

Rechnen mit Wahrscheinlichkeiten

Statistik (Biol./Pharm./HST) – FS 2015



TORQUE: etutoR

ETH zürich Sie sind angemeldet als [Markus Kalisch \(Logout\)](#) Sprache

→ Startseite → Meine Kurse → TORQUE → etutoR: Online Selbstlernkurs für die Statistik-Software R → etutoR

etutoR → Kursübersicht → Kursinformation → Discussion

Kurs: etutoR Bearbeiten einschalten

→ **Thema 1** Mein Bearbeitungsstand: 0/7 Forum: 1
Resource Discussions: 3
Tests: 3

Lektion 1: R und R-Studio

Um loslegen zu können, installieren wir zuerst R und R-Studio.

Lernziele: R und R-Studio installieren können

→ **Thema 2** Mein Bearbeitungsstand: 0/6 Forum: 1
Resource Discussions: 2
Tests: 2
Datei: 1

Lektion 2: Daten einlesen

Zuerst schauen wir an, wie wir Daten in R einlesen können.

Lernziele: Daten aus Excel exportieren und in R einlesen können.

→ **Thema 3** Mein Bearbeitungsstand: 0/8 Forum: 1
Resource Discussions: 3
Tests: 3
Datei: 1

Lektion 3: Daten bearbeiten

Jetzt geht es darum, die Daten in R entsprechend zu bearbeiten.

Lernziel: Datentypen kennen; Daten auswählen und zusammenführen können; mit fehlenden Werten umgehen können

→ **Thema 4** Mein Bearbeitungsstand: 0/5 Forum: 1
Resource Discussions: 2
Tests: 2

Lektion 4: Grafiken

Nun schauen wir, wie wir in R Grafiken erzeugen und exportieren können.

Lernziele: Grafiken erstellen, beschriften und für Berichte oder Vorträge exportieren können.

→ **Thema 5** Mein Bearbeitungsstand: 0/7 Forum: 1
Resource Discussions: 3
Tests: 3

Lektion 5: Zufallsvariablen

Mit R kann man bequem Wahrscheinlichkeiten etc. berechnen.

Lernziele: Wahrscheinlichkeiten von bekannten Verteilungen mit R berechnen und Zufallszahlen erzeugen können.

Thema 10 = Musterlösung

TORQUE: etutoR

ETH zürich Sie sind angemeldet als [Markus Kalisch \(Logout\)](#) Sprache ▾

→ Startseite → Meine Kurse → TORQUE → etutoR: Online Selbstlernkurs für die Statistik-Software R → etutoR → Thema 1

etutoR → Kursübersicht → Kursinformation → Discussion

Kurs: etutoR

Bearbeiten einschalten

Thema 1

[Thema 2 ▶](#)

Fortschritte 

 R und R-Studio

R und R-Studio - 1



[FAQ: R und R-Studio](#)

[↓ See All](#)

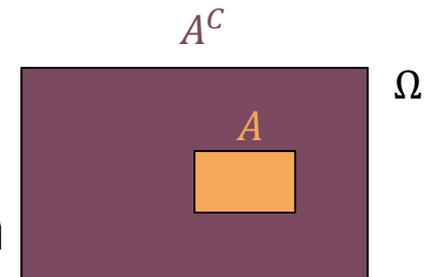
[? Ask a Question](#)

 R Studio 1

Wiederholung: Wa. Modell

- Grundraum Ω : Alle möglichen Elementarereignisse ω
Bsp: $\Omega = \{1,2,3,4,5,6\}$, $\omega = 3$
- Ereignis “A”: Teilmenge des Grundraums
Bsp: $A = \{2,4,6\}$ (“gerade Zahlen”)
- Wahrscheinlichkeit: Axiome von Kolmogorov
 - $P(A) \geq 0$
 - $P(\Omega) = 1$
 - $P(A \cup B) = P(A) + P(B)$ falls $A \cap B = \{\}$
- Mengenoperationen: Venn-Diagramm
- Wa. berechnen
 - Summe der Elementarereignisse
 - Günstige / Mögliche Fälle
 - Mengenoperationen / Venn Diagramm

Bsp: $P(A^c) = P(\Omega) - P(A) = 1 - P(A)$

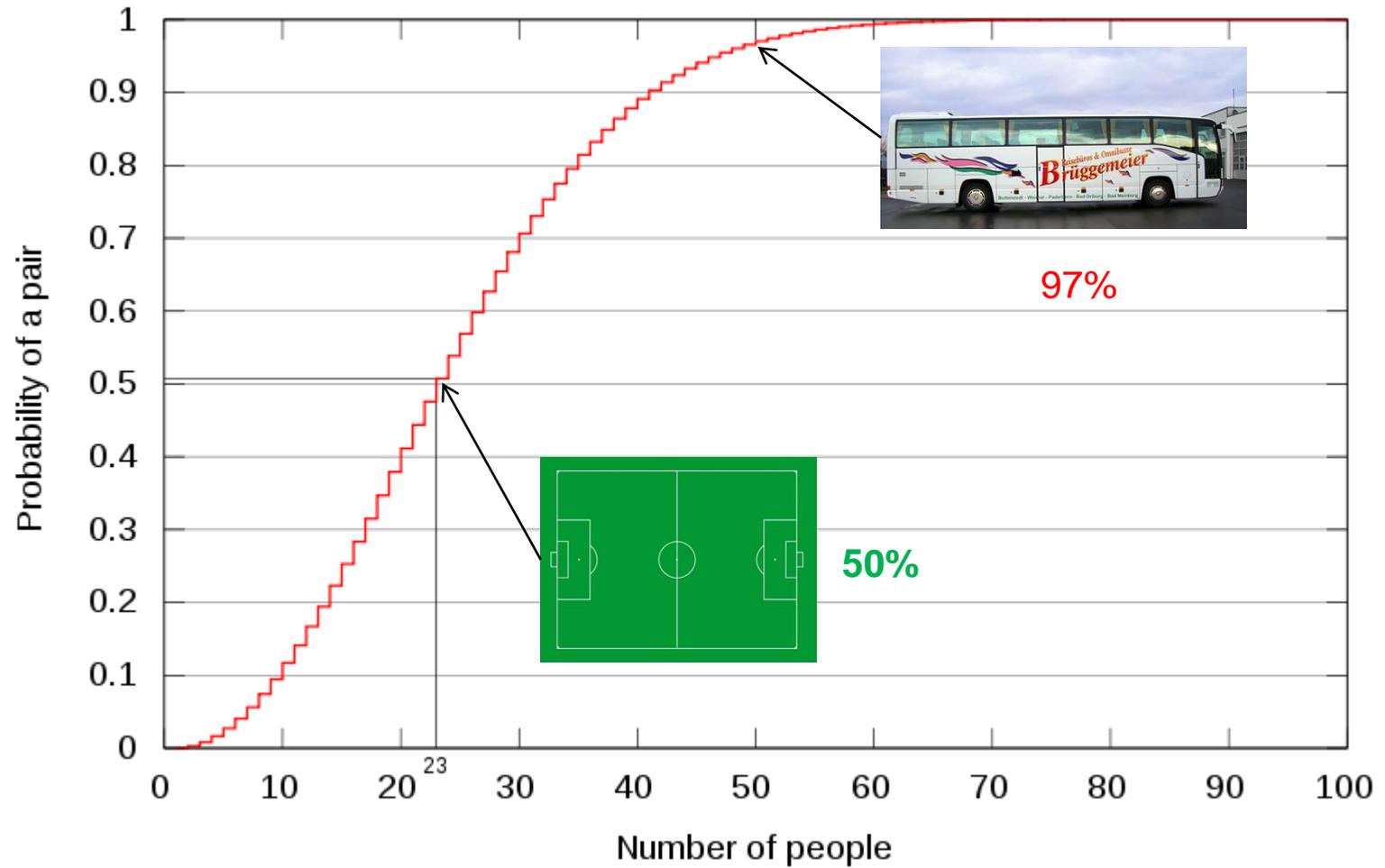


MAGIE

Geburtstagsparadox

- Wie gross ist die Wa., dass in einer Gruppe von n Personen mind. 2 Personen am gleichen Tag Geburtstag haben?
- Ann.: Alle Geburtstage sind gleich wahrscheinlich
- Ereignis A : Mind. 2 Personen haben am gleichen Tag Geburtstag
- $P(A) = 1 - P(A^C)$
$$P(A^C) = \frac{365 * 364 * \dots * (365 - n + 1)}{365^n}$$
- Bsp mit $n=40$: $P(A) = 1 - \frac{365*364*\dots*326}{365^n} \approx 0.89$

Geburtstagsparadox

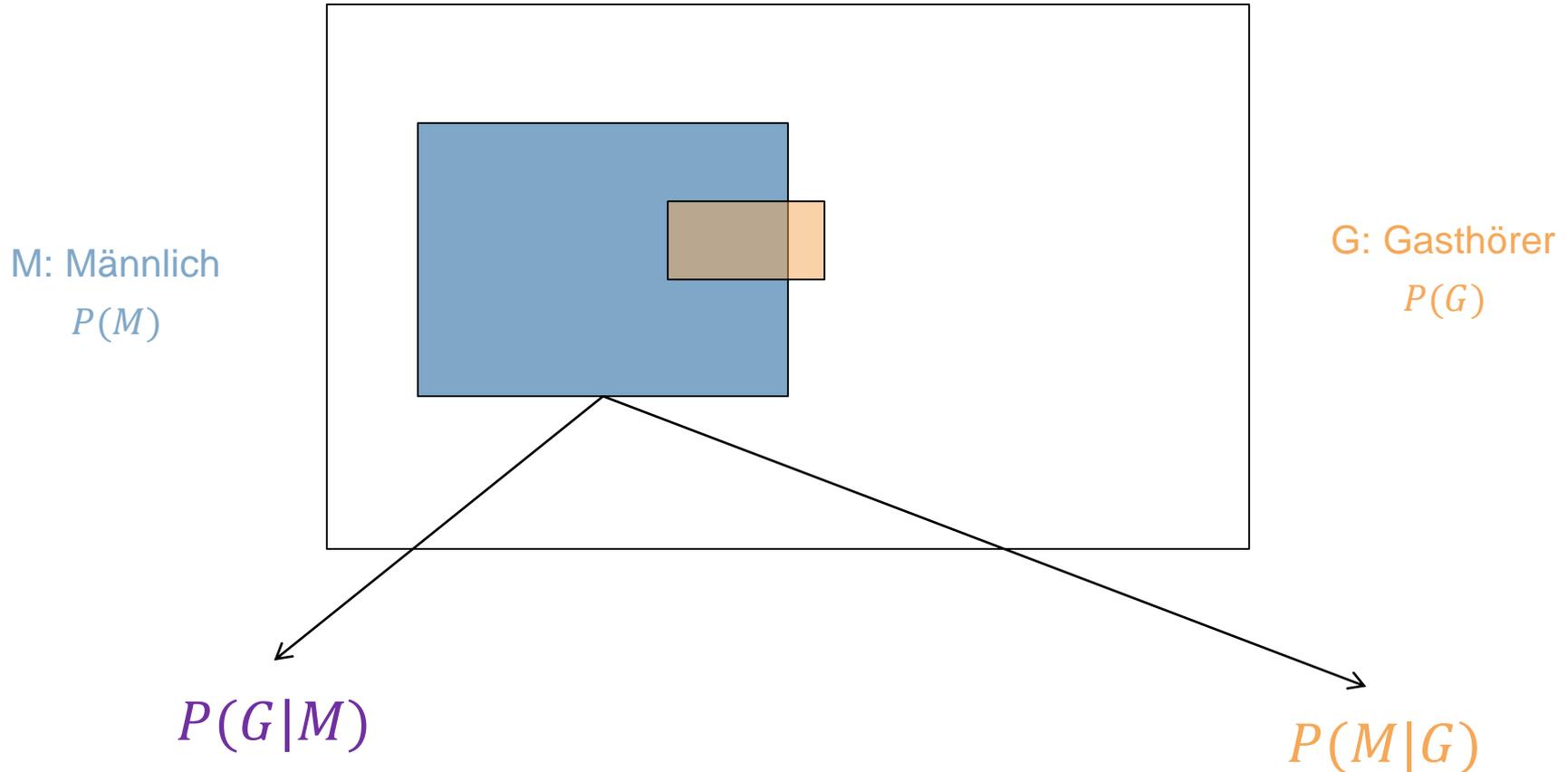


Abgezockt - Wollen Sie Informationen kaufen?

- Würfelspiel 100 mal (Runden):
 - Gerade AZ: Sie gewinnen 10 SFr
 - Ungerade AZ: Ich gewinne 10 SFr
- Jede Runde, nach dem Wurf und vor dem Aufdecken:
Sie dürfen 50 Rappen zahlen und
 - erfahren, ob $AZ \leq 3$ und
 - entscheiden, ob Sie diese Runde mitspielen wollen
- Lohnt es sich, für die Information zu zahlen?
- Enthält das Ereignis « $AZ \leq 3$ » Informationen über das Ereignis “Gerade AZ”?

Bedingte Wahrscheinlichkeit

Ω : Studenten dieser VL



Wa. für Gasthörer,
wenn ein Mann gewählt wurde

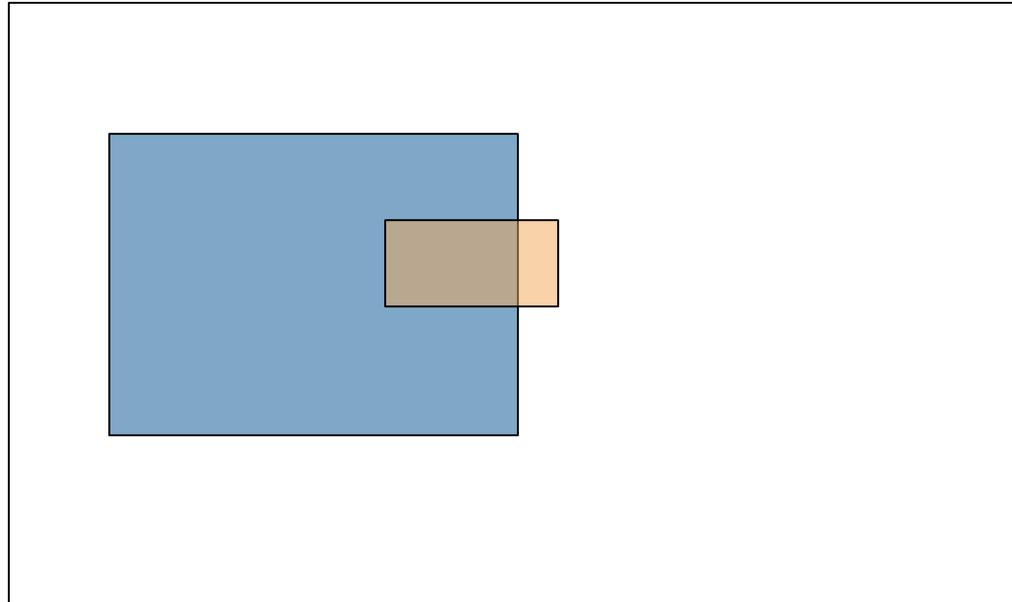
Wa. für Mann,
wenn Gasthörer gewählt wurde

Bedingte Wahrscheinlichkeit



Ω : Studenten dieser VL

M: Männlich
 $P(M)$



G: Gasthörer
 $P(G)$

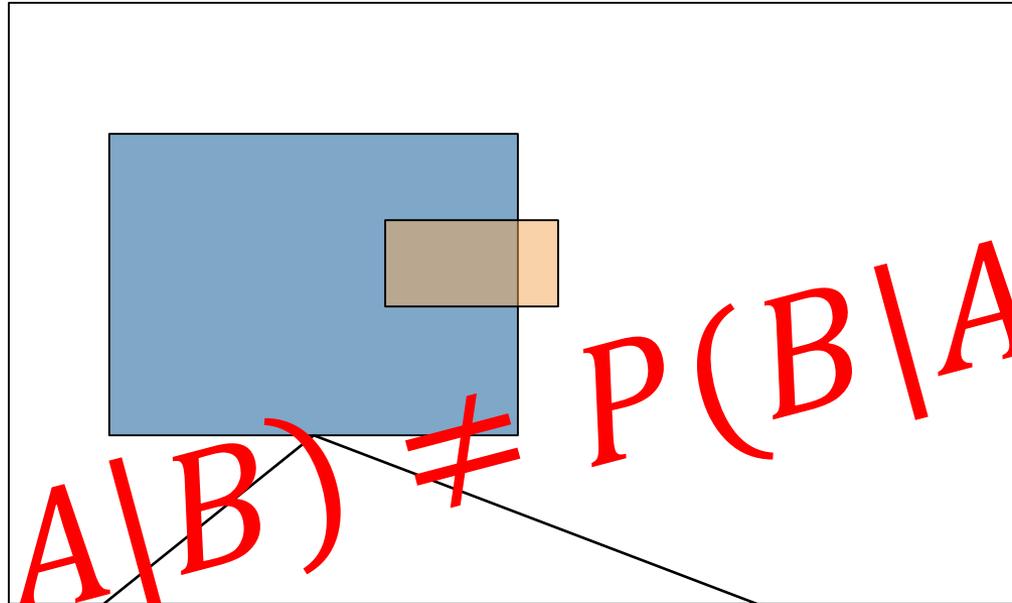
Welche Aussage ist korrekt?

1. $P(M|G) = P(G|M)$
2. $P(M|G) > P(G|M)$
3. $P(M|G) < P(G|M)$

Bedingte Wahrscheinlichkeit

Ω : Studenten dieser VL

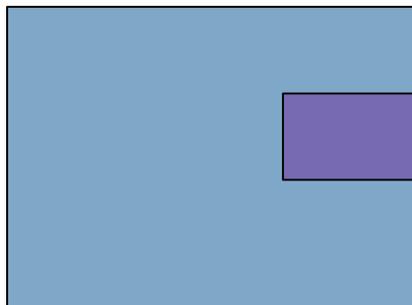
M: Männlich
 $P(M)$



G: Gasthörer
 $P(G)$

$P(A|B) \neq P(B|A)$

$P(G|M)$ klein



Neuer Grundraum

$P(M|G)$ gross

Neuer Grundraum



Brustkrebsstudie UK

- Seit 1988 in UK: Frauen über 50 machen Brustkrebstest
- Bsp: 2004/2005 wurden 1.7 Mio Tests gemacht
- Wa. für Brustkrebs: 0.01

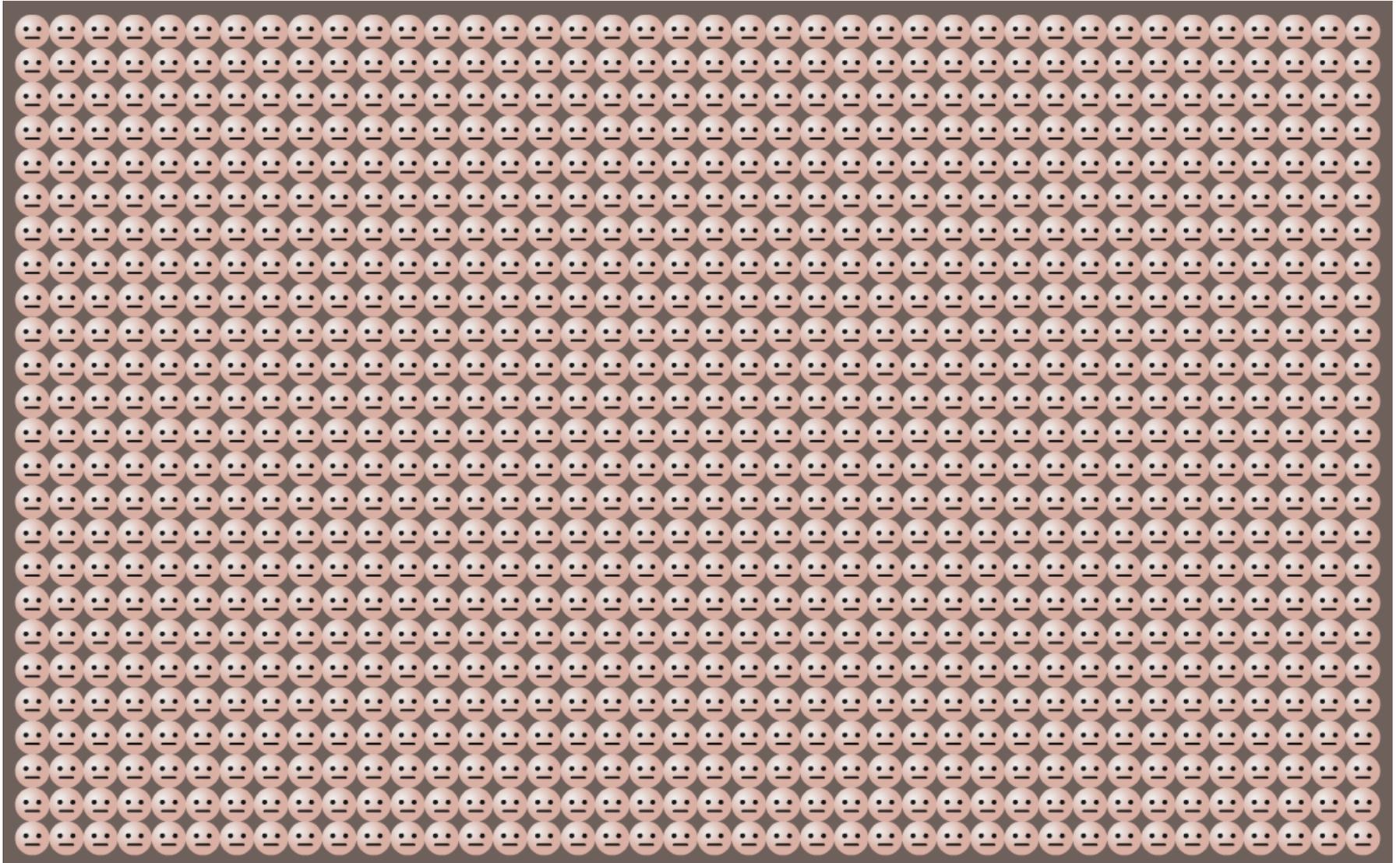
- Test positiv, falls krank: 0.85
- Test negativ, falls gesund: 0.9

- Angenommen, der Test von Frau Smith ist positiv.
Wie wahrscheinlich ist es, dass Frau Smith krank ist?

- <http://understandinguncertainty.org/node/182>

Brustkrebsstudie UK: Bsp mit 1000 Frauen

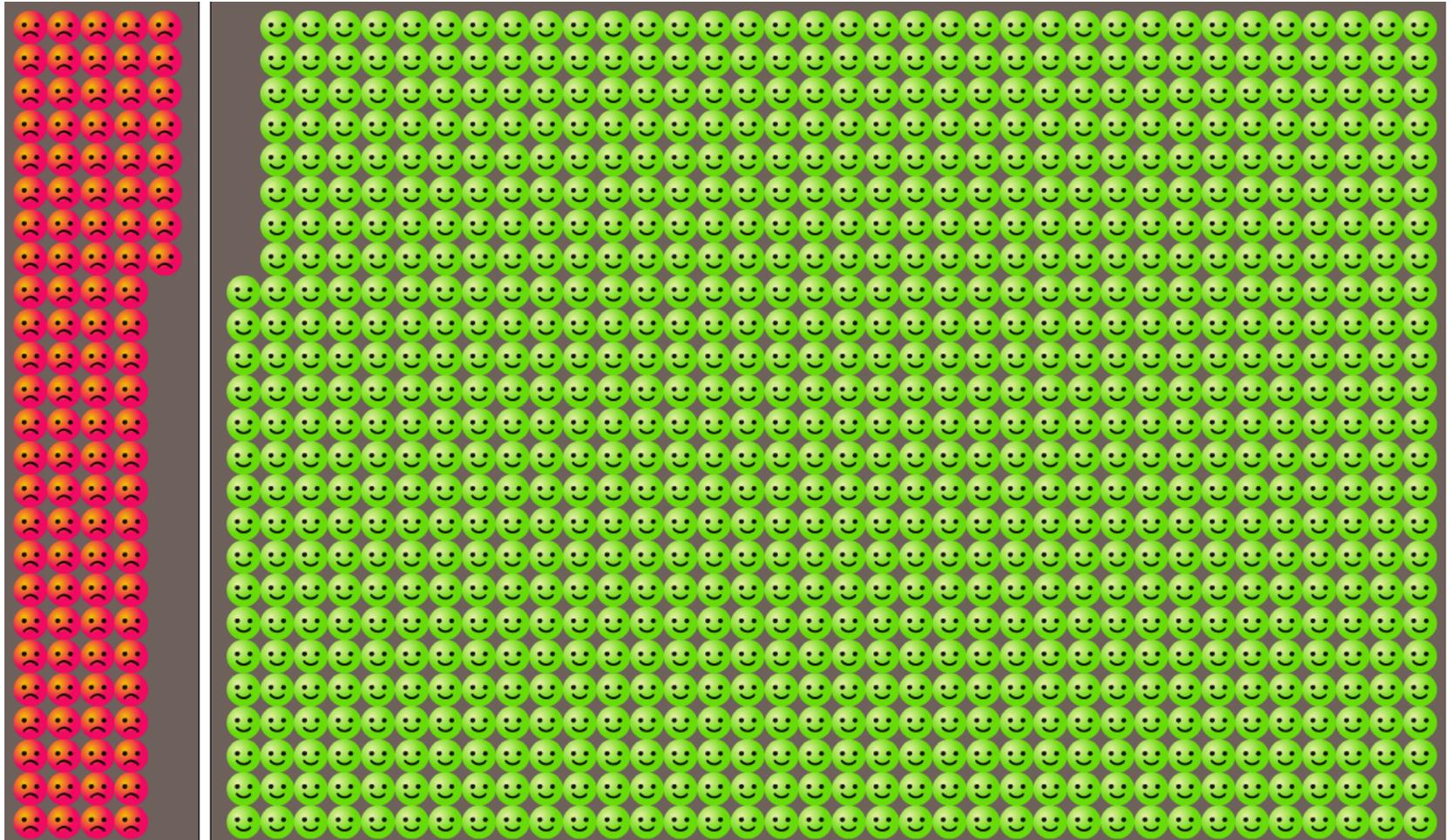
1000 Tests



Brustkrebsstudie UK: Bsp mit 1000 Frauen

108 positiv

892 negativ



Brustkrebsstudie UK: Bsp mit 1000 Frauen

108 positiv

892 negativ

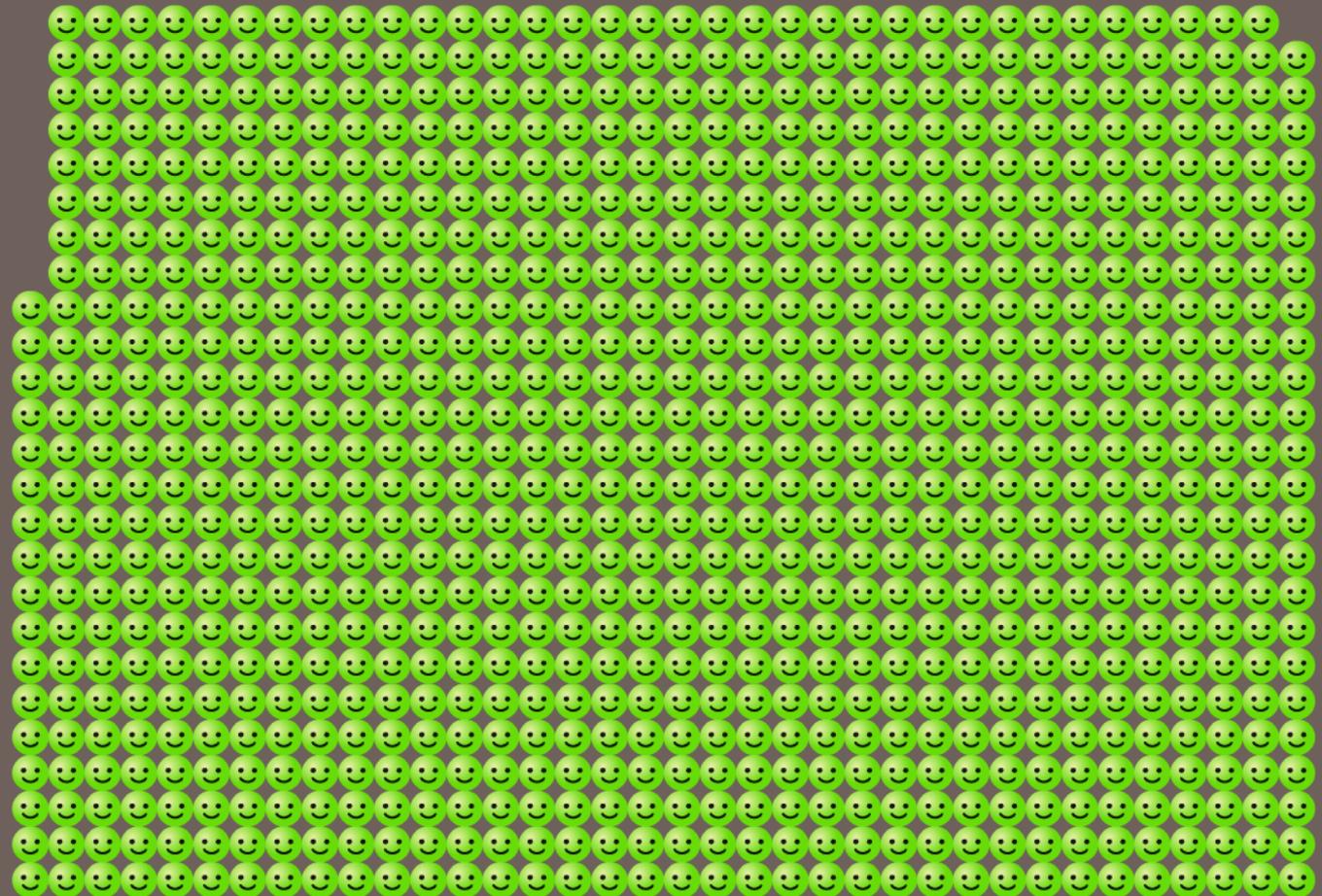
Von 108 positiv getesteten Frauen sind nur 9 krank (ca. 8%) !

krank



krank

gesund



gesund

Brustkrebsstudie UK: Erklärung

- K: Krank, T: Test positiv
- Gegeben: $P(K) = 0.01, P(T|K) = 0.85, P(T^c|K^c) = 0.9$
- Gesucht: $P(K|T)$

- Satz der totalen Wahrscheinlichkeit:

$$\begin{aligned} P(T) &= P(T|K)P(K) + P(T|K^c)P(K^c) = \\ &= 0.85 * 0.01 + 0.1 * 0.99 = 0.108 \end{aligned}$$

- Satz von Bayes:

$$P(K|T) = \frac{P(T|K)P(K)}{P(T)} = \frac{0.85 * 0.01}{0.108} = 0.08$$

- Gegeben der Test ist positiv, ist die Wa. wirklich krank zu sein nur 8% !

Publikationen verstehen: Odds und odds Ratio

- Rauchen und Lungenkrebs:
H.F. Dorn, «The Relationship of Cancer of the Lung and the Use of Tobacco», American Statistician, 1954

Two years ago cancer of the lung was an unfamiliar and little discussed disease outside the pages of medical journals. Today it is a common topic of discussion, apparently throughout the entire world. The newspapers, radio, television, popular magazines, business and trade journals vie with medical journals for the last word on the relationship of lung cancer and the use of tobacco. This topic has become the featured subject on the program of almost every up-to-date and alert organization interested in medical and health problems.

Datenlage: Studie von 1954

Aus: "Statistical Sleuth", Kap 18.1.3

	Lungenkrebs (L)	Kontrolle	Total
Raucher (R)	83	72	155
Nichtraucher	3	14	17
Total	86	86	172

Unfaire Auswahl?

- Alternative zu Wa.: odds, odds-Ratio (OR)

- $P(L|R) \approx \frac{83}{155} = 0.54$; $P(L|R^C) \approx \frac{3}{17} = 0.18$

- $odds(L|R) = \frac{P(L|R)}{P(L^C|R)} = \frac{P(L|R)}{1-P(L|R)} \approx \frac{0.54}{0.46} = 1.17$
 $odds(L|R^C) \approx 0.22$

Wie genau?

- $OR = \frac{odds(L|R)}{odds(L|R^C)} = \frac{1.17}{0.22} = 5.33$

"Odds ratio"

"Die odds an Lungenkrebs zu erkranken sind für Raucher ca. 5 mal grösser."

Rauchen und Lungenkrebs: Aktuelles Wissen

- Einfach verständliche Übersicht des aktuellen Wissens:
http://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm
- “Cigarette smoking is the **number one risk factor** for lung cancer. In the United States, **cigarette smoking causes about 90% of lung cancers.**” [aus obiger Quelle]