression	4
Dienstag, 21. April 2015	Wahrscheinlichkeiten	5
Dienstag, 28. April 2015	diskrete Zufallsvariablen: Binomial, Hypergeo, Poisson	6
Dienstag, 5. Mai 2015	stetige ZV: Gleichtvtlg., Normalvtlg.	7
Dienstag, 12. Mai 2015	Verteilungsparameter, Schätzfunktionen	8
Dienstag, 19. Mai 2015	Punktschätzer, Konfidenzintervalle	9
Dienstag, 26. Mai 2015	Pfingsten	
Dienstag, 2. Juni 2015	Tests	10
Dienstag, 9. Juni 2015	multivariate Verfahren	11
Dienstag, 16. Juni 2015	multivariate Verfahren	12
Dienstag, 23. Juni 2015	WH,Puffer, evtl. Fragen zu Probeklausur	13
Dienstag, 30. Juni 2015	Prüfungszeit	

Prof. Dr. Stefan Etschberger
Hochschule Augsburg

Regression: 4 eindimensionale Beispiele

- Berühmte Daten aus den 1970er Jahren:

i	x_{1i}	x_{2i}	x_{3i}	x_{4i}	y_{1i}	y_{2i}	y_{3i}	y_{4i}
1	10	10	10	8	8,04	9,14	7,46	6,58
2	8	8	8	8	6,95	8,14	6,77	5,76
3	13	13	13	8	7,58	8,74	12,74	7,71
4	9	9	9	8	8,81	8,77	7,11	8,84
5	11	11	11	8	8,33	9,26	7,81	8,47
6	14	14	14	8	9,96	8,10	8,84	7,04
7	6	6	6	8	7,24	6,13	6,08	5,25
8	4	4	4	19	4,26	3,10	5,39	12,50
9	12	12	12	8	10,84	9,13	8,15	5,56
10	7	7	7	8	4,82	7,26	6,42	7,91
11	5	5	5	8	5,68	4,74	5,73	6,89

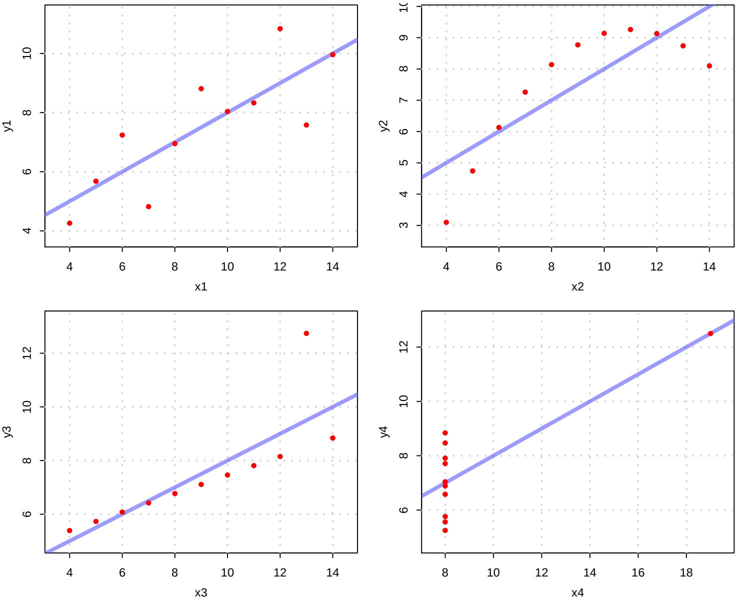
(Quelle: **anscombe**)

Regression: 4 eindimensionale Beispiele

- In folgender Tabelle: Jeweils Ergebnisse der linearen Regressionsanalyse
- dabei: x_k unabhängige Variable und y_k abhängige Variable
- Modell jeweils: $y_k = a_k + b_k x_k$

k	$\hat{a}_k$	$\hat{b}_k$	R_k^2
1	3,0001	0,5001	0,6665
2	3,0010	0,5000	0,6662
3	3,0025	0,4997	0,6663
4	3,0017	0,4999	0,6667

Plot der Anscombe-Daten

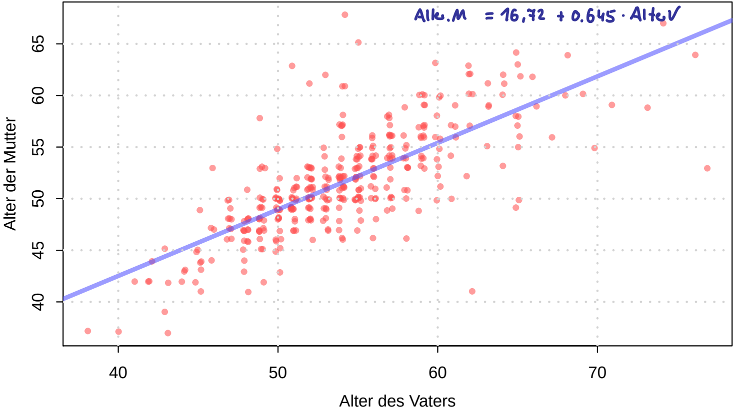


Beispieldaten


```r
meineRegression = lm(AlterM ~ AlterV)
meineRegression

plot(AlterV, AlterM,
     xlab="Alter des Vaters",
     ylab="Alter der Mutter")
abline(meineRegression)
```

```
##
## Call:
## lm(formula = AlterM ~ AlterV)
##
## Coefficients:
## (Intercept)       AlterV
##     16.7247       0.6447
```



Cook's Distanz

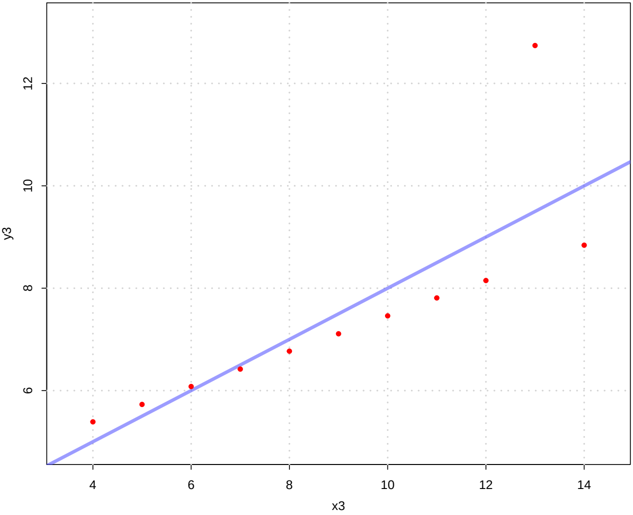
- Oft Kritisch: Einzelne Punkte, die Modell stark beeinflussen
- Idee: Was würde sich ändern, wenn solche Punkte weggelassen würden?
- Cook-Distanz: Misst den Effekt eines gelöschten Objekts
- Formel für ein lineares Modell mit einem unabh. Merkmal:

$$D_i = \frac{\sum_{j=1}^{n} (\hat{y}_j - \hat{y}_{j\,(\text{ohne } i)})^2}{\text{MSE}}$$

- Dabei bedeutet:
  - $\hat{y}_j$: Prognosewert des kompletten Modells für das j-te Objekt
  - $\hat{y}_{j\,(\text{ohne } i)}$: Prognosewert des Modells ohne Objekt i für das j-te Objekt
  - $\text{MSE} = \frac{1}{n} \cdot \sum (\hat{y}_i - y_i)^2$: Normierender Term (Schätzwert für Fehlerstreuung)

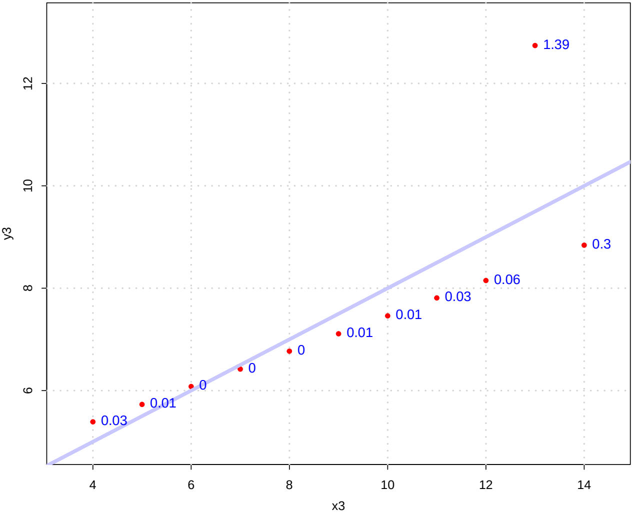
Ausreißer?

- Anscombe-Daten: Regressionsmodell Nr. 3



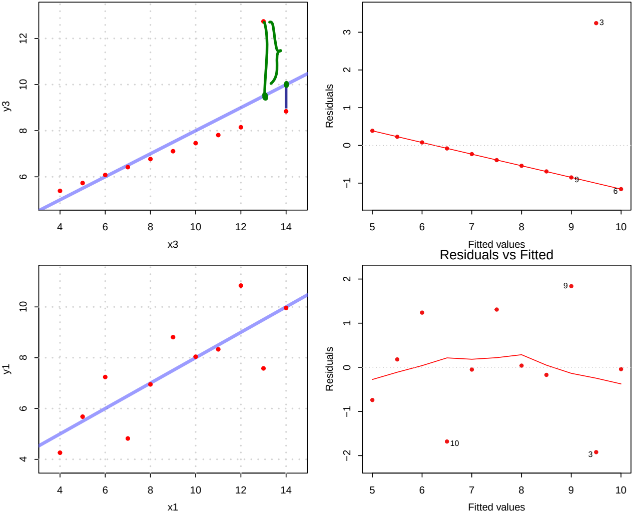
Ausreißer?

- Anscombe-Daten: Regressionsmodell Nr. 3
- Darstellung der Cook-Distanz neben Punkten
- Faustformel: Werte über 1 sollten genau untersucht werden



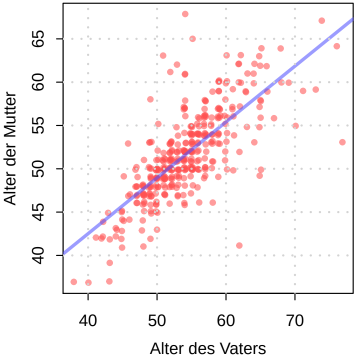
Residualanalyse

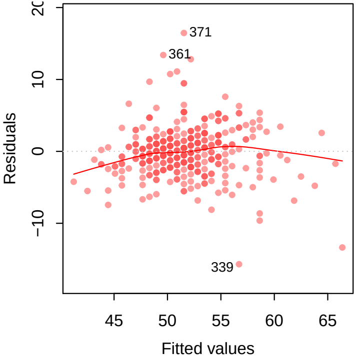
- Oft aufschlussreich: Verteilung der Residuen e_i
- Verbreitet: Graphische Darstellungen der Residuen
- Z.B.: e_i über $\hat{y}_i$



Wichtige Eigenschaften der Residuenverteilung

- Möglichst keine systematischen Muster
- Keine Änderung der Varianz in Abhängigkeit von $\hat{y}_i$ (Homoskedastizität)
- Nötig für inferentielle Analysen: Näherungsweise Normalverteilung der Residuen (q-q-plots)





Kausalität versus Korrelation

Exkurs: Kausalität vs. Korrelation

- Meist wichtig für sinnvolle Regressionsanalysen:
- Kausale Verbindung zwischen unabhängigem und abhängigem Merkmal
- Sonst bei Änderung der unabhängigen Variablen keine sinnvollen Prognosen möglich
- Oft: Latente Variablen im Hintergrund

3 Wahrscheinlichkeitstheorie

Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter

Kombinatorik

Anordnungen von n Elementen
≙ Permutationen

$$n!$$

Anordnungen von n Elementen bei denen es Gruppen von ununterscheidbaren Objekten gibt

Beispiel : MISSISSIPPI

$$\frac{11!}{4! \cdot 4! \cdot 2! \cdot 1!} = 34650$$

„I" „P" „M"

Vertauschung des „S"

allgemein : n Objekte in k Gruppen mit jeweils $n_1, \ldots n_k$ nicht unterscheidbaren Elementen

Anzahl der Permutationen
$$\frac{n!}{\prod_{i=1}^{k} n_i!}$$

Kombinationen ≙ Auswahl von k aus n Elementen

Beispiel : Tresor , 4 Räder , jeweils 0-9,a-z

$$\boxed{36} \cdot 36 \cdot 36 \cdot 36 = 36^4 = 1\,679\,616$$

allgemein : n^k mit Reihenfolge mit Wiederholung

Beispiel : 1. , 2. , 3. Semestersprecher sollen aus 250 Leuten ausgewählt werden

$$250 \cdot 249 \cdot 248 = \frac{250!}{247!} = \frac{250!}{(250-3)!}$$

$$= 250 \underbrace{nPr}_{TR} 3 = 15\,438\,000$$

allgemein : $\frac{n!}{(n-k)!}$ mit Reihenfolge ohne Wiederholung

Beispiel : 3 Leute für Semesterabschlussparty-organisation sollen aus 250 Leuten ausgewählt werden

$$\frac{250 \cdot 249 \cdot 248}{3 \cdot 2 \cdot 1} = \frac{250!}{(250-3)! \cdot 3!} = \binom{250}{3}$$

$$= 250 \ nCr \ 3 = 2\,573\,000$$

allgemein : $\binom{n}{k}$ ohne Reihenfolge ohne Wiederholung

Beispiel : 3 (n) Sorten Getränke (B, S, W)
10 (k) Flaschen sollen gekauft werden

Anzahl Möglichkeiten für Einkauf :

B B B B B | S S | W W W
| S S S S S S S S S | W
B B B B B B B B B B | |
| |
| |

$$\binom{12}{2} = \binom{\overset{n+k-1}{12}}{\underset{k}{10}} = 66$$

allgemein: $\boxed{\binom{n+k-1}{k}}$ mit Wiederholung ohne Reihenfolge

Kombinatorik: Anzahl von Kombinationen bei Auswahl

2-mal Würfeln, das heißt Auswahl von $k = 2$ aus $n = 6$ Zahlen.

$\rightarrow$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- mit WH, mit RF: alle Möglichkeiten, $6^2 = 36$
- ohne WH, mit RF: Diagonale entfällt, $36 - 6 = 30 = 6 \cdot 5 = \dfrac{6!}{(6-2)!}$
- ohne WH, ohne RF: Hälfte des letzten Ergebnisses: $\frac{30}{2} = 15 = \frac{6!}{4!2!} = \binom{6}{2}$
- mit WH, ohne RF: Letztes Ergebnis plus Diagonale, $15 + 6 = 21 = \binom{7}{2}$

Auswahl von k aus n Dingen	mit Wiederholung	ohne Wiederholung
mit Reihenfolge	n^k	$\dfrac{n!}{(n-k)!}$
ohne Reihenfolge	$\dbinom{n+k-1}{k}$	$\dbinom{n}{k}$

Zufallsvorgänge, Ereignisse und Wahrscheinlichkeiten

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

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

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

Ereignisse und Wahrscheinlichkeiten

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

$\omega = (1,3)$ $\quad$ $A = \{(1,3)\}$

$\omega \in \Omega$ $\quad$ $A \subset \Omega$

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

$$A \subset \Omega$$

- 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:**

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

funktioniert ✓

falls alle ω die gleiche „Chance" haben

- **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

HA

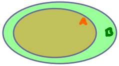
Rechenregeln für Wahrscheinlichkeiten

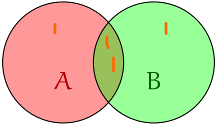
- Wichtige Rechenregeln:

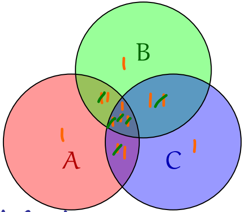
leere Menge

1. $P(A) \leqq 1$
2. $P(\emptyset) = 0$
3. $A \subset B \Rightarrow P(A) \leqq P(B)$
4. $P(\bar{A}) = 1 - P(A)$
5. $P(A \cup B) =$
   $P(A) + P(B) - P(A \cap B)$

A ist Teilmenge von B







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

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

- **Beispiel:**

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

Beispiel Gegenereignis

Der Fall Sally Clark

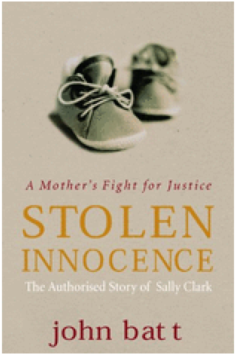
- Sally Clarks Söhne Christopher und Harry sterben 1996 und 1997 beide kurz nach der Geburt an plötzlichem Kindstod.
- Kinderarzt: „Wahrscheinlich Mord, da 2 maliger plötzlicher Kindstod sehr unwahrscheinlich!“ (ohne konkrete Hinweise)
- Gerichtliche Untersuchung
- Hauptargument der Anklage gestützt 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

I didn't kill my babies

In November 1999 solicitor Sally Clark, 36, was convicted of one of the most horrific crimes any mother could imagine – the murder of her baby sons, Christopher, aged 11 weeks, and Harry, eight weeks.



Beispiel Gegenereignis

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:

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

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.
- 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.

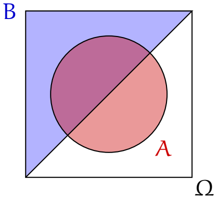
Bedingte Wahrscheinlichkeiten

- 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:

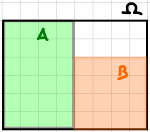
Wahrsch. von Ereignis A
unter Bedingung, dass
Ereignis B eintritt

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

- Im Venndiagramm:

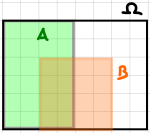


Bedingte Wahrscheinlichkeit: $P(A) = 0.5$, $P(B) = 1/3$



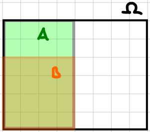
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= 0$$



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

$$= \frac{8/48}{16/48} = 1/2$$



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

$$= \frac{P(B)}{P(B)} = 1$$