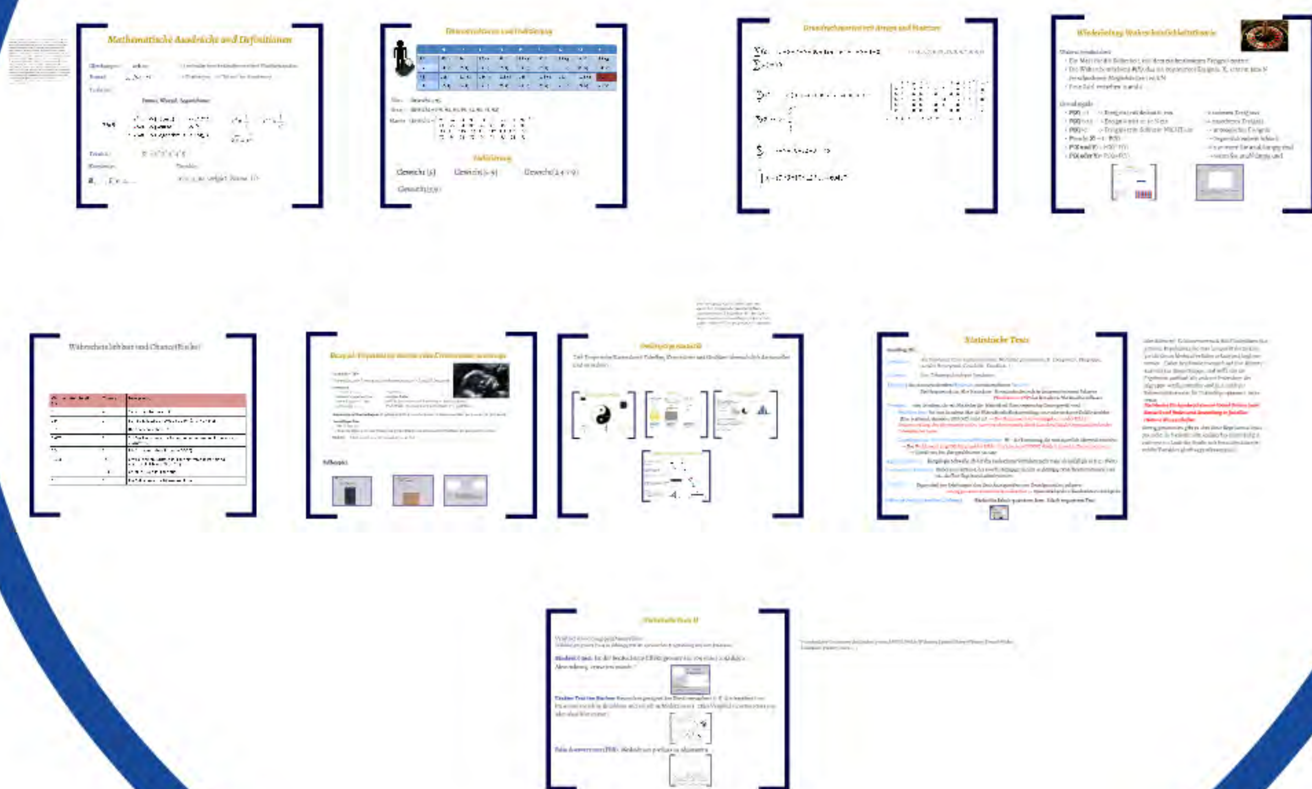
Biostatistik und Programmieren mit R

Auffrischung:

Grundlegende Konzepte der Mathematik und der Statistik



Mathematische Ausdrücke und Definitionen

Gleichungen: $a+b=c$ --> verbinden zwei Ausdrücke mit einem Gleichheitszeichen

Formel: $SD=\sqrt{\frac{\sum_{i=1}^{n}(x_i-m)^2}{n-1}}$ --> Gleichungen, ein "Rezept" zur Berechnung

Funktion:

Potenz, Wurzel, Logarithmus:

$2^3=8$

- $2^3=x$ -> (Potenz) -> $2*2*2$
- $x^3=8$ -> (Wurzel) -> $\sqrt[3]{8}$
- $2^x=8$ -> (Logarithmus) -> $\log_2 8$

$$x^{-1}=\frac{1}{x} \qquad x^{-n}=\frac{1}{x^n}$$

$$\sqrt[n]{x}=x^{\frac{1}{n}}$$

Fakultät: $5!=1*2*3*4*5$

Konstanten

$\mathbf{R}$, π, F, e, 2, ...

Variable

x, y, z, m, weight, Name, ID

Datenstrukturen und Indizierung

	t1	t 2	t3	t4	t5	t6	t7	t8	t 9
P 1	95 Kg	94 Kg	94 Kg	93 Kg	93 Kg	92 Kg	93 Kg	91 Kg	92 Kg
P 2	84 Kg	85 Kg	84 Kg	85 Kg	84 Kg	85 Kg	86 Kg	85 Kg	84 Kg
P 3	115 Kg	114 Kg	114 Kg	113 Kg	113 Kg	112 Kg	111 Kg	112 Kg	110 Kg
P 4	75 Kg	74 Kg	74 Kg	75 Kg	74 Kg	75 Kg	75 Kg	74 Kg	75 Kg

Wert: Gewicht = 95

Array: Gewicht = (95, 94, 94, 93, 92, 93, 91, 92)

Matrix: $$\text{Gewicht} = \begin{pmatrix} 95 & 94 & 94 & 93 & 93 & 92 & 93 & 91 & 92 \\ 84 & 85 & 84 & 85 & 84 & 85 & 86 & 85 & 84 \\ 115 & 114 & 113 & 113 & 113 & 112 & 111 & 112 & 110 \\ 75 & 74 & 74 & 75 & 74 & 74 & 75 & 74 & 75 \end{pmatrix}$$

Indizierung

Gewicht [5]

Gewicht[5-9]

Gewicht[2 4 7-9]

Gewicht[3,9]

Grundrechenarten mit Arrays und Matrizen

$$\sum(a) = (2+5+7+9+10+12+8+4+7+6+4+2)$$

$a = (2, 5, 7, 9, 10, 12, 8, 4, 7, 6, 4, 2)$

$$\sum(a) = 76$$

$$\sum_i (G_{i,j}) = \quad g = \begin{pmatrix} 2 & 8 & 12 & 16 & 25 & 27 & 24 & 21 & 26 & 11 & 6 & 7 \end{pmatrix}$$

$$\sum_j (G_{i,j}) = \quad g' = \begin{pmatrix} 76 \\ 23 \\ 0 \\ 84 \\ 2 \\ 0 \end{pmatrix}$$

$$G = \begin{pmatrix} 2 & 5 & 7 & 9 & 10 & 12 & 8 & 4 & 7 & 6 & 4 & 2 \\ 0 & 0 & 0 & 3 & 7 & 6 & 4 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 5 & 4 & 8 & 9 & 12 & 14 & 17 & 5 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 76 \\ 23 \\ 0 \\ 84 \\ 2 \\ 0 \end{pmatrix}$$

$$g = \begin{pmatrix} 2 & 8 & 12 & 16 & 25 & 27 & 24 & 21 & 26 & 11 & 6 & 7 \end{pmatrix}$$

$$\sum_{i=3}^{7} a_i = (7+9+10+12+8) = 46$$

$$\prod_{3}^{7} a_i = (7*9*10*12*8) = 60480$$

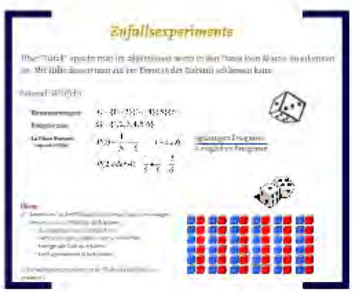
Wiederholung Wahrscheinlichkeitstheorie

Wahrscheinlichkeit

- Ein Mass für die Sicherheit, mit dem ein bestimmtes Ereignis eintritt
- Die Wahrscheinlichkeit ***P(X)***, das ein bestimmtes Ereignis, X, eintritt (aus N verschiedenen Möglichkeiten) ist 1/N
- Eine Zahl zwischen 0 und 1

Grundregeln:

- ***P(X)*** = 1 -> Ereignis tritt definitiv ein -> sicheres Ereigniss
- ***P(X)***= 0.5 - > Ereignis tritt zu 50 % ein -> unsicheres Ereignis
- ***P(X)***=0 -> Ereignis tritt definitiv NICHT ein -> unmögliches Ereignis
- **P**(nicht **X**) = **1** - **P**(X) -> Gegenwahrscheinlichkeit
- **P**(**X und Y**) = P(X)*P(Y) -> nur wenn Sie unabhängig sind
- **P(X oder Y)**= P(X)+P(Y) -> wenn Sie unabhängig und



Zufallsexperimente

Über "Zufall" spricht man im allgemeinen, wenn in den Daten kein Muster zu erkennen ist. Mit Hilfe dessen man auf ein Event in der Zukunft schliessen kann

Beispiel: Würfeln

Elementarereignis: $E = \{1\}\{2\}\{3\}\{4\}\{5\}\{6\}$

Ereignisraum: $\Omega = \{1,2,3,4,5,6\}$

La Place Formel:
(ungezinkte Würfel)

$$P(i) = \frac{1}{N} = \frac{1}{6} \qquad i = 1,\ldots 6$$

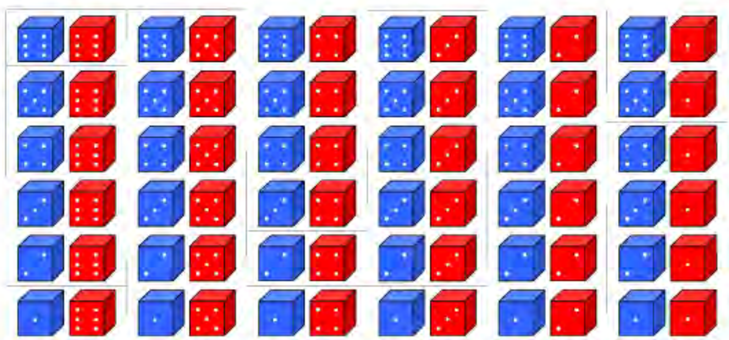
$$\frac{\#\text{günstigen Ereignisse}}{\#\text{ möglichen Ereignisse}}$$

$$P(2\ oder\ 4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

Übung:

1) Berechnen Sie die Wahrscheinlichkeiten durch einmaliges werfen von zwei Würfeln als Ergebnis
   - die Augensumme 5 zu berechnen
   - Nicht die Augensumme 5 oder 4 zu würfeln
   - eine gerade Zahl zu erhalten
   - die Augensumme 15 zu erhalten

2) Für welche Augensumme ist die Wahrscheinlichkeit am grössten ?

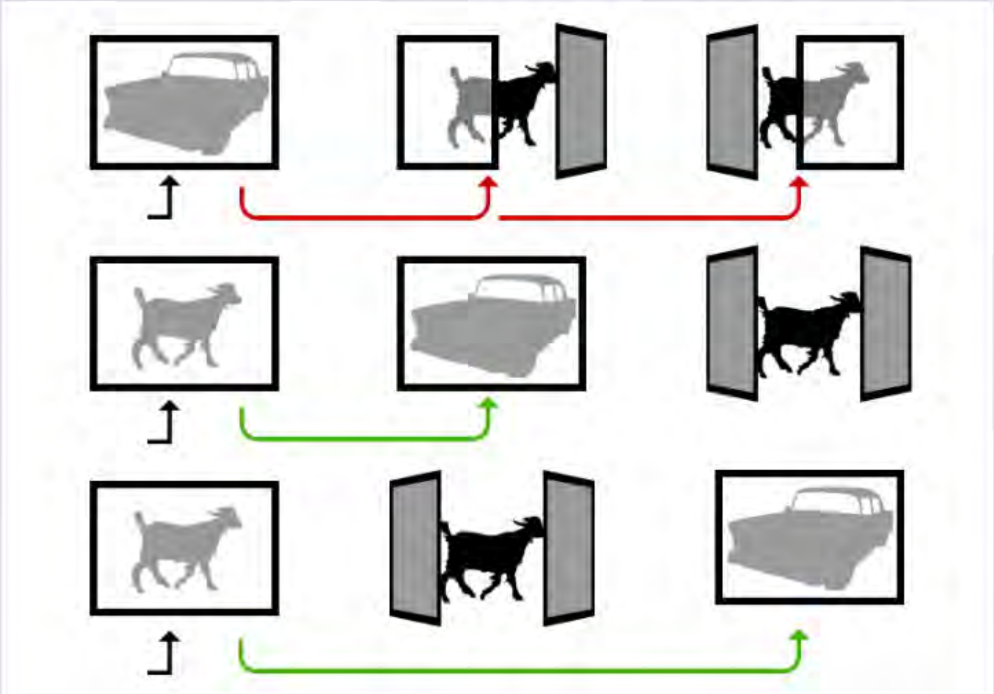


Das Ziegen Problem

Am 9. September 1990 wurde ein Rätsel aus der Wahrscheinlichkeitsrechnung allgemein bekannt (Marilyn vos Savant), das seit langem gelöst ist, doch bis heute die Gemüter bewegt .

Idee:

Lösung:



Frage:

- Hypothese 1) Man sollte bei seiner ursprünglichen Wahl bleiben, da sich durch einen Wechsel die Chancen verschlechtern würden
- Hypothese 2) Man sollte wechseln, weil die Wahrscheinlichkeit dadurch zu gewinnen steigt .
- Hypothese 3) Es ist egal, ob man wechselt oder nicht, da die Chancen 50:50 stehen.

http://www.userpages.de/ziegenproblem/index.php

Wahrscheinlichkeit und Chance(Risiko)

Wahrscheinlichkeit P(X)	Chance	Interpretation
1	∞	Sicheres eintreten von X
0.9	9	Das Ereignis wird zu 90% eintreffen (Chance = 9:1)
0.75	3	Die Chancen stehen 3:1
0.667	2	2:1 Das Eintreten von X ist doppelt so wahrscheinlich wie nicht einzutreten
0.5	1	Die Chancen stehen 1:1 (oder 50-50)
0.333	0.5	X wird in einem Drittel aller Fälle eintreten (X ist halb so wahrscheinlich wie "Nicht X")
0.1	0.1111	X tritt in 10% aller Fälle auf
0	0	Ein Auftreten von X ist ausgeschlossen

Beispiel: Tripeltest im Rahmen des Ersttrimester-Sceenings

Zeitpunkt: 4. SSM

Was wird gemacht ? Biochemischer Blutuntersuchung + Ultraschall Diagnostik

Bewertung:

- kleiner als 1:500 - unauffällig
- zwischen 1:250 und 1:500 - erhöhtes Risiko
- zwisch 1:250 und 1:150 - auffällig (Amniozentese/Chorionbiopsie wird angeboten)
- grösser als 1:150 - Hoch-Risiko (Amniozentese/Chorionbiopsie wird empfohlen)

Amniozentese/Chorionbiopsie: Ergebnis zu 100% sicher aber invasiv und mit einer Abort-Rate von ca. 2% , bzw ca. **6%**

Vorteil Tripel-Test:

- Nicht- Invasiv
- Kann beruhigen u. zu einer Entscheidung gegen Routine Amniozentese/Chorionbiopsie bei älteren Müttern führen

Nachteil: Falsch-Positiv (5 %) , Sensitivität bei ca. 40-70%

Fallbeispiel:

Fall 1: Mutter(31): gesund

SS stabil

T-Risiko: **1:3800**

=> Test gilt als klar negativ

Nicolai *2012

Fall 2: Mutter(33) gesund, leichte Komplikationen,
T-Risiko **1:18**
=> Test gilt als positiv/pathologisch

Risiko (T13) = 1:18
P(T13) = 0.04 => Risiko liegt bei 5,56 %

P(nicht T13) => trotzdem ein GESUNDES Kind zu bekommen, liegt bei 94,4 %

P(Abort unter Chorionbiopsie) => 6%
Risiko einer Fehlbildung (ohne Abort) steigt um den Faktor 3-5

Elina *2014

Fall 2: Mutter(33) gesund, leichte Komplikationen,
T-Risiko **1:18**
=> Test gilt als positiv/pathologisch

Risiko (T13) = 1:18
P(T13) = 0.04 => Risiko liegt bei 5,56 %

P(nicht T13) => trotzdem ein GESUNDES Kind zu bekommen, liegt bei 94,4 %

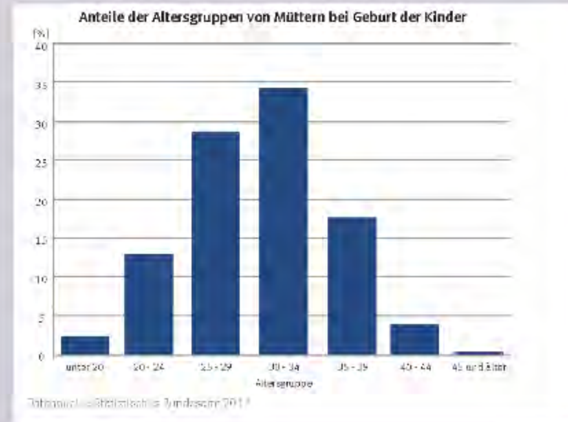
P(Abort unter Chorionbiopsie) => 6%
Risiko einer Fehlbildung (ohne Abort) steigt um den Faktor 3-5


Elina *2014

Rechenbeispiel in grösseren Dimensionen:

Hintergrundrisiko für Down Syndrom: 1:700
- stark abhängig vom Alter
(Sinn macht es nur zwischen 25 und 40 Jahren)

http://www.nackentransparenz-forum.de/risikotabelle-t21.html



Übung:

Angenommen, 400.000 der Schwangere machen
Zu wieviele der Tests wären falsch positiven (bei einer FPR von 5%) ?
Diese falsch positiven Tests werden anschliessend durch eine Fruchtwasseruntersuchung überprüft. Wie gross wird etwa die hierdurch verursachte Anzahl an Fehlgeburten sein (Abort-Risiko: 1%)?