DESKRIPTIVE STATISTIK

nominale Merkmale	lassen sich nicht in natürlicher Weise anordnen
ordinale Merkmale	lassen sich in natürlicher Reihenfolge anordnen
diskrete Merkmale	Wertebereich abzählbar, aber möglicherweise unendlich
stetige Merkmale	Merkmale, die prinzipiell jeden beliebigen Wert (in einem Intervall) annehmen

Grundgesamtheit: Ω	h_i: absolute Häufigkeit einer Merkmalsausprägung a_i
Grösse Stichprobe: n	$h_1 + h_2 + \cdots + h_m = n$
Objekte: $\omega_1, \omega_2 \ldots, \omega_n$	f_i: relative Häufigkeit, $f_i = \frac{h_i}{n}$, $(f_1 + f_2 + \cdots + f_m = 1)$
Merkmale: $X(\omega_i) = x_i$	Höhe eines Balkens/Stabes der PMF

KUMULATIVE HÄUFIGKEIT

$$H(x) = Anzahl\ aller\ Stichprobenwerte \leq X$$

empirische relative Summenhäufigkeit:

Klasse → $\frac{H(x)}{n}$ → rechter Wert Klasse

$$H(x) = \sum_{i:a_i \leq x} h_i \rightarrow absolut$$

$$F(x) = \frac{H(x)}{n} = \sum_{i:a_i \leq x} f_i \rightarrow relativ$$

ABSOLUTE HÄUFIGKEIT (RECHTECKSFLÄCHE)

Höhe Rechteck: $h(x) = \frac{h_i}{d_i}$ (abs. Häufigkeitsdichte) und $f(x) = \frac{f_i}{d_i}$ (rel. Häufigkeitsdichte)

d_i: Klassenbreite (Breite im Intervall, Säulenbreite im Histogramm)

RELATIVE HÄUFIGKEITSDICHTEFUNKTION (PDF)

$$f(x) = \frac{h(x)}{n}, f: \mathbb{R} \rightarrow [0,1]$$

KUMULATIVE VERTEILUNGSFUNKTION (CDF)

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, (Integration\ von\ PDF)$$

in Klassen:

$$F(x) = \frac{F(B)}{B - A} \cdot (X - A) + F(A)$$

LAGEMASSE

arithmetisches Mittel:

$$\bar{x} = \frac{1}{n}\sum_{1}^{m} x_i = \frac{1}{n}\sum_{1}^{m} h_i \cdot a_i = \sum_{1}^{m} f_i \cdot a_i$$

Median: 0.5 – Quantil = Median/Zentralwert → teilt in 2 gleich grosse Hälften

$$x_{med} = \begin{cases} \frac{1}{2} * \left(x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}\right) \rightarrow falls\ n\ gerade \\ x_{\left[\frac{n+1}{2}\right]} \rightarrow falls\ n\ ungerade \end{cases}$$

Modus x_{mod}:

- "Maximum der PMF/PDF"
- häufigster Stichprobenwert → Wert der am meisten vorkommt

rechtsschief	$x_{mod} < x_{med} < \bar{x}$
symmetrisch	$x_{mod} = x_{med} = \bar{x}$
linksschief	$x_{mod} > x_{med} > \bar{x}$

STREUMASSE

Varianz:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{m} h_i\,(a_i - \bar{x})^2 = \sum_{i=1}^{m} {a_i}^2 \cdot f_i - \bar{x}^2$$

korrigierte Varianz:

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

Standardabweichung:

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

korr. Standardabweichung:

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

→ ! Bei klassierten Daten Klassenmitte verwenden!

KENNGRÖSSEN

Interquartilsabstand (IQR): $3. Quartil - 1. Quartil = Q_3 - Q_1$

Mean absolute deviation MAD: $1.4826 \cdot median(|x_1 - \bar{x}|, |x_2 - \bar{x}|, \ldots, |x_n - \bar{x}|)$

Perzentil/Quantile q: Zahl R teilt Stichprobenwerte im Verhältnis $q: (1-q)$

$R_q = \frac{1}{2} \cdot (x_{[n \cdot q]} + x_{[n \cdot q + 1]})$ $\quad n \cdot q = ganze\ Zahl$

$R_q = x_{[n \cdot q]}$ $\quad n \cdot q \neq ganze\ Zahl, mit\ n \cdot q\ nächstgrössere\ Zahl$

Quartile: $Q_1: R_{0.25},\ Q_2 = Median: R_{0.5}, Q_3: R_{0.75} \rightarrow 1,2,3\ Quartil$

bei klassierten Daten mit Verteilfunktion: $Rq = F^{-1}(q)$

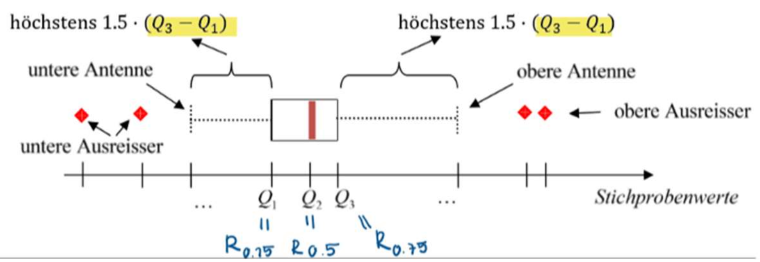
Klasse suchen in der q ist

$$\frac{F(b) - F(a)}{b - a} = \frac{q - F(a)}{R_q - a}$$

$$R_q = a + \frac{(b-a)}{F(b) - F(a)}(q - F(a))$$

$$\boxed{F(a) = \frac{\overbrace{absolut}^{Wert\ von\ nächster\ Klasse\ links\ nehmen}}{n}}$$

BOXPLOT



Interquartilsabstand (IQR): $3. Quartil - 1. Quartil = Q_3 - Q_1$

untere Antenne: $1.5 \cdot (Q_3 - Q_1)$ von $Q_1 \mid Q_1 - 1.5 \cdot IQR$

obere Antenne: $1.5 \cdot (Q_3 - Q_1)$ von $Q_3 \mid Q_3 + 1.5 \cdot IQR$

➔ IMPORTANT! Alles Werte!

KORRELATION

Pearson-Korrelation misst Stärke/Richtung des linearen Zusammenhangs

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x \cdot s_y}$$

$\bar{x} = \frac{1}{n}\sum_{i=1}^n x$ (aritm. Mittelwert)	$\overline{x^2} = \frac{1}{n}\sum_{i=1}^n x^2$
$\bar{y} = \frac{1}{n}\sum_{i=1}^n y$ (aritm. Mittelwert)	$\overline{y^2} = \frac{1}{n}\sum_{i=1}^n y^2$
$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\overline{x^2} - \bar{x}^2}$ (Standardabweichung)	$s_{xkorr} = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$ (korr. Standardabweichung)
$s_y = \sqrt{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2}\ \sqrt{\bar{y} - y^2}$ (Standardabweichung)	$s_{ykorr} = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2}$ (korr. Standardabweichung)
$s_{xy} = \overline{xy} - (\bar{x} \cdot \bar{y}) = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ (Kovarianz)	$s_{xykorr} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ (korr. Kovarianz)
$\overline{xy} = \frac{1}{n}\sum_{i=1}^n x_i \cdot y_i$	$\overline{x^2} = \frac{1}{n}\sum x_i \cdot x_i$ $\overline{y^2} = \frac{1}{n}\sum y_i \cdot y_i$

Korrelation: immer $-1 \leq r_{xy} \leq 1$	
$r_{xy} > 1$	**positive** Korrelation
$r_{xy} < 1$	**negative** Korrelation
$r_{xy} = 1$	Punkte genau auf Gerade mit **positiver** Steigung
$r_{xy} = -1$	Punkte genau auf Gerade mit **negativer** Steigung
$r_{xy}\ nah\ an\ \pm 1$	Punkte streuen eng um Gerade
r_{xy}	kein linearer Zusammenhang

Spearman Rangkorrelation misst Stärke/Richtung des monotonen Zusammenhangs

$$r_{sp} = \frac{\sum_{i=1}^{n}\left(rg(x_i) - \frac{n+1}{2}\right)\left(rg(y_i) - \frac{n+1}{2}\right)}{\sqrt{\sum_{i=1}^{n}\left(rg(x_i) - \frac{n+1}{2}\right)^2}\ \sqrt{\sum_{i=1}^{n}\left(rg(y_i) - \frac{n+1}{2}\right)^2}}$$

$$r_{sp} = 1 - \frac{6 \cdot \sum_{i=1}^{n} {d_i}^2}{n \cdot (n^2 - 1)}\ mit\ d_i = rg(x_i) - rg(y_i)$$

➔ 2.te nur wenn keine Bindungen!!

Kovarianz Rangliste:

$$\frac{1}{n}\sum_{i=1}^{n}(rg(x_i) - \overline{rg(x)})(rg(y_i) - \overline{rg(y)})$$

*insert absolute Häufigkeits, PDF, PMF, etc. examples

KOMBINATORIK

	Ohne Wiederholungen/ ohne zurücklegen	Mit Wiederholungen/ mit zurücklegen
Auswahl $\boldsymbol{k = n}$ Reihenfolge wichtig!	Permutation, ohne Wiederholung $z = n!$ *"ABC anordnen"*	Permutation mit Wiederholung $z = \frac{n!}{n_1! \cdot \ldots \cdot n_s!}$ *"ANNA anordnen"*
Auswahl k Reihenfolge wichtig! n: wird gezogen k: Anzahl Ziehungen	geordnete Stichprobe ohne zurücklegen Es gilt $k < n$ $z = \frac{n!}{(n-k)!}$ *"Nummer aus Schüssel behalten"*	geordnete Stichprobe mit zurücklegen k beliebig $z = n^k$ *"aus Schüssel & zurücklegen"*
Auswahl k Reihenfolge unwichtig!	ungeordnete Stichproben ohne zurücklegen Es gilt $k < n$ $z = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ *"3 aus 6"* Binomialkoeffizient	ungeordnete Stichprobe mit zurücklegen k beliebig $z = \binom{n+k-1}{k} = \frac{(n-1+k)!}{k!(n-1)!}$ *"3 aus 6"*

ELEMENTARE WAHRSCHEINLICHKEITSRECHNUNG

Ω: Menge aller möglichen Ereignisse: Ergebnisraum

$|\Omega|$: Mächtigkeit

Ereignis: Teilmenge $A\ von\ \Omega$, z.B. alle geraden Zahlen $A = \{2,4,6,..\}$

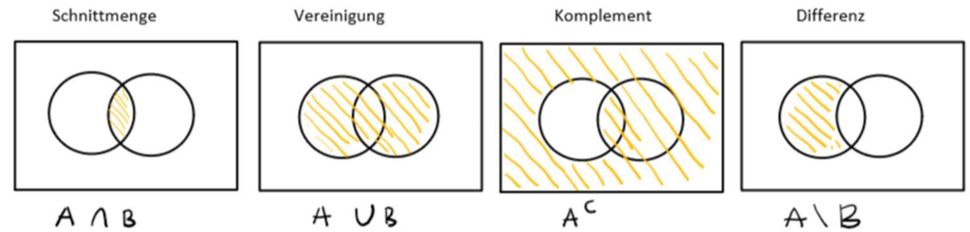
Elementarereignis $= \{\omega_i\}, \{\omega_2\}$ enthalten nur ein Ergebnis

Ereignisraum: Menge aller Teilmengen von Ω (Potenzmenge von Ω) → $2^{|\Omega|}$

Zähldichte: PDF $\rho: \Omega \to [0,1]$ → mit welcher Wahrscheinlichkeit Ergebnisse auftreten
Wahrscheinlichkeitsmass P ist eine Funktion $P: 2^{\Omega} \to [0,1]$ mit welcher Wahrscheinlichkeit das Ergebnis, das durch die Menge M beschrieben wird, eintritt.
Gegenereignis von $A = A^C\ mit\ A \cup A^C = \Omega$

Sicheres Ereignis $A = \Omega$

Unmögliches Ereignis $A = \{\}\ (leere\ Menge)$



Laplace Wahrscheinlichkeit

$$P(M) = \frac{|M|}{|\Omega|} = \frac{Anzahl\ \text{günstige}\ \text{Fälle}}{Anzahl\ \text{mögliche}\ \text{Fälle}}$$

Frequentistische Wahrscheinlichkeit

$$P(A) = v_A$$

Axiome von Kolgarov

Axiom 1:	$0 \leq P(A) \leq 1$
Axiom 2: sicheres Ereignis	$P(\Omega) = 1$
Axiom 3: Vereinigung	Falls $A \cap B = \{\}$ dann gilt $P(A \cup B) = P(A) + P(B)$
Unmögliches Ereignis	$P(\{\}) = 0$
Komplementäres Ereignis	$P(\Omega \backslash A) = P(\bar{A}) = P(A^C) = 1 - P(A)$
Sigma Additivität	$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots$ falls disjunkt

diskrete Zufallsvariabel: endlich, unendlich abzählbarer Wert

stetige Zufallsvariabel: x-beliebiger Wert von Intervall

Wahrscheinlichkeitsfunktion (PDF): $p_i = f(x_i) = P(X = x_i)$
$\sum_i p_i = 1$, wenn alle p_i gleich sind → diskrete Gleichverteilung
kumulative Verteilungsfunktion CDF: $F(x) = P(X \leq x)$

Erwartungswert: $E(x) = \mu = \sum x_i \cdot P(X = x) = \sum x_i \cdot p(x_i) = \sum x_i \cdot p_i$
Varianz: $Var(x) = E[(X - E[X]^2)] = \sum_{i=1}^{n}(x_i - E[X])^2\, p_i$

$$Var(X) = E[x^2] - (E[x])^2$$

$$Var(X) = \left(\sum_{i=1}^{n} {x_i}^2 p_i\right) - (E[x])^2$$

Ereignisraum

- Wahrscheinlichkeit am Ende des Pfades (Blatt) →Multiplikation
- von mehreren Pfaden beschriebene Wahrscheinlichkeit → Addition

bedingte Wahrscheinlichkeit → gegeben, dass

Prävalenz	$\boldsymbol{P(k^+) \to 1 - (Pk^+) \to P(k^-)}$
Sensitivität	$P(\boldsymbol{A}^+\|k^+) \to P(A^-\|P^-)$
Spezifität	$P(\boldsymbol{A}^-\|k^-) \to P(A^+\|k^-)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, B\ gegeben$$

SATZ VON BAYES

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

SATZ DER TOTALEN WAHRSCHEINLICHKEIT

$$P(A) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)$$

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{\left(P(A|B_i) \cdot P(B_i)\right)}{\sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)}$$

$$\boldsymbol{P(B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})}$$

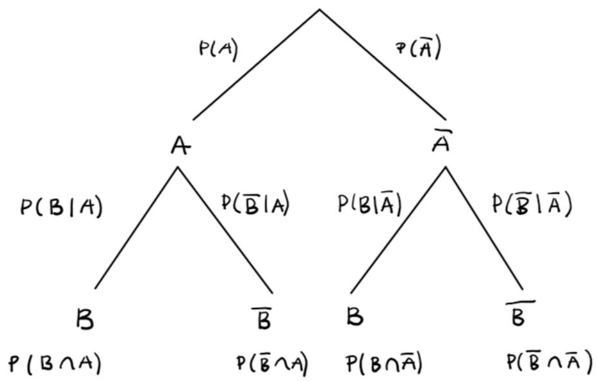
MULTIPLIKATIONSSATZ

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

Stochastische Unabhängigkeit:

$$P(A|B) = P(A)\ mit\ P(B) \neq 0$$

$P(A \cap B) = P(A) \cdot P(B)$



SPEZIELLE VERTEILUNGEN – DISKRETE VERTEILUNGEN

UNIFORMVERTEILUNG

Ergebnisraum: $\Omega = \{a, a+a, \ldots, b-1, b\}(endlich)$

Erwartungswert: $E[X] = \frac{a+b}{2}, x \sim Unif(\{a,b\})$

Varianz: $Var[X] = \frac{(b-a+2)(b-a)}{12} = \frac{(b-a+1)^2-1}{12}$

BERNOULLIVERTEILUNG

- Zufallsvariabel kann 2 Werte annehmen $P(X=1) = p$ oder $P(X=0) = 1-p$

$P = P(X=1)$ Erfolgswahrscheinlichkeit

Ergebnisraum: $\Omega = \{0,1\}, X \sim Bernoulli(p)$

Erwartungwert: $E[X] = 0 \cdot (1-p) + 1 \cdot p = p$

$$E[X^2] = 1^2 \cdot P(X=1) = 1 \cdot p = p$$

Varianz: $Var[X] = E[X^2] - E[X]^2 = (0-p)^2(1-p) + (1+p)^2 \cdot p = p(1-p) = p \cdot q$

BINOMINALVERTEILUNG

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, X \sim B(n,p)$$

Ergebnisraum: $\Omega = \{0,1,2,\ldots,n\}(endlich)$

Erwartungswert: $\mu = E[X] = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} p = n \cdot p$

Varianz: $\sigma^2 = V[X] = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(x_i) = n \cdot p(1-p)$

Standardabweichung: $\sigma = S(X) = \sqrt{npq}$

Symmetrie für $p = 0.5$ bei wachsenden $p \neq 0.5$ immer symmetrisch falls $n \cdot p(1-p) > 10$, gilt $Bin(n,p)$ als symmetrisch

POISSONVERTEILUNG

$$P(X=x) = \frac{\lambda^k \exp(-\lambda)}{x!}, X \sim Pois(\lambda)$$

Ergebnisraum: $\Omega = \{0,1,2,\ldots\}(abzählbar, unendlich)$

Erwartungswert: $\lambda = E[X]$ für kleine λ → stark rechtsschief

Varianz: $Var[X] = \lambda$ je grosser λ → symmetrie bei $\lambda > 10$

Standardabweichung:

HYPERGEOMETRISCHE VERTEILUNG

$$P(X=k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}, X \sim H(N,M,n)$$

N: Objekte Total

M: Merkmale Total

$N - M$: Objekte anderer Sorte

n: gezogene Objekte ohne Zurücklegen

Ergebnisraum: $\Omega = \{0,1,\ldots,N\}\ (endlich)$

Erwartungwert: $\mu = E[X] = n \cdot \frac{M}{N}$

Varianz: $\sigma^2 = V(x) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)$

Standardabweichung: $\sigma = S(X) = \sqrt{n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)} = \sqrt{V(X)}$

SPEZIELLE VERTEILUNGEN – STETIGE VERTEILUNGEN

STETIGE WAHRSCHEINLICHKEITSFUNKTION PDF

$f(x) \geq 0$ für alle $x \in \mathbb{R}$

f ist integrierbar

$$\int_{-\infty}^{\infty} f(u)\, du = 1$$

CDF

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$$

KENNZAHLEN STETIGER ZUFALLSVARIABELN

Erwartungswert: $\mu = E[X] = \int_{-\infty}^{\infty} f(x) \cdot x dx$

Varianz: $\sigma^2 = V(x) = \int_{-\infty}^{\infty} f(x) \cdot (x - E[X])^2\, dx$

Standardabweichung: $\sigma = \sqrt{V(x)}$

$$P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x) dx$$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a) = 1 - \int_{-\infty}^{a} f(x) dx = \int_{a}^{\infty} f(x) dx$$

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_{a}^{b} f(x) dx$$

	diskret	stetig
Wertebereich	abzählbar	beliebig
Wahrscheinlichkeitsfunktion	$p_i = P(X = x_i)$ $\sum_{i=1}^{n} p_i = 1$	$f(x) \geq 0, \int_{-\infty}^{\infty} f(z) dz = 1$ stetig
kumulative Verteilfunktion $F(x) = P(x \leq \lambda)$ Kennzahlen	$F(x) = \sum p_i$ $\mu = \sum x_i p_i$ $\sigma^2 = \sum (x_i - \mu)^2$	$F(x) = \int_{-\infty}^{x} f(z) dz$ $\mu = \int_{-\infty}^{\infty} x f(x) dx$ $\sigma^2 = \int_{-\infty}^{\infty} (\lambda - \mu)^2 f(x) dx$

STETIGE UNIFORMVERTEILUNG

Erwartungswert: $E[X] = \frac{a+b}{2}, X \sim Unif([a,b])$

Varianz: $Var(x) = \frac{(b-a)^2}{12}$

$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$	$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$

DICHTEFUNKTION NORMALVERTEILUNG PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right\} f\ddot{u}r\ X \in (-\infty, \infty)$$

$$X \sim N(\mu, \sigma^2)$$

- Dichtefunktion symmetrisch um $X = \mu$
- nähert sich asymptotisch der 0 an

KUMULATIVE VERTEILFUNKTION CDF

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2} \cdot \frac{(2-\mu)^2}{\sigma^2}\right\}$$

Standardnormalverteilung

$$\emptyset u = Fu$$

$$\emptyset(-u) = 1 - \emptyset(u)$$

beliebige Normalverteilung muss standardisiert werden

$$u = \frac{X - \mu}{\sigma}$$

innerhalb $[\mu - \sigma, \mu + \sigma] = 68\%$

innerhalb $[\mu - 2\sigma, \mu + 2\sigma] = 95\%$

innerhalb $[\mu - 3\sigma, \mu + 3\sigma] = 99.7\%$

Stetige Verteilungen		
Uniform $X \sim Unif([a,b])$	$\begin{cases} 0 & a < 0 \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & b > 0 \end{cases}$	$\mu = \frac{a+b}{2}$ $\sigma^2 = \frac{(b-a)^2}{12}$
Normal $X \sim \mathcal{N}(\mu, \sigma^2)$	$f(x) = \varphi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right\}$	$E[x] = \mu$ $Var(x) = \sigma^2$

Summe von 2 Zufallsvariabeln x, y

normalverteilt: $S \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

poissonverteilt: $S \sim Pois(\lambda_x + \lambda_y)$

binominalverteilt: $S \sim Bin(nx, ny, p)$

$$E[S] = E[X] + E[Y]$$

$$E[aX + b] = aE[X] + b$$

$$Var[S] = Var[x] + Var[y] + 2cov(xy)$$

cov: Kovarianz

Zentraler Grenzwertsatz

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow S_n \sim N(n\mu, n\sigma^2)\ mit\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Binominalverteilung: $\mu = n \cdot p$ $\quad \sigma^2 = npq$ $\quad q = (1-p)$

$P(a \leq x \leq b) = \sum_{x=a}^{b} P(X = x) \approx \emptyset_{\mu,\sigma^2}(b + \frac{1}{2}) - \emptyset_{\mu,\sigma^2}(a - \frac{1}{2})$ $\quad npq > 9$

Poissonverteilung: $\mu = \lambda$ $\quad \sigma^2 = \lambda$

$P(a \leq x \leq b) = \sum_{x=a}^{b} P(X = x) \approx \emptyset_{\mu,\sigma^2}(b + \frac{1}{2}) - \emptyset_{\mu,\sigma^2}(a - \frac{1}{2})$ $\quad \lambda > 9$

SUMMARY DISKRETE UND STETIGE VERTEILUNGEN

Diskrete und stetige Verteilungen

	diskrete Zufallsvariablen	stetige Zufallsvariablen
Dichtefunktion / PMF/PDF	$f(x) = P(X = x)$	$f(x) = F'(x) \neq P(X = x)$
Kumulative Verteilungsfunktion / CDF	$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$
Wahrscheinlichkeiten	$P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$ $P(a < X \leq b) = \sum_{a < x \leq b} f(x)$ $P(a < X < b) = \sum_{a < x < b} f(x)$ $P(a < X) = 1 - F(a)$	$\left.\begin{matrix} P(a \leq X \leq b) \\ P(a < X \leq b) \\ P(a < X < b) \end{matrix}\right\} = \int_a^b f(x)dx$ $P(a < X) = 1 - F(a)$
Graphische Darstellung von f	Stabdiagramm	Graph
Erwartungwert	$E(X) = \sum_{x \in \mathbb{R}} f(x) \cdot x$	$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x\, dx$
Varianz	$V(X) = \sum_{x \in \mathbb{R}} f(x) \cdot (x - E(X))^2$	$V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2\, dx$

Satz

Für diskrete und stetige Zufallsvariablen X und Y gelten die folgenden Regeln:

(1) Linearität des Erwartungswertes:

$$E(X + Y) = E(X) + E(Y) \text{ und } E(\alpha X) = \alpha E(X) \text{ mit } \alpha \in \mathbb{R}.$$

(2) Verschiebungssatz für die Varianz:

$$V(X) = E(X^2) - (E(X))^2$$

(3) $V(\alpha X + \beta) = \alpha^2 \cdot V(X)$ mit $\alpha, \beta \in \mathbb{R}$.

(4) Sind X und Y stochastisch unabhängig, so gilt:

$$V(X + Y) = V(X) + V(Y)$$

LINEARE REGRESSION

$$y = h(x) = d + m \cdot x$$

d: Achsenabschnitt m: Steigung x: erklärende Variabel

$$y_i = d + mx_i + \varepsilon_i$$

y_i: Zielvariabel d: Achsenabschnitt m: Steigung x: erklärende Variabel

ε_i: zufälliger Fehler Rest

METHODE DER KLEINSTEN QUADRATE

$$Q(d,m) = \sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} \left(y_i - (d + mx_i)\right)^2 = \underbrace{\sum_{i=1}^{n} (y_1 - \hat{y})^2}_{Summe\ der\ Residuen\ Quadrate}$$

$$\hat{y} = d + mx$$

$$\varepsilon_i = y_i - \hat{y}_i\ (Residuen)$$

$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{S_{xy}}{S_x^2} = \frac{Kovarianz}{Varianz}$$

$$d = \bar{y} - m\bar{x}$$

$$S_y^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}_{Totale\ Varianz} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}_{Fehlervarianz} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}_{erklärte\ Varianz} = S_\varepsilon^2 + S_{\hat{y}}^2$$

Residuenvarianz: $S_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - mx_i - d)^2 = S_y^2 - S_{\hat{y}}^2$

korrigierte Gesamtvarianz: $S_y^2 = \frac{1}{n-1}\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{Totale\ Varianz}$

korrigierte erklärte Varianz: $\underbrace{S_{\hat{y}}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{erklärte\ Varianz}$

Bestimmtsheitsmass: $R^2 = \frac{S_{\hat{y}}^2}{S_y^2} = \frac{S_{xy}^2}{S_x^2 \cdot S_y^2} = r_{xy}^2$

TRANSFORMATIONEN

muss wieder zurücktransformiert werden → $\log(m) \rightarrow m = e^{\cdots}$

$y = q \cdot x^m$	$\log(y) = \log(q) + m \cdot \log(x)$
$y = q \cdot m^x$	$\log(y) = \log(q) + x \cdot \log(m)$
$y = q \cdot e^{mx}$	$\log(y) = \log(q) + m \cdot x$
$y = q \cdot e^{q\frac{m}{x}}$	$\ln(y) = q - m \cdot u,\ u = \frac{1}{x}$
$y = \frac{1}{q + m \cdot x}$	$v = q + m \cdot x,\ v = \frac{1}{y}$
$y = q + m \cdot \ln(x)$	$y = q + m \cdot u,\ u = \ln(x)$
$y = \frac{1}{q \cdot m^x}$	$\log\left(\frac{1}{y}\right) = \log(q) + \log(m) \cdot x$

SCHLIESSENDE STATISTIK

SCHÄTZFUNKTIONEN

$$\mu = arithmetisches\ Mittel = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \rightarrow Schätzung$$

$$\sigma = empirische\ Standardabweichung$$

$$p = relative\ Häufigkeit$$

Erwartungstreue → μ $E(\Theta) = \theta$

Effizienz → σ effizienter wenn: $V(\Theta_1) < V(\Theta_2)$

Konsistenz → je grosser die Stichprobe, desto genauer der Schätzer

konsistent wenn: $E(\Theta) \rightarrow \theta\ und\ V(\Theta) \rightarrow \theta, für\ n \rightarrow \infty$

Schätzfunktionen für wichtige Parameter

	Schätzfunktion	**Schätzwert**
Erwartungswert Spezialfall: Anteilswert einer Bernoulli-Verteilung	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$	$\hat{\mu} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ $\hat{p} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{\text{Anzahl 1en}}{n}$
Varianz	$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
Standardabweichung	$S = \sqrt{S^2}$	$\hat{\sigma} = s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$

Satz
(1) $\bar{X}$ und S^2 sind erwartungstreu und konsistent.
(2) S ist konsistent, aber nicht erwartungstreu.

Schätzer/ Verteilung	Kenngrössen	Schätzer für Param.
Binominal $X \sim B(n,p)$	$E[X] = np$ $Var(X) = np(1-p)$	$\hat{p} = \bar{x}$
Poisson $X \sim Pois(\lambda)$	$E[X] = Var(X) = \lambda$	$\hat{\lambda} = \bar{x}$
Uniform $X \sim Unif([a,b])$	$\mu = \frac{a+b}{2}$ $\sigma^2 = \frac{(b-a)^2}{12}$	$\hat{a} = \bar{x} - Sd(X) \cdot \sqrt{3}$ $\hat{b} = \bar{x} - Sd(X) \cdot \sqrt{3}$ Sd: Standardabweichung
Normal $X \sim \mathcal{N}(\mu, \sigma^2)$	$E[x] = \mu$ $Var(x) = \sigma^2$	$\hat{\mu} = \hat{x}$ $\hat{\sigma} = Sd(x)$

Anpassungen von Verteilungen

Theoretische kumulative Wahrscheinlichkeitsfunktion

$$F(x) = P(X \leq x)(theoretische\ CDF)$$

Empirische kum. Wahrscheinlichkeitsfunktion

$$\hat{F}n = Anzahl\ Beobachtungen$$

Konfidenzintervall Normalverteilung

$$[\bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\frac{a}{2}}, \bar{X}_n - \frac{\sigma}{\sqrt{n}} \underbrace{q_{1-\frac{a}{2}}}_{wird\ aus\ der\ Tabelle\ bestimmt}]$$

VERTRAUENSINTERVALLE

	(1) Verteilung der Grundgesamtheit	(2) Param.	(3) Schätzfunktionen	(4) zugehörige standardisierte Zufallsvariable	(5) Verteilung und benötigte Quantile	(6) Zufallsvariablen für Intervallgrenzen
1	Normalverteilung (Varianz σ^2 bekannt)	μ	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$	$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	**Standardnormalverteilung** (Tabelle 2) $c = u_p$ mit $p = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \frac{\sigma}{\sqrt{n}}$ $\Theta_o = \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}$
2	Normalverteilung (Varianz σ^2 unbekannt und $n \leq 30$; sonst Fall 1 mit s als Schätzwert für σ)	μ	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	***t*-Verteilung** (Tabelle 4) mit $f = n - 1$ $c = t_{(p;f)}$ mit $p = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \frac{S}{\sqrt{n}}$ $\Theta_o = \bar{X} + c \cdot \frac{S}{\sqrt{n}}$
3	Normalverteilung	σ^2	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$Z = (n-1)\frac{S^2}{\sigma^2}$	**Chi-Quadrat-Verteilung** (Tabelle 3) mit $f = n - 1$ $c_1 = z_{(p_1;f)}$ mit $p_1 = \frac{1-\gamma}{2}$ $c_2 = z_{(p_2;f)}$ mit $p_2 = \frac{1+\gamma}{2}$	$\Theta_u = \frac{(n-1) \cdot S^2}{c_2}$ $\Theta_o = \frac{(n-1) \cdot S^2}{c_1}$
4	Bernoulli-Verteilung Anteilsschätzung (mit $n\hat{p}(1-\hat{p}) > 9$)	p	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ X_i 0/1-wertig mit $P(X_i = 1) = p$	$U = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}$	**Standardnormalverteilung** (näherungsweise), Tabelle 2 $c = u_q$ mit $q = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \sqrt{\frac{\bar{X} \cdot (1 - \bar{X})}{n}}$ $\Theta_o = \bar{X} + c \cdot \sqrt{\frac{\bar{X} \cdot (1 - \bar{X})}{n}}$
5	beliebig mit $n > 30$	μ, σ^2	wie im Fall 1 (gegebenenfalls mit s als Schätzwert für σ) bzw. im Fall 3			