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)

$\tilde{x} = \begin{cases} x_{m+1} & falls\ n = 2m+1 \\ \frac{1}{2}\llbracket(x)\rrbracket_m + x_{m+1}) & falls\ n = 2m \end{cases}$	quantile(A, 0.5) → = 4

P-Quantil

$\tilde{x}_p = \begin{cases} x_{\lceil np \rceil} & falls\ np \notin \mathbb{N} \\ \frac{1}{2}\llbracket(x)\rrbracket_{np} + x_{np+1}) & falls\ np \in \mathbb{N} \end{cases}$	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

Interquatilabstand

$I = \tilde{x}_{0.75} - \tilde{x}_{0.25}$	quantile(A, 0.75) - quantile(A, 0.25) → = 2,25

(empirische) Standardabweichungen

$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$	std(A) → 2,25
$s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2}$	std(B) → 12,83

(empirische) Kovarianz

var(A)

$s_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$	~~std~~(A,B) → cov	5.36	-27.25 -27.25	164.5
	$s_{x,y}$ = -27,25	var(B)

Möglicherweise, könnte ein Multiplikator vor der Tabelle stehen. (Bsp.: 1.0e+05)

5.36	-27.25
-27.25	164.5

(empirischer) Korrelationskoeffizient

$r_{x,y} = \frac{s_{x,y}}{s_x s_y}$	corrcoef(A,B) → $r_{x,y}$ = -0,92

1	-0.92
-0.92	1

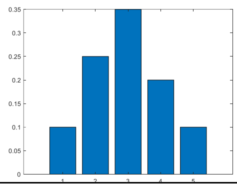
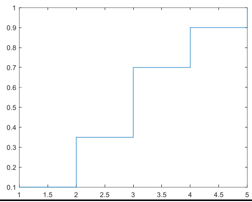
r(x,y) = 1 Perfekte linae abhänigkeit r(x,y) > 0 Positiv korrieliert	r(x,y) < 0 Negativ korrieliert r(x,y) = 0 unkorrieliert

Bsp.: k = 2, n = 7	
Permutation (ohne Wiederholung)	
$V(n) = n!$	factorial(n) → =5040
Möglichkeiten für Anordnungen von 10 CDs (n) im Regal	factorial(10) = 3.628.800
Permutation (mit Wiederholung) (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
Variation (ohne Wiederholung)	
$V(n,k) = \frac{n!}{(n-k)!}$	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)
Variationen (mit Wiederholung)	
$V(n,k) = n^k$	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
Kombination (ohne Wiederholung)	
$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$	nchoosek(n,k) → = 21
beim Loto werden 6 Zahlen (k) aus 49 Zahlen (n) ausgewählt	nchoosek(49,6) = 13.983.816
Kombination (mit Wiederholung)	
$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$	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 → X = [1, 2, 3, 4, 5]
Wahrscheinlichkeiten zu diesen Zufallsvariablen → P = [0.1, 0.25, 0.35, 0.2, 0.1]

Verteilung	**bar(X,P)**  $P(X \cancel{\le} x) = \begin{cases} 0 & x \notin T \\ 0.1 & x = 1 \\ 0.25 & x = 2 \\ 0.35 & x = 3 \\ 0.2 & x = 4 \\ 0.5 & x = 5 \end{cases}$
Verteilungsfunktion $\overline{F}(x) = \sum_{i:a_i \le x} f_i$	**stairs(cumsum(P))**  **(sieht ein bisschen Falsch aus, da die y-Achse bei 0,1 beginnt, ist es aber nicht)** $P(X \le x) = \begin{cases} 0 & x < 1 \\ 0.1 & 1 \le x < 2 \\ 0.35 & 2 \le x < 3 \\ 0.7 & 3 \le x < 4 \\ 0.9 & 4 \le x < 5 \\ 1 & 5 \le x \end{cases}$
Erwartungswert E(X) $\mathrm{E}(\mathrm{x}) = \sum xP(X = x)$	**sum(X.*P) → =2,95**
Varianz $Var(x) = E(X - \mu)^2$	**var(X,P) → =1,25**
Standard $\sigma_x = \sqrt{Var(X)}$	**std(P,X)** std(x,P) **→ =0,09**

Bernouli-Verteilung
$X \sim Ber(p)$

Verteilung	$P(X = x) = \begin{cases} q \; falls \; x = 0 \\ p \; falls \; x = 1 \\ 0 \; sonst. \end{cases}$
Verteilungsfunktion	$P(X \le x) = F(x) = \begin{cases} q \; falls \; x < 0 \\ p \; falls \; 0 \le x < 1 \\ 0 \; falls \; x \ge 1 \end{cases}$
Erwartungswert:	$E[X] = p$
Varianz:	$Var[X] = pq$

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

Poisson-Verteilung $X \sim Po(\lambda)$	
Verteilung	$P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} \; falls \; x \in \mathbb{N}_0 \\ 0 \; sonst. \end{cases}$
Verteilungsfunktion	$P(X \leq x) = \begin{cases} 0 \; falls \; x < 0 \\ e^{-\lambda} \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k}{k!} \; falls \; 0 \leq x \end{cases}$
Erwartungswert	$E[X] = \lambda$
Varianz	$Var[X] = \lambda$
Bsp.: Schrauben in einer Kiste	
durchsch. 2(λ) deffekt Wahrschk., dass genau 5 kaputt (x) sind	poisspdf(x, λ) → y = 0,036
durchsch. 2 deffekt(λ) Wahrschk., dass höchstens/bis zu/max 2 kaputt (x) sind	poisscdf(x, λ) → y = 0,68
durchsch. 2 deffekt(λ) Wahrschk., dass mehr als/min 2 kaputt (x) sind	1-poisscdf(x, λ) → y = 0,32

Eponentialverteilung $X \sim exp(\lambda)$	
Verteilungsdichte	$f(x) = \begin{cases} \lambda e^{-\lambda x} \; für \; x \geq 0 \\ 0 \; sonst \end{cases}$
Verteilungsfunktion	$P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} \; für \; x \geq 0 \\ 0 \; sonst \end{cases}$
Erwartungswert	$E[X] = \frac{1}{\lambda}$
Varianz	$Var[X] = \frac{1}{\lambda^2}$
Bsp.: 10 Anrufe pro Stunde (λ)	
Mit welcher Wk. liegen zwischen zwei Anrufen mehr als/min als 6 Minuten → 0.1 Stunden (x)?	1-expcdf(x*λ)
Mit welcher Wk. liegen zwischen zwei Anrufen weniger als 6 Minuten → 0.1 Stunden (x)?	expcdf(x*λ)

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, $[x_1-x_2]$)
Bereich/Integral plotten ezplot(F, min, max)
Ableiten: diff(f)
Stammfunktion bilden: F = int(f, x)
→ f: zu intigrierende Funktion x:
In Matlab wird das "+ C" nach der aufleitung weggelassen!!!
Integral bestimmen integral(f, min, max)

wichtiger Kommentar hiezu

Für Funktionen F(t) = $1-e^{-\lambda t}$
→ λ = syms y →t = syms x → F = @(x) 1-exp(-y*x)
F'(x) = diff(f,x) →y*exp(-y*x)

Für unbekannte Variablen werden einfach weitere symbolische Variablen angelegt

Bei einer Funktion wie f=@(x) λ .*exp(-C*x)
wird fehlerweise bein der Integration (int(f,x)) die 1 weggellassen
statt "1 – exp(-λ*x)" wird "– exp(-λ*x)"angezeigt !!!

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

Providerbeispiel: λ = 0.5
f=@(x)1-exp(-0.5*x)

Wk,dass zwischen zwei Ankünften **höchstens** 1 Sekunde liegt $P(A \leq 1) = \; A(1) = 1 - e^{-0.5}$	integral(f,0,1) → 0,394
Wk,dass zwischen zwei Ankünfte **zwischen** 1 und 2 Sekunden… $P(1 < A < 2) = \; A(2) - A(1)$	integral(f,1,2) → 0,239
Wk,dass zwischen zwei Ankünften **mehr** als 2 Sekunde liegen $P(A > 2) = \; 1 - A(2) = 1 - e^{-0.5}$	1-integral(f,0,2) → 0,368
Wk,dass zwischen zwei Ankünften **genau** als 2 Sekunde liegen	geht nicht → 0

Erwartungswert E(X) oder μ_x Kein expliziete Funktion gefunden ! Bsp.: f(x) = 2x	
$E(X)=\int_{-\infty}^{\infty} xf(x)dx$	x muss expliziet zu f(x) dazu multipliziert werden! f(x)*x → t=@(x)2x.^2 integral(t,min,max)
Varianz Kein expliziete 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 → t=@(x)2x.^3 integral(t, min, max) – $(\mu_x)^2$
Standardabweichung 6$_x$ Kein expliziete Funktion gefunden !	
$\sqrt{Var(X)}$	sqrt(Varianz)
Verteilungsdichte	
$f(x)=F'(x)$ $P(X=x)=0$	
Verteilungsfunktion	
$F(x)=\int_{-\infty}^{x} f(t)dt$	int(f,x)

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

Normalverteilung

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

Bei weiteren Zufallsvariablen, die dazu addiert werden muss:

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

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

4.Schließende Statistik

Schätzfunktionen X~U(μ, σ)	
Stichprobe = X = [x1,x2,x3,….]	
Arithmetisches Mittel $\widehat{\mu} = \overline{x}$	mean(X)
Empirische Varianz $\widehat{\sigma^2} = s^2$	var(X)
Relative Häufigkeit $\widehat{\pi_x} = \overline{p_x}$	mean(X<x)

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