Beschreibende Statistik

Lageparameter

Arithmetisches Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} (x_1 + \dots + a_n)$$

Das Arithmetische Mittel \bar{x} minimiert die Funktion

$$g(t) = \sum_{i=1}^{n} (x_i - t)^2$$

Geometrisches Mittel

$$\bar{x}_{geom} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Median

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} &, ungerade \\ \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) &, gerade \end{cases}$$

Der Median \tilde{x} minimiert die Funktion

$$g(t) = \sum_{i=1}^{n} |x_i - t|$$

${\bf Streungsmaße}$

(empirische) Varianz

$$var = \sigma^2 = s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$
alternativ
$$var = \sigma^2 = \frac{n}{n-1} \cdot (\overline{x^2} - \bar{x}^2)$$

Standardabweichung

$$\sigma = s_n = \sqrt{\sigma^2}$$
$$\sigma = s_n = \sqrt{s_n^2}$$

mittlere absolute Abweichung

$$\frac{1}{n} \sum_{i=1}^{n} |x_i - \tilde{x}| \text{ für Median}$$

 $\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$ für arithmetisches Mittel

Kovarianz und Korrelationskoeffizient

Kovarianz

$$cov(x,y) = S_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$
alternativ

$$cov(x,y) = S_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y})$$

Korrelationskoeffizent

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

Der Korrelationskoeffizent liegt immer zwi-

schen $-1 \le r \le +1$. Je näher r_{xy} bei -1 (negative Korellation/Steigung), oder +1 (positive Steigung/Korrelation) liegt, desto genauer schmiegen sich die Messwerte an eine Gerade an. Bei r_{xy} nahe 0 gibt es keinen linearen Zusammenhang zwischen den Merkmalen.

Regressionsrechnung

Regressionsgerade

Variante 1
$$y = \bar{y} + \frac{S_{xy}}{\sigma_x^2} \cdot (x - \bar{x})$$
Variante 2
$$y = b + a \cdot x$$

$$a = \frac{S_{xy}}{\sigma_x^2} \text{ und } b = \bar{y} - a \cdot \bar{x}$$

Kleinste quadratische Abweichung

Die Parameter a, b, c, \dots werden so gewählt,

$$Q(a, b, c, ...) = \sum_{i=1}^{\text{dass}} (f_{a,b,c,...}(x_i) - y_i)^2$$

minimal ist $f_{a,b,c...}(x_i)$ ist die Funktion dessen Parameter gesucht werden Nullsetzen der partiellen Ableitungen:

$$\frac{\partial}{\partial a}Q(a,b) = 0$$

$$\frac{\partial}{\partial b}Q(a,b) = 0$$

Über die Ableitungen lassen sich die Parameter finden welche die vorgegebene

Funktion am besten annähern

Vergleich ermittelter Kurven

Um Kurven zu vergleichen, einfach die ermittelten Parameter in die Q(a, b, c, ...) Funktion eingeben und Wert berechnen. Je kleiner der Wert desto besser passt die Kurve

Wahrscheinlichkeitstheorie

Wahrscheinlichkeitsräume

Der Wahrscheinlichkeitsbegriff

$$Ergebnismenge = \Omega$$
 Beispiel Würfel $\Omega = \{1, 2, 3, 4, 5, 6\}$ Ein Ereignis ist eine Teilmenge der Ergebnismenge $\varnothing \subseteq \Omega \cong$ unmögliches Ereignis $\Omega \subseteq \Omega \cong$ sicheres Ereignis

 $A = \{1, 2, 3\}$ Ereignis

 $\bar{A} = \{4, 5, 6\}$ Gegenereignis

Elementarereignis

einelementige Teilmenge von Ω Ereignis, eine 3 werfen

$$B = \{3\} P(\{3\}) = \frac{1}{6}$$

Laplace-Versuch

Jedes Elementarereignis ist gleich

wahrscheinlich
$$P(\{\omega_i\}) = \frac{1}{|\Omega|}$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$$

Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit

Wahrscheinlichkeit für A unter der

Bedingung B

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

$$P(\bar{A}|B) = 1 - P(A|B)$$

Formel von Bayes

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

Satz der totalen Wahrscheinlichkeit

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

Viel Felder Tafel				
	A	$ig ar{A}$	\sum	
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	P(B)	
\bar{B}	$P(A \cap \bar{B})$	$P(\bar{A}\cap \bar{B})$	$P(\bar{B})$	
\sum	P(A)	$P(\bar{A})$	1	
ъ.	D. 1 . 1.	1. 0	' 1	

Die Ränder sind immer die Summen der zugehörigen Zeilen oder Spalten

Allgemeine Regeln

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

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\overline{A}) = 1 - P(A)$$

$$P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$$

$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$
Wenn A und B unabhängig, dann gilt

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

Zufallsvariablen

Eine Zufallsvariable ist eine Zuordnungsvorschrift die jedem möglichen Ergebnis eines Zufallsexperiments eine

Größe zuordnet
$$X = k \stackrel{\frown}{=} \{\omega \in \Omega | X(\omega = k) \}$$

$$X = 3 \stackrel{\frown}{=} \{\omega \in \Omega | X(\omega = 3) \}$$

$$X < k \stackrel{\frown}{=} \{\omega \in \Omega | X(\omega < k) \}$$

Diskrete Verteilungen

Binomialverteilung

Mit zurücklegen, Wahrscheinlichkeit für jedes Ereignis gleich

$$X \sim B(n, p)$$

n =: Stichprobenumfang

p =: Wahrscheinlichkeit

(p muss bei Binomial verteilung fest bleiben)

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

$$P(X \le k) = \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1 - p)^{n - i}$$

$$P(X > k) = 1 - P(X \le k)$$
Eingabe Taschenrechner
$$\binom{n}{i} \stackrel{?}{=} n |nCr| |k|$$

Binomialverteilung approximieren

Die Binomialverteilung kann mit der Poisson Verteilung approximiert werden, dann gilt $\lambda = n \cdot p$

Die Binomialverteilung kann auch mit der Normalverteilung approximiert werden,

bedingung ist

$$X \sim B(n, p) \approx N(n \cdot p, n \cdot p \cdot (1 - p))$$

falls gilt
 $n \cdot p \cdot (1 - p) > 9$

Bei der approximation mit der Normalverteilung kann man eine Stetigkeitskorrektur verwenden um ein

besseres Ergebnis zu erhalten

$$P(X \le k) \approx F_N(R+0,5)$$

$$P(X < k) \approx F_N(R - 0, 5)$$

$$P(a \le X \le b) \approx F_N(b+0,5) - F_N(a-0,5)$$

Hypergeometrische Verteilung

Ohne zurücklegen, Wahrscheinlichkeit ändert sich nach jedem Ereignis $X \sim H(N, M, n)$ n =: Stichprobenumfang N =: Gesamtzahl

M =: Anzahl der Elemente mit der Eigenschaft $P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$ $P(X \le k) = \sum_{i=0}^{k} \frac{\binom{M}{i} \cdot \binom{N-M}{n-i}}{\binom{N}{n}}$ $P(X > k) = 1 - P(X \le k)$

Hypergeometrische Vert. approximieren

Die hypergeometrische Verteilung kann mit der **Binomialverteilung** approximiert werden. Dabei muss folgende Bedingung gelten

$$\frac{\text{gelten}}{\frac{n}{N}} < 0,05$$

Poisson Verteilung

Schlüsselwörter sind **Ereignisse pro Zeiteinheit**, zum Beispiel Anrufe innerhalb bestimmter Zeitspanne

$$X \sim Pois(\lambda)$$

$$P(X = k) = \pi_{\lambda}(k) = \frac{\lambda^{k}}{k!} \cdot e^{-\lambda}$$

Geometrische Verteilung

$$X \sim Geom(n, p)$$
$$P(X = n) = (1 - p)^{n-1} \cdot p$$

Beispiel: Ein Würfel wird so lange gewürfelt bis eine 6 Auftritt. Die Zufallsvariable X ist gleich Anzahl der Würfe

Stetige Verteilungen

Dichtefunktion

Die Dichtefunktion ist ein Hilfsmittel zur

Beschreibung einer stetigen Wahrscheinlichkeitsverteilung

Bedingungen der Dichtefunkion

$$\int_{-\infty}^{\infty} f(x) \ge 0$$
$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Die Dichtefunktion muss **nicht** stetig sein Die Dichtefunktion ist die Ableitung der Verteilungsfunktion F(x)

Verteilungsfunktion

Eine Verteilungsfunktion ist eine Funktion F, die jedem x einer Zufallsvariable X genau eine Wahrscheinlichkeit $P(X \le x)$ zuordnet $F(x) \to P(X \le x)$

Bedingungen der Verteilungsfunktion Die Verteilungsfunktion **muss** stetig sein Die Verteilungsfunktion **muss** monoton steigend sein

$$\lim_{\substack{x \to \infty \\ \lim_{x \to -\infty}}} F(x) = 1$$

Normalverteilung

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

Ist $X \sim N(0,1)$ dann heißt sie

Standardnormalverteilt Jede Normalverteilung kann standardisiert werden, das heißt die Mitte der Kurve wird auf den Nullpunkt gesetzt

Wenn $X \sim N(\mu, \sigma^2)$ verteilt ist dann ist die standardiserte Zufallsvariable $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$ standardnormalverteilt Ist die Zufallsvariable standardverteilt kann die Wahrscheinlichkeit aus der Tabelle abgelesen werden

abgelesen werden
$$P(X \le k) = \Phi(k)$$

$$P(X = k) = 0 \text{ ("Integral ohne Breite!")}$$

$$P(X \le -k) = 1 - \Phi(+k)$$
 allgemein gilt
$$X \sim N(\mu, \sigma^2)$$

$$P(X \le k) = \Phi(\frac{k-\mu}{\sigma})$$

$$P(a \le X \le b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

Quantile der Normalverteilung

Tabelliert ist das β -Quantil z_{β} der Normalverteilung N(0,1) $P(X \leq z_{\beta}) = \beta$ $z_{1-\beta} = -z_{\beta}$ Beispiel

$$\beta = 0.9 = > z_{\beta} = 1.28155$$

Exponentialverteilung

Eine exponentialverteilte Zufallsvariable T hat die Dichte

$$f(t) = \begin{cases} \lambda \cdot e^{-\lambda \cdot t} & , t \ge 0 \\ 0 & , t < 0 \end{cases}$$

und daraus eribt sich die Verteilungsfunktion $F(x) = P(T \le x) =$

$$= \int_{-\infty}^{x} f(t)dt = \begin{cases} 1 - e^{-\lambda \cdot x} &, x \ge 0 \\ 0 &, x < 0 \end{cases}$$

Die Exponentialverteilung ist Gedächtnislos

Gleichverteilung(Rechteckverteilung)

$$f(t) = \begin{cases} \frac{1}{b-a} & , t \in [a,b] \\ 0 & , sonst \end{cases}$$

$$F(t) = \begin{cases} 0 & ,t < a \\ \frac{t-a}{b-a} & ,t \in [a,b] \\ 1 & ,t > b \end{cases}$$

Erwartungswert und Varianz

Erwartungswert

Erwartungswert und Mittelwert sind prinzipiell gleichwertig, der Erwartungswert entspricht der theoretischen Erwartung, der Mittelwert entspricht den tatsächlichen

Werten

Zufallsvariable mit diskreter Verteilung

$$\mu = E(X) = \sum_{i=0}^{n} (x_i \cdot p_i)$$

Zufallsvariable mit Dichtefunktion f

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

Exponentialverteilung mit Zufallsvariable T

$$E(T) = \sigma_T = \frac{1}{\lambda}$$

Für Binomialverteilung

$$\mu = E(X) = n \cdot p$$

Für geometrische Verteilung

$$\mu = E(X) = \frac{1}{p}$$

Für Poissonverteilung

$$\mu = E(X) = \lambda$$

Für Hypergeometrischeverteilung

$$E(S_n) = E(X_1 + \dots + X_n) = n \cdot E(X_1) = n \cdot \frac{M}{N}$$

Für Gleichverteilung(Rechteckverteilung)

$$E(T_i) = \frac{a+b}{2}$$

Allgemeine Regeln für den Erwartungswert

$$a, b \in \mathbb{R}$$

$$E(aX + b) = a \cdot E(X) + b$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$$

Varianz

Zufallsvariable mit diskreter Verteilung

$$\sigma^2 = Var(X) = \sum (x_i - \mu)^2 \cdot p_i$$

Zufallsvariable mit Dichtefunktion f

$$Var(X) = E(X^{2}) - (E(X))^{2}$$

Varianz aus Erwartungswert berechnen

$$\sigma^2 = E(X^2) - \mu^2$$

Exponentialverteilung mit Zufallsvariable T

$$Var(T) = \frac{1}{\lambda^2}$$

Für Binomialverteilung

$$\sigma^2 = n \cdot p \cdot (1 - p)$$

Für geometrische Verteilung

$$\sigma^2 = \frac{1}{p^2} - \frac{1}{p}$$

Für Poissonverteilung

$$\sigma^2 = Var(X) = E(X^2) - E(X)^2 = \lambda$$

Für Hypergeometrischeverteilung

$$Var(S_n) = n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}$$

Für Gleichverteilung(Rechteckverteilung)

$$Var(T_i) = \frac{(b-a)^2}{12}$$

Allgemeine Regeln für Varianz

$$Var(aX + b) = a^{2} \cdot Var(X)$$

$$Var(X + Y) =$$

$$Var(X) + Var(Y) + 2 \cdot cov(X, Y)$$
wobei gilt:
$$Cov(X, Y) = E((X - \mu_{X})(Y - \mu_{Y})) =$$

$$E(X \cdot Y) - \mu_{X}\mu_{Y}$$

bei unabhängigen Zufallsvariablen X und Y gilt Cov(X,Y) = 0 siehe unten.

Unabhängige Zufallsvariablen

Allgemeine Regeln

$$Var(X + const) = Var(X)$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

$$Var(X + Y) = Var(X) + Var(Y)$$

$$E(aX + b) = aE(X) + b$$

Wichtige Sätze der Stochastik

Zentraler Grenzwertsatz

n groß (Anzahl der Zufallsvariablen) $n \ge 30$ X_i unabhängig und identisch verteilt

â haben die gleiche Verteilung

$$E(X_i) = \mu$$

$$Var(X_i) = \sigma^2$$

$$\sum X_i \sim N(n \cdot \mu, n \cdot \sigma^2)$$

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} = \overline{x} \sim N(\mu, \frac{\sigma^2}{n})$$

Manche Verteilungen verhalten sich in der Summe anders, zum Beispiel die Rechteckverteilung ist nicht mehr R-Verteilt. Dann wird der Zentrale Grenzwertsatz verwendet

Induktive Statistik - Schätztheorie

Schätzfunktionen

Maximum-Likelihood-Schätzer

$$L(x_1,\ldots,x_n,\alpha) = \prod_{i=1}^n f(x_i)$$

 $f(x_i)$ muss eine Dichtefunktion sein $\frac{\partial \ln L(x_1,\dots,x_n,\alpha)}{\partial \alpha} = 0$ Die Funktion nach dem Parameter α

ableiten und Nullsetzen Das Ergebnis ist der Maximum-Likelihood-Schätzer

Konfidenzintervalle

Intervall für E(X) einer Normalverteilung

Ist $X \sim N(\mu, \sigma^2)$ verteilt, dann ist $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$

Bei bekannter Standardabweicheung σ

erwartungsstreuer Schätzer bei nunabhängigen Stichproben $\alpha = \text{Signifikanzwahrscheinlichkeit}$ (Irrtumswahrscheinlichkeit)

 $1 - \alpha = Vertrauensniveau$ Ist α gegegeben, berechne das Quantil

$$z_{1-(\alpha/2)}$$

Bei **unbekannter** Standardabweicheung σ

$$\left[\bar{x} - t_{1-(\alpha/2)} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{1-(\alpha/2)} \cdot \frac{s}{\sqrt{n}}\right]$$

Anstatt σ^2 wird der erwartungstreue

Schätzer
$$s$$
 verwendet $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

beziehungsweise

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

Induktive Statistik - Hypothesentest

Tests für Lageparameter

Gauß-Test

noch zu füllen

t-Test

noch zu füllen

Tests für Streuungsmaße

 χ^2 - Anpassungstest

Der χ^2 -Anpassungstest überprüft ob eine unbekannte Wahrscheinlichkeitsverteilung einem bestimmten Verteilungsmodell folgt

$$T = \frac{1}{n} \left(\sum_{i=1}^{r} \frac{N_i^2}{p_i} \right) - n$$
alternativ
$$T = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{n \cdot p_i}$$

Das Ergebnis T mit $\chi^2_{r-1:1-\alpha}$ Wert aus der Tabelle vergleichen

 $T < \chi^2$ = Hypothese wird nicht verworfen

Übersicht: Induktive Statistik

GEGEBENENFALLS FOLGENDES NOCH ZU DEN EINZELNEN POSITIONEN VERSCHIEBEN

Konfidenzbereich/Test für Erwartungswert

Varianz σ^2 bekannt

Zweiseitige Konfidenzintervalle

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0, 1)$$

ist zu gegebenen Konfidenzniveau $1-\alpha$ das Konfidenzintervall gleich

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Einseitige Konfidenzintervalle

$$\left[-\infty, \bar{x} + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right] \text{ bzw.}$$
$$\left[-\infty, \bar{x} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right[$$

Zweiseitige Tests

Für
$$\mu_0$$
 sei $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

 $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ $H_0: \mu = \mu_0 \text{ Bereich: } -z_{1-\frac{\alpha}{2}} \le Z \le z_{1-\frac{\alpha}{2}}$

Einseitige Tests

 $H_0: \mu < \mu_0$ Bereich: $Z < z_{1-\alpha}$ $H_0: \mu > \mu_0$ Bereich: $Z > -z_{1-\alpha}$ Varianz σ^2 unbekannt

Zweiseitige Konfidenzintervalle

Allgemeine Matheregeln

Potenzen und Logarithmen

Potenzgesetze

$$a^0 = 1$$

$$a^1 = a$$

$$a^m \cdot a^n = a^{m+n}$$

$$(a^n)^m = a^{n \cdot m}$$

$$\frac{a^n \cdot b^n = (a \cdot b)^n}{\frac{a^n}{a^m} = a^{n-m}}$$

$$a^{\frac{b}{n}} = \sqrt[n]{b}$$

$$\prod_{i=1}^{n} a^{x_i} = a^{\sum_{i=1}^{n} x_i}$$

Logarithmusregeln

$$x = \log_a y \Leftrightarrow y = a^x$$

$$\log 1 = 0$$

$$\log x \cdot y = \log x + \log y$$

$$-\log x = \log \frac{1}{x}$$

$$\log \frac{x}{y} = \log x - \log y$$

$$\log x^n = n \cdot \log x$$

$$\log_a x = \frac{\log x}{\log a}$$

$$\log\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \log x_i$$

Ableitungen und Integrale			
Grundlegende Ableitungsregeln			
f(x)	f'(x)		
c = const	0		
x^n	$n \cdot x^{n-1}$		
\sqrt{x}	$\frac{1}{2\sqrt{x}}$		
e^x	e^x		
a^x	$\ln a \cdot a^x$		
$\ln x$	$\frac{1}{x}$		
$\log_a x$	$\frac{1}{\ln a \cdot x}$		
$\sin x$	$\cos x$		
$\cos x$	$-\sin x$		
$\tan x$	$\frac{1}{\cos^2 x}$		
$\cot x$	$\frac{1}{\sin^2 x}$		
Verknüpfte Ableitungsregeln			

f(x)	f'(x)
(f(x) + g(x))	(f'(x) + g'(x))
$(f(x) \cdot g(x))$	$(f'(x) \cdot g(x)) + (f(x) \cdot g'(x))$
$\frac{f(x)}{g(x)}$	$\frac{(f'(x)\cdot g(x)) - (f(x)\cdot g'(x))}{g(x)^2}$
f(g(x))	$f'(g(x)) \cdot g'(x)$

wichtige Stammfunktionen

f(x)	F(x)
$x^n, n \neq 1$	$\frac{1}{n+1} \cdot x^{n+1} + c$
$\frac{1}{x}, x \neq 0$	
\sqrt{x}	$\frac{2}{3} \cdot x^{\frac{3}{2}} + c$
e^x	$e^x + c$

Bestimmte Integrale

$$\int_{a}^{b} f(x)dx = [F(x) + C]_{a}^{b} = F(b) - F(a)$$