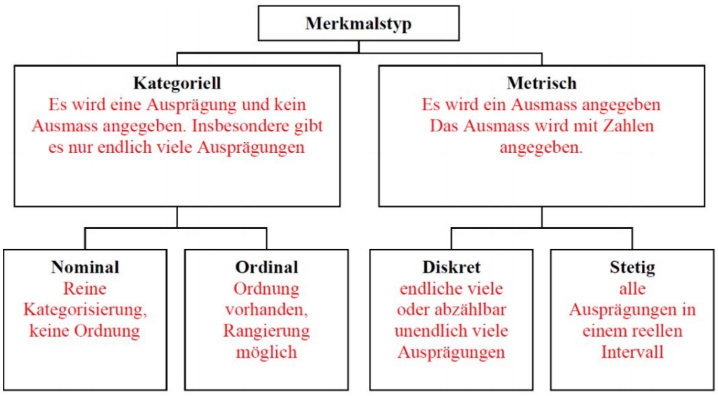
STS Zusammenfassung

Joël Plambeck – plambjoe@students.zhaw.ch – Version 0.1

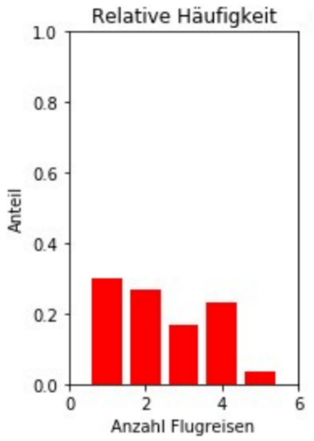
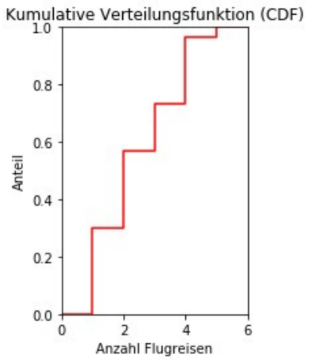
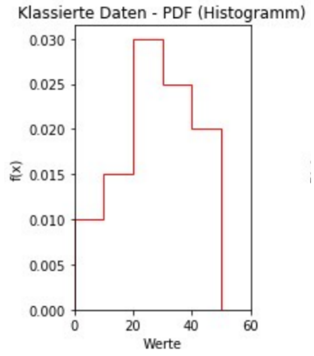
Table of Contents

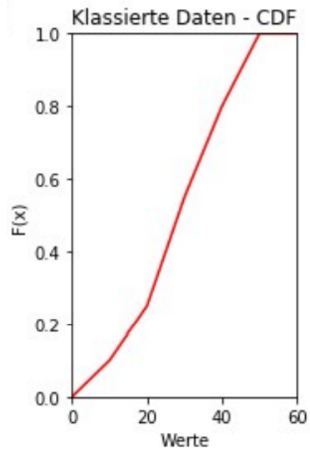
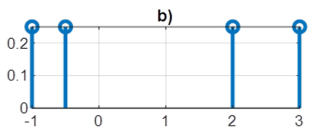
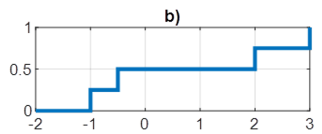
1. Stichproben



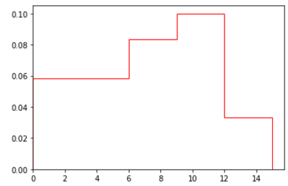
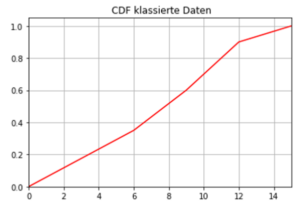
2. PDF & CDF

Nichtklassiert	Klassiert

<table>
<tr><th colspan="2">Nichtklassiert</th><th colspan="2">Klassiert</th></tr>
<tr><th>PDF</th><th>CDF</th><th>PDF</th><th>CDF</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>

$$\text{höhe} = \frac{\text{rel. Hkeit.}}{\text{breite}}$$</td>
<td>
</td>
</tr>
<tr>
<td>
</td>
<td>
</td>
<td>

Klasse	Kl
[0,6)	7
[6,9)	5
[9,12)	6
[12,15)	2

</td>
<td>
</td>
</tr>
</table>

PDF Probability density function

CDF Cumulative distribution function

3. Kennzahlen

	nicht Klassiert	Klassiert
Quantil	$X_q = \begin{cases} \frac{1}{2}(X_{nq} + X_{nq+1}) & n \cdot q \quad \text{ganzzahlig} \\ \lceil X_{nq} \rceil & n \cdot q \quad \text{nicht ganzzahlig} \end{cases}$	$\text{F(A)} \leq x \leq \text{F(B)}$ $\frac{F(b) - F(a)}{b - a} = \frac{q - F(a)}{R_q - a}$ $R_q = a + (b - a) \cdot \frac{q - F(a)}{F(b) - F(a)}$ $F(x) = \text{CDF an Stelle} \quad x$ $R_q = 425 \quad (\text{Bsp}), \quad q = 0.3 \quad (\text{Bsp})$
arith. Mittel	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$	$\bar{x} = \sum_{i=1}^{m} M_i \cdot f_i$
Varianz	$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$	$S^2 = \sum_{i=1}^{n} (M_i - \bar{x})^2 \cdot f_i$

	nicht Klassiert	Klassiert
Varianz (korr)	$S_{\text{korr}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$S_{\text{korr}}^2 = \frac{n}{n-1}S^2$
Standardabweichung	$S = \sqrt{S^2}$	$S = \sqrt{S^2}$
Standardabweichung (korr)	$S_{\text{korr}} = \sqrt{S_{\text{korr}}^2}$	$S_{\text{korr}} = \sqrt{S_{\text{korr}}^2}$

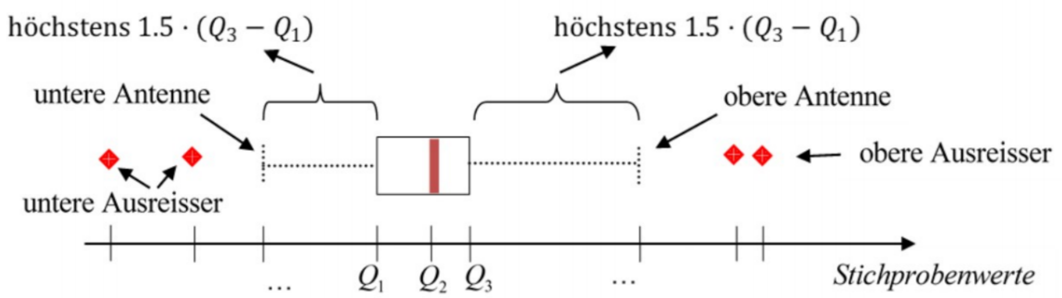
$$\frac{n}{n-1}S^2 = S_{\text{korr}}^2$$

$$\sqrt{\frac{n}{n-1}}S = S_{\text{korr}}$$

3.1. Boxplot

Box: Zeigt das erste Quartil , den Median und das dritte Quartil.

Quartilsabstand: , d.h. die Breite der Box.



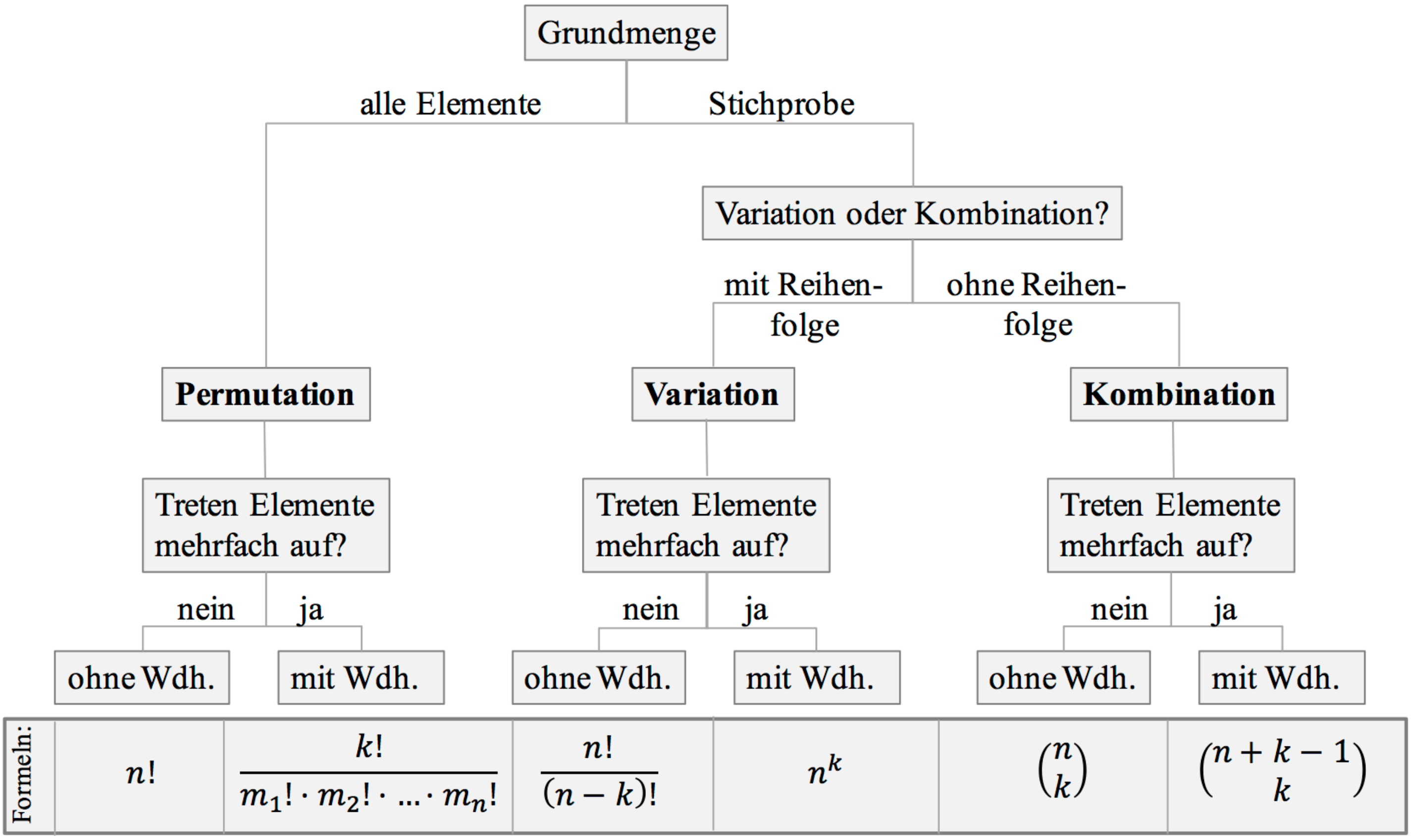
4. Korrelation

Gegeben sind die Wertepaare $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Bezeichnung	**Formel**
Standardabweichung	$S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{S_x^2}$ $S_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{S_y^2}$
Varianz	$S^2 = \overline{x^2} - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Kovarianz	$S_{xy} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})} = \overline{xy} - \bar{x} \cdot \bar{y}$
Korrelationskoeffizient	$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \text{ und } \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

5. Kombinatorik



Binomialkoeffizient: $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$ $\quad\frac{n!}{(n-k)!} = \binom{n}{k} \cdot k!$

Zahlenschlossproblem

Das Zahlenschloss hat 6 Zahlenkränze mit den Zahlen 0 bis 9. An jedem Kranz können unabhängig von den anderen Kränzen 10 verschiedene Zahlen eingestellt werden. Dabei kann jede Zahl mehrfach vorkommen.

10^6

Schwimmwettkampf

Platzierung 1-3 mit 10 Schwimmern

$$10 \cdot 9 \cdot 8 = \frac{10!}{(10-3)!} = 720$$

Lotto

Bei der Ziehung 6 aus 49 spielt die Reihenfolge der gezogenen Zahlen keine Rolle, allerdings kann jede der Zahlen nur einmal vorkommen, da die Kugeln nicht mehr zurückgelegt werden. Käme es auf die Reihenfolge an, so hätte man wie im Problem mit den Schwimmerinnen 6 Plätze für mögliche 49 Zahlen zu besetzen.

$$\frac{49!}{(49-6)!}$$

Bitproblem

Binärzahl mit 64 Stellen:

$$2^{64}$$

Zahnarztproblem

Eine Zahnärztin erlaubt den Kindern, nach der Behandlung zur Belohnung 3 Spielzeuge aus 5 Töpfen auszusuchen. Die 5 Töpfe sind jeweils mit einer Art Spielzeug befüllt. Wie viele verschiedene Möglichkeiten hat ein Kind? Das komplizierteste der hier beschriebenen Abzählprobleme wird mit einem Ersatzproblem gelöst. Dabei handelt es sich um ein äquivalentes Problem, welches einfacher zu lösen ist. Jede Auswahl von 3 Objekten aus den 5 Schalen kann eindeutig durch 3 Striche III und 4 Kreuze XXXX beschrieben werden:

$$\underbrace{\ }_{k_1=0} X \underbrace{\ }_{k_2=0} X \underbrace{I}_{k_3=1} X \underbrace{\ }_{k_4=0} X \underbrace{II}_{k_5=2}$$

$$\binom{7}{3} = \binom{7}{4} = 35$$

Fussballmannschaft

Aus 20 Studierenden sollen 11 Personen ausgewählt werden. Wie beim Lotto 6 aus 49 ist das Problem hier 11 aus 20.

$$\binom{20}{11} = 167'960$$

Buchstabenproblem

Mit 10 verschiedenen Buchstaben Worte von 5 Zeichen bilden.

Tellschiessen

Wilhelm Tell schiesst mit drei Pfeilen auf eine Zielscheibe, welche in 10 ringförmige Bereiche unterteilt ist. Wenn man Wilhelm die Möglichkeit zugesteht daneben zu schiessen, so hat er 3 Versuche mit jeweils 11 verschiedenen Resultaten, wobei Resultate mehrfach vorkommen können. 3 Objekte aus 11 möglichen Sorten zu wählen.

$$\binom{11-1+3}{3} = \binom{13}{3} = 286$$

Napoleon

Napoleon schart seine 10 Generäle an einen kreisrunden Tisch mit 11 Plätzen. Bei der Sitzordnung kommt es nur auf die Reihenfolge der Personen an. Man könnte den Tisch auch drehen und die Sitzordnung bleibt dieselbe.

$$10! = 3628800$$

Gruppen

Wie viele verschiedene Personengruppen kann man aus einer Klasse mit 20 Studierenden bilden? Die so beschriebene Gruppe kann mit dem folgenden Bitmuster abgebildet werden: 1 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 1 1 0

$$2^{20} - 1$$

Teilmengen

Eine Möglichkeit die Teilmengen der Menge {1,2,3,4} zu kategorisieren (Potenzmenge)

$$2^n = 2^4 = 16$$

Variation (mit Reihenfolge)		Kombination (ohne Reihenfolge)	
Mit Wiederholung	Ohne Wiederholung	Mit Wiederholung	Ohne Wiederholung
Zahlenschloss (1) Bitproblem (4) Buchstabenproblem (7b) n^k	Schwimmwettkampf (2) Buchstabenproblem (7a) Napoleon (9) $\frac{n!}{(n-k)!} = \binom{n}{k} \cdot k!$	Zahnarztproblem (5) Tellschiessen (8) $\binom{n-1+k}{k}$	Lotto (3) Fussballmannschaft (6a) Teilmengenproblem (11) $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

6. Elementare Wahrscheinlichkeitsrechnung

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

$$E(X) = \sum_{x \in \mathbb{R}} (x \cdot f(x))$$

$$V(X) = E\Big((X - E(X))^2\Big) = \sum_{x \in \mathbb{R}} \Big((x - E(X))^2 \cdot f(x)\Big)$$

Linearität des Erwartungswertes

- $E(X + Y) = E(X) + E(Y)$
- $E(\alpha X) = \alpha E(X)$

Verschiebungssatz für die Varianz

- $V(X) = E(X^2) - E(X)^2 = \left(\sum_{x \in \mathbb{R}} P(X = x) \cdot x^2\right) - E(X)^2$
- $V(\alpha X + \beta) = \alpha^2 \cdot V(X) \quad \text{mit} \quad \alpha, \beta \in \mathbb{R}$

7. Spezielle Verteilungen

Eine **diskrete** Zufallsvariable nimmt nur bestimmte Werte an

Eine **stetige** Zufallsvariable kann jedem beliebigen Wert in einem Intervall annehmen

7.1. Diskrete und stetige Verteilungen

	diskrete Zufallsvariablen	**stetige Zufallsvariablen**
Graphische Darstellung von f	Stabdiagramm	Graph
Dichtefunktion / PDF	$f(x) = P(X = x)$	$f(x) = F'(x)P(X = x)$
Kummulative Verteilungsfunktion / CDF	$F(x) = P(X \leq x) = \sum_{x \leq X} f(x)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$

	diskrete Zufallsvariablen	**stetige Zufallsvariablen**
Wahrscheinlichkeiten	$P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$	$P(a \leq X \leq b) = \int_a^b f(x)dx$
Erwartungswert	$E(X) = \sum_{x \in \mathbb{R}} f(x) \cdot x$ $E(X^2) = \sum_{x \in \mathbb{R}} f(x) \cdot x^2$	$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x dx$ $E(X^2) = \int_{-\infty}^{\infty} f(x) \cdot x^2 dx$
Varianz	$V(X) = \sum_{x \in \mathbb{R}} f(x) \cdot (x - E(X))^2$	$V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2$

Verschiebungssatzu für die Varianz = $V(X) = E(X^2) - E(X)^2$

7.2. Diskrete Verteilungen

Verteilungen:	**Hypergeometrisch**	**Binomial**	**Poisson**
Funktion	$X \sim H(N, M, n)$	$X \sim B(n, p)$	$X \sim P(\lambda)$
Erwartungswert $E(x) = \mu$	$n \cdot \frac{M}{N}$	$n \cdot p$	λ
Varianz $V(X) = \sigma^2$	$n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$	$n \cdot p \cdot (1-p)$	λ
Standardabweichung $S(X) = \sigma$	$\sqrt{n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}}$	$\sqrt{n \cdot p \cdot (1-p)}$	$\sqrt{\lambda}$
Dichtefunktion $P(X = k)$	$\frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$	$\binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$	$\frac{\lambda^k}{k!} \cdot e^{-\lambda}$

7.2.1. Hypergeometrische Verteilung

Wir betrachten eine Urne mit N Objekten. Darunter sind M Objekte einer bestimmten Sorte, wir nennen sie Sorte M, die Merkmalsträger und $N - M$ andersartige Objekte. Es wird zufällig eine Stichprobe von n Objekten aus der Urne entnommen. Das Ziehen kann auf einmal passieren oder auch nacheinander, wichtig aber ist, dass ohne Zurücklegen der Objekte gezogen wird

$X \sim H(N, M, n)$

N Total/Alle Lose

M Merkmalsträger/Gewinne

n Stichproben/Ziehungen

7.3. Bernoulliverteilung

Bernoulli-Experimente sind Zufallsexperimente mit nur zwei möglichen Ergebnissen. Wir bezeichnen diese Ergebnisse mit 1 und 0.

$P(X=1)=p$ $\quad E(x)=p$ $\quad V(x)=p\cdot(1-p)$

$P(X=0)=1-p$ $\quad E(x^2)=p$

7.3.1. Binomialverteilung

Die Binomialverteilung ist eine Bernoulliverteilung die n mal durchgeführt wird.

$X \sim B(n,p)$

n Anzahl Wiederholungen

p Wahrscheinlichkeit für ein Ergebnis 1, $p=\frac{M}{N}$

7.3.2. Poissonverteilung

$X \sim P(\lambda)$

λ $\lambda>0$, durchscnittliche Anzahl Ereignisse pro betratetes Zeitintervall

Die Poisson Verteilung wird immer dort als stochastisches Modell benutzt, wenn es um die Wahrscheinlichkeit für das Eintreten einer bestimmten Anzahl gleichartiger Ereignisse geht, welche in einem gegebenen Bereich beliebig oft auftreten können.

8. Standardnormalverteilung

$P(|x-\mu|\leq e)$

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

Gaussnormalverteilung: $\mu=0 \quad , \quad \sigma=1$

$$N(\mu,\sigma) \quad \text{und} \quad P(T\leq x)\rightarrow \phi_{\mu,\sigma}(x)=\phi\left(\frac{x-\mu}{\sigma}\right)$$

PDF

$$\varphi_{\mu,\sigma}(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

CDF

$$\phi_{\mu,\sigma}(x)=\int_{-\infty}^{x}\varphi_{\mu,\sigma}(t)dt=P(X\leq x)$$

8.1. Zentraler Grenzwertsatz

$$U_n=\frac{S_n-n\mu}{\sqrt{\mu}\cdot\sigma}=\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}$$

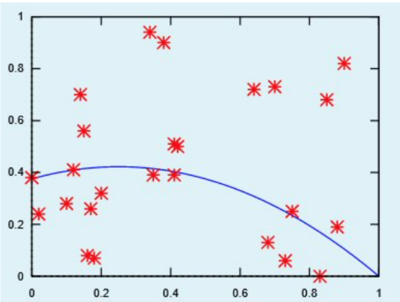
$$S_n=X_1+X_2+...+X_n$$

$$\overline{X}_n=\frac{1}{n}(X_1+X_2+...+X_n)$$

$$E(X_i)=\mu$$

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

	Summe	**Arithmethisches Mittel**	**Gleichverteilung: ganze Zahlen [z,...,w]**
Erwartungswert	$E(S_n) = \mu \cdot n$	$E(\overline{X}_n) = \mu$	$E(x) = \frac{z+w}{2}$
Varianz	$V(S_n) = \sigma^2 \cdot n$	$V(\overline{X}_n) = \frac{\sigma^2}{n}$	$V(x) = \frac{(w-z)\cdot(w+2-z)}{12}$
Verteilung	$N(\mu \cdot n, \sqrt{n} \cdot \sigma)$	$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	



$$E(\beta x) = \int_0^1 f(x)dx$$

$$V(\alpha x) = \alpha \cdot E(\beta x) \cdot (1 - E(\beta x))$$

8.2. Approximaiton durch Normalverteilung

1. Geschlossenes Intervall zu offenes intervall: $P(? < X < ?) \to P(? + 1 \le X \le ? - 1)$
2. Intervall um 0.5 auf beiden Seiten vergrössern: $P(? - 0.5 \le Y \le ? + 0.5)$

9. KQM

KQM

							Arithmetisches Mittel	
x_i	4	0	-2	-4	-5	2	$-\frac{5}{6}$	$= \overline{x}$
y_i	3	-1	-3	-3	1	0	$-\frac{3}{6}$	$= \overline{y}$
x_i^2	16	0	4	16	25	4	$\frac{65}{6}$	$= \overline{x^2}$
y_i^2	9	1	9	9	1	0	$\frac{29}{6}$	$= \overline{y^2}$
$x_i y_i$	12	0	6	12	-5	0	$\frac{25}{6}$	$= \overline{xy}$
$\hat{y}_i$	$\frac{94}{73}$	$-\frac{14}{73}$	$-\frac{68}{73}$	$-\frac{122}{73}$	$-\frac{149}{73}$	$\frac{40}{73}$	$-\frac{3}{6}$	$= \overline{\hat{y}}$
$\hat{y}_i^2$	1.658	$\frac{196}{5329}$	$\frac{4624}{5329}$	2.793	$\frac{22201}{5329}$	$\frac{1600}{5329}$	1.646	$= \overline{\hat{y}^2}$
ε_i	$\frac{125}{73}$	$-\frac{59}{73}$	$-\frac{151}{73}$	$-\frac{97}{73}$	$\frac{222}{73}$	$-\frac{40}{73}$	0	$= \overline{\varepsilon_i}$
ε_i^2	2.932	$\frac{3481}{5329}$	4.279	1.766	9.248	$\frac{1600}{5329}$	3.196	$= \overline{\varepsilon_i^2}$

Residuenvarianz	$S_{\varepsilon}^2 = \overline{\varepsilon^2} - \bar{\varepsilon}^2 = S_y^2 - \frac{S_{xy}^2}{S_x^2}$
Erklärtevarianz	$S_{\hat{y}}^2 = \overline{\hat{y}^2} - \bar{\hat{y}}^2$
Totalevarianz	S_y^2
Bestimmtheistmass	$R^2 = \frac{S_{\hat{y}}^2}{S_y^2}$
Pearson Korrelationskoeffizient	$r_{xy} = \sqrt{\frac{S_{xy}^2}{S_x^2 \cdot S_y^2}} = \frac{S_{xy}}{S_x \cdot S_y}$

$$S_x^2 = \overline{x^2} - \bar{x}^2$$

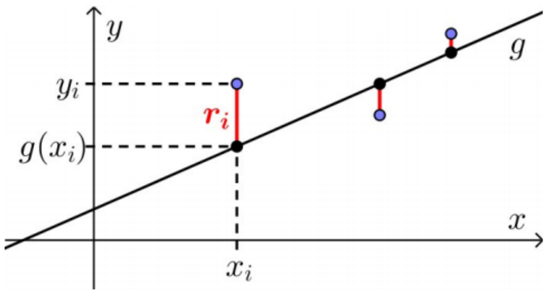
$$S_y^2 = \overline{y^2} - \bar{y}^2$$

$$S_{xy} = \overline{xy} - (\bar{x} \cdot \bar{y})$$

$$\hat{y}_i = g(x_i) = m \cdot x_i + d$$

$$\varepsilon_i = y_i - \hat{y}_i$$

9.1. Regressionsgerade



Normalengleichung

$$A^T \cdot A \cdot \vec{x} = A^T \cdot \vec{y} \quad \Leftrightarrow \quad A^T \cdot A \cdot \begin{pmatrix} m \\ q \end{pmatrix} = A^T \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\vec{x} = \left(A^T \cdot A\right)^{-1} \cdot A^T \cdot \vec{y}$$

nach y	nach x
$y(x) = m \cdot x + d$ $m = \frac{S_{xy}}{S_x^2}$ $d = \bar{y} - m \cdot \bar{x}$	$x(y) = m \cdot y + d$ $m = \frac{S_{xy}}{S_y^2}$ $d = \bar{x} - m \cdot \bar{y}$

10. Schliessende Statistik

10.1. Grösse des Vertrauensintervall (VI)

- Je grösser das γ, desto grösser das VI
- Je grösser der Stichprobenumfang n, desto kleiner das VI (bei konsistenten Schätzfunktionen)
- Bei unbekannter Varianz wird ds VI grösser

7.3.3 Übersicht über verschiedene Vertrauensintervalle zum Niveau γ

	(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			

11. Tabelle

Tabellen Verteilung

Version 0.1
Last updated 2021-01-18 11:51:59 +0100