Beschreibende Statistik

Ω: *Grundgesamtheit*

ω: *Element der Grundgesamtheit*

$X: \Omega \to M$: *Merkmal*

$X: (\omega) = x$: *Ausprägung des Merkmals*

Modalwert(e) $\boldsymbol{x_{mod}}$		Am häufigsten auftretende Ausprägung
Mittelwert $\boldsymbol{\bar{x}}$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	
Median $\boldsymbol{x_{0.5}}$	$x_{0.5} = \begin{cases} x_{\frac{n+1}{2}} & , n\ ungerade \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & , n\ gerade \end{cases}$	Liegt in der Mitte der sortierten Daten x_i
Spannweite	$max\, x_i - min\, x_i$	Distanz zw. größtem und kleinstem Messwert
Stichprobenvarianz $\boldsymbol{s^2}$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $= \frac{n}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$	Gemittelte Summe der quadratischen Abweichungen vom Mittelwert
Stichproben-standardabweichung s	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{s^2}$	Streuungsmaß mit gleicher Einheit wie beobachete Daten x_i
p-Quantil $\boldsymbol{x_p}$	$x_p = \begin{cases} x_{floor(np)+1}, & n \cdot p \notin \mathbb{N} \\ \frac{1}{2}\left(x_{np} + x_{np+1}\right), & n \cdot p \in \mathbb{N} \end{cases}$	Teilt sortierte Daten im Verhältnis p : (1 - p)
Empirische Kovarianz $\boldsymbol{s_{xy}}$	$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ $= \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}\right)$	Für $s_{xy} > 0$ hat Punktewolke steigende, für $s_{xy} <$ 0 fallende Tendenz
Empirischer Korrelationskoeffizient r	$r = \frac{s_{xy}}{s_x \cdot s_y}$	Näherungsweise linearer Zusammenhand zw. x und y falls $\lvert r \rvert \approx 1$
Regressionsgerade	$y = mx + t$ mit $m = r \cdot \frac{s_y}{s_x}$ und $t = \bar{y} - m \cdot \bar{x}$	

Elementare Wahrscheinlichkeitsrechnung

$\Omega: Ergebnisraum, \qquad Menge\ aller\ möglichen\ Ergebnisse\ eines\ Experiments$

$\omega \in \Omega: Elementarereignis, \qquad Einzelnes\ Element\ von\ \Omega$

$E \cup F: Ereignis\ E\ oder\ Ereignis\ F\ treten\ ein\ (mindestens\ eins)$

$E \cap F: Ereignis\ E\ und\ Ereignis\ F\ treten\ ein$

$E \cap F = \emptyset: Paarweise\ disjunkte\ Ereignisse$

De Morgan'sche Regeln	$\overline{\bigcup_{\imath=1}^{n} E_\imath} = \bigcap_{i=1}^{n} \bar{E}_i$ $\overline{\bigcap_{\imath=1}^{n} E_\imath} = \bigcup_{i=1}^{n} \bar{E}_i$	
Axiome von Kolmogorov	1. $0 \leq P(E) \leq 1$ 2. $P(\Omega) = 1$ 3. $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=n}^{\infty} P(E_i)$, falls $E_i \cap E_j = \emptyset\ für\ i \neq j$	
Folgen aus den Axiomen	$P(\bar{E}) = 1 - P(E)$ $P(E \cup F) = P(E) + P(F) - P(E \cap F)$	
Laplace Experiment	$P(E) = \frac{Anz\ der\ für\ E\ günstigen\ Ergebnisse}{Anz\ der\ möglichen\ Ergebnisse} = \frac{\lvert E \rvert}{n}$	Zufallsexperiment mit n gleichwahrscheinlichen Elementarereignissen
Bedingte Wahrscheinlichkeit	$P(E \mid F) = P_F(E) = \frac{\lvert E \rvert}{\lvert F \rvert} = \frac{P(E \cap F)}{P(F)}$	Die Wahrscheinlichkeit für das Eintreten von E, wenn F eingetreten ist.
Regeln für bedingte Wahrscheinlichkeit	Für $E, F \neq \emptyset$ 1. $P(E \cap F) = P(E \mid F) \cdot P(F)$ 2. $P(E \cap F) = P(F \mid E) \cdot P(E)$	
Satz der totalen Wahrscheinlichkeit	$\Omega = \bigcup_{i=1}^{n} E_i\ mit\ E_i \cap E_j = \emptyset, dann\ gilt:$ $P(F) = \sum_{i=1}^{n} P(F \cap E_i) = \sum_{i=1}^{n} P(F \mid E_i) \cdot P(E_i)$	Alle E müssen disjunkt sein. Allgemein die Wahrscheinlichkeit, dass F Eintritt. Man nimmt nicht an, dass E_i eingetreten ist.
Vierfeldertafel	$Spezialfall: \Omega = E \cup \bar{E}$ $P(F) = P(F \cap E) + P(F \cap \bar{E})$ $P(\bar{F}) = P(\bar{F} \cap E) + P(\bar{F} \cap \bar{E})$	
Formel von Bayes	$\mathrm{P}(E_k \mid F) = \frac{P(F \mid E_k) \cdot P(E_k)}{\sum_{i=1}^{n} P(F \mid E_i) \cdot P(E_i)}$	Hilfreich, wenn man $\mathrm{P}(F \mid E_i)$ kennt, nicht aber $\mathrm{P}(E_k \mid F)$
Unabhängigkeit	Gilt, wenn $P(E \mid F) = P(E)\ bzw. P(E \cap F) = P(E) \cdot P(F)$ Falls E, F unabhängig, gilt auch $\bar{E}, F$ bzw. $\bar{E}, \bar{F}$ unabh	

Kombinatorik

Ermittlung der Mächtigkeit von Ereignissen

Allgemeines Zählprinzip	Anzahl der Möglichkeiten, für ein k-stufiges Zufallsexperiment mit n_i Varianten im i-ten Schritt: $n_1 \cdot n_2 \dots n_k$	Baum, der je nach Ebene unterschiedlich viele Kinder haben kann
Anzahl der Permutationen einer n-elementigen Menge	n-maliges Ziehen **ohne Zurücklegen mit Beachtung der Reihenfolge** - für n unterscheidbare Elemente: $n! = n \cdot (n-1) \dots \cdot 2 \cdot 1$ - für k Klassen mit je n_i nicht unterscheidbaren Elementen: $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$	
k-maliges ziehen aus einer n-elementigen Menge	(siehe Tabelle unten)	

	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge
ohne Zurücklegen $k \leq n$	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)!k!}$
mit Zurücklegen $k > n$ möglich	n^k	$\binom{n+k-1}{k}$

Zufallsvariablen

Abbildung $X: \Omega \rightarrow \mathbb{R},\ \omega \rightarrow X(\omega) = x$, heißt Zufallsvariable

x: Realisation der ZV X

Diskrete ZV: $X(\Omega) = \{x_1, x_2, \dots, x_n\}(n \in \mathbb{N})$

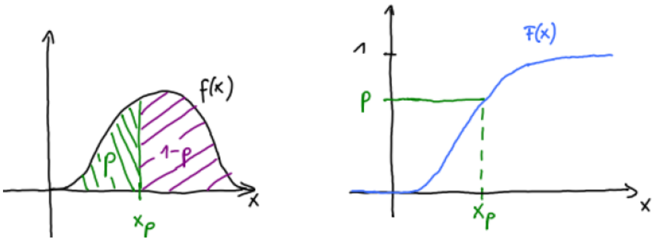
Stetige ZV: $X(\Omega) \subseteq \mathbb{R}$

Eindimensionale ZV: $X: \Omega \rightarrow \mathbb{R}$

Mehrdimensionale ZV: $X: \Omega \rightarrow \mathbb{R}^p$

Verteilungsfunktion	$x \in \mathbb{R}, F: \mathbb{R} \rightarrow [0,1]$ $$F(x) = P(X \leq x)$$ Eigenschaften: - $0 \leq F(x) \leq 1$ - $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$ - rechtsseitig stetig, also: $\lim_{x \rightarrow b+} F(x) = F(b)$ - monoton wachsend - $P(X > x) = 1 - F(x)$ - $P(a < X \leq b) = F(b) - F(a)$	
Diskrete Zufallsvariablen	<u>Wahrscheinlichkeitsverteilung:</u> $p(x) = \begin{cases} P(X = x_i), falls\ x = x_i \ \in X(\Omega) \\ 0, \quad sonst \end{cases}$ *<u>Eigenschaften Wahrscheinlichkeitsverteilung:</u>* - $0 \leq p(x) \leq 1$ - $\sum_{i=1}^{n} p(x_i) = 1$ Es gilt: - $F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$ - $F(x)$ist eine rechtsseitig stetige Treppenfunktion mit Sprüngen bei den Realisationen x_i	
Stetige Zufallsvariablen	<u>Wahrscheinlichkeitsdichte:</u> $$P(a < X < b) = \int_a^b f(x)\, dx$$ <u>Eigenschaften Wahrscheinlichkeitsdichte:</u> - $f(x) \geq 0$ - $\int_{-\infty}^{\infty} f(x)\, dx = 1$ Es gilt: - $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \ und\ F'(x) = f(x)$ - $F(x)$ ist stetig und $P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$	Wahrscheinlichkeitsdichte ist die Wahrscheinlichkeit, dass ein Zufallsexperiment ein Ergebnis in einem bestimmten Bereich liefert

Erwartungswert	$E[X] = \mu$ Für diskrete ZV: $$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$$ Für stetige ZV: $$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$ Sei g(X) eine Funktion der ZV X, dann gilt: Für diskrete ZV: $$E[g(X)] = \sum_{i=1}^{n} g(x_i) \cdot p(x_i)$$ Für stetige ZV: $$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$$ <u>Eigenschaften des Erwartungswerts</u>: - $E[b] = b$ - $E[aX + b] = aE[X] + b$ - $E[X_1 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i]$	Der Erwartungswert ist der Schwerpunkt der Verteilung einer Zufallsvariable
Varianz	$$\sigma^2 = Var[X] = E[(X - \mu)^2]$$ <u>Standardabweichung</u>: $$\sigma = \sqrt{V[X]}$$ <u>Verschiebungssatz</u>: $$Var[X] = E[X^2] - (E[X])^2$$ <u>Eigenschaften der Varianz</u>: - $Var[b] = 0$ - $Var[aX + b] = a^2 Var[X]$ - $Var[X_1 + \cdots + X_n] = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov[X_i, X_j]$ - $Var[X_1 + X_2] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2]$ - Falls X_i, X_j *paarweise unabhängig*: $$Var[X_1 + \cdots + X_n] = \sum_{i=1}^{n} Var[X_i]$$	Die Varianz ist ein quadratisches Streuungsmaß einer ZV X
Kovarianz	$$Cov[X, Y] = E[(X - E[X])\,(Y - E[Y])]$$ Wenn X, Y stochastisch unabhängig ⇒ $Cov[X, Y] = 0$ <u>Verschiebungssatz</u>: $$Cov[X, Y] = E[XY] - E[X] \cdot E[Y]$$ <u>Eigenschaften der Kovarianz</u>: - $Cov[X, Y] = Cov[Y, X]$ - $Cov[X, X] = Var[X]$ - $Cov[aX, Y] = aCov[X, Y]$	Die Kovarianz beschreibt die Abhängigkeit zweier ZV X und Y. Je stärker diese korrelieren, desto größer die Kovarianz.

Quantile	Das p-Quantil ist der kleinste Wert $x_p \in \mathbb{R}$ für den gilt: $$F(x_p) \geq p$$  Abbildung: p-Quantil einer stetigen ZV mit streng monoton wachsendem $F(x)$: $$x_p = F^{-1}(p)$$	
Chebyshev-Ungleichung	$X: ZV$ $\mu = E[X]$ $\sigma^2: Var[X]$ Es gilt für jedes beliebige $k \in \mathbb{R}^+ \setminus \{0\}$: $$P(\lvert X - \mu \rvert \geq k \cdot \sigma) \leq \frac{1}{k^2}$$	
Schwaches Gesetz der großen Zahlen	Mit Chebyshev und $\mu = E[X_i] \;\; (i = 1, \dots, n) \;\; \Rightarrow \;\; \mu = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = E[\bar{X}]$ folgt: $X_1, X_2, \dots$ ist eine Folge von unabhängigen, identisch verteilten ZV mit EW $E[X_i] = \mu$ und Varianz $Var[X_i] = \sigma^2$. Dann gilt für ein beliebig kleines $\epsilon > 0$: $$P(\lvert \bar{X} - \mu \rvert > \epsilon) \to 0 \quad für\; n \to \infty$$ d.h. der MW $\bar{X}$ konvergiert stochastisch gegen den EW μ.	

Diskrete Verteilungen

P(X=x): Wahrscheinlichkeit | **E[X]**: Erwartungswert

F(x): Verteilungsfunktion | **Var[X]**: Varianz

Bernoulliverteilung	$P(X=1)=p, \quad P(X=0)=1-p$ $E[X]=p$	Indikatorvariable mit den Werten 1 bei Erfolg und 0 bei Misserfolg
Binomialverteilung	$P(X=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}, \quad k\in\{0,1,\dots,n\}$ $E[x]=np$ $Var[X]=np(1-p)$ $dbinom(k,n,p)=P(X=k)$ $pbinom(k,n,p)=F(k)$ *q-Quantil*: $qbinom(q,n,p)$ *k binomialverteilte Zufallszahlen*: $rbinom(k,n,p)$	Anzahl der Erfolge bei n-maligem Ziehen mit Zurücklegen
Hypergeometrische Verteilung	$P(X=k)=\frac{\binom{M}{k}\cdot\binom{N}{n-k}}{\binom{M+N}{n}}, k\in\{0,1,\dots,\min\{n,M\}\}$ $E[X]=n\cdot\frac{M}{M+N}$ $Var[X]=n\cdot\frac{M}{M+N}\left(1-\frac{M}{M+N}\right)\frac{M+N-n}{M+N-1}$ $dhyper(k,M,N,n)=P(X=k)$ $phyper(k,M,N,n)=F(k)$	Anzahl der Erfolge beim n-maligen Ziehen ohne Zurücklegen aus einer Menge mit M Elementen, die Erfolg bedeuten und N Elementen, die Misserfolg bedeuten
Poisson-Verteilung	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, k\in\{0,1,\dots\} mit \sum_{k=0}^{\infty}P(X=k)=1$ $E[X]=\lambda$ $Var[X]=\lambda$ $dpois(k,\lambda)=P(X=k)$ $ppois(k,\lambda)=F(k)$	„Verteilung der seltenen Ereignisse". Häufigkeit punktförmiger Ereignisse in einem Kontinuum. Die durchschnittlich zu erwartende Anzahl der Erfolge λ pro Maßeinheit (i.a. Zeiteinheit) sei bekannt.
Gleichverteilung	$P(X=x_k)=\frac{1}{n}$ $E[X]=\frac{1}{n}\sum_{k=1}^{n}x_k=\bar{x}$ $Var[x]=\frac{1}{n}\sum_{k=1}^{n}x_k^2-\bar{x}^2$ $sample(1{:}N,n){:}\, n\ Zufallszahlen\ zw\ 1\ und\ N$	Alle Werte $\{x_1,\dots,x_n\}$ einer ZV X sind gleichwahrscheinlich.

Stetige Verteilungen

f(x): Dichte | **E[X]=μ**: Erwartungswert | **σ**: Standardabweichung=$\sqrt{Var[X]}$

F(x): Verteilungsfunktion | **Var[X]**: Varianz

Stetige Gleichverteilung	$f(x) = \frac{1}{b-a}\ für\ x \in [a,b]$ $E[X] = \frac{a+b}{2}$ $Var[X] = \frac{(b-a)^2}{12}$ $dunif(x,a,b) = f(x)$ $punif(x,a,b) = F(x)$ $runif(n):\ n\ Zufallszahlen\ zw\ 0\ und\ 1$	Zufallszahlen aus einem Intervall [a,b]
Normalverteilung	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}$ $E[X] = \mu$ $Var[X] = \sigma^2$ $dnorm(x,\mu,\sigma) = f(x)$ $pnorm(x,\mu,\sigma) = F(x)$ $qnorm(q,\mu,\sigma)$: q-Quantil	Beschreibt viele reale Situationen, insbesondere Grenzverteilung unabhängiger Summen
Standardnormal-verteilung		
Exponentialverteilung	$f(x) = \lambda e^{-\lambda x}\ \ (x \geq 0)$ $F(x) = 1 - e^{-\lambda x}$ $E[X] = \frac{1}{\lambda}$ $Var[X] = \frac{1}{\lambda^2}$ $dexp(x,\lambda) = f(x)$ $pexp(x,\lambda) = F(x)$	Modellierung von Lebensdauern, Wartezeiten. Sei $Y_t \sim P_{\lambda t}$ im Intervall [0,t] von t Zeiteinheiten, dann beschreibt die Exponentialverteilung die Wartezeit X bis zum Eintreten eines Ereignisses
Chiquadrat-Verteilung	$E[X] = n$ $Var[X] = 2n$ $dchisq(x,n) = f(x)$ $pchisq(x,n) = F(x)$	$Z_1, \dots, Z_n$ seien unabhängige, standardnormalverteilte ZV => $X = Z_1^2 + \cdots + Z_n^2$ hat Chiquadratverteilung mit n Freiheitsgraden. Summen unabhängiger, standardnormalverteiler ZV
t-Verteilung	$E[Y] = 0\ für\ n > 1$ $Var[Y] = \frac{n}{n-2}\ für\ n > 2$ $dt(y,n) = f(y)$ $pt(y,n) = F(y)$	$Z \sim N_{0,1}$ und $X \sim x_n^2 \Rightarrow Y = \frac{Z}{\frac{x}{\sqrt{n}}}$ ist t-verteilt mit n Freiheitsgraden Schätz- und Testverfahren bei unbekannter Varianz.