Christian-Albrechts-Universität zu Kiel

08_verteilungen

Wahrscheinlichkeitsrechnung und Verteilungen

Daten, Variablen, Merkmale

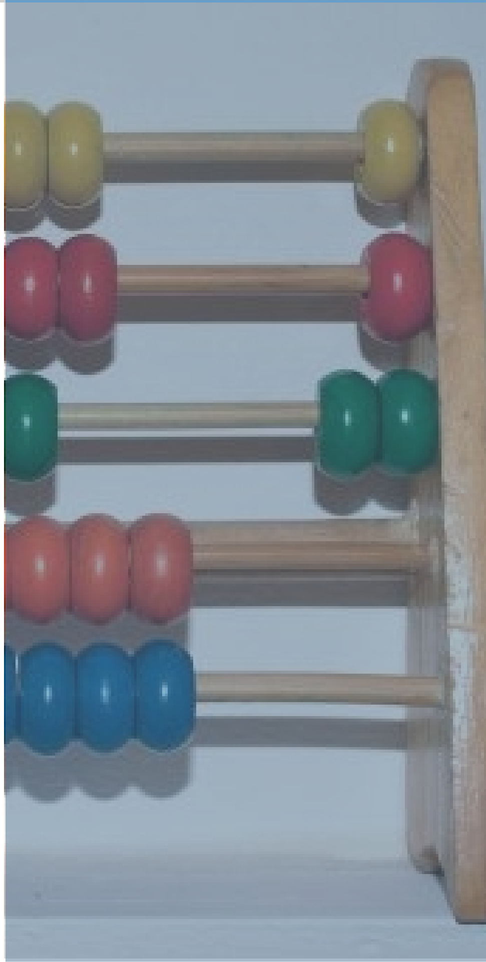
Variable oder Merkmal:
Das, was gemessen oder untersucht werden soll
Bsp: Körpergröße

Merkmalsträger
Das, dessen Merkmal gemessen wird.
Bsp: Ich als Besitzer einer Körpergröße, Gräber, Personen...

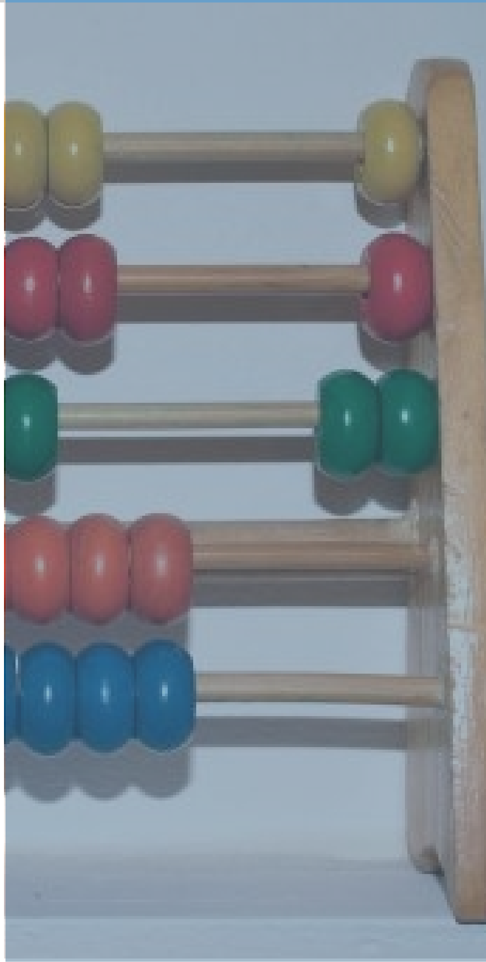
Variablenausprägen oder Merkmalswerte (oder einfach Werte):
Tatsächlich gemessene Eigenschaften.
Bsp: Meine Körpergröße beträgt 1.81 m.

Diskrete Variablen:
Variable, die nur bestimmte Werte ohne Zwischenwerte annehmen können.
Bsp: Einkommen, Anzahl von Keramikobjekten, Geschlecht (?)

Stetige Variablen:
Variablen, die jeden Wert und jeden Zwischenwert annehmen können.
Bsp: Angaben wie Körpergröße, Temperaturen, Prozentangaben

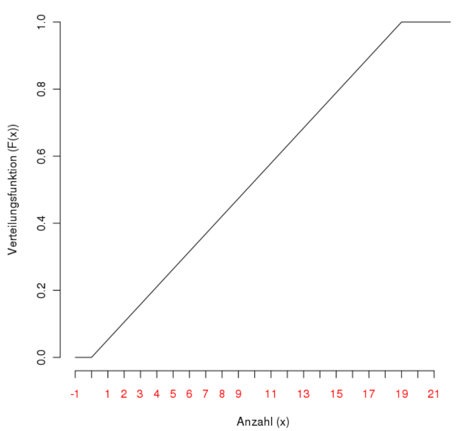


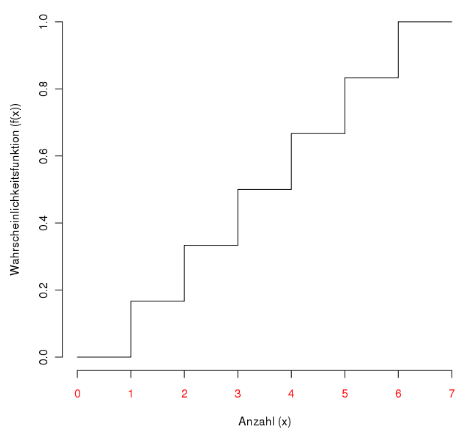
Verteilungen und Stetigkeit



Stetige Verteilung
Verteilung, bei denen für eine Variable X alle Zwischenstufen fließend definiert sind.

Diskrete Verteilung
Verteilung, die „Sprünge“ aufweist

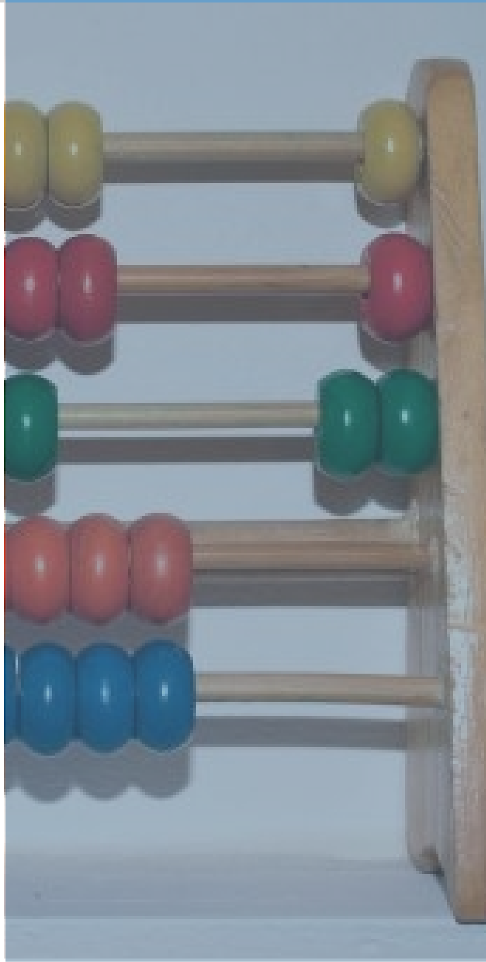




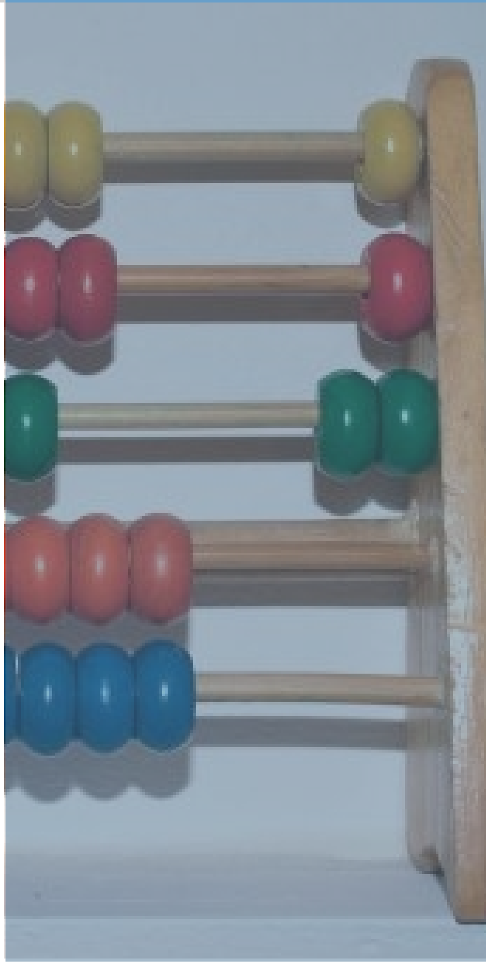
Verteilungen und Stetigkeit

Stetige Verteilung
Verteilung, bei denen für eine Variable X alle Zwischenstufen fließend definiert sind.

Diskrete Verteilung
Verteilung, die „Sprünge“ aufweist



Theoretische Verteilungen



Bestimmte Klassen von Ereignissen weisen bestimmte Verteilungsmuster auf

Binominalverteilung
Poissonverteilung

Gleichverteilung
Normalverteilung

Verteilungen sind Modelle, denen Prozesse mit bestimmten Eigenschaften zu Grunde liegen.

Sie dienen dazu, diese Eigenschaften bei echten Verteilungen zu prüfen und diese echten Verteilungen ggf. in Bereichen zu schätzen, für die keine Daten vorliegen

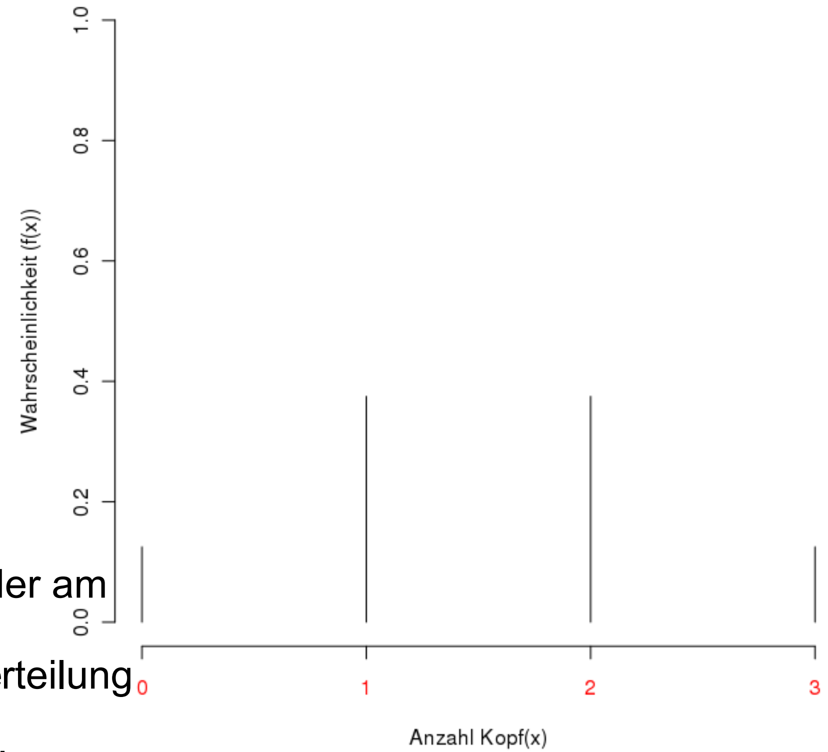
Christian-Albrechts-Universität zu Kiel

Wahrscheinlichkeitsfunktion (Dichtefunktion)

Ordnet jedem Ereignis der Zufallsvariable eine Wahrscheinlichkeit zu

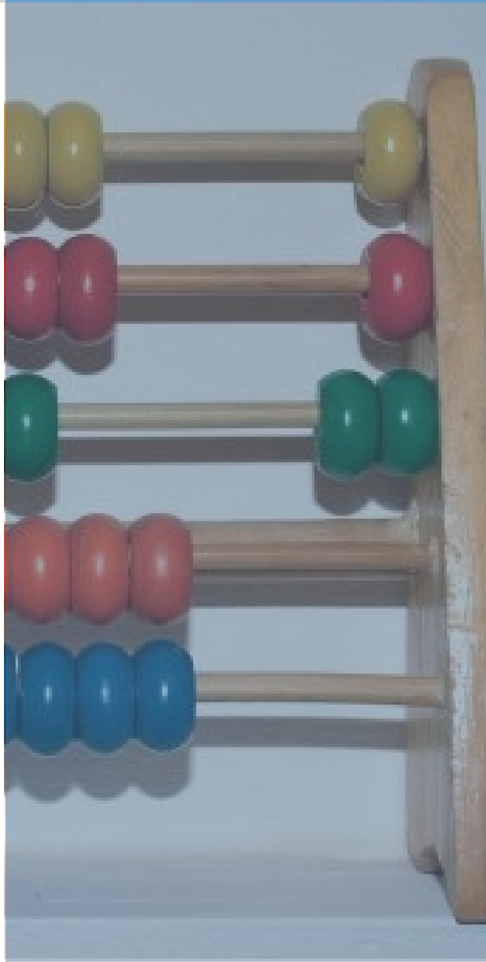
Beispiel Münzwurf

$$f(x_i)=\begin{Bmatrix} P(x_i=0)=\frac{1}{8} \\ P(x_i=1)=\frac{3}{8} \\ P(x_i=2)=\frac{3}{8} \\ P(x_i=3)=\frac{1}{8} \end{Bmatrix}$$



2 typische Eigenschaften
Erwartungswert: Der Wert, der am wahrscheinlichsten ist
Streuung: Die Varianz der Verteilung

weitere: Schiefe und Kurtosis"

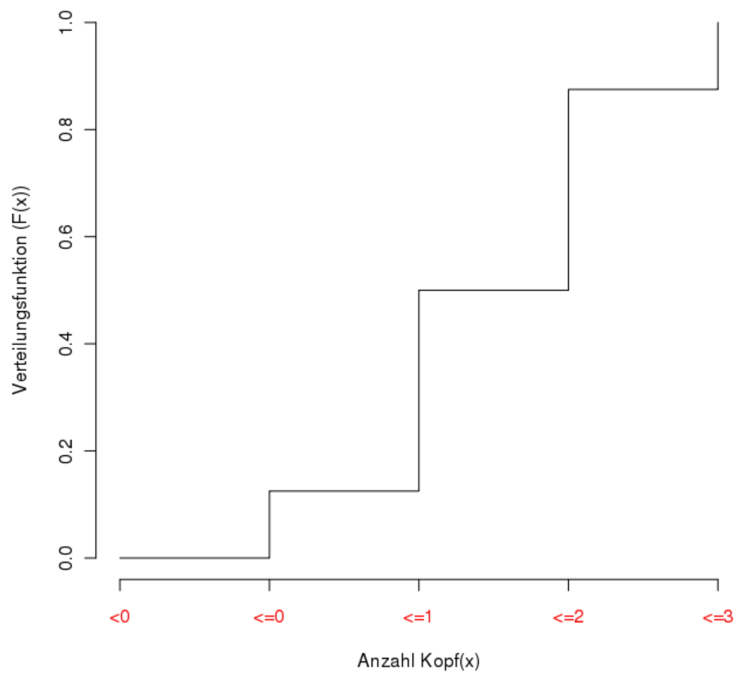


Verteilungsfunktion

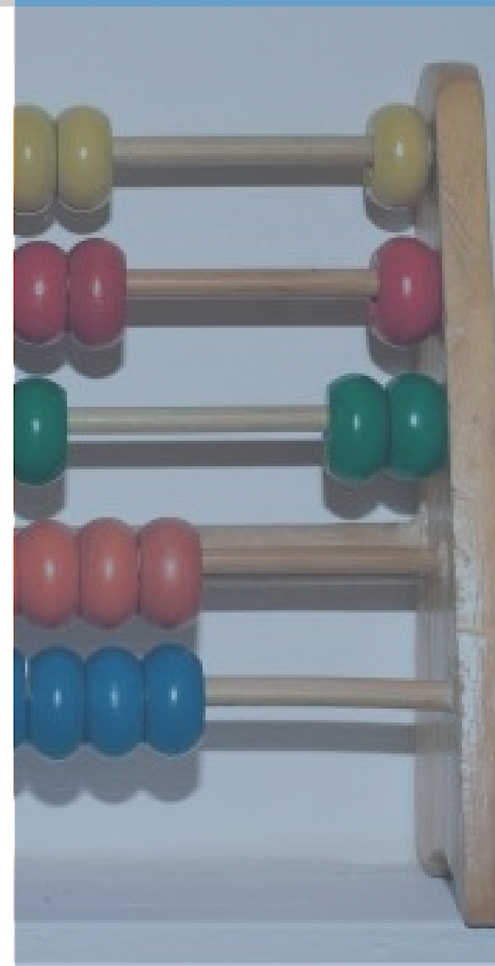
Ist die Summenfunktion der Wahrscheinlichkeitsfunktion

dh. „Wie hoch ist die Wahrscheinlichkeit, das bis zu 2x Kopf fällt?“

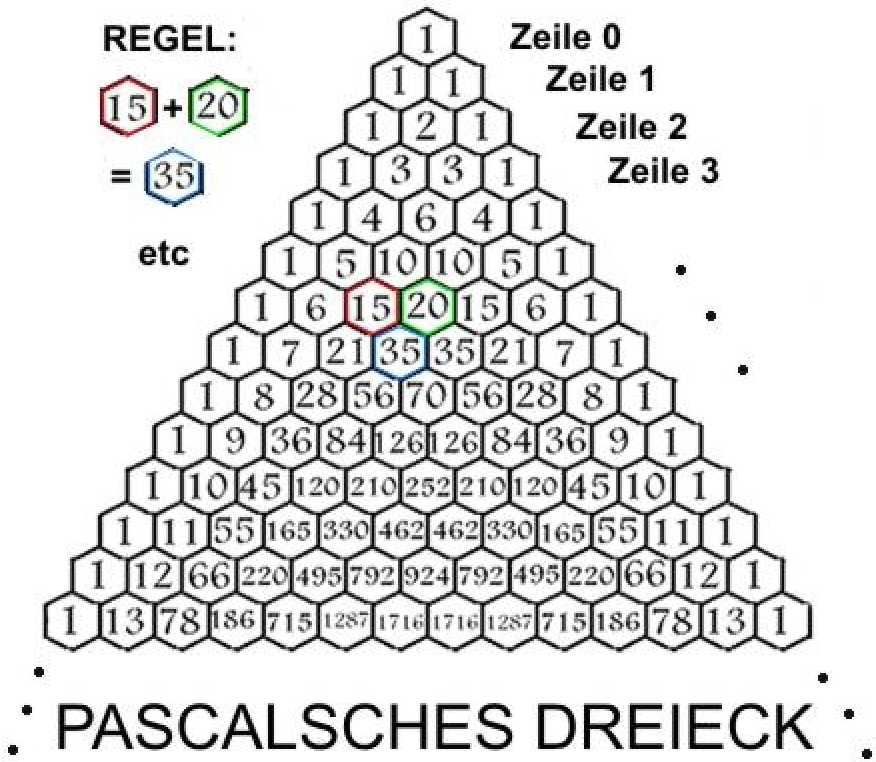
$$F(x_i) = \begin{cases} P(x_i < 0) = 0 \\ P(x_i \leq 0) = \frac{1}{8} \\ P(x_i \leq 1) = \frac{4}{8} \\ P(x_i \leq 2) = \frac{7}{8} \\ P(x_i \leq 3) = 1 \end{cases}$$



Eigenschaften:
$0 <= F(x) <= 1$
F(x) ist monoton nicht fallend
$F(x_1) <= F(x_2) ... <= F(x_n)$

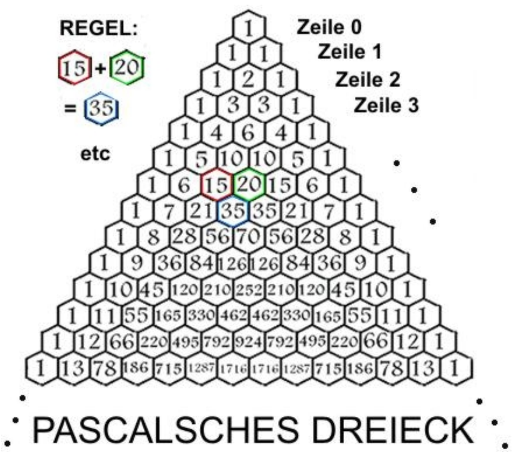


Pascalsches Dreieck

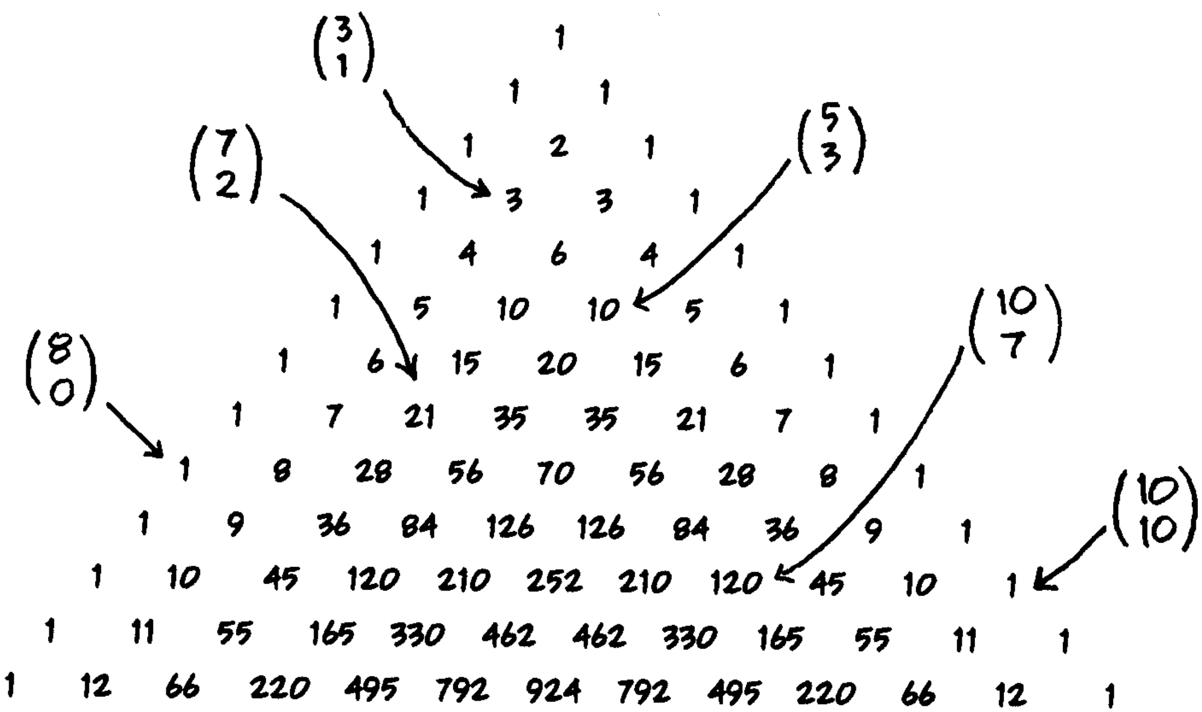


Quelle: http://www.wolfram-stanek.de/pascal_dreieck.jpg

Pascalsches Dreieck

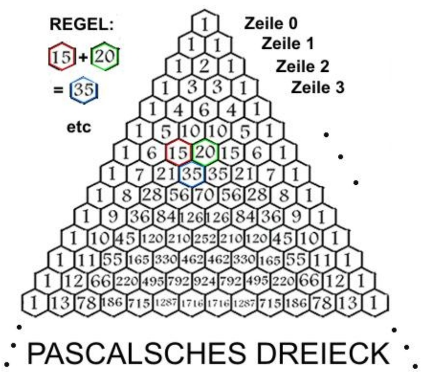


Quelle: http://www.wolfram-stanek.de/pascal_dreieck.jpg

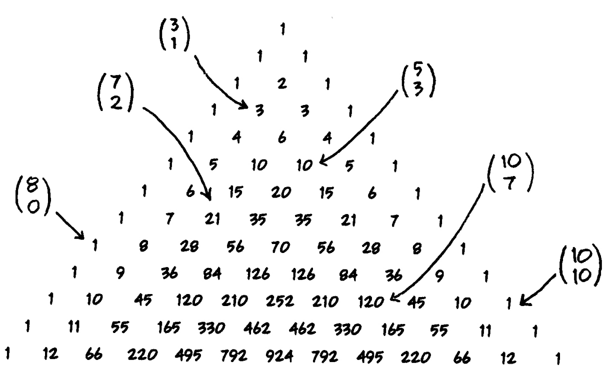


Quelle: C. Rinne

Pascalsches Dreieck



Quelle: http://www.wolfram-stanek.de/pascal_dreieck.jpg



Quelle: C. Rinne

Wahrscheinlichkeit aus dem Pascalschen Dreieck

Für einzelne Zelle:

Zahl der Zelle/2^(Zeile)

ZB: $\binom{7}{2} = 21$
Zeile 7: $n^k = 2^7 = 128$

21/128 = 0.1640625

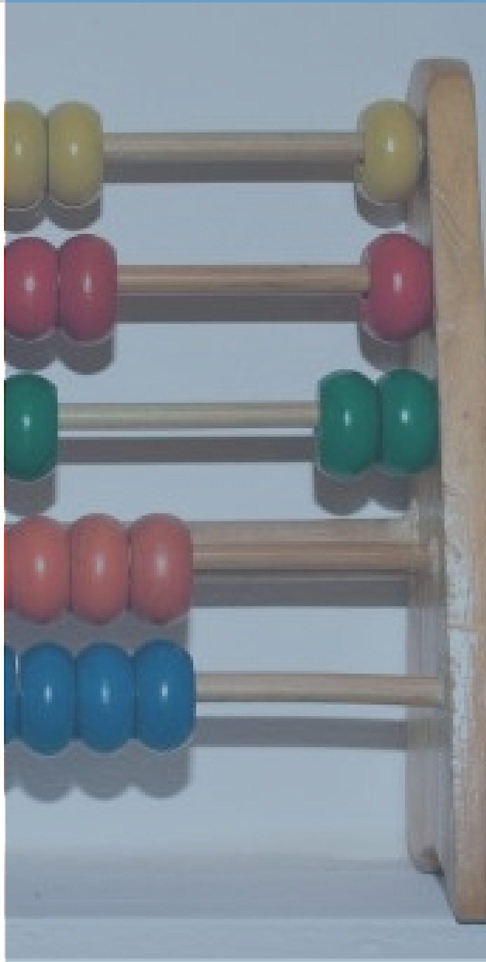
Für kumulative Zellen (unter oder gleich 2 von 7 Fällen)

Summe der Zahlen aller eingeschlossener Fälle (1, 7, 21) / 2^(Zeile)

(1+7+21)/2^7=29/128 = 0.2265625

Christian-Albrechts-Universität zu Kiel

Binominalverteilung [1]



Beschreibt Zufallexperimente “mit Zurücklegen” wie Münzwurf

Bedingungen:

Es gibt genau zwei mögliche, sich gegenseitig ausschließende Ergebnisse (z.B. Kopf oder Zahl)

Die Wahrscheinlichkeiten für beide Ausgänge sind konstant.

Die einzelnen Versuche sind voneinander unabhängig.

Eigenschaften:

$$\text{Allgemein: } B_{n;k;p} = \binom{n}{k} * p^k * (1-p)^{n-k} : \text{Formel für die Binominalverteilung}$$

$$n : Zahl\ der\ Beobachtungen$$
$$k : Zahl\ der\ günstigen\ Ausgänge$$
$$p : Wahrscheinlichkeit\ \ für\ einen\ günstigen\ Ausgang$$

$$Erwartungswert\,('\,Mittelwert\,',\ Wahrscheinlichster\ Wert) : E(x) = \mu = n * p$$
$$Varianz\,(Streuung) : var(x) = \sigma^2 = n * p * q = n * p * (1-p)$$
$$Standardabweichung : sd(x) = \sqrt{var(x)}$$

Binominalverteilung [1]

Beschreibt Zufallexperimente “mit Zurücklegen” wie Münzwurf

Bedingungen:

Es gibt genau zwei mögliche, sich gegenseitig ausschließende Ergebnisse (z.B. “männlich” oder “weiblich”)

Die Wahrscheinlichkeiten für beide Ausgänge sind konstant. Insgesamt wissen wir aus der Kultur, das Männerbestattungen unterrepräsentiert sind (70% Frauengräber, 30% Männergräber)

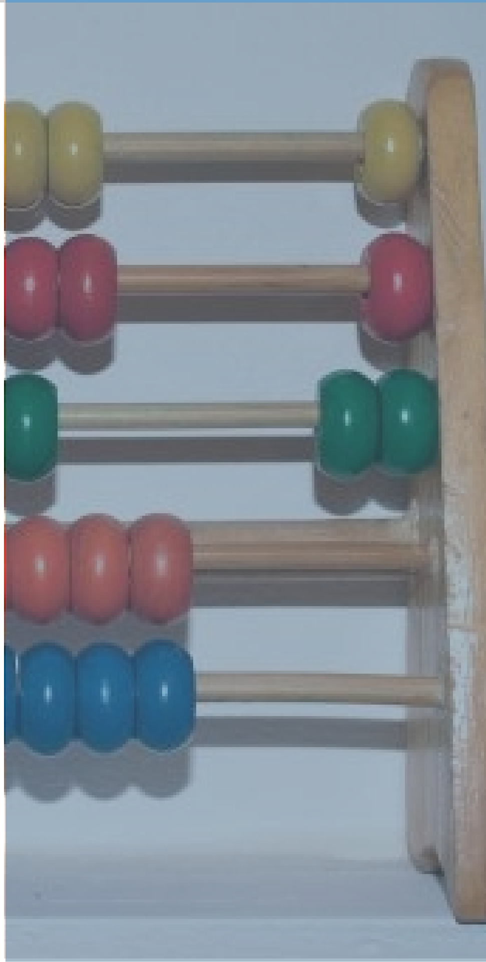
Die einzelnen Versuche sind voneinander unabhängig.

Frage: Wie wahrscheinlich ist genau ein Satz von Ergebnissen (z.B. 2x Männer, 5x Frauen) (P(W)=0.7, P(M)=1-P(W)=0,3

$$\frac{7!}{5!-(7-5)!}=21=\binom{7}{5}$$

Bei 2x Männer und 5x Frauen (Gesamt 7) gibt es 21 Möglichkeiten, diese anzuordnen (Kombination ohne Zurücklegen, Kombinatorik)

Da sich die Ereignisse gegenseitig ausschließen, jede der Kombination für uns aber ein „günstiges Ergebnis“ ist, können wir die Wahrscheinlichkeiten addieren.



Binominalverteilung [2]

Da sich die Ereignisse gegenseitig ausschließen, jede der Kombination für uns aber ein „günstiges Ergebnis“ ist, können wir die Wahrscheinlichkeiten addieren.

$$P(WWMMMMM) = P(W)*P(W)*P(M)*P(M)*P(M)*P(M)*P(M)$$

$$P(WWMMMMM) = 0.7*0.7*0.3*0.3*0.3*0.3*0.3 = 0.7^2(Frau)*(1-0.7)^{7-2}(Mann),$$

da alle anderen Ausgänge genauso wahrscheinlich sind:

$$P(WWMMMMM) = 21*0.7^2*(1-0.7)^{7-2} = 0{,}0250047$$

Allgemein: $B_{n;k;p} = \binom{n}{k} * p^k * (1-p)^{n-k}$: Formel für die Binominalverteilung

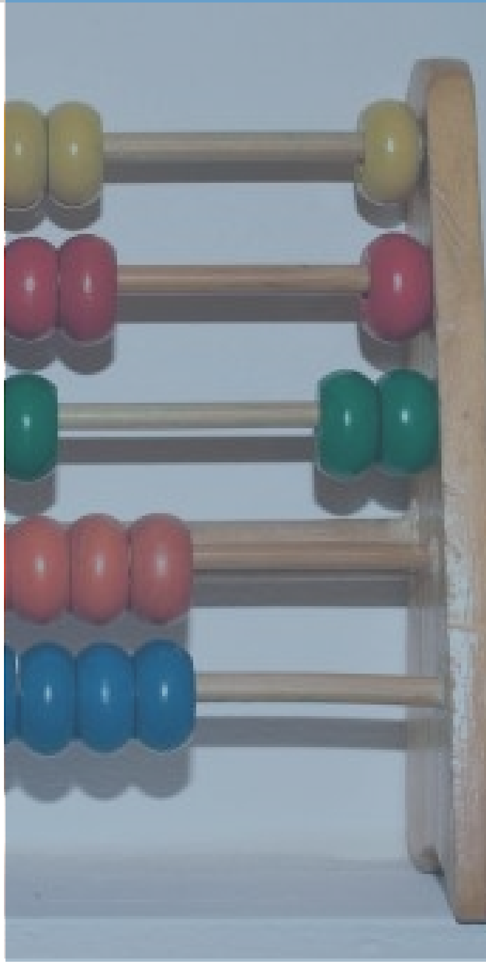
Erwartungswert: $E(x) = n*p,\ hier\ E(x) = 7*0{,}7 = 4{,}9$

Varianz: $Var(x) = n*p*(1-p),\ hier\ Var(x) = 7*0{,}7*(1-0{,}7) = 1{,}47,$

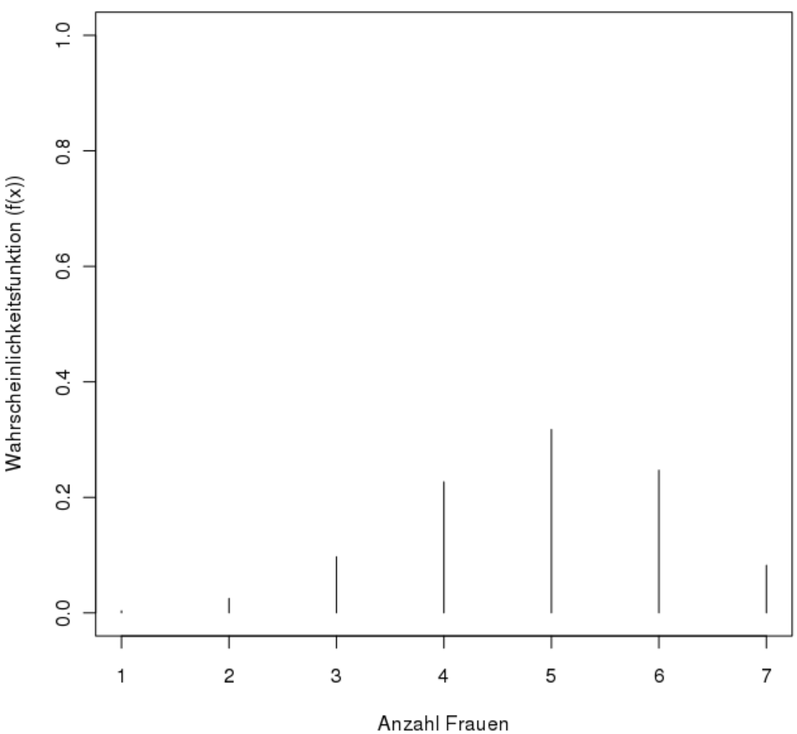
$$daher\ sd(x) = \sqrt{Var(x)} = 1{,}212435565$$

In R:

```
> dbinom(x=2, size=7, prob=0.7)
[1] 0.0250047
```



Binominalverteilung [2]



n, jede der Kombination
n wir die

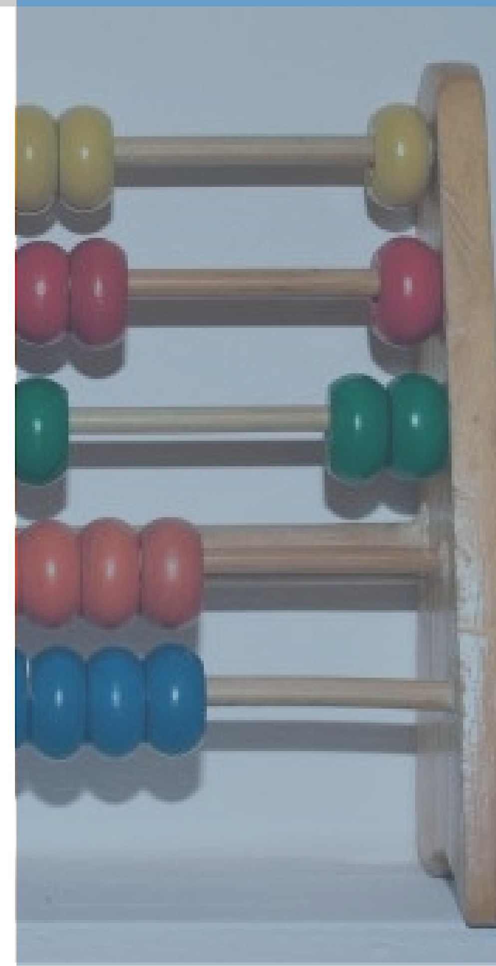
$P(M)*P(M)*P(M)$
$(Frau)*(1-0.7)^{7-2}(Mann),$

$=0{,}0250047$

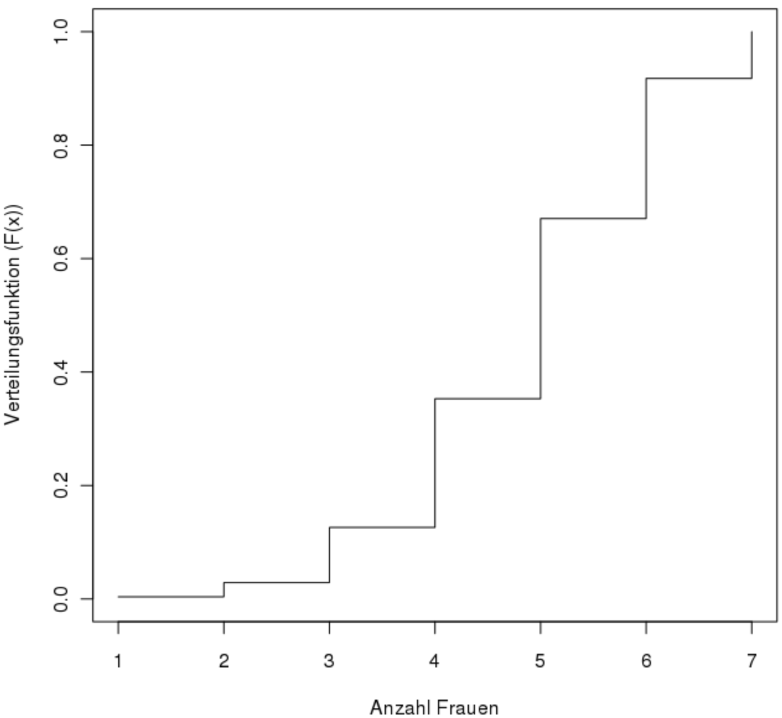
ıominalverteilung

$0{,}7)=1{,}47,$
5565

```
plot(1:7,dbinom(1:7,7,0.7),type="h",xlab="Anzahl
Frauen",ylab="Wahrscheinlichkeitsfunktion (f(x))",ylim=c(0,1))
```



Binominalverteilung [2]



ßen, jede der Kombination
hen wir die

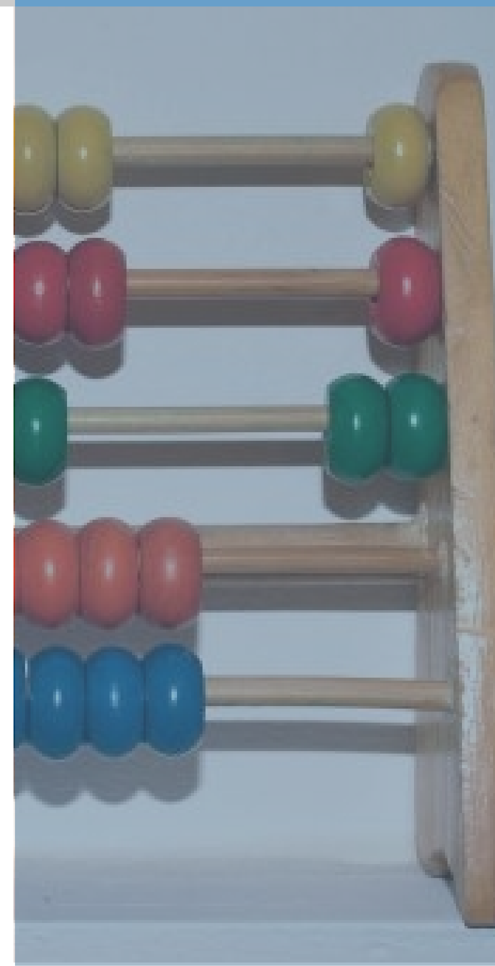
$)*P(M)*P(M)*P(M)$
$.7^2(Frau)*(1-0.7)^{7-2}(Mann),$

$^{7-2}=0{,}0250047$

Binominalverteilung

$-0{,}7)=1{,}47,$
435565

```
plot(1:7,pbinom(1:7,7,0.7),type="s",xlab="Anzahl
Frauen",ylab="Verteilungsfunktion (F(x))",ylim=c(0,1))
```



Binominalverteilung [2]

Da sich die Ereignisse gegenseitig ausschließen, jede der Kombination für uns aber ein „günstiges Ergebnis“ ist, können wir die Wahrscheinlichkeiten addieren.

$$P(WWMMMMM) = P(W) * P(W) * P(M) * P(M) * P(M) * P(M) * P(M)$$

$$P(WWMMMMM) = 0.7 * 0.7 * 0.3 * 0.3 * 0.3 * 0.3 * 0.3 = 0.7^2 (Frau) * (1 - 0.7)^{7-2} (Mann),$$

da alle anderen Ausgänge genauso wahrscheinlich sind:

$$P(WWMMMMM) = 21 * 0.7^2 * (1 - 0.7)^{7-2} = 0{,}0250047$$

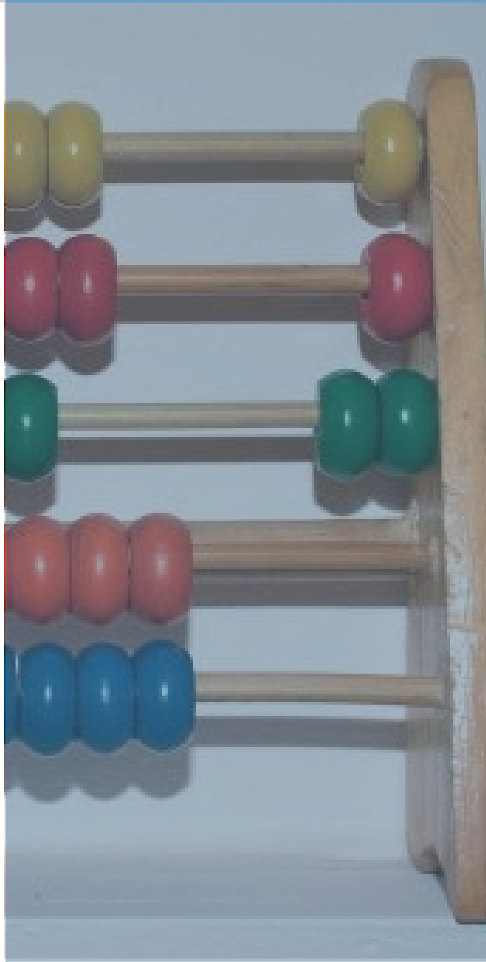
$$P(WMMMMMM) = 7 * 0.7^1 * (1 - 0.7)^{7-1} = 0.0035721$$

$$P(MMMMMMM) = 1 * 0.7^0 * (1 - 0.7)^{7-0} = 0.0002187$$

$$kumulative\ Wahrscheinlichkeit:$$

$$P(WWMMMMM) + P(WMMMMMM) + P(MMMMMMM) = 0.0287955$$

Eher unwahrscheinliche Verteilung für diese Kultur! Auf einem Niveau von 0.05% nicht zufällig.

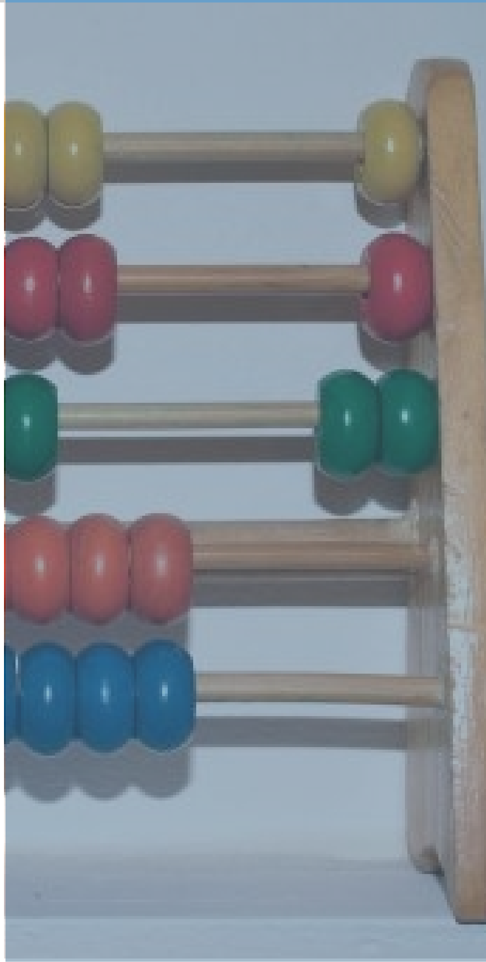


Poissonverteilung [1]

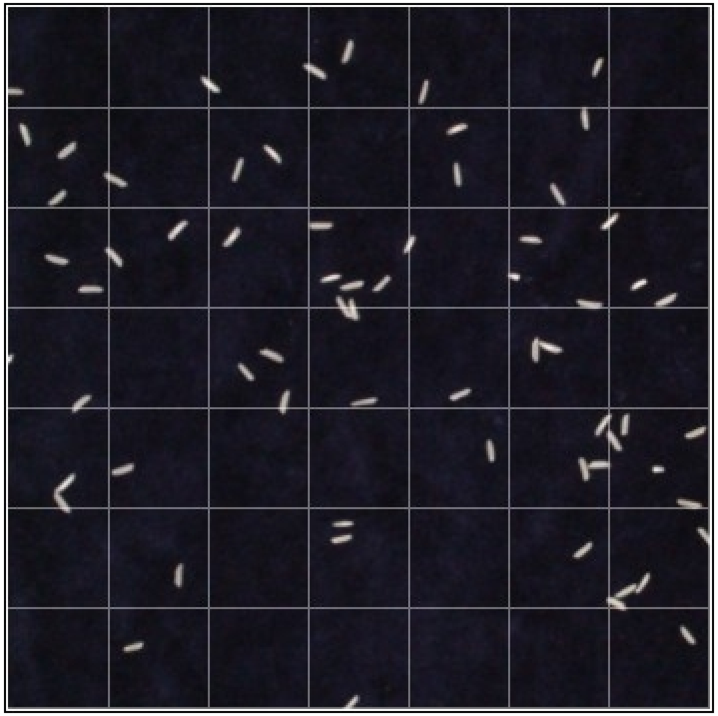
Beschreibt Zufallsexperimente mit seltenen Ereignissen und einer großen Grundgesamtheit ($n \rightarrow \infty$, $p \rightarrow 0$)
Wir betrachten dabei einen bestimmten Ausschnitt (Raum, Zeit)

Frage: Wieviele Ereignisse treten in einem gegebenen Ausschnitt der Realität auf, gegeben eine konstante Eintrittswahrscheinlichkeit?

Klassisches Beispiel: Reiskörner auf einem Schachbrett.



Poissonverteilung



Quelle: wikipedia

Reiskorn-Beispiel (nach Wikipedia)

Wir haben 7x7 Felder (ok, es ist nicht wirklich ein Schachbrett), n=49 Quadrate
Verstreut werden N=66 Reiskörner

Durchschnittliche Anzahl je Feld => 66/49 = 1.34693
Dies ist der Erwartungswert (λ).

Anzahl Körner je Feld	Anzahl der entsprechenden Felder
0	16
1	14
2	10
3	6
4	1
5	2

Christian-Albrechts-Universität zu Kiel

Poissonverteilung [1]

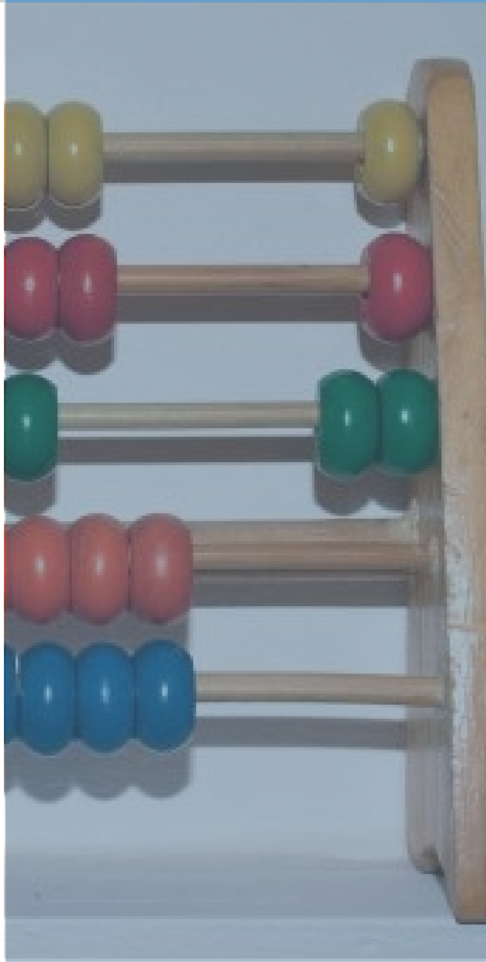
Beschreibt Zufallsexperimente mit seltenen Ereignissen und einer großen Grundgesamtheit (n → ∞, p → 0)
Wir betrachten dabei einen bestimmten Ausschnitt (Raum, Zeit)

Frage: Wieviele Ereignisse treten in einem gegebenen Ausschnitt der Realität auf, gegeben eine konstante Eintrittswahrscheinlichkeit?

Klassisches Beispiel: Reiskörner auf einem Schachbrett.

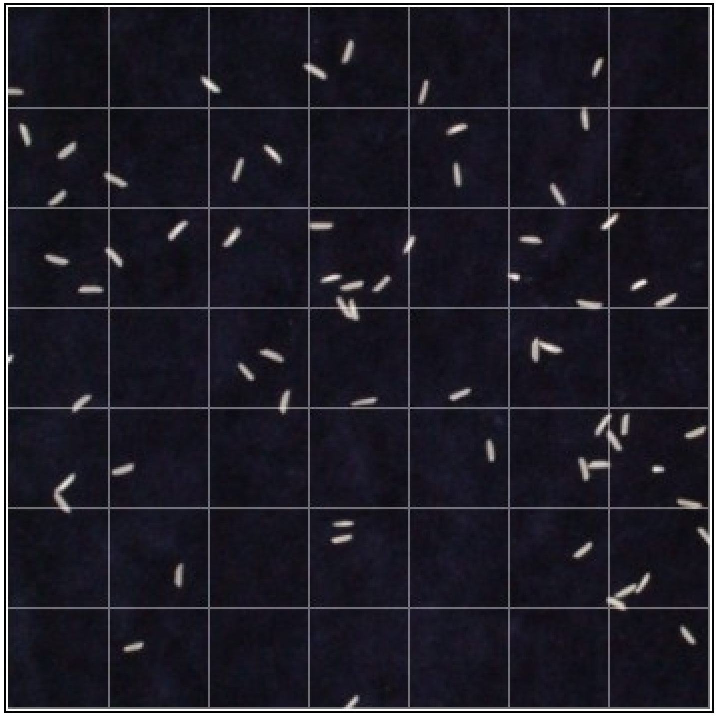
$$P_\lambda(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

$$e: Eulersche\,Zahl\,, e=\sum_{i=o}^{\infty}\frac{1}{k!}=\frac{1}{0}!+\frac{1}{1}!\frac{1}{2}!\ldots=2{,}7182818284$$



Poissonverteilung

$$P_\lambda(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$



Quelle: wikipedia

Reiskorn-Beispiel (nach Wikipedia)

Wir haben 7x7 Felder (ok, es ist nicht wirklich ein Schachbrett), n=49 Quadrate
Verstreut werden N=66 Reiskörner

Durchschnittliche Anzahl je Feld => 66/49 = 1.34693
Dies ist der Erwartungswert (λ).

$$P_\lambda(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

$$P_{66/49}(X=0)=\frac{(66/49)^0}{0!}e^{-66/49}=0.260035069$$

$$26\,Prozent\ oder\ p*n=12.74\ Felder\ sollten\ ohne\ Reiskorn\ bleiben$$

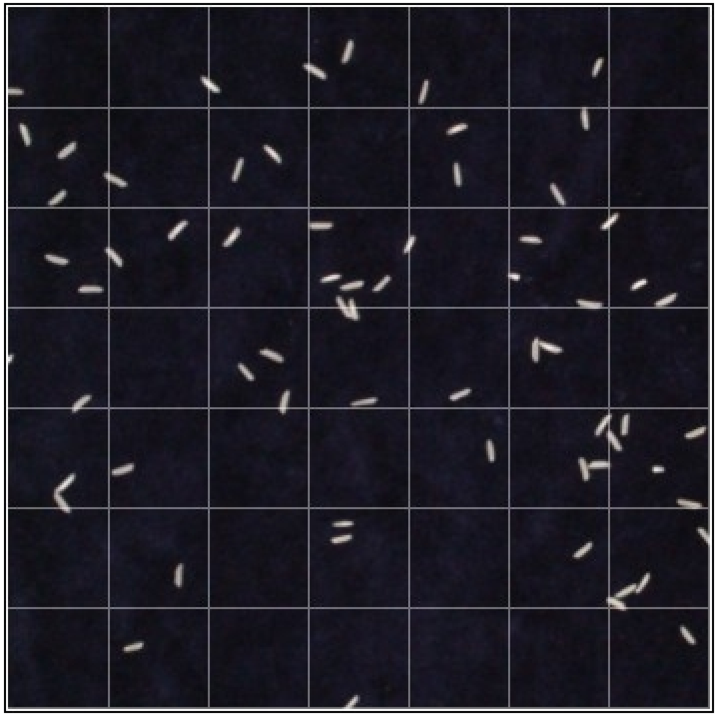
$$P_{66/49}(X=1)=\frac{(66/49)^1}{1!}e^{-66/49}=0.350251318$$

$$35\,Prozent\ oder\ p*n=17.15\ Felder\ sollten\ ein\ Reiskorn\ haben$$

$$P_{66/49}(X=2)=\frac{(66/49)^2}{2!}e^{-66/49}=0.23588354$$

$$23{,}6\,Prozent\ oder\ p*n=11.56\ Felder\ sollten\ zwei\ Reiskörner\ haben$$

Poissonverteilung



Quelle: wikipedia

Reiskorn-Beispiel (nach Wikipedia)

Wir haben 7x7 Felder (ok, es ist nicht wirklich ein Schachbrett), n=49 Quadrate
Verstreut werden N=66 Reiskörner

Durchschnittliche Anzahl je Feld => 66/49 = 1.34693
Dies ist der Erwartungswert (λ).

Anzahl Körner je Feld	Anzahl der entsprechenden Felder	Nach Poisson
0	16	12,74
1	14	17,16
2	10	11,55
3	6	5,19
4	1	1,75
5	2	0,47

Beziehung zur Binomialverteilung

$$\text{Allgemein: } B_{n;k;p} = \binom{n}{k} * p^k * (1-p)^{n-k} : \text{Formel für die Binominalverteilung}$$

$$n : Zahl\ der\ Beobachtungen$$
$$k : Zahl\ der\ günstigen\ Ausgänge$$
$$p : Wahrscheinlichkeit\ \ für\ einen\ günstigen\ Ausgang$$

Das gleiche in Binomial

n=49 Quadrate, k=0...5, p=λ/n= 66/49/49 = 0.027488367 (da wir den gleichen Erwartungswert wollen, und dieser bei Binomial in Beziehung zu p steht: E=n*p > p=E/n)

Die Poissonverteilung ist eine gute Näherung für die Binomialverteilung (je größer n, um so besser, Faustregel: n>=50, p<=0.05)

Anzahl Körner je Feld	Anzahl der entsprechenden Felder	Nach Poisson	Nach Binom
0	16	12,74	12.50
1	14	17,16	17.32
2	10	11,55	11.75
3	6	5,19	5.20
4	1	1,75	1.69
5	2	0,47	0.43

Poissonverteilung [2]

Beschreibt Zufallsexperimente mit seltenen Ereignissen und einer großen Grundgesamtheit ($n \rightarrow \infty$, $p \rightarrow 0$)
Wir betrachten dabei einen bestimmten Ausschnitt (Raum, Zeit)

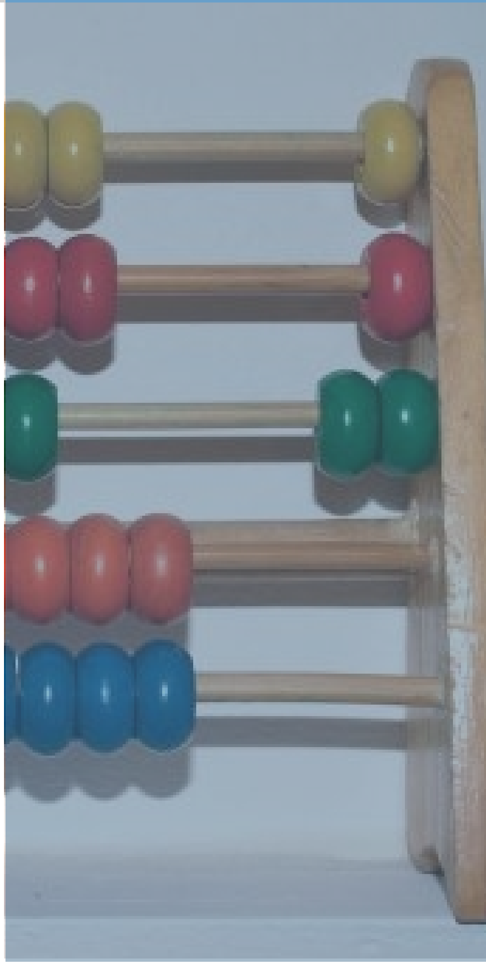
Beispiel nach Ihm, vereinfacht: Dechsel in Körpergräbern der Aldenhovener Platte

Theoretisch gibt es eine (unbestimmbar) sehr große Anzahl von Bestattungen, wir betrachten Ausschnitt

Hier:
n = 94
Dechsel: 38

Durchschnittliche Dechselzahl:
0,404255319

Anzahl Dechsel	Gräber mit der Anzahl Dechsel
0	65
1	23
2	4
3	1
4	1
5	0
6	0



Christian-Albrechts-Universität zu Kiel

Poissonverteilung [2]

Parameter: x, λ=Durchschnittliche Häufigkeit

$$P_{\lambda}, x = \frac{\lambda^x * e^{-\lambda}}{x!}$$

e = Eulersche Zahl (2,718281828459...)

für x=0 bei $\lambda=0{,}404255319$:

$$P_{0{,}404255319;0} = \frac{0{,}404255319^0 * e^{-0{,}404255319}}{0!} = 0{,}667473681$$

$$94 * 0{,}667473681 = 62{,}742526014$$

für x=1 bei $\lambda=0{,}404255319$:

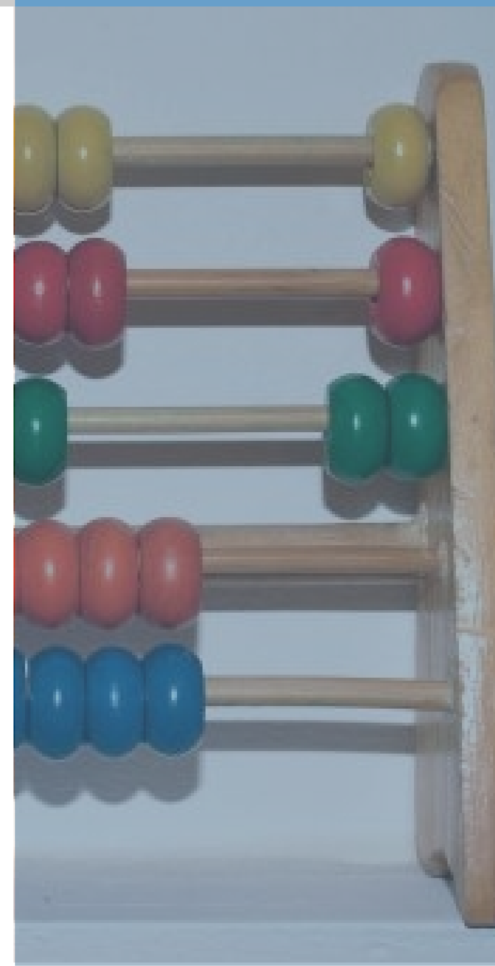
$$P_{0{,}404255319;1} = \frac{0{,}404255319^1 * e^{-0{,}404255319}}{1!} = 0{,}269829786$$

$$94 * 0{,}269829786 = 25{,}363999861 \; etc.$$

Erwartungswert: $E(x)=\lambda=0{,}404255319$

Streuung: $Var(x)=\lambda=0{,}404255319$

Anzahl Dechsel	Gräber mit der Anzahl Dechsel	Anzahl nach Poisson-verteilung
0	65	62.74253
1	23	25.364
2	4	5.126766
3	1	0.6908408
4	1	0.06981902
5	0	0.005644942
6	0	0.0003803330



Beispiele für poissonverteilte Ereignisse:
Wegwerfprozesse (räumliche Verteilung); Radioaktiver Zerfall (zeitliche Verteilung)
In R:

```
> dpois(0,0.404255319)
[1] 0.6674737
```

Grundlegende statistische Verfahren für archäologische Datenanalyse in R

Christian-Albrechts-Universität zu Kiel

Poissonverteilung [2]

Paramete

e = Eulersch

für x=0 bei λ

$P_{0,404255319;0}=$

94

für x=1 bei λ

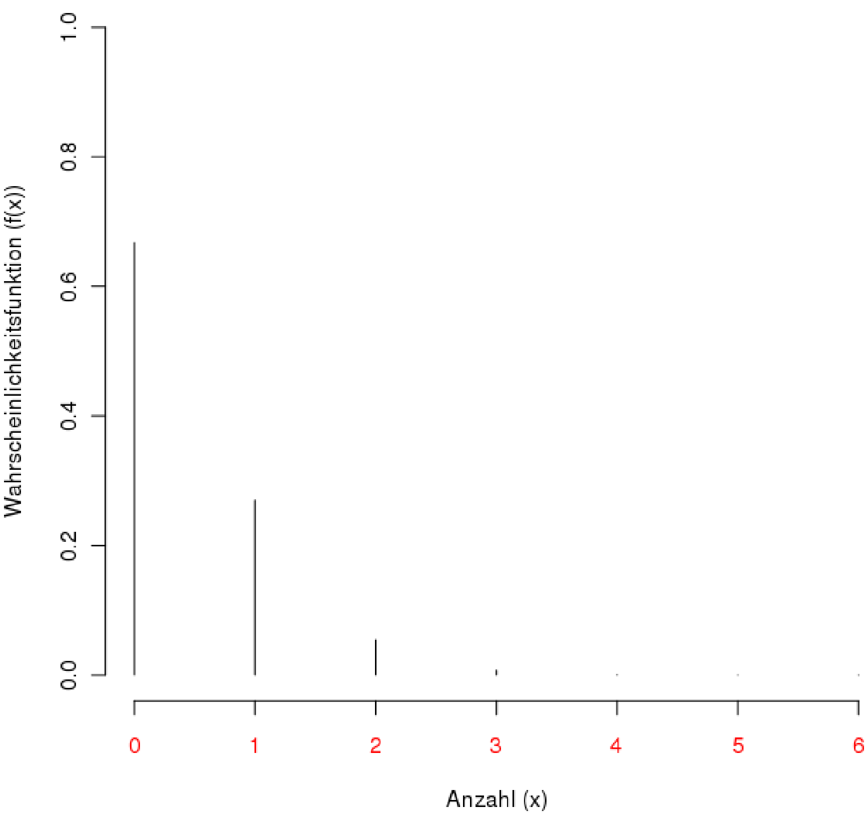
$P_{0,404255319;1}=$

94 ∗

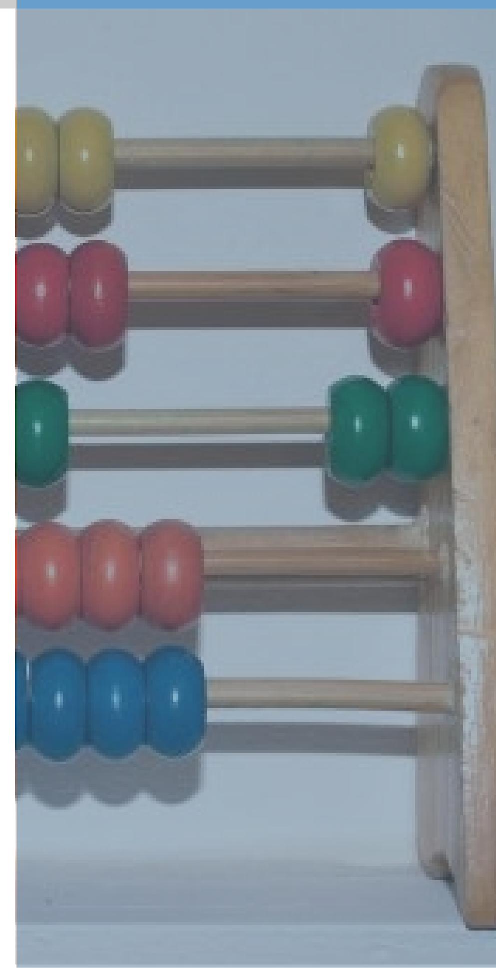
Erwartungsw

Streuung: *Va*

Beispiele
Wegwerfp
Radioaktiv



nzahl nach oisson- erteilung
2.74253
5.364
.126766
.6908408
.06981902
.005644942
.0003803330



Poissonverteilung [2]

Parame

e = Eulers

für x=0 be

$P_{0,404255319;}$

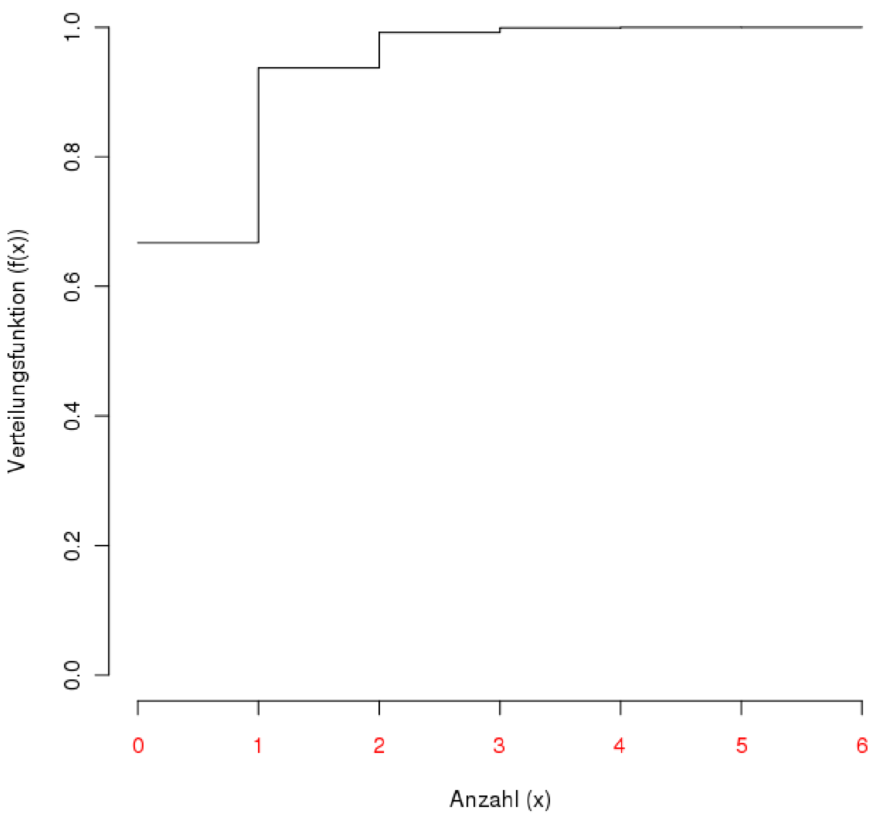
für x=1 be

$P_{0,404255319;}$

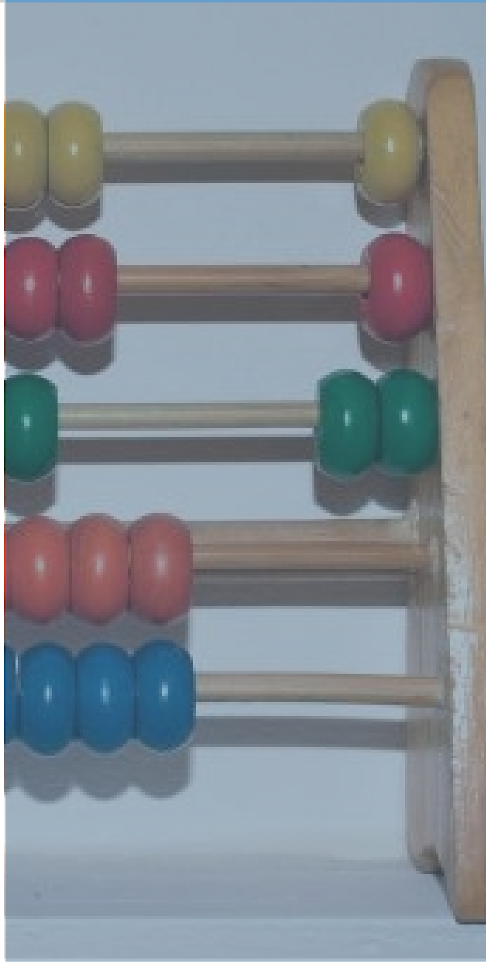
94

Erwartung
Streuung:

Beispiel
Wegwer
Radioak



Anzahl nach Poisson-verteilung
62.74253
25.364
5.126766
0.6908408
0.06981902
0.005644942
0.0003803330



Gleichverteilung (diskrete Variante)

Wenn alle Ergebnisse gleich verteilt sind
Bsp. Würfelwurf mit einem Würfel

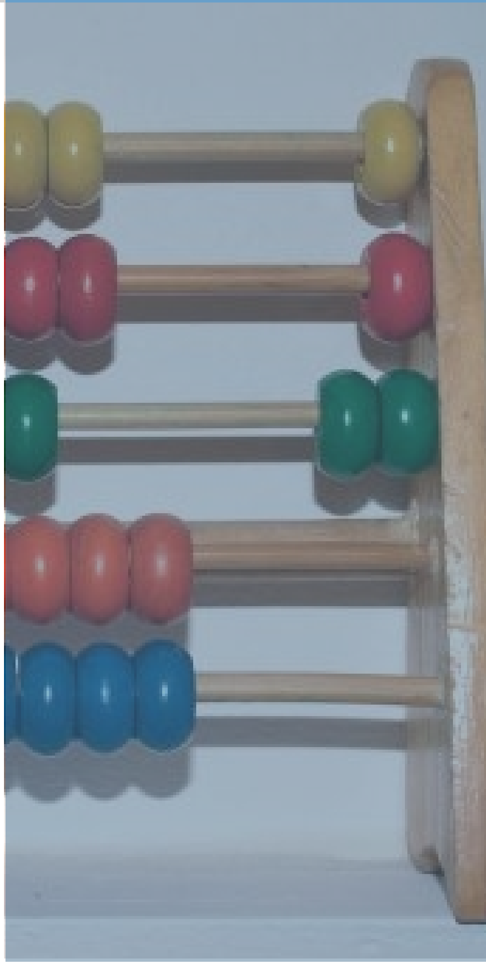
gg. ist ein Intervall [1..6], n=6, in dem die Ergebnisse gleich verteilt wahrscheinlich sind

$$f_G(x) = \begin{cases} \frac{1}{n} & \textit{für } x = x_i (1 \le i \le n) \\ 0 & \textit{sonst} \end{cases}$$

$$F_G(x) = \begin{cases} 0 & \textit{für } x < x_1 \\ \frac{i}{n} & \textit{für } 1 \le i < n \\ 1 & \textit{für } i \ge n \end{cases}$$

$$f_G(x) = \begin{cases} \frac{1}{6} & \textit{für } x = x_i (1 \le i \le 6) \\ 0 & \textit{sonst} \end{cases}$$

$$F_G(x) = \begin{cases} 0 & \textit{für } x < 1 \\ \frac{i}{6} & \textit{für } 1 \le i < 6 \\ 1 & \textit{für } i \ge 6 \end{cases}$$



Gleichverteilung (diskrete Variante)

Wenn a

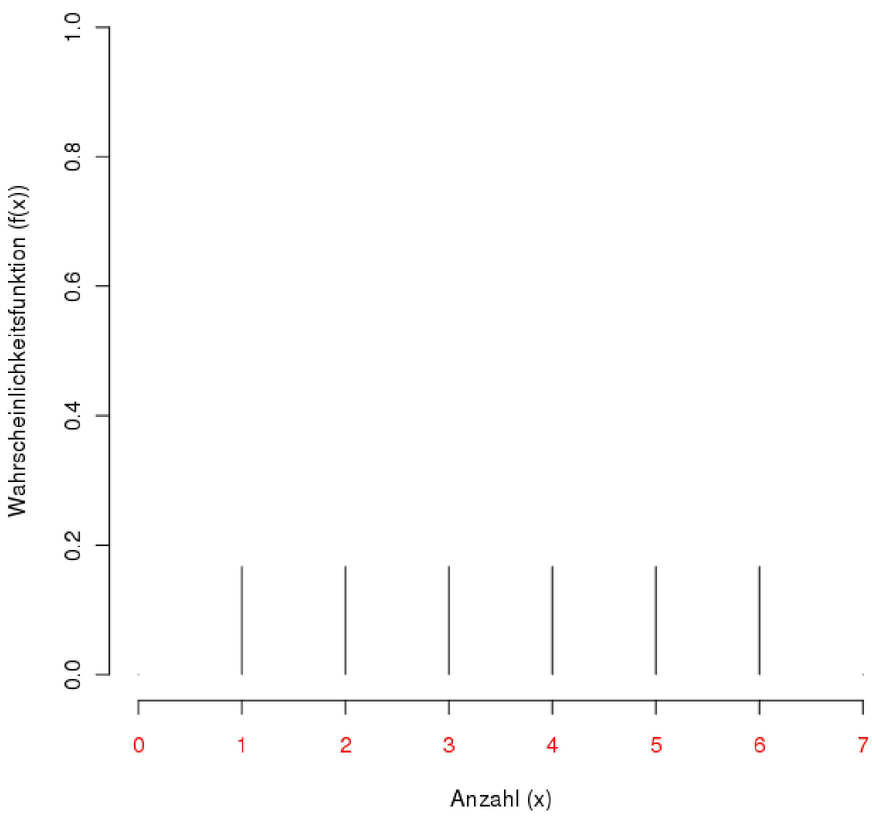
Bsp. W

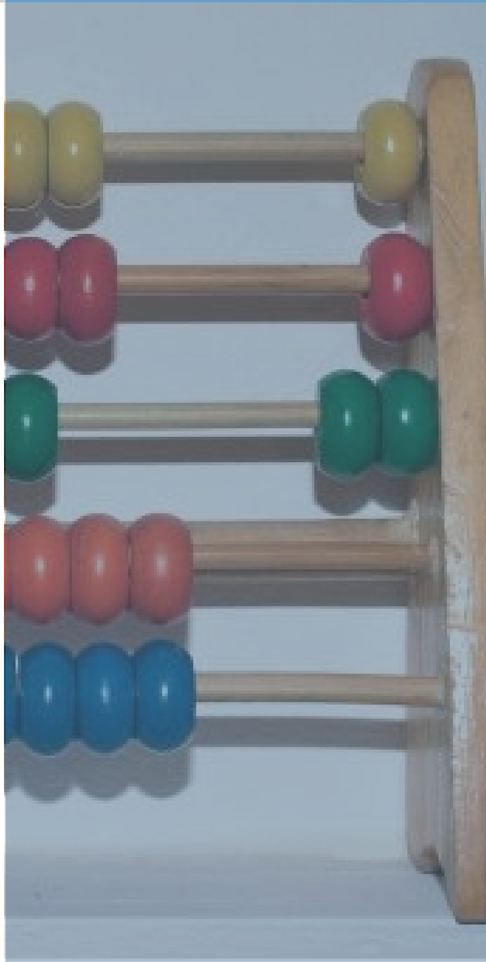
gg. ist e ... 'erteilt

wahrsc

$f_G(x) =$ ... 6)

$F_G($





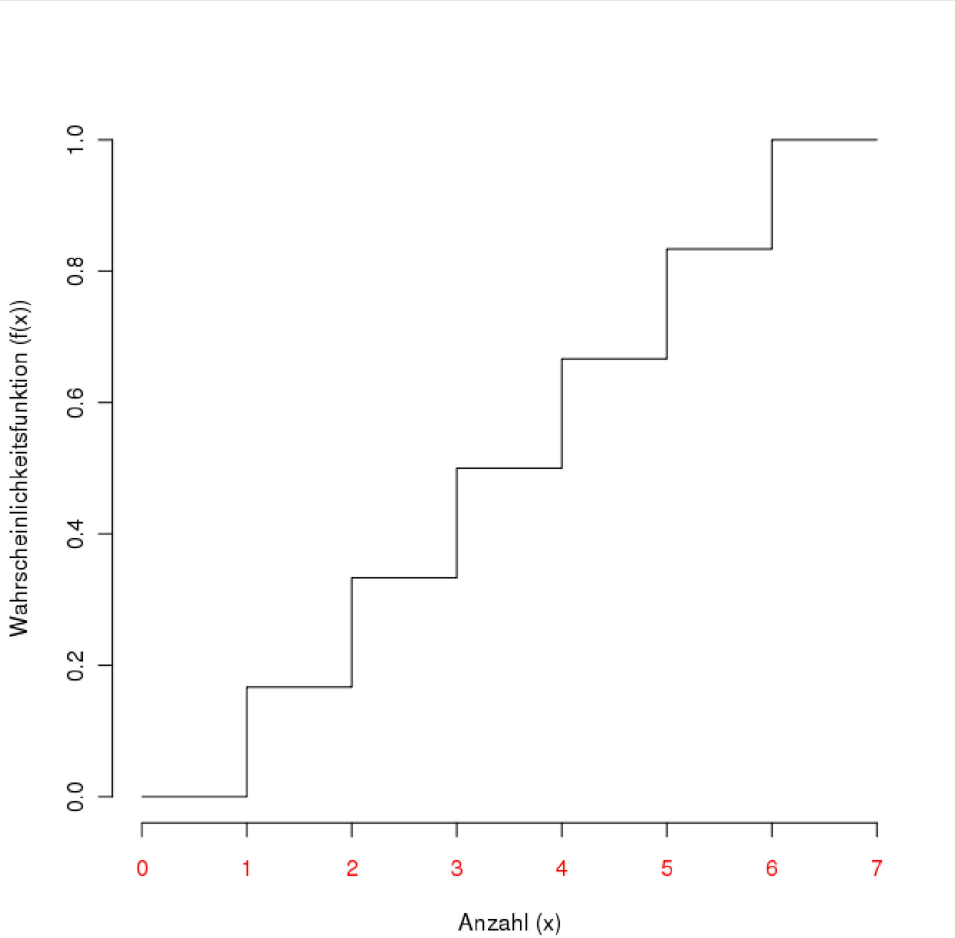
Gleichverteilung (diskrete Variante)

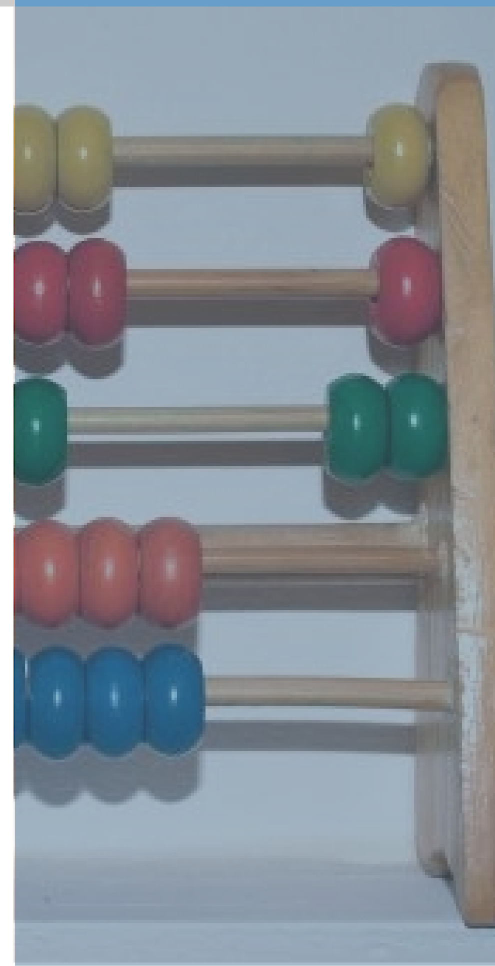
Wenn
Bsp. W

gg. ist verteilt
wahrs

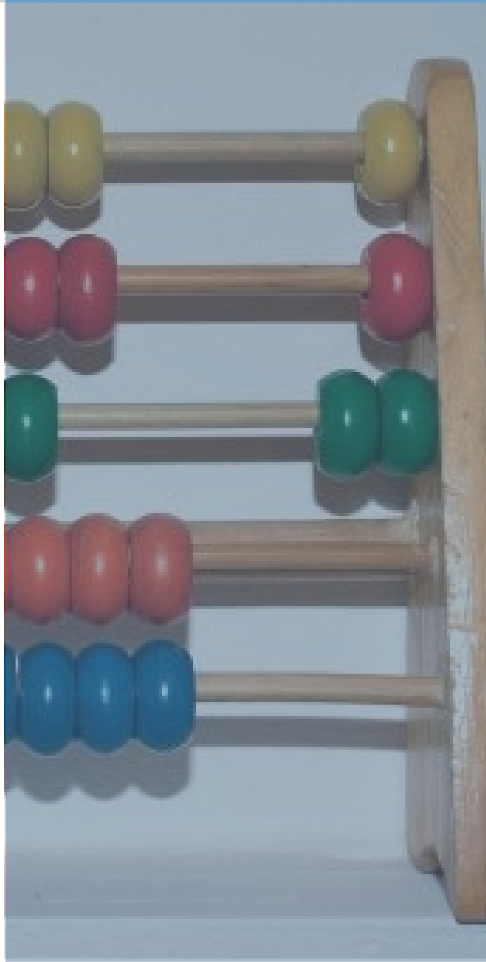
$f_G(x)$ $\leq 6)$

F_G





Gleichverteilung (stetige Variante)



Wenn alle Ergebnisse gleich wahrscheinlich sind, aber eine stetige Variable vorliegt

$$f_G(x) = \begin{cases} \dfrac{1}{max - min} & \text{für } min \leq x \leq max \\ 0 & sonst \end{cases}$$

$$F_G(x) = \begin{cases} 0 & \text{für } x < min \\ \dfrac{x - min}{max - min} & \text{für } min \leq x \leq max \\ 1 & \text{für } x > max \end{cases}$$

Bsp. Wartezeit auf den Bus
Busse fahren alle 20 min, wir wissen bloß nicht, wann. Innerhalb der nächsten 20 min besteht zu jedem Zeitpunkt die gleiche Wahrscheinlichkeit, dass der Bus kommt. Wie hoch ist die Wahrscheinlichkeit, das er in genau 5 min kommt?

Gleichverteilung (stetige Variante)

Wenn alle Ergebnisse gleich wahrscheinlich sind, aber eine stetige Variable vorliegt

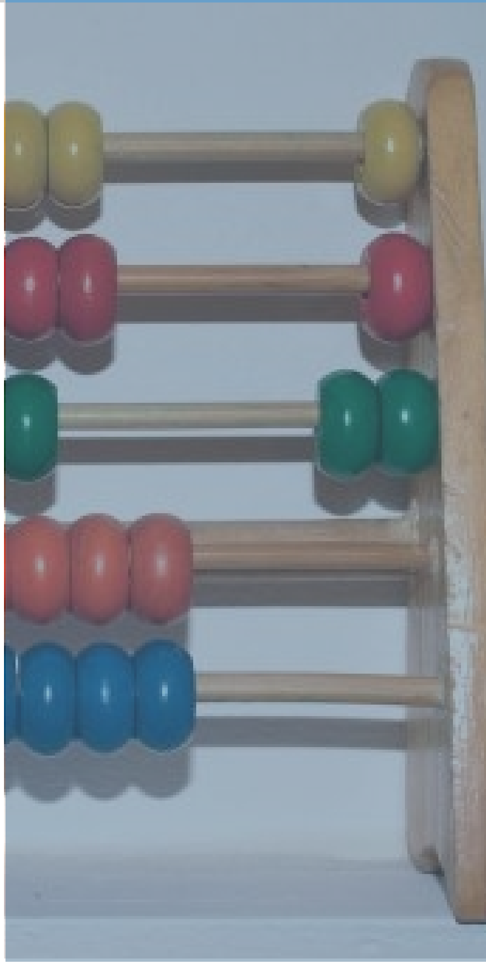
$$f_G(x)=\begin{cases}\dfrac{1}{max-min} & \text{für } min \leq x \leq max \\ 0 & sonst\end{cases}$$

$$F_G(x)=\begin{cases}0 & \text{für } x < min \\ \dfrac{x-min}{max-min} & \text{für } min \leq x \leq max \\ 1 & \text{für } x > max\end{cases}$$

Bsp. Wartezeit auf den Bus
Busse fahren alle 20 min, wir wissen bloß nicht, wann. Innerhalb der nächsten 20 min besteht zu jedem Zeitpunkt die gleiche Wahrscheinlichkeit, dass der Bus kommt. Wie hoch ist die Wahrscheinlichkeit, das er in genau 5 min kommt?

$1/\infty$

Wir haben unendlich viele mögliche Fälle, davon ist einer (in genau 5 min) unser günstiger Fall



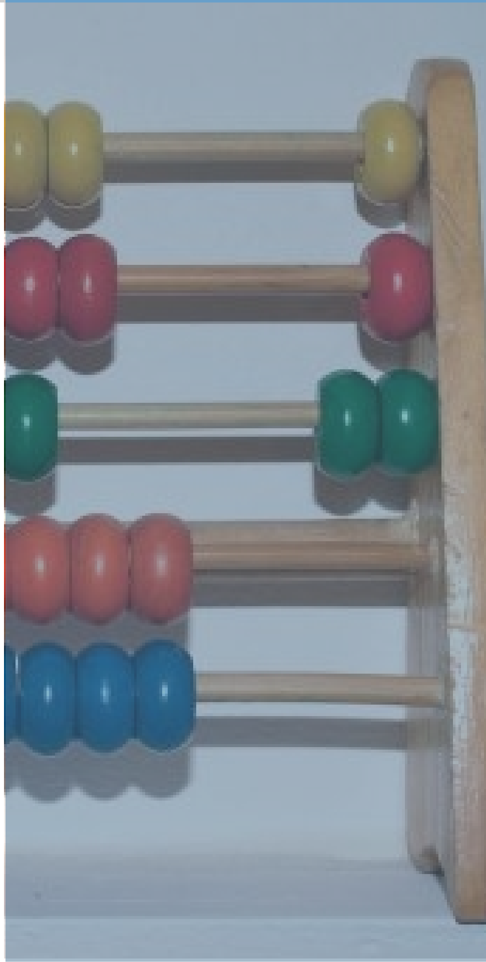
Gleichverteilung (stetige Variante)

Wenn alle Ergebnisse gleich wahrscheinlich sind, aber eine stetige Variable vorliegt

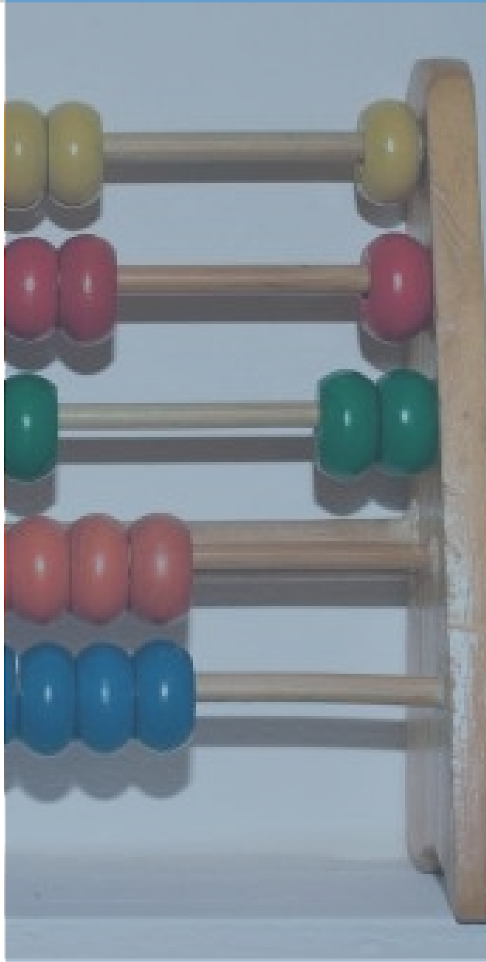
$$f_G(x)=\begin{cases} \dfrac{1}{max-min} & \text{für } min \leq x \leq max \\ 0 & sonst \end{cases}$$

$$F_G(x)=\begin{cases} 0 & \text{für } x < min \\ \dfrac{x-min}{max-min} & \text{für } min \leq x \leq max \\ 1 & \text{für } x > max \end{cases}$$

Bsp. Wartezeit auf den Bus
Busse fahren alle 20 min, wir wissen bloß nicht, wann. Innerhalb der nächsten 20 min besteht zu jedem Zeitpunkt die gleiche Wahrscheinlichkeit, dass der Bus kommt. Wie hoch ist die Wahrscheinlichkeit, das er in 3 bis 5 min kommt?



Gleichverteilung (stetige Variante)



Wenn alle Ergebnisse gleich wahrscheinlich sind, aber eine stetige Variable vorliegt

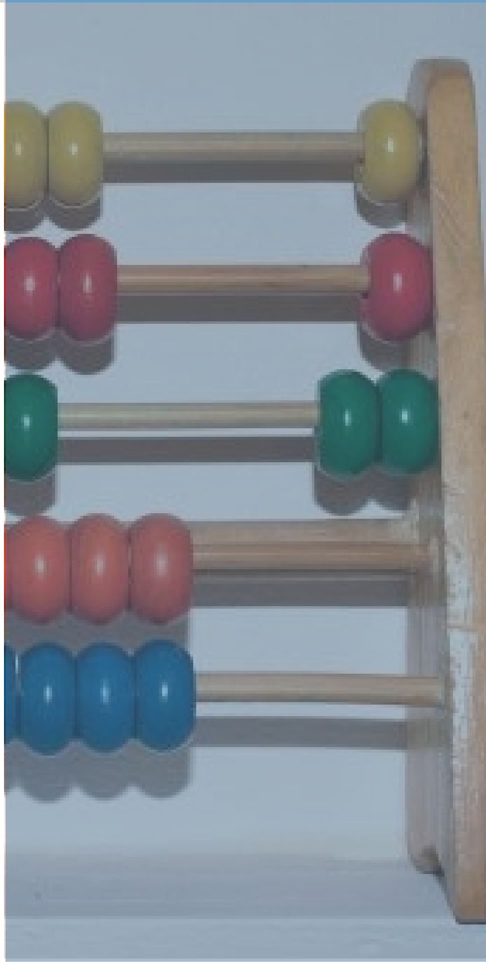
$$f_G(x) = \begin{cases} \frac{1}{max - min} & \text{für } min \leq x \leq max \\ 0 & sonst \end{cases}$$

$$F_G(x) = \begin{cases} 0 & \text{für } x < min \\ \frac{x - min}{max - min} & \text{für } min \leq x \leq max \\ 1 & \text{für } x > max \end{cases}$$

Bsp. Wartezeit auf den Bus
Busse fahren alle 20 min, wir wissen bloß nicht, wann. Innerhalb der nächsten 20 min besteht zu jedem Zeitpunkt die gleiche Wahrscheinlichkeit, dass der Bus kommt. Wie hoch ist die Wahrscheinlichkeit, das er in 3 bis 5 min kommt?

Achtung: Da wir keine diskreten Ereignisse haben, können wir nicht die Wahrscheinlichkeitsfunktion benutzen (es müsste das Integral über den Zeitraum gebildet werden. Die Wahrscheinlichkeit, das der Bus genau in 5 min. kommt, geht gegen 0). Daher Berechnung mit der Verteilungsfunktion.

Gleichverteilung (stetige Variante)



Bsp. Wartezeit auf den Bus, alle 20 min, in 3-5 min (= 2 min Intervall)?

Idee: Wahrscheinlichkeit für „in den nächsten 5 min“ - Wahrscheinlichkeit „in den nächsten 3 min“ = Wahrscheinlichkeit für „in 3-5 min“

$$F_G(3)=\left\{\frac{3-0}{20-0}=0{,}15 \quad F_G(5)=\left\{\frac{5-0}{20-0}=0{,}25 \quad F_G(5)-F_G(3)=0{,}10 \quad p(3\leq x\leq 5)=0.1=10\%\right.\right.$$

In R:

```
> punif(5,min=0,max=20)-punif(3,min=0,max=20)
[1] 0.1
```

Grundlegende statistische Verfahren für archäologische Datenanalyse in R

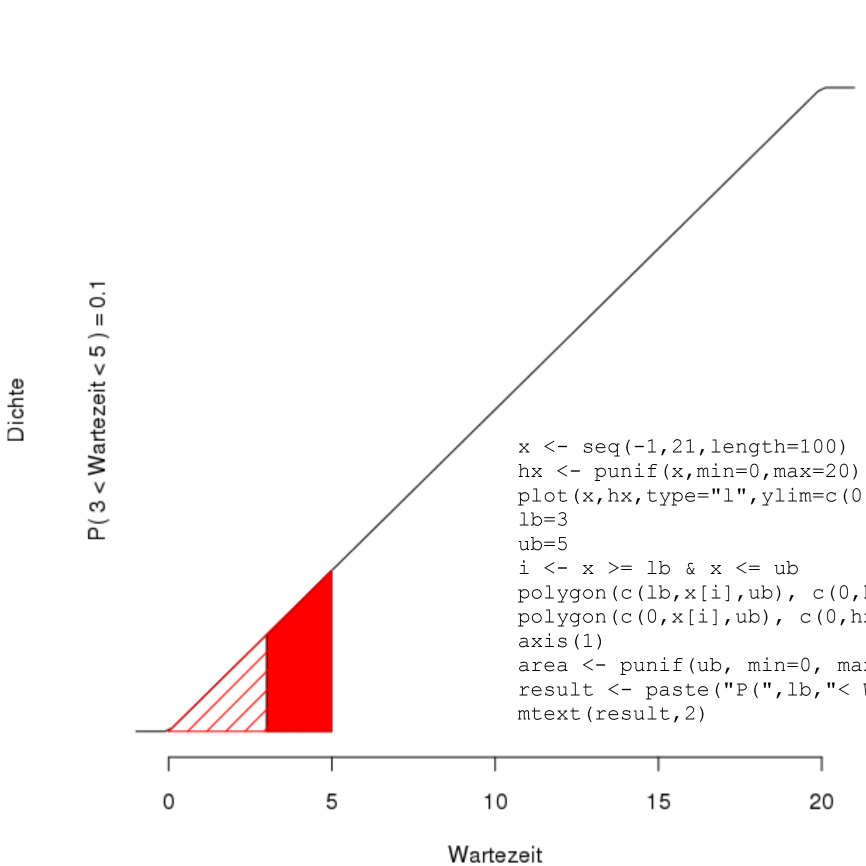
Christian-Albrechts-Universität zu Kiel

Gleichverteilung (stetige Variante)

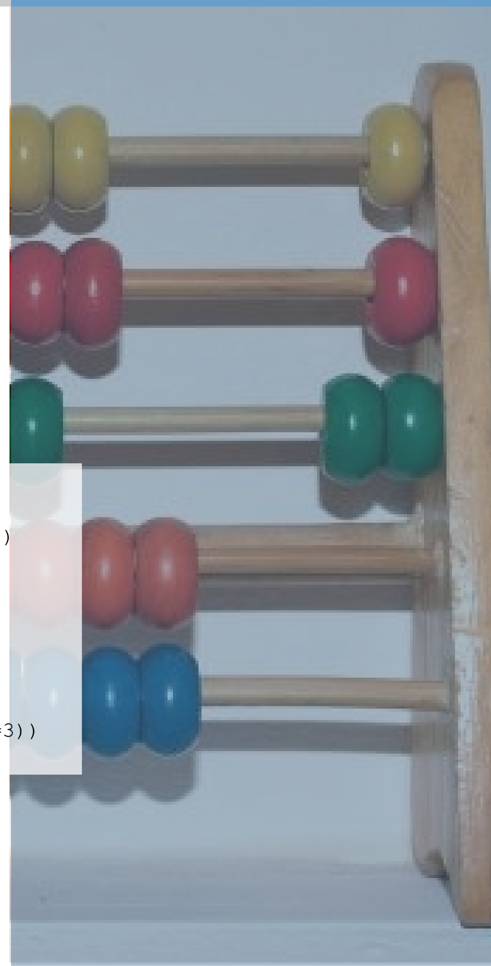
n?

in“ - Wahrscheinlichkeit in 3-5 min“

$$p(3 \leq x \leq 5) = 0.1 = 10\%$$

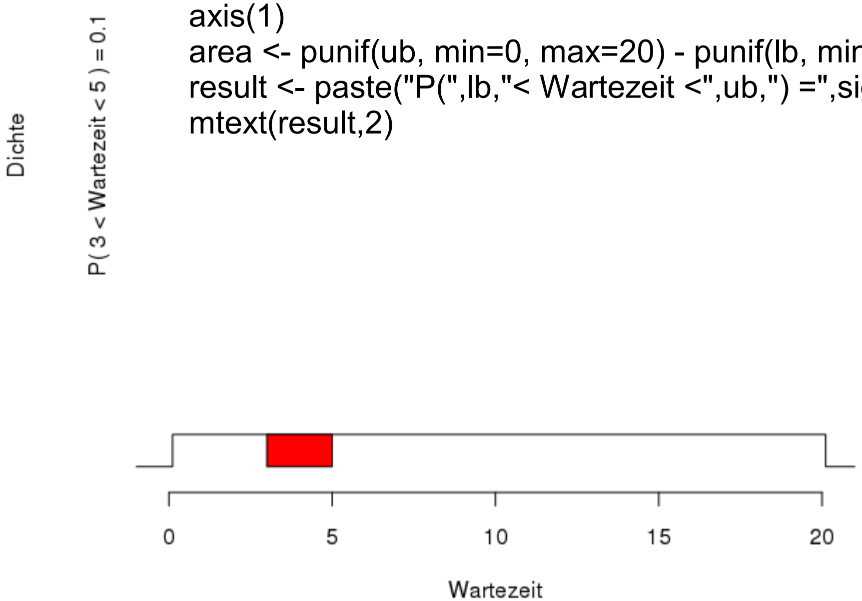


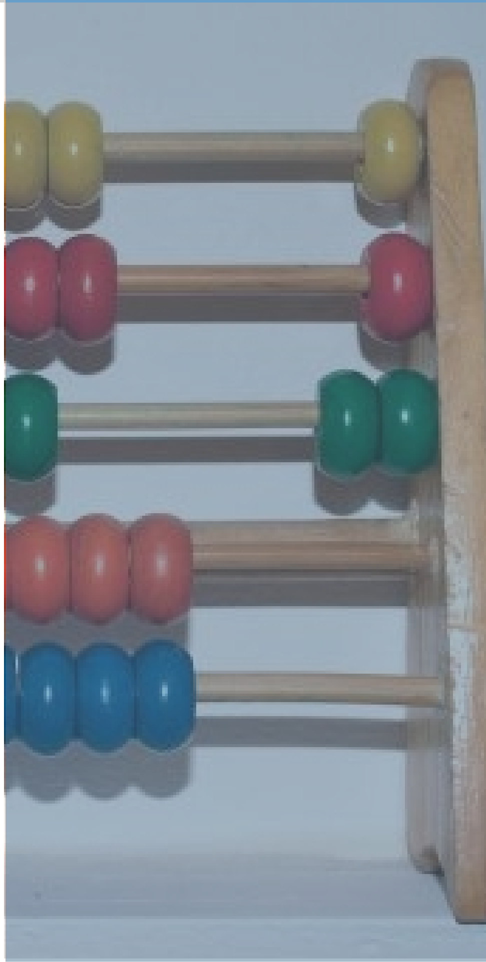

```
x <- seq(-1,21,length=100)
hx <- punif(x,min=0,max=20)
plot(x,hx,type="l",ylim=c(0,1),axes=F,xlab="Wartezeit", ylab="Dichte")
lb=3
ub=5
i <- x >= lb & x <= ub
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
polygon(c(0,x[i],ub), c(0,hx[i],0), density=10,col="red")
axis(1)
area <- punif(ub, min=0, max=20) - punif(lb, min=0, max=20)
result <- paste("P(",lb,"< Wartezeit <",ub,") =",signif(area, digits=3))
mtext(result,2)
```



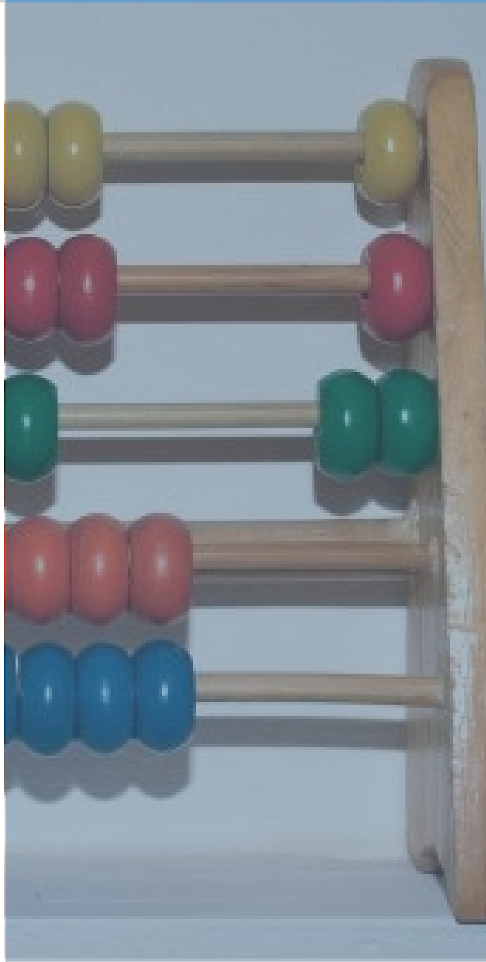
Gleichverteilung (stetige Variante)

```
x <- seq(-1,21,length=100)
hx <- dunif(x,min=0,max=20)
plot(x,hx,type="s",ylim=c(0,1),axes=F,xlab="Wartezeit", ylab="Dichte")
lb=3
ub=5
i <- x >= lb & x <= ub
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
axis(1)
area <- punif(ub, min=0, max=20) - punif(lb, min=0, max=20)
result <- paste("P(",lb,"< Wartezeit <",ub,") =",signif(area, digits=3))
mtext(result,2)
```





(stetige) Normalverteilung [1]



Total normal

Die häufigste (und damit wichtigste) Verteilung. Viele Prozesse, vor allem solche, die durch verschiedene Parameter gleichzeitig beeinflußt werden, folgen der Normalverteilung.

$$\text{Dichtefunktion: } f_{x,\mu,\sigma} = \frac{1}{\sigma * \sqrt{2 * \pi}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\mu = \mathit{arithmet.\ Mittel}\ ;\ \sigma = \mathit{Standardabweichung}\ ;\ e = \mathit{Eulersche\ Zahl} = 2{,}718281828459$$

$$\text{Erwartungswert: } E(X) = \mu$$

$$\text{Varianz: } Var(X) = \sigma$$

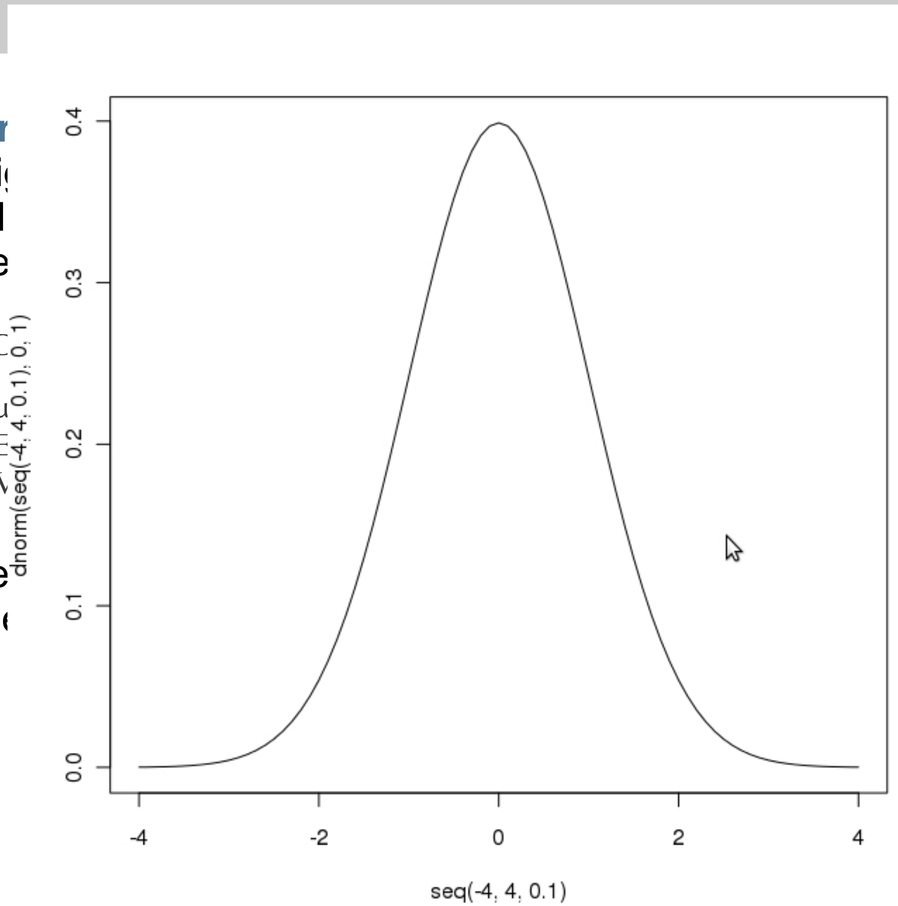
Die Verteilungsfunktion kann nicht als einfacher mathematischer Term angegeben werden (sie ist das Integral über die Dichte).

$$\text{Wahrscheinlichkeitsfunktion: } f_{x,\mu,\sigma} = \frac{1}{\sigma * \sqrt{2 * \pi}} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}}\, \delta t$$

(stetige) Normalverteilung [1]

Total nor

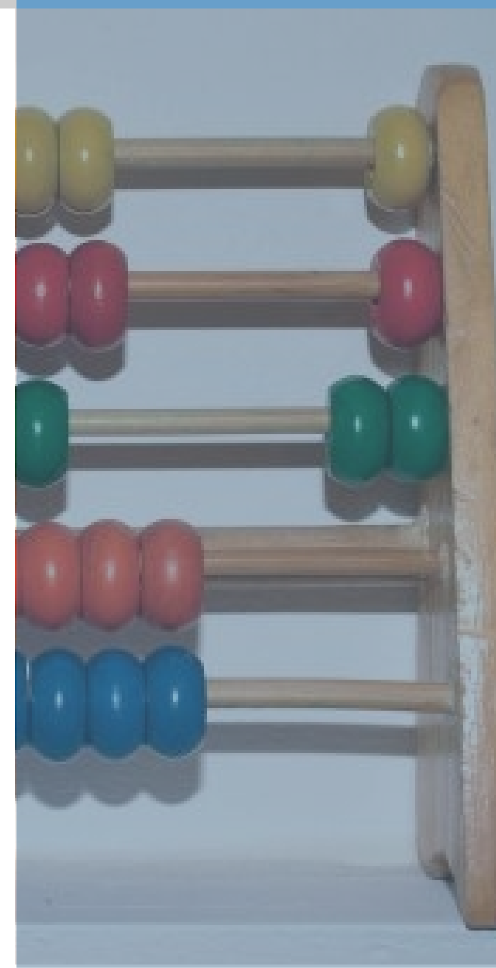
Die häufi ... se, vor allem solche, d ... fluß werden, folgen de



18281828459

Die Verte ... cher Term angegeb

```
plot(seq(-4,4,0.1),dnorm(seq(-4,4,0.1),0,1),type="l")
```



(stetige) Normalverteilung [1]

Total normal

Die häufigste (und damit wichtigste) Verteilung. Viele Prozesse, vor allem solche, die durch verschiedene Parameter gleichzeitig beeinflußt werden, folgen der Normalverteilung.

$$\text{Dichtefunktion: } f_{x,\mu,\sigma} = \frac{1}{\sigma * \sqrt{2 * \pi}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\mu = \mathit{arithmet.\ Mittel}\ ;\ \sigma = \mathit{Standardabweichung}\ ;\ e = \mathit{Eulersche\ Zahl} = 2{,}718281828459$$

$$\text{Erwartungswert: } E(X) = \mu$$

$$\text{Varianz: } Var(X) = \sigma$$

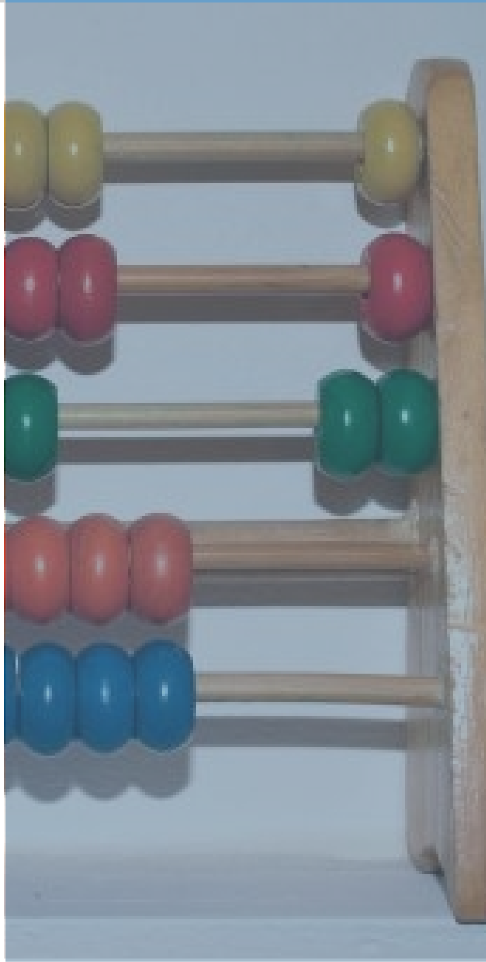
Die Verteilungsfunktion kann nicht als einfacher mathematischer Term angegeben werden (sie ist das Integral über die Dichte).

Bei einer normalverteilten Variable gibt die Standardabweichung σ jenen Bereich an, innerhalb dessen die Werte symmetrisch um den Mittelwert streuen :

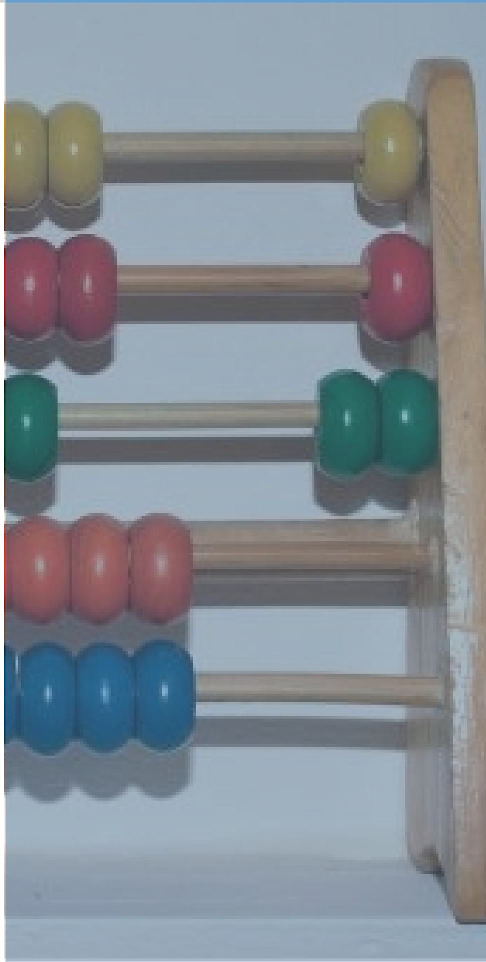
Im Bereich ±1 σ liegen 68,27 % aller Beobachtungen ("ca. 2 Drittel").

Im Bereich ±2 σ liegen 96,45 % aller Beobachtungen.

Im Bereich ±3 σ liegen 99,73 % aller Beobachtungen.



(stetige) Normalverteilung [2]



Noch normaler als Normal

Da sich mit integrierten Funktionen (Flächen unterhalb der Kurve) schlecht rechnen läßt, gibt es eine Standardnormalverteilung mit

Arithm. Mittel von 0: $\mu=0$
Standardabweichung von 1: $\sigma=1$

Für diese sind für die Werte die entsprechenden Flächeninhalte in Tabellen angegeben. Man kann jede normalverteilte Verteilung auf die Standardnormalverteilung normieren, indem man deren Mittelwert abzieht und durch die Standardabweichung teilt (sogenannter z-Wert).

Beispiel nach Shennan:
Länge von Pfeilspitzen mit einem arithm. Mittel von 110 und einer Standardabweichung von 20. Wieviel Prozent der Längen liegen zwischen 110 und 140?

$$Z=\frac{x-\mu}{\sigma}=\frac{140-110}{20}=1.5$$

(stetige) Normalverteilung [3]

Noch normaler als Normal

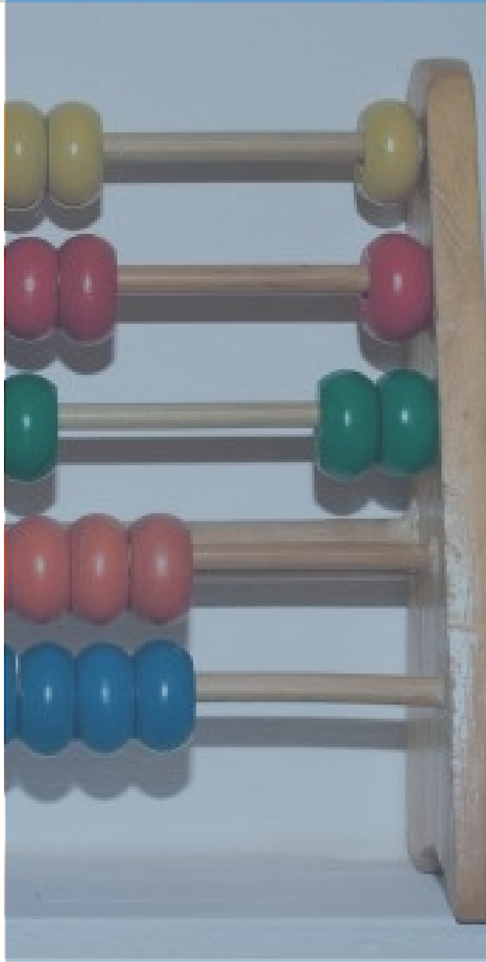
Beispiel nach Shennan:
Länge von Pfeilspitzen mit einem arithm. Mittel von 110 und einer Standardabweichung von 20. Wieviel Prozent der Längen liegen zwischen 110 und 140?

$$Z = \frac{x - \mu}{\sigma} = \frac{140 - 110}{20} = 1.5$$

Nachschlagen in der Tabelle bei Shennan ergibt:
für den Bereich von 140 bis zum Ende der Verteilung (geht gegen ∞) liegt die Wahrscheinlichkeit (Fläche unter der Kurve) bei 0,06681
Uns interessiert aber der Bereich vom arithm. Mittel (0.5 der Fläche unter der Kurve) bis 140, daher
0,5-0,06681=0,433... ~ 43,3% der Pfeilspitzen haben eine Länge von 110 bis 140.

In R:

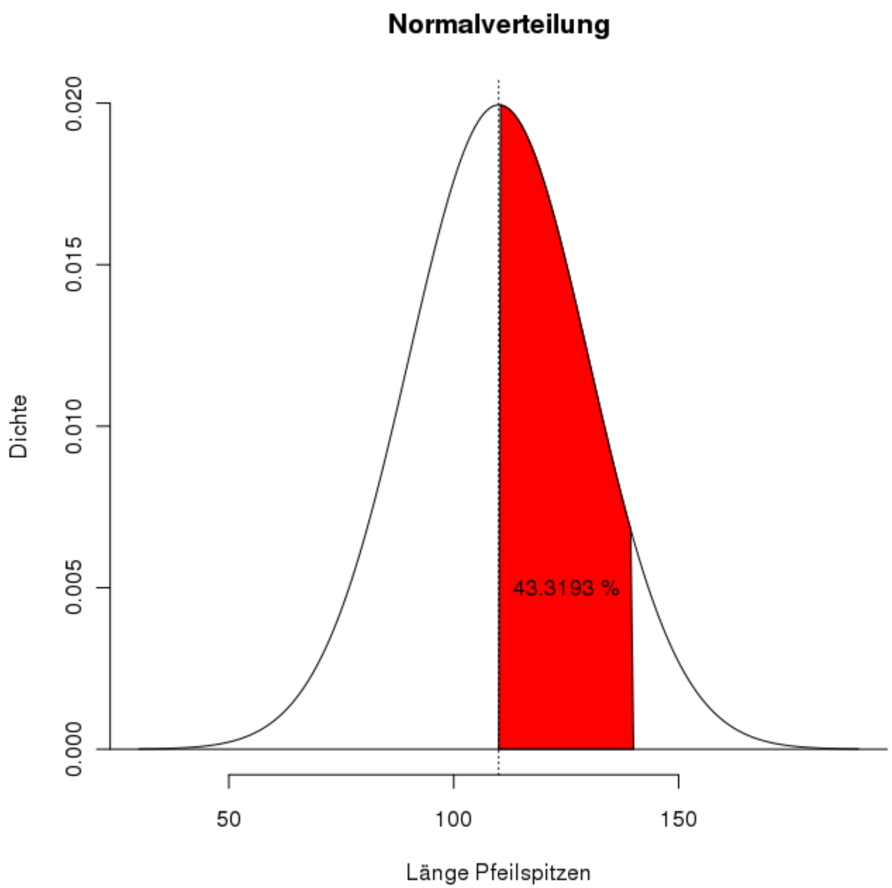
```
> pnorm(110,110,20)-(1-pnorm(140,110,20))
[1] 0.4331928
```



(stetige) Normalverteilung [3]

Noch n

Beispie
Länge \
Standaı
zwische



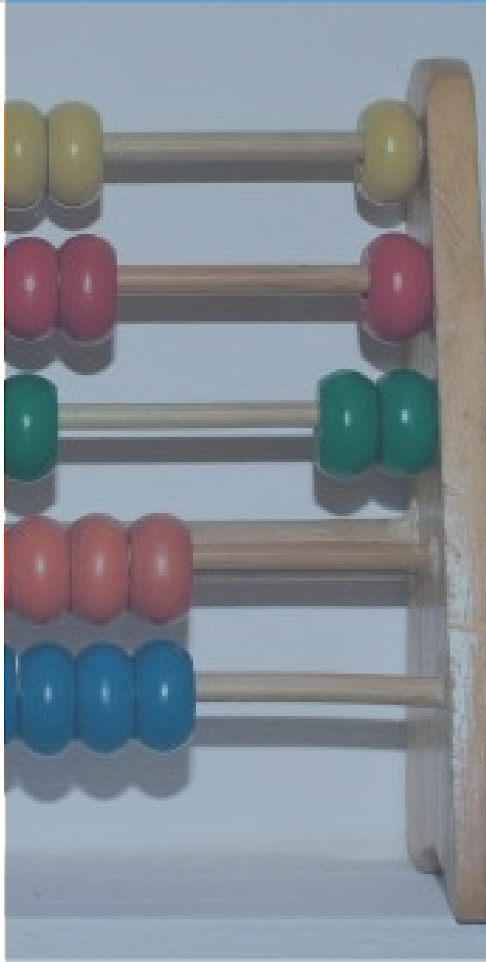
einer
gen

Nachsc
für den
die Wal
Uns inte
der Kur
0,5-0,06
bis 140

egen ∞) liegt

Fläche unter

ınge von 110



Konfidenzintervall [1]

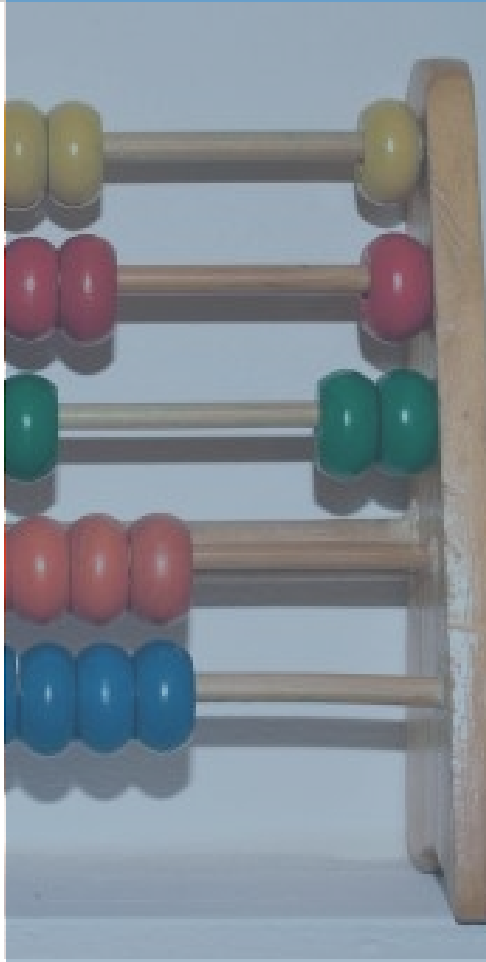
Mit Hilfe von Verteilungen lassen sich Brücken von Grundgesamtheit zu Sample schlagen
Bsp. Normalverteilung

Frage: Wie groß ist der Mittelwert einer Grundgesamtheit bei einer gg. Stichprobe?

Beispiel Shennan: Wir haben eine zufällige Stichprobe von 50 Pfeilspitzen mit Längen arithm. Mittel von 22,6 mm und einer Standardabweichung von 4,2 mm in Normalverteilung.

Für Normalverteilung wissen wir: Im Rahmen 1,96 Standardabweichungen befinden sich ziemlich genau 95% der Werte.

```
> pnorm(1.96,mean=0,sd=1)-(pnorm(-1.96,mean=0,sd=1))
[1] 0.9500042
```



Konfidenzintervall [1]

Mit Hilfe von Verteilungen lassen sich Brücken von Grundgesamtheit zu Sample schlagen
Bsp. Normalverteilung

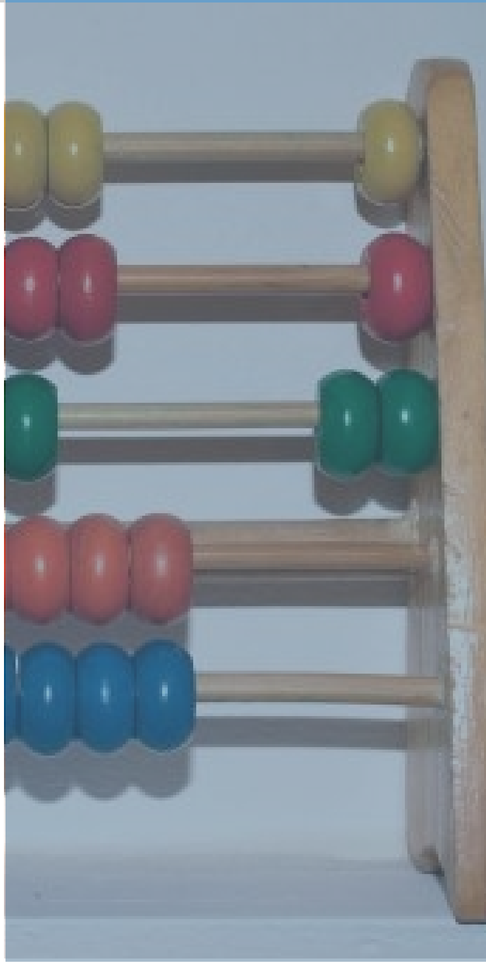
Frage: Wie groß ist der Mittelwert einer Grundgesamtheit bei einer gg. Stichprobe?

Beispiel Shennan: Wir haben eine zufällige Stichprobe von 50 Pfeilspitzen mit Längen arithm. Mittel von 22,6 mm und einer Standardabweichung von 4,2 mm in Normalverteilung.

Für Normalverteilung wissen wir: Im Rahmen 1,96 Standardabweichungen befinden sich ziemlich genau 95% der Werte.

Der Standardfehler einer Einschätzung der Mittelwerte einer Population aus einem Sample hängt von der Größe des Samples und von dessen Variabilität ab. Daher wird er berechnet:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4{,}2}{\sqrt{50}} = 0{,}594$$



Konfidenzintervall [2]

Beispiel Shennan: Wir haben eine zufällige Stichprobe von 50 Pfeilspitzen mit Längen arithm. Mittel von 22,6 mm und einer Standardabweichung von 4,2 mm in Normalverteilung.

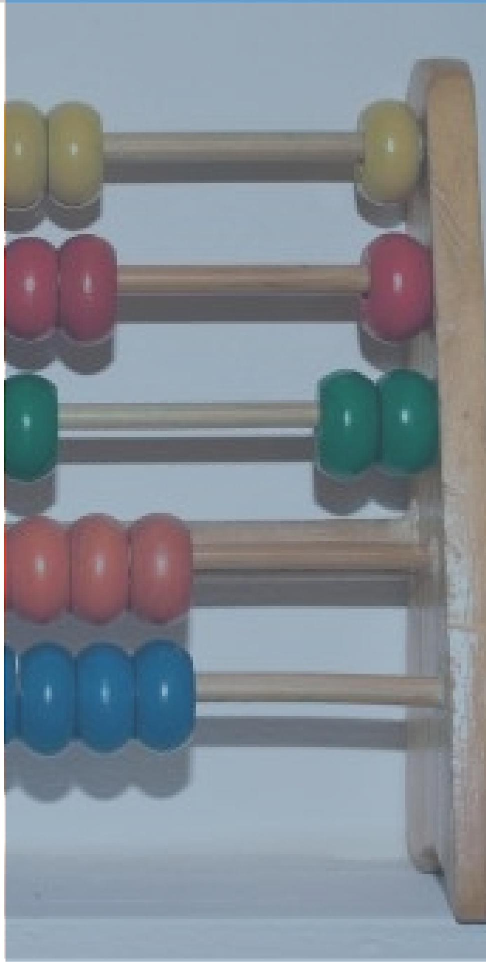
Standardfehler: 0,594

Das wahre arithm. Mittel der Grundgesamtheit befindet sich mit 95% Wahrscheinlichkeit innerhalb 1.96 Standardfehlern um den Mittelwert des Samples

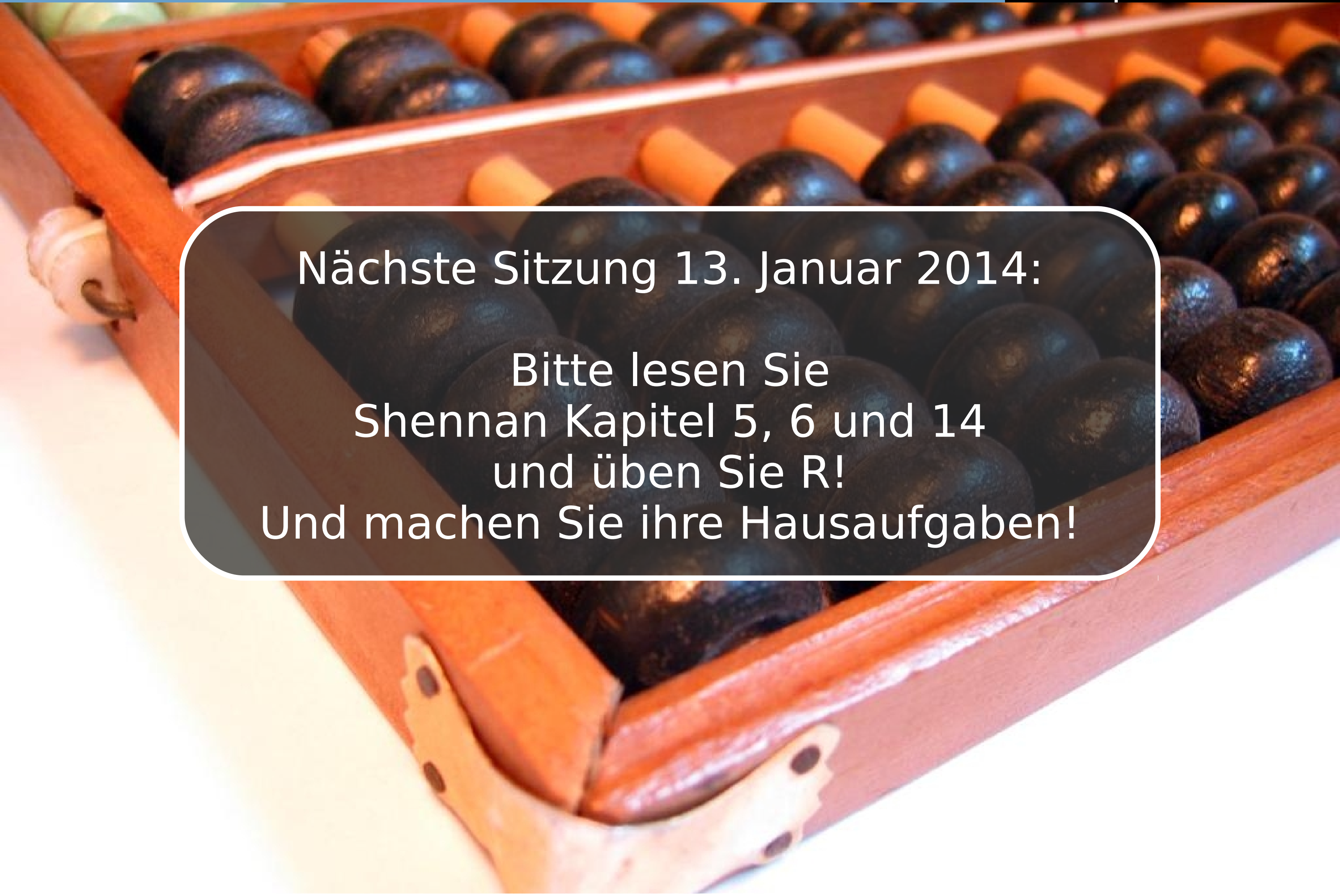
$$\mu = \bar{x} \pm (s_{\bar{x}} * 1{,}96) = 22{,}6 \pm (0{,}594 * 1{,}96) = 22{,}6 \pm 1{,}164$$

Ähnliche Verfahren werden angewand bei der Berechnung des Standardfehlers für ^{14}C-Daten.

Mit diesem Wert kann man jetzt zwei Samples vergleichen und feststellen, ob ihre Mittelwerte auf die selbe Population hindeuten
→ t-Test



ende_06_chi-quadrat/zusammenhangsmaße



Nächste Sitzung 13. Januar 2014:

Bitte lesen Sie
Shennan Kapitel 5, 6 und 14
und üben Sie R!
Und machen Sie ihre Hausaufgaben!

Hypergeometrische Verteilung [1]

Beschreibt Zufallexperimente ohne Zurücklegen

Bedingungen:

Es gibt genau zwei mögliche, sich gegenseitig ausschließende Ergebnisse (z.B. Schwarz oder Weiss)

Frage: Wie wahrscheinlich ist es, aus N+M=50 Kugeln, von denen M=20 weiss (und daher N=30 schwarz) sind, bei k=10 gezogenen Kugeln genau x=3 weisse zu erwischen?

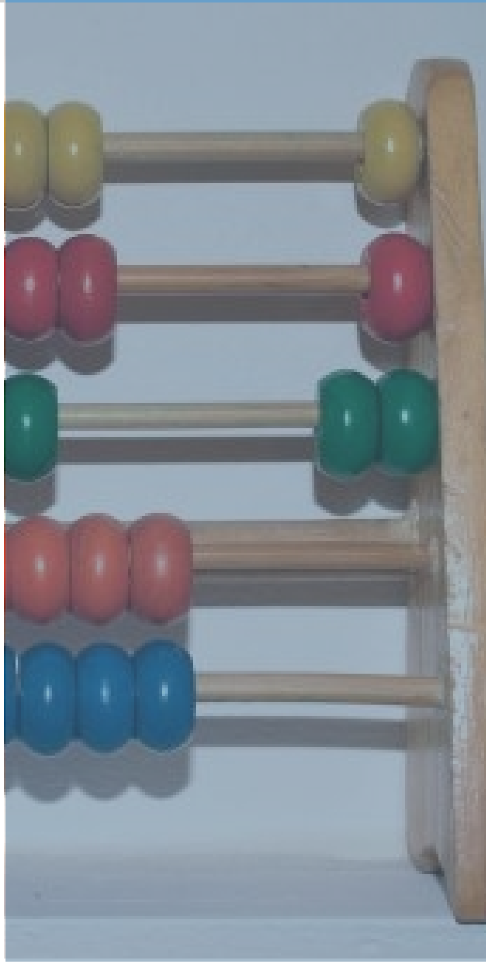
$$H_{N;M;k;x} = \frac{\binom{M}{x} * \binom{N}{k-x}}{\binom{N+M}{k}} = \frac{\binom{20}{3} * \binom{30}{10-3}}{\binom{30+20}{10}} = \frac{1140 * 2035800}{10272278170} = 0{,}225929629 = 22{,}6\,\%$$

$$\text{Erwartungswert: } E(x) = \frac{k * M}{N} = \frac{10 * 20}{50} = 4$$

$$\text{Streuung: } Var(x) = \frac{k * M}{N} * \left(1 - \frac{M}{N}\right) * \left(\frac{N+M-k}{N+M-1}\right) = \frac{10 * 20}{50} * \left(1 - \frac{20}{50}\right) * \left(\frac{50+20-10}{50+20-1}\right) = 2{,}086956522$$

In R:

```
> dhyper(x=3,m=20,n=30,k=10)
[1] 0.2259296
```



Hypergeometrische Verteilung [2]

Beschr

Beding

Es gibt ... e

Ergebni

Frage: \ ... enen M=20

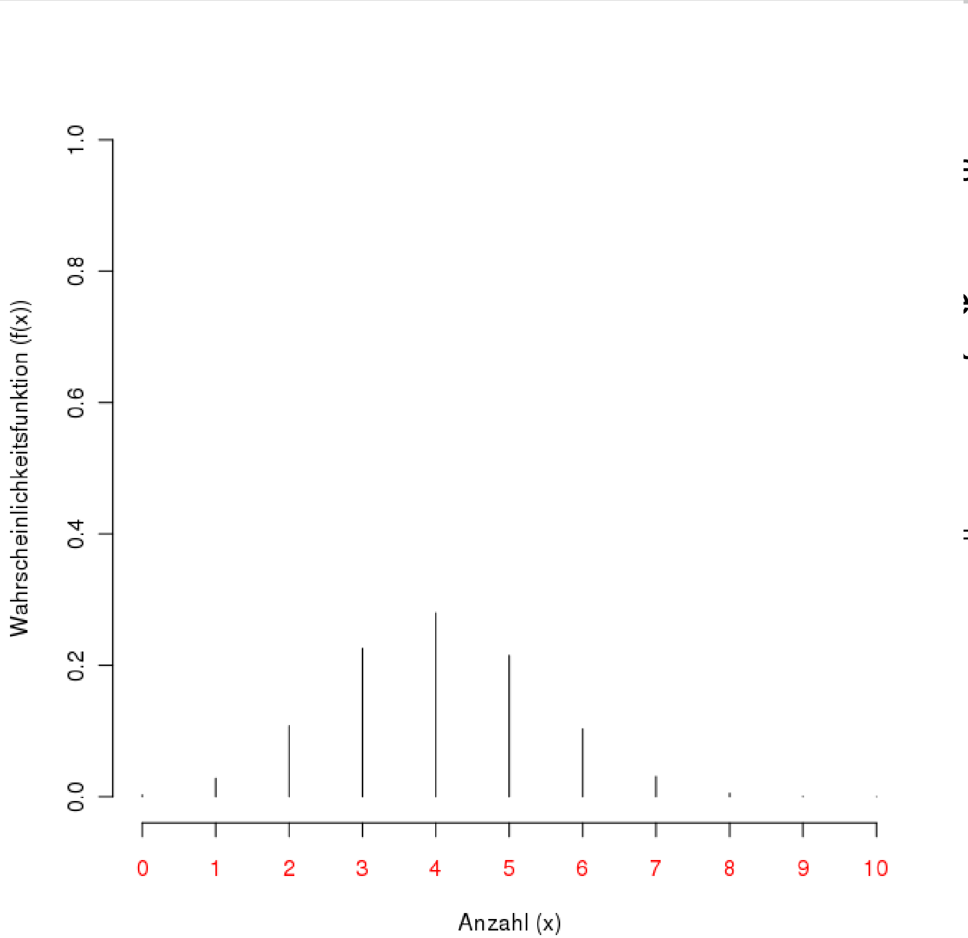
weiss (u ... ugeln

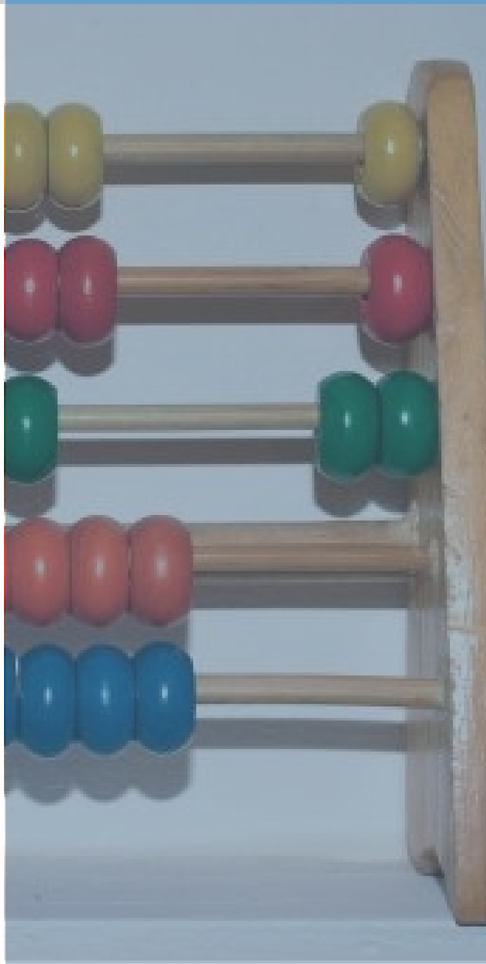
genau x

$= 22,6\,\%$

In R:

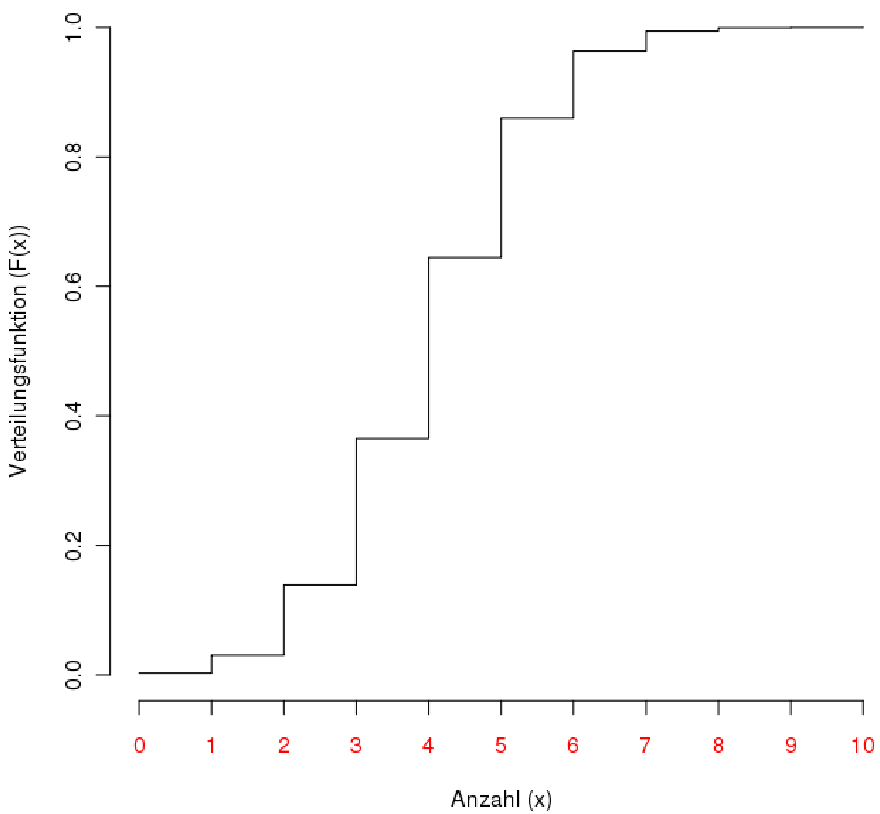
```
> dhyp
[1] 0.
```





Hypergeometrische Verteilung [2]

Besch
Beding
Es gibt
Ergebr



de

Frage:
weiss (
genau

denen M=20
Kugeln

29=22,6%

In R:

```
> dhy
[1] 0
```