

# 1. Beschreibende Statistik

Werte werden in ein Array geschrieben um bestimmte Methoden anwenden zu können.

A = [2, 3, 4, 5, 5, 9, 1, 3, 5]      B = [193, 180, 183, 185, 170, 156, 200, 188, 183]

Bei Funktionsfragen : doc functionname

## Arithmetisches Mittel

$$\frac{1}{n} \sum_{i=1}^n x_i$$

mean(A) → = 4,11

## Median(50 % Quantil)

$$\tilde{x} = \begin{cases} x_{m+1} & \text{falls } n = 2m + 1 \\ \frac{1}{2} [(x)_m + x_{m+1}] & \text{falls } n = 2m \end{cases}$$

quantile(A, 0.5) → = 4

## P-Quantil

$$\tilde{x}_p = \begin{cases} x_{[np]} & \text{falls } np \notin \mathbb{N} \\ \frac{1}{2} [(x)_{np} + x_{np+1}] & \text{falls } np \in \mathbb{N} \end{cases}$$

quantile(A, np)  
quantile(A, 0.25) → = 2.75  
quantile(A, 0.75) → = 5

## Meisvorkommende Zahl

mode(A) → = 5

## empirische Varianz

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

var(A) → = 5,36

## empirische Std-Abweichung

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

std(A) → = 5,36

## Spannweite

$$R = x_{\max} - x_{\min}$$

max(A) - min(A) → = 8

## Interquartilabstand

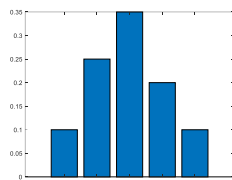
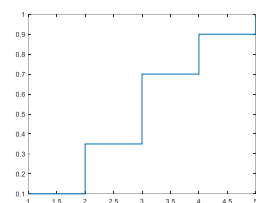
$$I = \tilde{x}_{0.75} - \tilde{x}_{0.25}$$

quantile(A, 0.75) - quantile(A, 0.25)  
→ = 2,25

(empirische) Standardabweichungen							
$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$	std(A) → 2,25						
$s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$	std(B) → 12,83						
(empirische) Kovarianz							
$s_{x,y} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$	std(A,B) → Cov	<table><tr><td>5.36</td><td>-27.25</td></tr><tr><td>-27.25</td><td>164.5</td></tr></table>	5.36	-27.25	-27.25	164.5	<div>var(A) ↓</div> <div>var(B) ↑</div>
5.36	-27.25						
-27.25	164.5						
Möglicherweise, könnte ein Multiplikator vor der Tabelle stehen. (Bsp.: 1.0e+05 )							
(empirischer) Korrelationskoeffizient							
$r_{x,y} = \frac{s_{x,y}}{s_x s_y}$	corrcoef(A,B) →	<table><tr><td>1</td><td>-0.92</td></tr><tr><td>-0.92</td><td>1</td></tr></table>	1	-0.92	-0.92	1	
1	-0.92						
-0.92	1						
$r_{x,y} = -0,92$							
$r(x,y) = 1$ Perfekte lineae abhängigkeit $r(x,y) > 0$ Positiv korreliert	$r(x,y) < 0$ Negativ korreliert $r(x,y) = 0$ unkorreliert						

<b>Bsp.: k = 2, n = 7</b>	
<b>Permutation (ohne Wiederholung)</b>	
$V(n) = n!$	factorial(n) → =5040
Möglichkeiten für Anordnungen von 10 CDs (n) im Regal	factorial(10) = 3.628.800
<b>Permutation (mit Wiederholung)</b> (nicht behandelt)	
$V(n, k) = \frac{n!}{k!}$	factorial(n)/factorial(k) → = 2520
Möglichkeiten für Anordnungen von 10 CDs (n) im Regal bei der 3 CDs (k) gleich sind	factorial(10)/factorial(3) = 604.800
<b>Variation (ohne Wiederholung)</b>	
$V(n, k) = \frac{n!}{(n-k)!}$	nchoosek(n,k)*factorial(k) → = 42
passwort mit einer Länge von 6(k) und mit dem Zeichen a-z (26(n)) die sich nicht wiederholen dürfen	nchoosek(n,k)*factorial(k)
<b>Variationen (mit Wiederholung)</b>	
$V(n, k) = n^k$	$n^k$ → = 49
Pin mit einer Länge von 4(k) und mit den Zahlen 0-9 (10(n))	$10^4 = 10.000$
Pin mit einer Länge von 4(k) und mit dem Zahlen 0-9 (10(n)) doch die erste Zahl darf keine 0 sein	$10^3 * 9 = 9.000$
<b>Kombination (ohne Wiederholung)</b>	
$\binom{n}{k} = \frac{n!}{k! (n-k)!}$	nchoosek(n,k) → = 21
beim Loto werden 6 Zahlen (k) aus 49 Zahlen (n) ausgewählt	nchoosek(49,6) = 13.983.816
<b>Kombination (mit Wiederholung)</b>	
$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k! (n-1)!}$	nchoosek(n+k-1,k) → = 28
beim Loto werden 6 Zahlen (k) aus 49 Zahlen (n) ausgewählt doch die Zahlen können sich wiederholen	nchoosek(49+6-1,6) = 25.827.165

## 2. Diskrete Wahrscheinlichkeitstheorie

Zufallsvariablen	
Zufallsvariablen $\rightarrow X = [1, 2, 3, 4, 5]$ Wahrscheinlichkeiten zu diesen Zufallsvariablen $\rightarrow P = [0.1, 0.25, 0.35, 0.2, 0.1]$	
Verteilung	<div> <math>\text{bar}(X, P)</math>  </div> $P(X \leq x) = \begin{cases} 0 & x \notin T \\ 0.1 & x = 1 \\ 0.25 & x = 2 \\ 0.35 & x = 3 \\ 0.2 & x = 4 \\ 0.5 & x = 5 \end{cases}$
Verteilungsfunktion $\bar{F}(x) = \sum_{i: a_i \leq x} f_i$	<div> <math>\text{stairs}(\text{cumsum}(P))</math>  </div> <p>(sieht ein bisschen falsch aus, da die y-Achse bei 0,1 beginnt, ist es aber nicht)</p> $P(X \leq x) = \begin{cases} 0 & x < 1 \\ 0.1 & 1 \leq x < 2 \\ 0.35 & 2 \leq x < 3 \\ 0.7 & 3 \leq x < 4 \\ 0.9 & 4 \leq x < 5 \\ 1 & 5 \leq x \end{cases}$
Erwartungswert $E(X)$ $E(x) = \sum x P(X = x)$	$\text{sum}(X.*P) \rightarrow = 2,95$
Varianz $\text{Var}(x) = E(X - \mu)^2$	$\text{var}(X, P) \rightarrow = 1,25$
Standard $\sigma_x = \sqrt{\text{Var}(X)}$	$\text{std}(P, X) \rightarrow = 0,09$
Bernouli-Verteilung	
$X \sim \text{Ber}(p)$	
<b>Verteilung</b>	$P(X = x) = \begin{cases} q & \text{falls } x = 0 \\ p & \text{falls } x = 1 \\ 0 & \text{sonst} \end{cases}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = F(x) = \begin{cases} q & \text{falls } x < 0 \\ p & \text{falls } 0 \leq x < 1 \\ 0 & \text{falls } x \geq 1 \end{cases}$
<b>Erwartungswert:</b>	$E[X] = p$
<b>Varianz:</b>	$\text{Var}[X] = pq$

<b>Geometrische Verteilung</b> $X \sim \text{geom}(p)$	
<b>Verteilung</b>	$P(X = x) = \begin{cases} pq^{[x]} & \text{falls } x \in \mathbb{N} \\ 0 & \text{sonst.} \end{cases}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = \begin{cases} 1 - q^{[x]} & \text{falls } x \geq 1 \\ 0 & \text{sonst.} \end{cases}$
<b>Erwartungswert</b>	$E[X] = \frac{1}{p}$
<b>Varianz</b>	$\text{Var}[X] = \frac{q}{p^2}$
Bsp.: $p = 0.2 \rightarrow$ Erfolg $q = 0.8 \rightarrow$ Misserfolg	
3 Versuche bis Erfolg	$\text{geopdf}(x-1, p) \rightarrow = 0,1024$
höchstens/bis zu/max 4 Versuche	$\text{geocdf}(x-1, p) \rightarrow = 0,590$
mehr als/min 4 Versuche	$1 - \text{geocdf}(x-1, p) \rightarrow = 0,409$
<b>Binominal-Verteilung</b> $X \sim \text{Bin}(n, p)$	
<b>Verteilung</b>	$P(X = x) = \begin{cases} \binom{n}{x} p^x * q^{n-x} & \text{falls } x \in \mathbb{N}_0 \\ 0 & \text{sonst.} \end{cases}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = F(x)$ $= \begin{cases} 0 & \text{falls } x < 0 \\ e^{-\lambda} \sum_{k=0}^{[x]} \binom{n}{k} p^k * q^{n-k} & \text{falls } 0 \leq x \leq n \\ 1 & \text{falls } x > n \end{cases}$
<b>Erwartungswert</b>	$E[X] = np$
<b>Varianz</b>	$\text{Var}[X] = npq$
Wahrscheinlichkeit, für 3 Erfolge (x) mit 4 Versuche (n) $p =$ Wk, für Erfolg	$\text{binopdf}(x, n, p) \rightarrow = 0,026$
Wahrscheinlichkeit, für höchstens/bis zu/max 3 Erfolge (x) mit 4 Versuche (n) $P(X \leq x)$	$\text{binocdf}(x, n, p) \rightarrow = 0,998$
Wahrscheinlichkeit, für mehr als/min 3 Erfolge (x) mit 4 Versuche (n) $P(X \geq x)$	$1 - \text{binocdf}(x, n, p) \rightarrow = 0,002$
Wahrscheinlichkeit, für 7 (x <sub>1</sub> ) oder 12 Erfolge(x <sub>2</sub> ) mit n Versuche $P(X_1 = x_1) + P(X_2 = x_2)$	$\text{binopdf}(x_1, n, p) + \text{binopdf}(x_2, n, p)$

<b>Poisson-Verteilung</b> $X \sim Po(\lambda)$	
<b>Verteilung</b>	$P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{falls } x \in \mathbb{N}_0 \\ 0 & \text{sonst.} \end{cases}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = \begin{cases} 0 & \text{falls } x < 0 \\ e^{-\lambda} \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k}{k!} & \text{falls } 0 \leq x \end{cases}$
<b>Erwartungswert</b>	$E[X] = \lambda$
<b>Varianz</b>	$Var[X] = \lambda$
Bsp.: Schrauben in einer Kiste	
durchsch. 2 ( $\lambda$ ) defekt Wahrschk., dass <b>genau</b> 5 kaputt ( $x$ ) sind	$\text{poisspdf}(x, \lambda) \rightarrow y = 0,036$
durchsch. 2 defekt ( $\lambda$ ) Wahrschk., dass <b>höchstens/bis zu/max</b> 2 kaputt ( $x$ ) sind	$\text{poisscdf}(x, \lambda) \rightarrow y = 0,68$
durchsch. 2 defekt ( $\lambda$ ) Wahrschk., dass <b>mehr als/min</b> 2 kaputt ( $x$ ) sind	$1 - \text{poisscdf}(x, \lambda) \rightarrow y = 0,32$
<b>Eponentialverteilung</b> $X \sim exp(\lambda)$	
<b>Verteilungsdichte</b>	$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$
<b>Erwartungswert</b>	$E[X] = \frac{1}{\lambda}$
<b>Varianz</b>	$Var[X] = \frac{1}{\lambda^2}$
Bsp.: 10 Anrufe pro Stunde ( $\lambda$ )	
Mit welcher Wk. liegen zwischen zwei Anrufen <b>mehr als/min</b> als 6 Minuten $\rightarrow 0.1$ Stunden ( $x$ )?	$1 - \text{expcdf}(x * \lambda)$
Mit welcher Wk. liegen zwischen zwei Anrufen <b>weniger</b> als 6 Minuten $\rightarrow 0.1$ Stunden ( $x$ )?	$\text{expcdf}(x * \lambda)$

### 3. Kontinuierliche Wahrscheinlichkeitstheorie

#### Stetige Funktionen

##### Wichtig!!

\*Symbolische Variable anlegen:

`syms x`

(auch möglich: `syms x y z`)

\*Funktionen eingeben:

`f(x)=x^2`

→ `f=@(x) x.^3 + 3*x.^2`

Vor jedem Operanten (\*, /, ^..) am besten ein "." setzen

Funktion plotten:

`fplot(f)`

`fplot(f, [x1-x2])`

Bereich/Integral plotten

`ezplot(F, min, max)`

Ableiten:

`diff(f)`

Stammfunktion bilden:

`F = int(f, x)`

→ f: zu integrierende Funktion x:

In Matlab wird das "+ C" nach der aufleitung weggelassen!!!

Integral bestimmen

`integral(f, min, max)`

Für Funktionen  $F(t) = 1 - e^{-\lambda t}$

→  $\lambda = \text{syms } y$  →  $t = \text{syms } x$  →  $F = @(x) 1 - \exp(-y*x)$

$F'(x) = \text{diff}(f,x)$  →  $y * \exp(-y*x)$

Für unbekannte Variablen werden einfach weitere symbolische Variablen angelegt

Bei einer Funktion wie  $f=@(x) \lambda * \exp(-C*x)$

wird fehlerweise bei der Integration (`int(f,x)`) die 1 weggelassen

statt " $1 - \exp(-\lambda*x)$ " wird " $-\exp(-\lambda*x)$ " angezeigt !!!

Dies Betrifft jedoch keine Abläufe mit der Funktion `integral(f,min,max)`!!!

Providerbeispiel:  $\lambda = 0.5$

$f=@(x) 1 - \exp(-0.5*x)$

<p>Wk, dass zwischen zwei Ankünften <b>höchstens</b> 1 Sekunde liegt</p> <p><math>P(A \leq 1) = A(1) = 1 - e^{-0.5}</math></p>	<p><code>integral(f,0,1)</code> → 0,394</p>
<p>Wk, dass zwischen zwei Ankünfte <b>zwischen</b> 1 und 2 Sekunden...</p> <p><math>P(1 &lt; A &lt; 2) = A(2) - A(1)</math></p>	<p><code>integral(f,1,2)</code> → 0,239</p>
<p>Wk, dass zwischen zwei Ankünften <b>mehr</b> als 2 Sekunde liegen</p> <p><math>P(A &gt; 2) = 1 - A(2) = 1 - e^{-0.5}</math></p>	<p><code>1-integral(f,0,2)</code> → 0,368</p>
<p>Wk, dass zwischen zwei Ankünften <b>genau</b> als 2 Sekunde liegen</p>	<p>geht nicht → 0</p>

wichtiger  
kommentar  
hierzu



<b>Erwartungswert E(X) oder <math>\mu_x</math></b> Kein explizite Funktion gefunden ! Bsp.: $f(x) = 2x$	
$E(X) = \int_{-\infty}^{\infty} x f(x) dx$	x muss expliziet zu f(x) dazu multipliziert werden! $f(x) * x \rightarrow t = 2x^2$ integral(t,min,max)
<b>Varianz</b> Kein explizite Funktion gefunden !	
$Var(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$	$(x - \mu)^2$ muss expliziet zu f(x) dazu multipliziert werden! $f(x) * x^2 \rightarrow t = 2x^3$ integral(t, min, max) - $(\mu_x)^2$
<b>Standardabweichung <math>\sigma_x</math></b> Kein explizite Funktion gefunden !	
$\sqrt{Var(X)}$	sqrt(Varianz)
<b>Verteilungsdichte</b>	
$f(x) = F'(x)$ $P(X = x) = 0$	
<b>Verteilungsfunktion</b>	
$F(x) = \int_{-\infty}^x f(t) dt$	int(f,x)

<b>Gleichverteilung</b> $X \sim U(a, b)$	
<b>Verteilungsdichte</b>	$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = F(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b \\ 1 & \text{falls } x > b \end{cases}$
<b>Erwartungswert</b>	$E[X] = \frac{(a+b)}{2}$
<b>Varianz</b>	$Var[X] = \frac{(b-a)^2}{12}$



## Normalverteilung

$$X \sim N(\mu, \sigma)$$

**Bei weiteren Zufallsvariablen, die dazu addiert werden muss:**

$$X + Y \sim N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$$

<b>Verteilungsdichte</b>	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
<b>Verteilungsfunktion</b>	$P(X \leq x) = F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$
<b>Erwartungswert</b>	$E[X] = \mu$
<b>Varianz</b>	$Var[X] = \sigma^2$
<b>Standardabweichung</b>	$\sqrt{Var[X]} = \sigma$
<b>Annahme, dass Größe ~ N(180.3, 7.17)</b>	
Wk, zufälliger man <b>höchstens</b> 1.75m $P(X) \leq 175$	<b>normcdf(x, μ, σ)</b>  normcdf(175, 180.3, 7.17) = 0,23
Welche Größe wird von 95% <b>nicht überschritten</b>	<b>norminv(p, μ, σ)</b>  norminv(0.95, 180.3, 7.17) = 192,09 cm
68-95-99.7-Regel	
68% liegt zwischen $x_1$ und $x_2$	$x_1 = \text{norminv}(0.1585, \mu, \sigma)$ $x_2 = \text{norminv}(0.8385, \mu, \sigma)$
95% liegt zwischen $x_1$ und $x_2$	$x_1 = \text{norminv}(0.0235, \mu, \sigma)$ $x_2 = \text{norminv}(0.9735, \mu, \sigma)$
99,7% liegt zwischen $x_1$ und $x_2$	$x_1 = \text{norminv}(0.0, \mu, \sigma)$ $x_2 = \text{norminv}(0.997, \mu, \sigma)$

## 4.Schließende Statistik

Schätzfunktionen $X \sim U(\mu, \sigma)$	
Stichprobe = $X = [x_1, x_2, x_3, \dots]$	
<b>Arithmetisches Mittel</b> $\hat{\mu} = \bar{x}$	mean(X)
<b>Empirische Varianz</b> $\hat{\sigma}^2 = s^2$	var(X)
<b>Relative Häufigkeit</b> $\hat{\pi}_x = \bar{p}_x$	mean(X < x)

(Wurde nicht behandelt)	
<b>Hypergeometrische Verteilung</b>	
Bsp.: Kiste mit 50 Schrauben(M) 10% sind defekt $\rightarrow x = M \cdot 0.1 \rightarrow K = M - x$ 8(n) Schrauben werden mit einem Griff entnommen Wahrscheinlichkeit, dass 3(n-x) defekt sind	
$y = f(x M, K, n) = \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}$	$\text{hygepdf}(x, M, K, n) \rightarrow y = 0,023$