

- Lücken-Skript -

Mathematik 3 TWIE

Studiengang Wirtschaftsingenieurwesen

Prof. Dr.-Ing. Stephan Sauter
Dipl.-Wirt.-Ing. Stefanie Kaub



Inhaltsverzeichnis

Vorwort	1
1 Einführung	2
1.1 Stochastik	2
1.2 Wahrscheinlichkeitsrechnung und Statistik	3
1.3 Datenmissbrauch in der Statistik	4
1.3.1 Skalen	4
1.3.2 Falsche Präzision	6
1.3.3 Prozente und Berechnungsbasis	7
1.3.4 Mittelwerte	7
1.3.5 Scheinkorrelationen	8
1.4 Statistiksoftware R(-Studio)	10
2 Grundlagen der Wahrscheinlichkeitsrechnung	11
2.1 Kombinatorik	11
2.1.1 Mathematische Grundlagen	11
2.1.2 Permutation	13
2.1.3 Kombination	15
2.1.4 Variationen	17
2.2 Kombinatorik Formeln	18
2.3 Wahrscheinlichkeitsrechnung	20
2.3.1 Grundbegriffe	20
2.3.2 Verknüpfung von Ereignissen	23
2.3.3 Zusammengesetzte Versuche	25
2.4 Wahrscheinlichkeit	26
2.4.1 Der Additionssatz	27
2.4.2 Laplace-Experiment	28
2.4.3 Mehrstufige Zufallsexperimente	29
2.4.4 Wahrscheinlichkeit mit Pfadregeln	31
2.4.5 Galton-Brett	32
2.4.6 Bedingte Wahrscheinlichkeit	34
2.4.7 Stochastische Unabhängigkeit	35
3 Wahrscheinlichkeitsverteilungen	37
3.1 Zufallsvariable	37
3.2 Verteilungsfunktionen	38
3.3 Diskrete Verteilungen	39
3.3.1 Kennwerte	40
3.3.2 Bernoulliverteilung <code>[dbern(x,prob)]</code>	41
3.3.3 Binomialverteilung <code>[dbinom(x,size,prob)]</code>	42
3.3.4 Hypergeometrische Verteilung <code>[dhyper(x,m,n,k)]</code>	46
3.3.5 Poisson-Verteilung <code>[dpois(x,lambda)]</code>	48
3.3.6 Überblick: Zufallsexperimente mit zwei Ausgängen	50

3.4	Stetige Verteilungen	51
3.4.1	Gaußsche Normalverteilung [<code>dnorm(x,mean,sd)</code>]	53
3.4.2	Wahrscheinlichkeiten bei normalverteilten Zufallsvariablen	55
3.4.3	Transformation zur Standardnormalverteilung	56
3.4.4	Exponentialverteilung [<code>dexp(x,rate)</code>]	58
3.4.5	Weibullverteilung [<code>dweibull(x,shape,scale)</code>]	60
3.5	Vergleich der Konzepte: Diskrete vs. stetige Verteilungen	63
4	Statistik	64
4.1	Deskriptive Statistik	64
4.1.1	Grundbegriffe	65
4.1.2	Kennwerte einer Stichprobe	66
4.1.3	Quantile und Quartile	70
4.1.4	Boxplot	71
4.2	Schließende Statistik	73
4.2.1	Parameterschätzungen	73
4.2.2	Lineare Regressionsanalyse	75
4.2.3	Gewinnung von Schätzfunktionen	80
4.3	Hypothesentests	89
4.3.1	Einseitiger Hypothesentest	90
4.3.2	Zweiseitiger Hypothesentest	91
4.3.3	Fehler 1. und 2. Art	92
4.3.4	Durchführung Hypothesentest	94
4.4	Arten von Parametertests	95
4.4.1	Zweiseitiger Hypothesentest mit unbekanntem Mittelwert μ einer Normalverteilung bei bekannter Varianz σ^2	95
4.4.2	Test für den unbekannten Mittelwert μ einer Normalverteilung bei unbekannter Varianz σ^2	97
4.4.3	Einseitiger Hypothesentest mit unbekanntem Mittelwert μ einer Normalverteilung bei bekannter bzw. unbekannter Varianz σ^2	99
4.4.4	Tests für einen unbekannten Anteilswert p	100
4.5	Statistische Signifikanz und fachliche Relevanz	103
4.6	Statistische Qualitätskontrolle unter Verwendung von Kontrollkarten	104
5	Anhang	105
5.1	Wahrscheinlichkeitsbegriff	105
5.2	Binomialverteilung	105
5.3	Gaußsche Normalverteilung	106
5.4	t-Test	107
5.5	Statistiksoftware R	108
5.5.1	Daten einlesen mit R	108
5.5.2	Grundlegende Syntax und Befehle	109
5.5.3	Datenaufbereitung und grafische Darstellung	110
5.5.4	Befehle zum Erzeugen von Zufallszahlen	111
5.5.5	Wahrscheinlichkeitsverteilungen	111
5.5.6	Interner Datensatz	112
5.5.7	Graphische Darstellungen	113

Vorwort

Das vorliegende Skript soll vorlesungsbegleitend dem Hörer das Abzeichnen bzw. Abschreiben der Inhalte ersparen. Falls eine Vorlesungsstunde versäumt wurde, kann der Hörer anhand des Skriptes ersehen, welcher Stoff z.B. mit einem Buch nachgeholt werden sollte.

Bei allen Betrachtungen steht eine anschauliche Darstellung im Vordergrund. Es soll versucht werden, dem Leser Hinweise zu geben, die ihm bei der Lösung der anstehenden Problemstellungen nützlich sind.

Insbesondere wird darauf hingewiesen, dass für die Prüfung das selbständige Lösen der Übungsaufgaben nicht nur empfohlen, sondern vorausgesetzt wird!

- Bronstein u.a.: Taschenbuch der Mathematik Edition Harri Deutsch
- Wilhelm Leupold u.a. : Mathematik - ein Studienbuch für Ingenieure. Band 2 Carl Hanser Verlag
- L. Papula: Mathematik für Ingenieure und Naturwissenschaftler. Band 3 Verlag Vieweg
- Fetzer, Fränkel: Mathematik 2 Springer Verlag
- B. Neumayer, S. Kaup: Mathematik für Ingenieure 2 Shaker Verlag Aachen
- www.wolframalpha.com

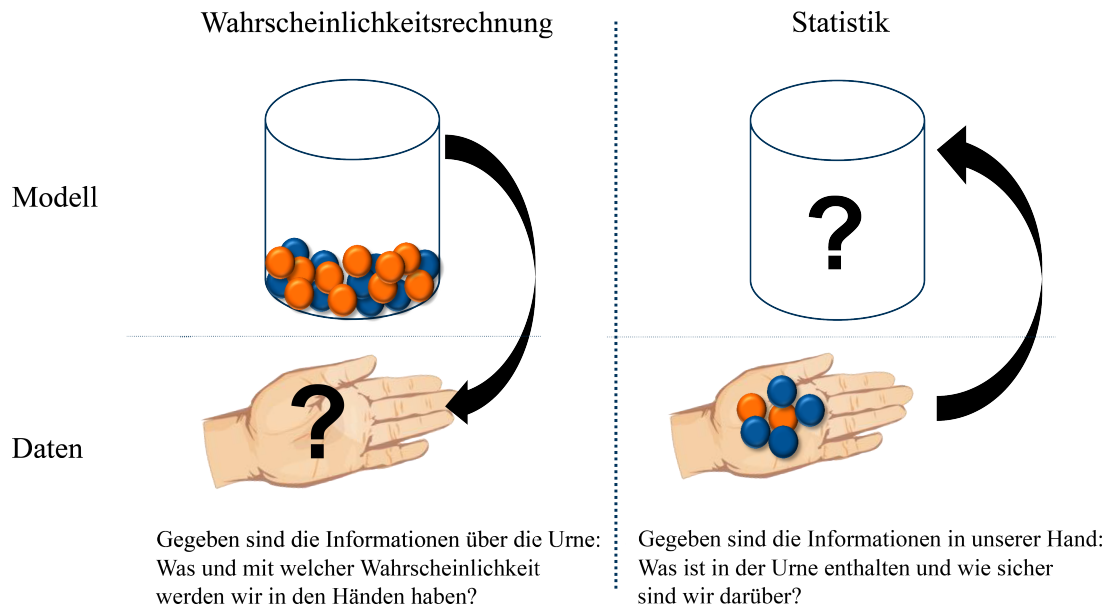
Musterlösungen für die Übungsaufgaben, Formelsammlungen, Skript und Link zum Download der Statistiksoftware R:

- www.freiwilligschlauwerden.de



1.2 Wahrscheinlichkeitsrechnung und Statistik

In der **Wahrscheinlichkeitsrechnung** geht man von einem Modell aus (man beschreibt sozusagen einen datengenerierenden Prozess) und leitet davon entsprechende Eigenschaften ab.



In der **Statistik** geht es darum, aus vorhandenen Daten (Stichproben) auf den datengenerierenden Mechanismus (das Modell) zu schließen.

Man denkt also gerade "in die andere Richtung". Mit ein paar (wenigen) Datenpunkten (z.B. Wasserstandsmessungen) versucht man mit diesem beschränkten Wissen auf ein gutes Modell zu schließen um u.a. Aussagen über die Zukunft zu treffen.

Auch wenn wir Experimente durchführen, erhalten wir Daten die entsprechend adäquat ausgewertet werden müssen. Wenn Sie also einen Fachartikel beurteilen sollen, dann kommt darin wohl fast immer auch eine **Datenanalyse** vor.

Um entsprechende Fehlschlüsse zu durchschauen (was auch ein Grund für den schlechten Ruf der Statistik ist) benötigen Sie das nötige Rüstzeug (z.B. dieses Skript).

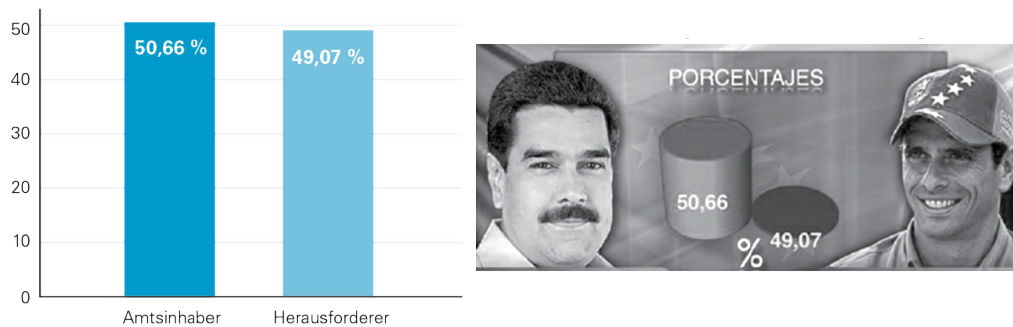
1.3 Datenmissbrauch in der Statistik

So lange es Datenerhebung gibt (bereits im 18. Jahrhundert kommen die ersten Datengrafiken in England auf), wurden Statistiken verwendet um bestimmte Interessen zu verfolgen. Heute mehr denn je, in Zeiten von Fake News und Social Media kann mittlerweile jeder Einzelne Informationen ungefiltert um den Globus verbreiten.

1.3.1 Skalen

Ein extremes Beispiel für manipulative Darstellung von Daten liegt bei folgender Auswertung nach der Wahl zum Präsidenten von Venezuela vor und soll stellvertretend für die Macht der grafischen Aufbereitung Daten aller Art an den Anfang gesetzt werden.

Die linke Seite zeigt das Ergebnis objektiv und nur ein sehr knapper Wahl-Sieg verhalf dem Präsidenten (Nicolás Maduro) zur Wiederwahl. Auf der rechten Seite wird das Ergebnis subjektiv von (vermutlich) regierungsnahen Medien als “haushoher” Sieg durch Abschneiden der Säulen dargestellt.



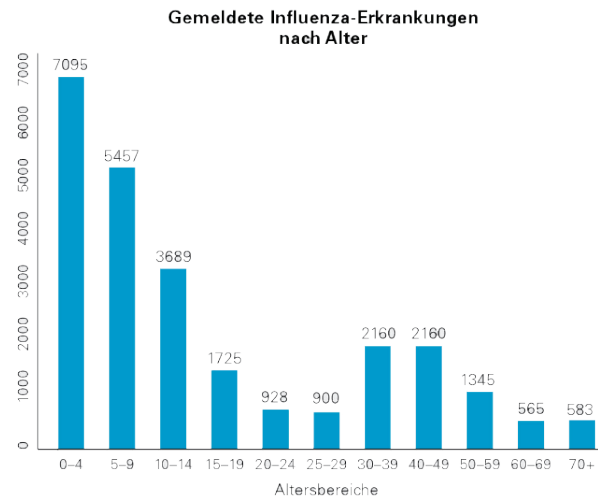
Quelle: Krämer, Walter, So lügt man mit Statistik (German Edition), Campus Verlag.

Immer trifft man diese Darstellung bei Daten an, die subjektiv kaum Veränderung zeigen, der Autor allerdings seine Thesen untermauert sehen möchte.

Dies gilt für die Abszisse (horizontale bzw. x-Achse) ebenso wie für Ordinate (y- bzw. senkrechte Achse).

Ein signifikantes Beispiel, bei welchem auf der Abszisse die Aussage manipulativ verändert wurde liegt in folgendem (medizinischen) Diagramm vor. Durch plötzliches Ändern der Altersbereiche (von 5 Jahren Abstand in jungen Jahren, bis 10 ab 30 Jahre und älter) ergibt sich (ganz natürlich) eine Verzerrung.

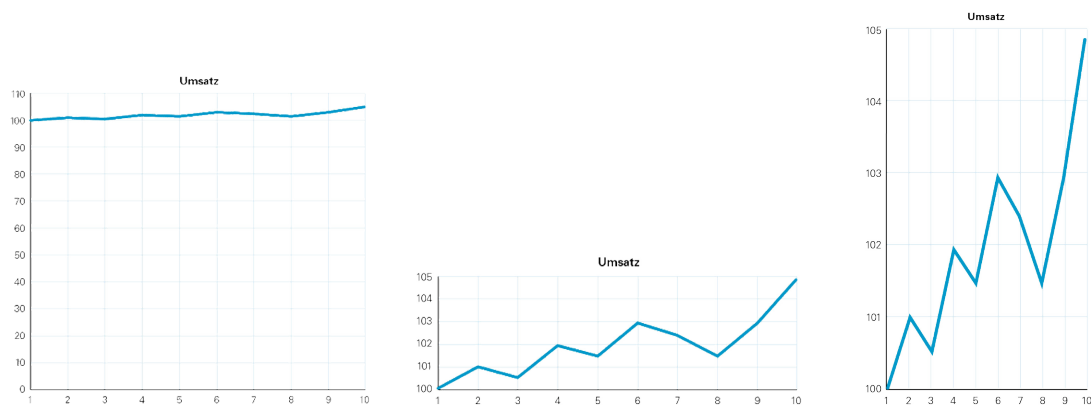
Der Grund könnte eine vermeintliche Rechtfertigung für die Anhebung der Krankenkassenbeiträge in dieser Altersgruppe sein.



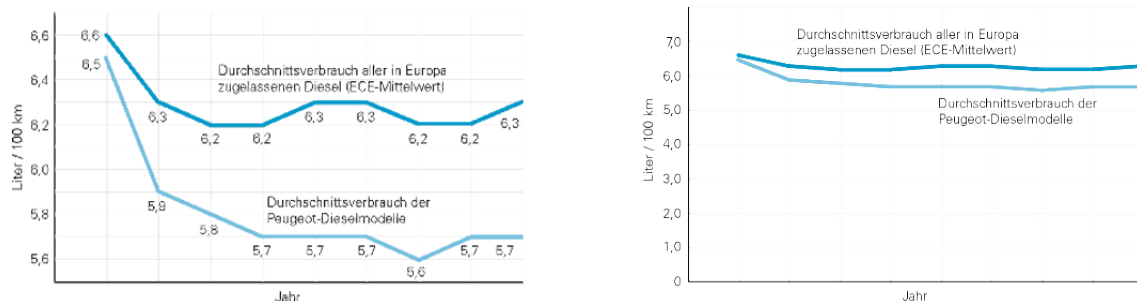
Quelle: Krämer, Walter, So lügt man mit Statistik (German Edition), Campus Verlag.

Sehr häufig treten verzerrte Skalenänderungen in der Wirtschaft bzw. speziell in der Finanzwirtschaft bei Fonds oder Aktien auf.

Beispiele die für sich selbst sprechen:

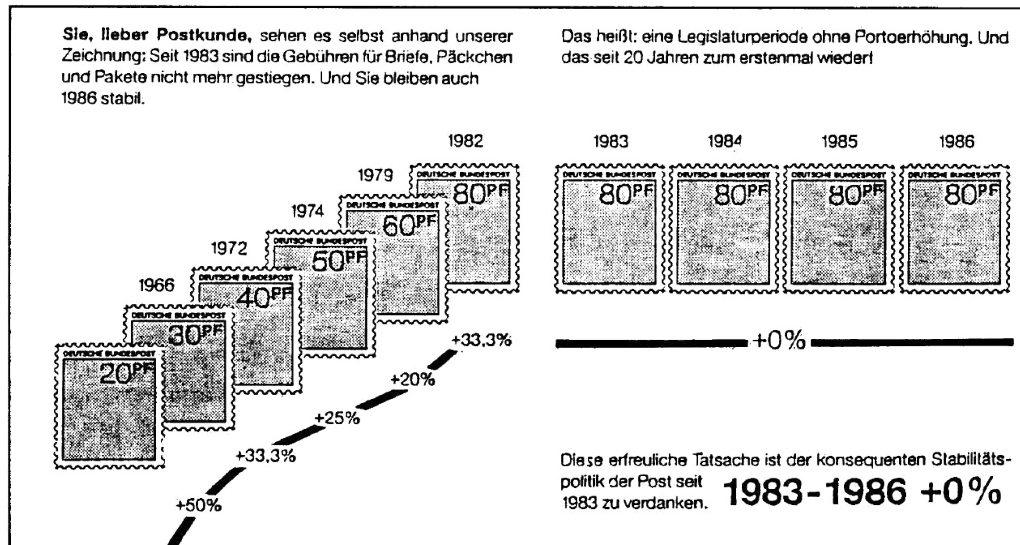


Quelle: Krämer, Walter, So lügt man mit Statistik (German Edition), Campus Verlag.



Quelle: Krämer, Walter, So lügt man mit Statistik (German Edition), Campus Verlag.

Seit 1983 stabile Gebühren



Quelle: Krämer, Walter, So lügt man mit Statistik (German Edition), Campus Verlag.

1.3.2 Falsche Präzision

Nehmen Sie einmal an, Sie sollen für jemanden einen Lebensmittel-Einkauf machen und haben vergessen den Einkaufszettel mitzunehmen. Sie wissen natürlich nicht mehr wieviel das gekostet hat. Wenn Sie dann sagen, 20 Euro hat das gekostet, dann bleibt ein Widerspruch im Raum (kann ja gar nicht sein...).

Sagen Sie allerdings 18.58 Euro, dann wird das wahrscheinlich ohne zu hinterfragen hingenommen. Ein Fall von falscher Präzision.

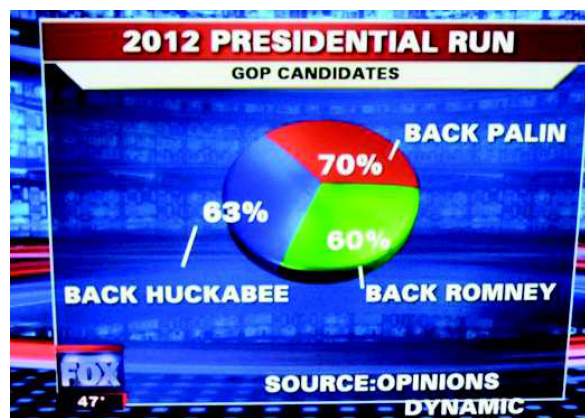
METZGEREI SACK PFINTALSTR. 13			
A00	U04	5628	
0.03.15	0.12.03	B16	
K	€/kg	€	
FLEISCHWURST			
24.7% FETTGEHALT			
0.202	18.80	3.80	
TH. ROTWURST			
14% FETTGEHALT			
0.182	23.80	4.33	
ZIGEUNER-KASSELER			
0.104	31.80	3.31	
MORTADELLA BOLOGNA			
0.091	22.90	2.08	
HAUSMACHER LEBERW.			
0.187	17.80	3.33	
PFEFFERBRATEN			
0.096	32.80	3.15	
6 **		20.00	
N.181			
ES BEDIENTE SIE:			
FRAU LÖFFLER			

CIVILIANS			
(a)	World War I	-	Not known
(b)	World War II		
Allied			
United Kingdom	..	..	60,595
Belgium	..	..	90,000
China	..	..	An enormous number
Denmark	..	..	Unknown
France	..	..	152,000
Netherlands	..	..	242,000
Norway	..	..	3,638
U.S.S.R.	..	..	6,000,000
			6,548,233

Quelle: Krämer, Walter, So lügt man mit Statistik (German Edition), Campus Verlag.

1.3.3 Prozente und Berechnungsbasis

- Norderneyer Badezeitung: „Fuhr vor einigen Jahren noch jeder zehnte Autofahrer zu schnell, so ist es heute jeder fünfte. Doch auch 5% sind zu viele, und so wird weiterhin kontrolliert, und die Schnellfahrer haben zu zahlen.“
- Ein Einzelhändler bezieht ein Produkt zu 100 € und verkauft es für 200 €. Hat er eine Gewinnspanne von 50% oder 100%?
- Die Hälfte aller Todesfälle ereignen sich in Krankenhäusern.
Also: Krankenhäuser sind lebensgefährlich?
- Nur 40% aller durch Autounfälle Gestorbenen hatten keinen Sicherheitsgurt angelegt.
Also: Keinen Gurt anlegen ist sicherer
- Aussage?



1.3.4 Mittelwerte

- „Zwei Männer sitzen im Wirtshaus, der eine verdrückt eine Kalbshaxe, der andere trinkt zwei Maß Bier. Statistisch gesehen ist das für jeden eine Maß Bier und eine halbe Haxe, aber der eine hat sich überfressen und der andere ist betrunken.“
- Das arithmetische Mittel verschleiert oft eine große Ungleichheit - es schweigt sich zur Streuung um den Mittelwert völlig aus.

Beispiel Besitztum/Kapital:

Wenn es in einem Dorf zehn Bauern gibt, von denen einer 40 Kühe hat und alle anderen haben nichts, so hat im Mittel jeder vier.

Für die neun armen Bauern ist das aber nur ein schwacher Trost. Offenbar macht es einen Unterschied, ob sich die Werte dicht um das Mittel sammeln oder ob sie in alle Winde streuen, aber diesen Unterschied sieht man dem (arithmetischen) Mittelwert nicht an.

Hier benötigt man zusätzlich die **Varianz** (siehe "Kennwerte einer Wahrscheinlichkeitsverteilung" in Kapitel 3).

- Eine weitere Konfusion (Verschleierung) betrifft die Zahl, durch die man beim arithmetischen Mittel die Merkmalsumme teilt.

Beispiel Verkehrssicherheit:

Womit ist man sicherer unterwegs, mit dem Flugzeug oder mit der Bahn?

Bezogen auf die zurückgelegte **Strecke**:

Bahn: 20 Verkehrstote pro 100 Milliarden Passagier-Kilometer

Flugzeug: 10 Verkehrstote pro 100 Milliarden Passagier-Kilometer

Bezogen auf die **Beförderungszeit**:

Bahn: 7 Verkehrstote pro 100 Millionen Passagier-Stunden

Flugzeug: 24 Verkehrstote pro 100 Millionen Passagier-Stunden

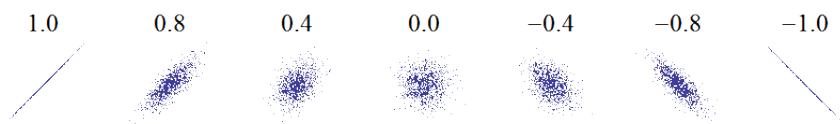
1.3.5 Scheinkorrelationen

Korrelationskoeffizient:

Der Korrelationskoeffizient, auch Produkt-Moment-Korrelation oder **Bravais-Pearson-Koeffizient** genannt, ist ein dimensionsloses Maß für den Grad des linearen Zusammenhangs zwischen zwei mindestens intervallskalierten Merkmalen. Er kann Werte zwischen -1 und +1 annehmen.

Bei einem Wert von +1 (bzw. -1) besteht ein vollständig positiver (negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen.

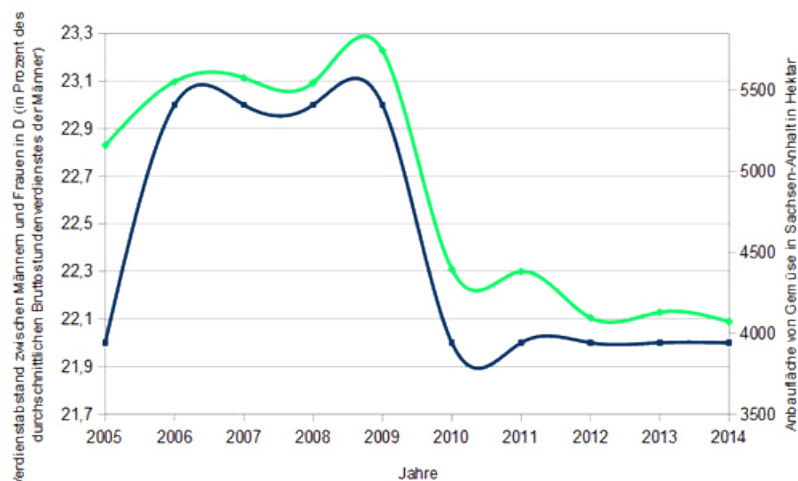
Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab.



Allerdings können diese ungeachtet dessen in nichtlinearer Weise voneinander abhängen. Damit ist der Korrelationskoeffizient **kein** geeignetes Maß für die (reine) stochastische Abhängigkeit von Merkmalen.

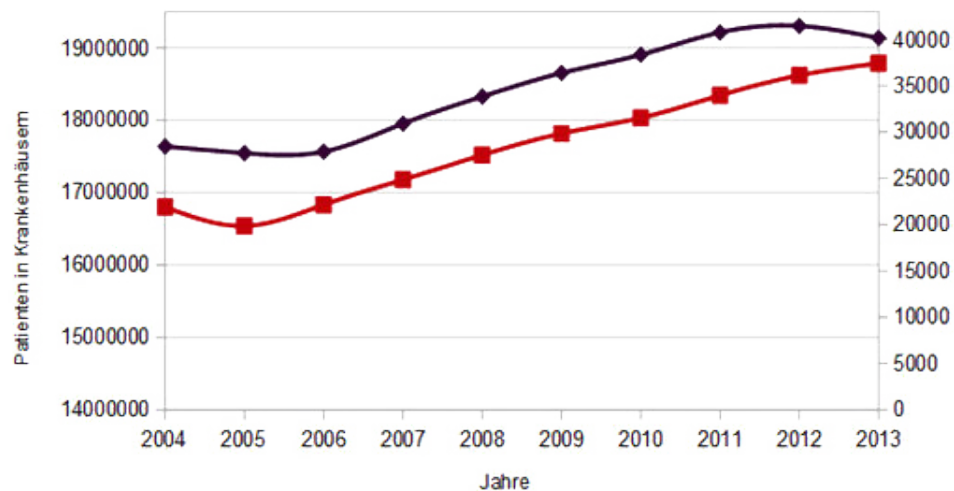
Beispiele für Scheinkorrelationen:

- Verdienstabstand zwischen Männern und Frauen (blau) und Anbaufläche von Gemüse in Sachsen-Anhalt (grün): Korrelation: 0.8982



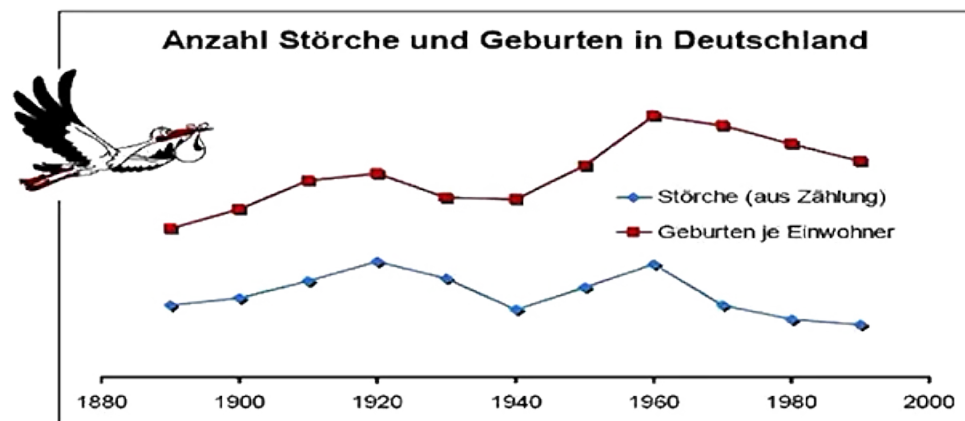
Quelle: Statistisches Landesamt Sachsen-Anhalt, www.scheinkorrelation.jimdo.com

- Patienten in Krankenhäusern (rot) und Studienanfänger im ersten Hochschul-Semester im Studienjahr in Berlin und Brandenburg (schwarz): Korrelation: 0,9803



Quelle: Statistisches Bundesamt & Amt für Statistik Berlin-Brandenburg, www.scheinkorrelation.jimdo.com

- Ein Klassiker der Statistik... Man kann tatsächlich nachweisen, dass in Regionen mit mehr Störchen auch mehr Kinder „auf die Welt kommen“.



Quelle: Statistisches Bundesamt

Ist damit bewiesen, dass Störche Kinder bringen?

Medizinisch-biologische Erkenntnisse sprechen dagegen - es gibt andere, gut belegte Theorien dazu, wie Kinder entstehen und von wo sie „gebracht“ werden. So weit, so klar - kann man das auch statistisch zeigen?

Ja - mit Regressionsanalysen und Drittvariablenkontrolle.

Entscheidender Schritt:

Es kommt eine Theorie ins Spiel, wie diese Signifikanz zu erklären ist.

Sowohl die Anzahl der Störche (x) als auch die Kinderzahl (y) werden vom Industrialisierungsgrad bzw. damit von der Bevölkerungsdichte (z) beeinflusst.

Berücksichtigt man diese gemeinsame Ursache beider Variablen (x) und (y), so verschwindet der scheinbare Zusammenhang.

Signifikanz bei $z \Rightarrow x$ und $z \Rightarrow y$;

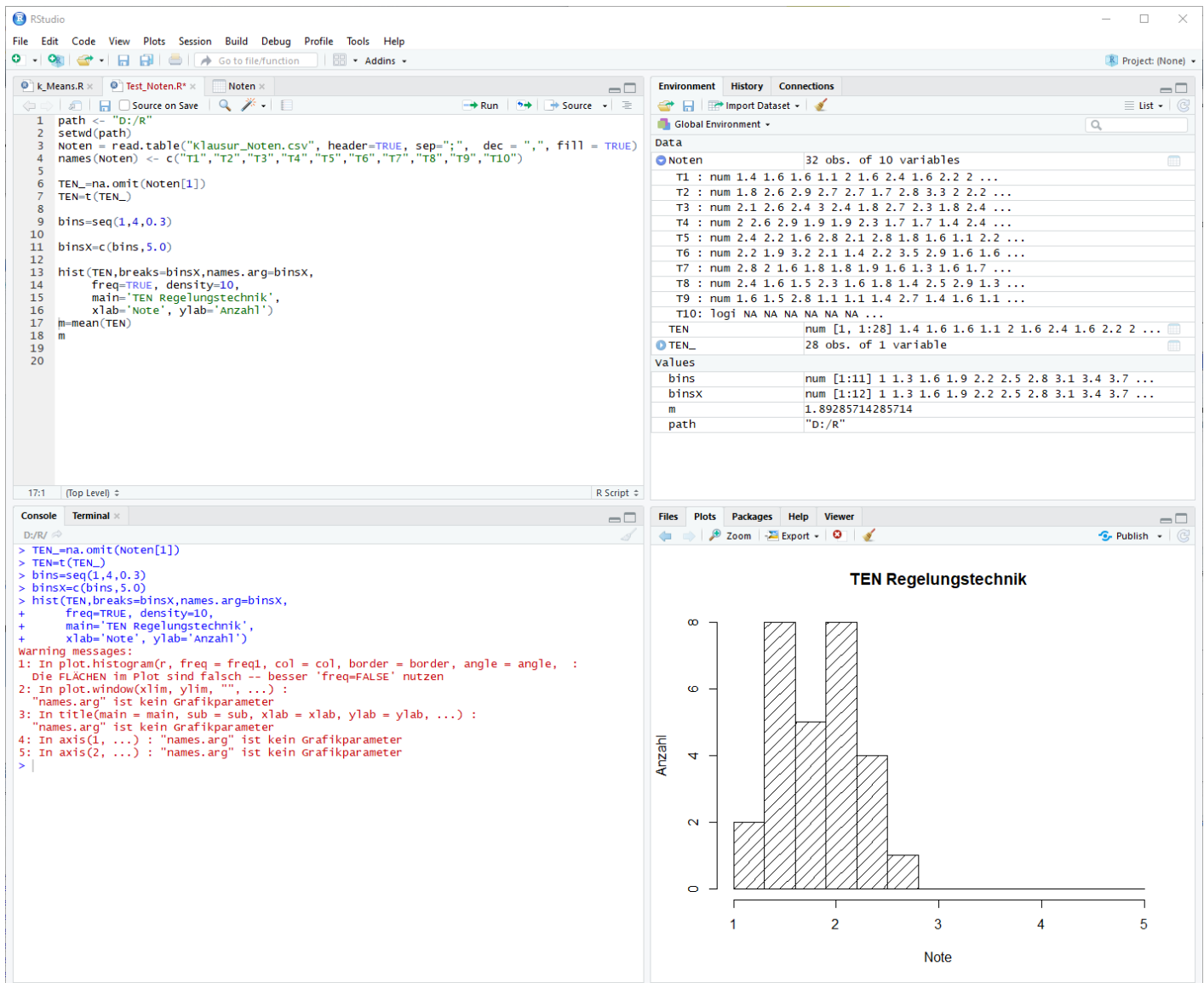
Der Zusammenhang $x \Rightarrow y$ ist allerdings keine Kausalbeziehung.

1.4 Statistiksoftware R(-Studio)

In dieser Statistik-Vorlesung wird ausschließlich die freie Software R verwendet. R-Inhalte sind ebenso klausurrelevant wie die Mathematik-Inhalte.

- R ist ein freies Open-Source Softwarepaket zu Statistik und Datenanalyse
- R ist sehr mächtig und weit verbreitet in Wissenschaft und Industrie (sogar von mehr Benutzern verwendet als z.B. SPSS (IBM))
- Ursprung von R: 1993 an der Universität Auckland von Ross Ihaka and Robert Gentleman entwickelt
- Seitdem: Viele Leute haben R verbessert mit tausenden von Paketen für viele Anwendungen
- Vor & Nachteil: Kein Point and Click Tool

R-Studio Oberfläche zum komfortablen Arbeiten mit R:



Weitere Softwaretools sind z.B.: SPSS, Matlab, Excel, Minitab, GeoGebra...

KAPITEL 2

Grundlagen der Wahrscheinlichkeitsrechnung

2.1 Kombinatorik

Ein wichtiges Hilfsmittel bei der Lösung von Aufgaben aus der Wahrscheinlichkeitsrechnung oder der Statistik ist die Kombinatorik. Hierbei geht es um Abzählmethoden von endlichen oder abzählbaren Objekten. Häufig werden diese anhand des Urnenmodells veranschaulicht.

Die Kombinatorik liefert Methoden zur Berechnung der Anzahl möglicher Anordnungen oder Auswahlen. Abhängig von Eigenschaften ob die einzelnen Objekte unterscheidbar sind, oder eine Reihenfolge der Objekte zu beachten ist, spricht man von Permutationen, Kombinationen oder Variationen.

Diese Begriffe werden im Folgenden definiert und anhand von Beispielen veranschaulicht.

2.1.1 Mathematische Grundlagen

1. Fakultät [`factorial(n)`]

Die Fakultät ist für alle Zahlen $n \in \mathbb{N}$ definiert durch

$$0! = 1 \quad (\text{für } n = 0) \quad \text{und} \quad n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n \quad \text{für } n \in \mathbb{N} \setminus \{0\}$$

sprich: n Fakultät

n	$n!$
0	1
1	1
2	$1 \cdot 2 = 2$
3	$1 \cdot 2 \cdot 3 = 6$
4	$1 \cdot 2 \cdot 3 \cdot 4 = 24$
5	$1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$
6	$1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$
10	3.628.800
15	1.307.674.368.000

Die Fakultät $n!$ ist also eine Schreibweise für das Produkt aller Zahlen $1, 2, 3, \dots, n$. Sie wird vor allem in der Kombinatorik oft verwendet, da $n!$ die Anzahl der Möglichkeiten angibt, eine beliebige Menge mit n Elementen zu ordnen.

Beispiel:

Es gibt $3! = 1 \cdot 2 \cdot 3 = 6$ Möglichkeiten, wie sich drei Personen für ein Foto aufstellen können.

2. Binomialkoeffizient [**choose**(**n**,**k**)]

Unter dem Binomialkoeffizient $\binom{n}{k}$ mit $n \geq k$ und $n, k \in \mathbb{N}$ versteht man

$$\binom{n}{0} = 1 \quad (\text{für } k = 0) \quad \text{und} \quad \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n \cdot \dots \cdot (n-k+1)}{1 \cdot \dots \cdot k} \quad \text{für } k \in \mathbb{N} \setminus \{0\}.$$

spricht: **n über k**

Die Bezeichnung "Binomialkoeffizient" wird von der Regelmäßigkeit der Vorfaktoren bei den Binomischen Formeln abgeleitet (**Pascal'sches Dreieck**):

$$\begin{array}{ll}
 (a+b)^0 & n=0 \qquad \qquad \qquad 1 \qquad \qquad \qquad k=0 \\
 (a+b)^1 & n=1 \qquad \qquad \qquad 1a + 1b \qquad \qquad \qquad k=1 \\
 (a+b)^2 & n=2 \qquad \qquad \qquad 1a^2 + 2ab + 1b^2 \qquad \qquad \qquad k=2 \\
 (a+b)^3 & \vdots \qquad \qquad \qquad 1a^3 + 3a^2b + 3ab^2 + 1b^3 \qquad \qquad \qquad \vdots \\
 (a+b)^4 & \qquad \qquad \qquad 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4 \\
 (a+b)^5 & \qquad \qquad \qquad 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5 \\
 (a+b)^6 & \qquad \qquad \qquad 1a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + 1b^6 \\
 & \vdots \qquad \qquad \qquad \dots
 \end{array}$$

			$\binom{0}{0} =$			
		$\binom{1}{0} =$		$\binom{1}{1} =$		
	$\binom{2}{0} =$		$\binom{2}{1} =$		$\binom{2}{2} =$	
$\binom{3}{0} =$		$\binom{3}{1} =$		$\binom{3}{2} =$		$\binom{3}{3} =$

Allgemein gilt der **Binomische Lehrsatz**

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \quad \text{für } n, k \in \mathbb{N}.$$

Der Binomialkoeffizient kann auch als mathematische Funktion gesehen werden, mit der sich einer der Grundaufgaben der Kombinatorik lösen lässt:

Er gibt an, auf wie viele verschiedene Arten man k Objekte aus einer Menge von n verschiedenen Objekten auswählen kann (ohne Zurücklegen, ohne Beachtung der Reihenfolge).

Deswegen auch oft genannt: **k aus n**

Beispiel:

Möglichkeiten im Lotto 6 Zahlen aus 49 unterschiedlich anzukreuzen: $\binom{n}{k} = \binom{49}{6} \approx 14 \text{ Mio.}$

2.1.2 Permutation

Definition:

Unter einer **Permutation** (von lateinisch permutare = vertauschen) versteht man eine Anordnung von Objekten in einer **bestimmten Reihenfolge**. Abhängig davon, ob manche Objekte mehrfach auftreten dürfen spricht man von

Permutation **mit** Wiederholung, oder
Permutation **ohne** Wiederholung

Einsatzbereiche von Permutationen kommen z.B. vor in der:

- Analysis: Umordnung von Reihen
- Kryptographie: Verschlüsselungsverfahren
- Informatik: Sortiervverfahren
- Quantenmechanik: Pauli-Prinzip (Besetzung der Energieniveaus mit Elektronen im Atom)

Anzahl von Permutationen von n verschiedenen Kugeln

Man stellt sich vor es sind n Plätze vorhanden, auf die die n verschiedenen Kugeln verteilt werden können.

Den ersten Platz kann man mit n verschiedenen Kugeln belegen.

Ist diese Stelle einmal besetzt, so bleiben nur mehr $n - 1$ Kugeln um den 2. Platz zu belegen.

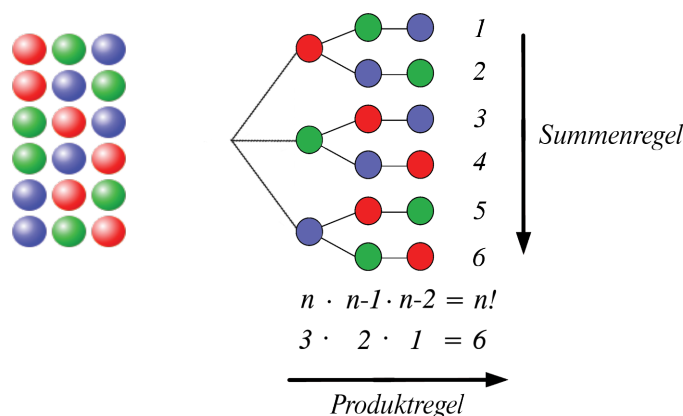
Ist auch dieser Platz besetzt, so bleiben noch $n - 2$ Kugeln für den 3. Platz usw.

Für die Besetzung der Plätze 1, 2, ..., n gibt es daher der Reihe nach $n, n - 1, \dots, 1$ Möglichkeiten:

$$n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 1 = n!$$

Beispiel:

Man habe 3 verschiedenfarbige Kugeln und soll angeben auf wie viele verschiedene Arten sich diese Kugeln anordnen lassen.



Anhand der Anordnung und des Baumdiagramms erkennt man, dass sich drei verschiedenfarbige Kugeln auf sechs verschiedene Arten anordnen lassen.

Permutation ohne Wiederholung

Die Anzahl der Permutationen von n verschiedenen Kugeln beträgt: $n!$

Beispiel:

Man habe 5 verschiedene Bücher und soll angeben, auf wie viel verschiedene Arten sich diese Bücher (auf 5 Plätze) im Regal anordnen lassen.

Wir schauen uns jedes Buch einzeln an, d.h. wir überlegen uns, wie viel Möglichkeiten das erste Buch hat, dann wie viel Möglichkeiten das Zweite hat, dann das Dritte, ...

Das erste Buch hat 5 Plätze zur Auswahl, es hat also 5 Möglichkeiten. Bei dem zweiten Buch, gibt es nur noch 4 freie Plätze, damit 4 Möglichkeiten.

Das dritte Buch hat dementsprechend noch 3 Möglichkeiten, usw...

Es gibt also $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$ Möglichkeiten die Bücher im Regal anzuordnen.

Permutation mit Wiederholung

Die Anzahl der Permutationen von n Kugeln, von denen $k_1, k_2, \dots, k_n$ jeweils gleich sind beträgt: $\frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!}$

Die Ganzzahligkeit des Ergebnisses ist durch die Ganzzahligkeit der Binomialkoeffizienten gewährleistet.

Beispiel:

Man habe 5 Kugeln von denen 3 gleich sind und soll angeben, auf wie viel verschiedene Arten sich diese Kugeln anordnen lassen.

Werden die gleichen Kugeln untereinander vertauscht, was für genau $3!=6$ verschiedene Arten möglich ist, so entstehen dabei keine neuen Anordnungen. Es gibt also nicht $5!$ Permutationen, sondern nur $\frac{5!}{3!} = 20$ Möglichkeiten die Kugeln anzuordnen.

Aufgaben:

1. Wie viele Möglichkeiten gibt es, sechs verschiedenfarbige Kugeln in einer Reihe anzuordnen?
2. Vier Damen und vier Herren passieren nacheinander eine Drehtür. Auf wie viele Arten können sie dies tun?
3. In einer Urne befinden sich drei blaue und zwei rote Kugeln. Wie viele Möglichkeiten gibt es, die Kugeln in einer Reihe anzuordnen?

4. Wie viele verschiedene sechsziffrige Zahlen gibt es, die zweimal die 1, dreimal die 2 und einmal die 4 enthalten?

5. Auf wie viele Arten kann man die Buchstaben des Wortes MISSISSIPPI anordnen?

6. Berechnen Sie $20!$ sowie $\binom{9}{7}$ mit R!

2.1.3 Kombination

Definition:

Eine **Kombination** ist eine Auswahl von Objekten aus einer gegebenen Grundmenge, die (im Gegensatz zur Permutation) **nicht alle** Objekte der Grundmenge enthalten muss und bei der (ebenfalls im Gegensatz zur Permutation) die Reihenfolge unberücksichtigt bleibt. (von lateinisch *combinatio* "Zusammenfassung")

Darf jedes Objekt nur genau einmal auftreten, spricht man von einer **Kombination ohne Wiederholung**.

Können Objekte dabei mehrfach ausgewählt werden, so spricht man von einer **Kombination mit Wiederholung**.

Kombinationen ohne Wiederholung

Einer Urne mit n verschiedenen Kugeln entnimmt man k Kugeln ohne Zurücklegen. Die Reihenfolge der gezogenen Kugeln soll dabei ohne Bedeutung sein. Eine solche ungeordnete Stichprobe von k Elementen heißt eine Kombination k -ter Ordnung ohne Wiederholung (siehe auch Binomialkoeffizient: n über k , bzw. k aus n).

Die Anzahl der Möglichkeiten aus n verschiedenen Kugeln k Kugeln **ohne** Zurücklegen und ohne Berücksichtigung der Reihenfolge zu ziehen beträgt $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Beispiele:

1. Befinden sich in einer Urne drei verschiedenfarbige Kugeln (eine weiße, eine graue und eine schwarze) und ziehen wir nacheinander wahllos zwei Kugeln ohne Zurücklegen, und spielt dabei die Reihenfolge keine Rolle, so sind folgende 3 Ziehungen möglich:

$$\Rightarrow \binom{n}{k} =$$

2. Lotto 6 aus 49: $\binom{n}{k} =$

Kombinationen mit Wiederholung

Einer Urne mit n verschiedenen Kugeln entnimmt man k Kugeln mit Zurücklegen. Die Reihenfolge der gezogenen Kugeln soll dabei ohne Bedeutung sein. Eine solche ungeordnete Stichprobe von k Elementen heißt eine Kombination k -ter Ordnung mit Wiederholung.

Die Anzahl der Möglichkeiten aus n verschiedenen Kugeln k Kugeln **mit** Zurücklegen ohne Berücksichtigung der Reihenfolge zu ziehen:
$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Beispiel:

Befinden sich in einer Urne drei verschiedenfarbige Kugeln (eine weiße, eine graue und eine schwarze) und ziehen wir nacheinander wahllos zwei Kugeln mit Zurücklegen, und spielt dabei die Reihenfolge keine Rolle, so sind folgende 6 Ziehungen möglich:

$$\Rightarrow \frac{(n+k-1)!}{k!(n-1)!} =$$

Aufgaben:

1. Einer Warenlieferung von 12 LED-Lampen soll zu Kontrollzwecken eine Stichprobe von 3 LED entnommen werden.
Wie viele verschiedene Stichproben sind dabei möglich?

2. Für eine Parallelschaltung von 3 Widerständen stehen 5 verschiedene ohmsche Widerstände R_1, R_2, R_3, R_4, R_5 zur Verfügung.
Wieviele verschiedenen Schaltmöglichkeiten gibt es, wenn jeder der 5 Widerstände höchstens einmal verwendet werden darf?

3. In einer Tüte mit Gummibärchen befinden sich ausreichend viele Gummibärchen in 5 verschiedenen Farben.
Wieviele Kombinationen gibt es, wenn man aus der Tüte 5 Gummibärchen herausholt?

4. Aus einer Urne mit fünf nummerierten Kugeln wird dreimal eine Kugel gezogen und jeweils wieder zurückgelegt. Man kann also bei allen drei Ziehungen immer aus fünf Kugeln auswählen.
Wenn man die Reihenfolge der gezogenen Zahlen nicht berücksichtigt...

2.1.4 Variationen

Definition:

Eine Variation oder geordnete Stichprobe ist eine Auswahl von Objekten in einer bestimmten Reihenfolge.

Darf jedes Objekt nur einmal auftreten, spricht man von einer **Variation ohne Wiederholung**.

Können Objekte dabei mehrfach ausgewählt werden, so spricht man von einer **Variation mit Wiederholung**.

Eine **Variation** ist eine Auswahl von k Objekten aus einer Menge von n Objekten, wobei die Reihenfolge der Auswahl eine Rolle spielt. Werden alle verfügbaren Objekte ausgewählt, gilt also $k = n$, so spricht man von einer **Permutation**.

Spielt bei der Auswahl der Objekte die Reihenfolge keine Rolle, spricht man von einer **Kombination**.

Variation ohne Wiederholung

Die Anzahl der Möglichkeiten aus n verschiedenen Kugeln, k Kugeln **ohne** Zurücklegen aber unter Berücksichtigung der Reihenfolge zu ziehen beträgt

$$\frac{n!}{(n-k)!} = n(n-1) \cdot \dots \cdot (n-k+1)$$

Beispiel:

In einer Urne befinden sich eine weiße, eine graue und eine schwarze Kugel. Zieht man nacheinander zwei Kugeln ohne Zurücklegen aber mit Berücksichtigung der Reihenfolge, so ergeben sich folgende 6 Möglichkeiten:

$$\Rightarrow \frac{n!}{(n-k)!} =$$

Variation mit Wiederholung

Die Anzahl der Möglichkeiten aus n verschiedenen Kugeln, k Kugeln **mit** Zurücklegen aber unter Berücksichtigung der Reihenfolge zu ziehen beträgt n^k

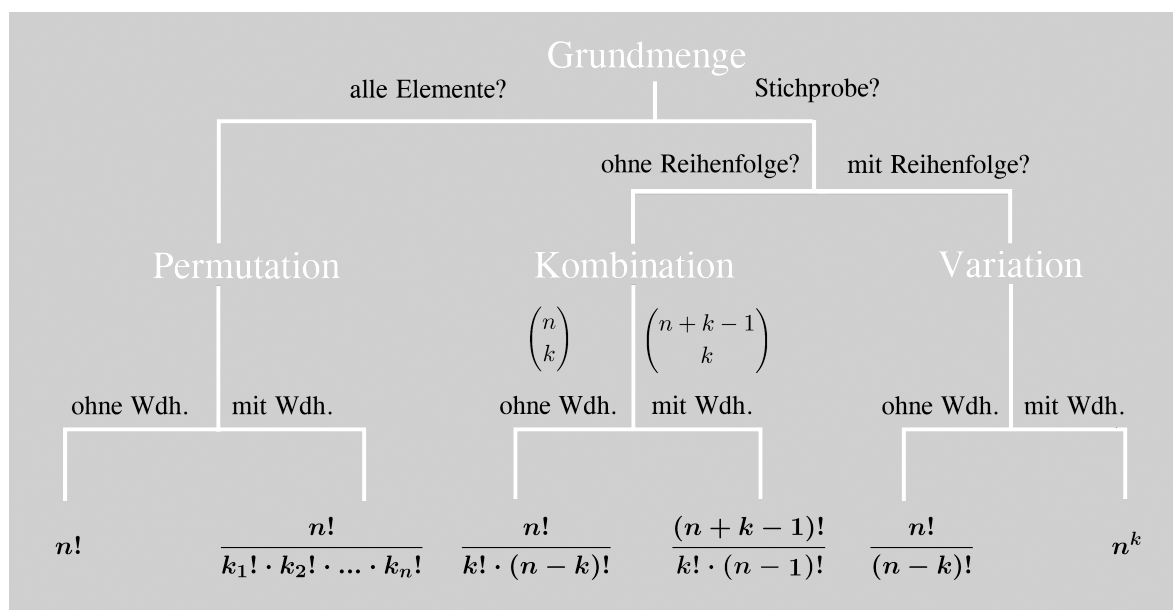
Beispiel:

In einer Urne befinden sich eine weiße, eine graue und eine schwarze Kugel. Zieht man nacheinander zwei Kugeln mit Zurücklegen aber mit Berücksichtigung der Reihenfolge, so ergeben sich folgende 9 Möglichkeiten:

$$\Rightarrow n^k =$$

Beispiel:

Wie viele verschiedene Würfe sind generell mit drei Würfeln möglich?

2.2 Kombinatorik Formeln**Aufgaben:**

1. Auf einem Regal sollen 10 verschiedene Bücher angeordnet werden.
Wie viele Möglichkeiten gibt es dafür?
2. Auf einem Regal sollen 10 Bibeln angeordnet werden, davon sind 5 Bibeln von der selben Ausgabe.
Wie viele Möglichkeiten gibt es dafür?

3. In einer Urne liegen $n = 3$ verschieden farbige Kugeln. Es werden wahllos $k = 2$ Kugeln gezogen
- ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge der Ziehung.
Wieviele Möglichkeiten gibt es?
 - die Kugeln werden jeweils zurückgelegt, aber die Reihenfolge der Ziehung wird nicht beachtet.
Wieviele Möglichkeiten gibt es?
 - ohne Zurücklegen werden sie in der Reihenfolge der Ziehung angeordnet.
Wieviele Möglichkeiten gibt es?
 - die Kugeln werden jeweils zurückgelegt, und ihre Reihenfolge der Ziehung beachtet.
Wieviele Möglichkeiten gibt es?
4. Beim Pferdetoto muss in der sogenannten Dreierwette der Zieleinlauf der ersten $k = 3$ Pferde in der richtigen Reihenfolge vorausgesagt werden.
Wieviele verschiedene Dreier-Wetten sind möglich, wenn $n = 10$ Pferde starten?
5. Beim gleichzeitigen Wurf zweier unterschiedlich gekennzeichnete Würfel interessieren die verschiedenen Augenpaare.
- a) Wie viele Augenpaare sind insgesamt möglich?
 - b) Wie viele Augensummen sind insgesamt möglich?

2.3 Wahrscheinlichkeitsrechnung

2.3.1 Grundbegriffe

Begriff	Beschreibung
Zufallsexperiment	Ein Zufallsexperiment ist ein Vorgang, der nach einem festgelegten Plan abläuft, dessen Ergebnis aber nicht vorhersehbar ist und bei dem alle möglichen Ergebnisse im Voraus bekannt sind.
Ergebnis ω	Der mögliche Ausgang eines Zufallsexperiments heißt Ergebnis. Z.B. könnte das Ergebnis eines einzelnen Würfelwurfs $\omega = 1$ sein.
Ergebnismenge Ω	Enthält alle möglichen Ergebnisse eines modellierten Zufallsexperiments. Z.B. ist die Ergebnismenge des Würfelwurfs: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Mächtigkeit $ \Omega $	Die Mächtigkeit der Ergebnismenge ist die Anzahl ihrer Elemente.
Ereignis E (Großbuchstaben)	Jede Teilmenge der Ergebnismenge Ω heißt Ereignis und ist abhängig von der Fragestellung. Man sagt, ein Ereignis "tritt ein", wenn das Ergebnis eines Zufallsexperimentes ein Element dieses Ereignisses ist. Manchmal können auch mehrere Ereignisse gleichzeitig eintreten, z.B. dass man eine gerade Zahl würfelt, $E = \{2; 4; 6\}$.
Elementarereignis $\{\omega\}$	Ein Elementarereignis ist ein Ereignis mit einem Element.
Ereignismenge Σ	Die Ereignismenge ist die Menge aller Ereignisse abhängig von der Fragestellung und des Experiments (und nicht das gleiche wie die Ergebnismenge!). Ein Ereignis kann nämlich auch unmöglich sein und ist deshalb nicht Teilmenge von Ω .
Ereignisraum $\mathcal{P}(\Omega)$	Menge aller prinzipiell möglichen Ereignisse , unabhängig von der konkreten Fragestellung (Hasse-Diagramm).
Maximale Mächtigkeit $ \mathcal{P}(\Omega) $	Die maximale Mächtigkeit lässt sich berechnen durch: 2 hoch die Anzahl an möglichen Ergebnissen. $ \mathcal{P}(\Omega) = 2^{ \Omega }$.

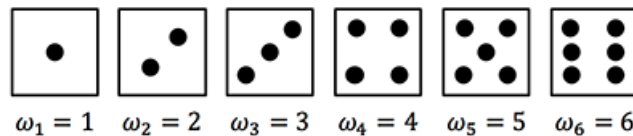
Anmerkungen:

- Die leere Menge $\emptyset, \{\}$ ist ebenfalls ein Ereignis. Es tritt nie ein und heißt daher **unmögliches Ereignis**.
- Andererseits ist auch Ω selbst ein Ereignis. Es tritt immer ein und heißt **sicheres Ereignis**.
- Die Potenzmenge von Ω , geschrieben als $\mathcal{P}(\Omega)$, ist eine Menge, deren Elemente alle Teilmengen von Ω sind.
- Insbesondere ist auch die leere Menge $\emptyset$ und die Ergebnismenge Ω selbst Teil von $\mathcal{P}(\Omega)$.
- $\mathcal{P}(\Omega)$ ist verwandt mit dem Binomialkoeffizienten.

Die Anzahl der Mengen mit **k Elementen** in der Potenzmenge, einer Menge mit **n Elementen** ist gleich dem **Binomialkoeffizient** $\binom{n}{k}$.

Beispiele:1. Zufallsexperiment: **Wurf eines homogenen Würfels**

Ergebnisse	$\omega = 1, 2, 3, 4, 5, 6$
Ergebnismenge	$\Omega = \{1, 2, 3, 4, 5, 6\}$
Mächtigkeit	$ \Omega = 6$
Ereignisse	$E = \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 3, 5\}, \{2, 4, 6\}, \dots, \{\Omega\}$
Elementarereignisse	$\{\omega\} = \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$
Ereignisraum	$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 3, 5\}, \{2, 4, 6\}, \dots, \{\Omega\}\}$
Maximale Mächtigkeit	$ \mathcal{P}(\Omega) = 2^6 = 64$

2. Zufallsexperiment: **Wurf einer Münze**

Ergebnisse	$\omega =$
Ergebnismenge	$\Omega =$
Mächtigkeit	$ \Omega =$
Ereignisse	$E =$
Elementarereignisse	$\{\omega\} =$
Ereignisraum	$\mathcal{P}(\Omega) =$
Maximale Mächtigkeit	$ \mathcal{P}(\Omega) =$

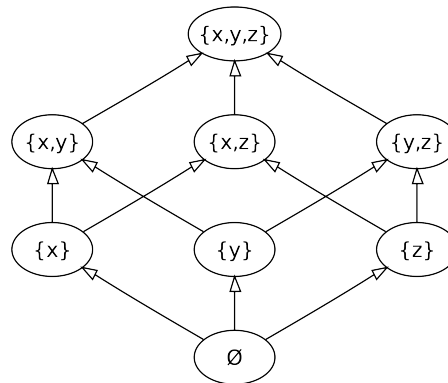
3. Zufallsexperiment: **Ziehen einer Kugel aus 3 Kugeln, die mit 1, 2 und 3 beschriftet sind:**

Ergebnisse	$\omega =$
Ergebnismenge	$\Omega =$
Mächtigkeit	$ \Omega =$
Ereignisse	$E =$
Elementarereignisse	$\{\omega\} =$
Ereignisraum	$\mathcal{P}(\Omega) =$
Maximale Mächtigkeit	$ \mathcal{P}(\Omega) =$

4. Sei X eine Menge, die wie folgt definiert ist: $X = \{x, y, z\}$.

Dann besteht der Ereignisraum (Potenzmenge) $\mathcal{P}(X)$ aus folgenden Elementen:

1. $\emptyset$
2. $\{x\}$
3. $\{y\}$
4. $\{z\}$
5. $\{x, y\}$
6. $\{x, z\}$
7. $\{y, z\}$
8. $\{x, y, z\}$



Hasse-Diagramm

$$\mathcal{P}(X) = \left\{ \emptyset, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\} \right\}$$

Mit $|X| = 3$ folgt, dass die Menge $2^3 = 8$ Elemente enthält.

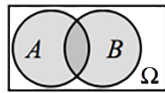
Die Verwandtschaft zum Binomialkoeffizient mit $\binom{n}{k}$ wird leicht ersichtlich:

- $\binom{3}{0} = 1$ Menge mit 0 Elementen (leere Menge)
 - $\binom{3}{1} = 3$ Menge mit 1 Elementen
 - $\binom{3}{2} = 3$ Menge mit 2 Elementen
 - $\binom{3}{3} = 1$ Menge mit 3 Elementen
-
- $\Sigma = 8$

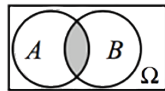
2.3.2 Verknüpfung von Ereignissen

Durch Verknüpfung von Ereignissen entstehen zusammengesetzte Ereignisse. Diese werden häufig anhand von Venn-Diagrammen veranschaulicht. Letztere bestehen aus einem Rechteck, in dem die Ausgangsereignisse (Mengen $A, B, \dots$) als Kreise oder Ellipsen dargestellt sind.

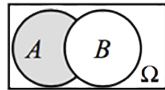
Für zwei Ereignisse A und B gibt es folgende Verknüpfungen:



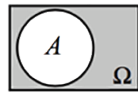
- Vereinigungsereignis: $A \cup B := \{x | x \in A \vee x \in B\}$
A tritt ein oder B tritt ein.



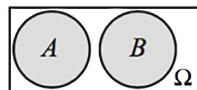
- Durchschnittsereignis: $A \cap B := \{x | x \in A \wedge x \in B\}$
A tritt ein und B tritt ein.



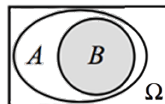
- Differenzereignis: $A \setminus B := \{x | x \in A \wedge x \notin B\}$
A tritt ein, aber B tritt nicht ein.



- Komplementärereignis: $\bar{A} = \Omega \setminus A$
A tritt nicht ein. (auch Gegenereignis genannt.)



- Disjunkte Ereignisse: $A \cap B := \emptyset$
Entweder tritt A oder B ein oder keins von beiden.



- Teilmenge: $B \subset A$ (B ist Teilmenge von A)

Beispiele:

1. Eine Zahl größer als 3 zu würfeln, wird mit dem Ereignis $A = \{4, 5, 6\}$ und eine ungerade Zahl zu würfeln mit dem Ereignis $B = \{1, 3, 5\}$ gekennzeichnet.

Das Ereignis eine Zahl größer als 3 oder eine ungerade Zahl zu würfeln, ist demzufolge:

$$A \cup B =$$

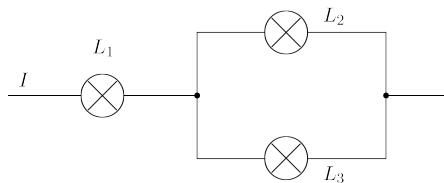
2. Eine Zahl kleiner als 4 zu würfeln, wird mit dem Ereignis $A = \{1, 2, 3\}$ und eine Zahl größer als 2 zu würfeln mit dem Ereignis $B = \{3, 4, 5, 6\}$ gekennzeichnet.

Das Ereignis eine Zahl kleiner als 4 und größer als 2 zu würfeln, ist demzufolge:

$$A \cap B =$$

3. Beim Wurf einer Münze sei A das Ereignis Zahl liegt oben, dann ist die Komplementärmenge $\bar{A}$ das Ereignis Kopf liegt oben bzw. $\bar{A} = \Omega - A$.

4. Der angegebene Stromkreis enthält drei Glühlämpchen L_1 , L_2 , L_3 .



Mit A_i werde das Ereignis bezeichnet, dass das i -te Glühlämpchen durchbrennt. Wie lässt sich dann das Ereignis B (Unterbrechung des Stromkreises durch A_i) beschreiben?

$B =$

5. **Zufallsexperiment:** Eine Münze wird drei Mal geworfen:

Wir wählen

$$\Omega = \{KKK, KKZ, KZK, ZKK, KZZ, ZKZ, ZZK, ZZZ\}.$$

Wobei ein K (Z) an Position j anzeigt, dass beim j -ten Wurf Kopf (Zahl) fällt.

Es sei A_j das Ereignis "Im j -ten Wurf fällt Kopf".

Als Teilmenge von Ω ist dann Beispielsweise $A_1 = \{KKK, KKZ, KZK, KZZ\}$.

Die Menge $A_1 \cup A_2 \cup A_3$ ist das Ereignis "Es fällt mindestens einmal Kopf", während

$A_1 \cap A_2 \cap A_3$ das einelementige Ereignis $\{KKK\}$ bezeichnet.

Das Ereignis "Es fällt mindestens zweimal Kopf" lässt sich schreiben als

$$(A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3)$$

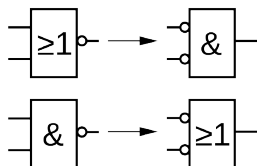
De Morgansche Gesetze

Für zwei beliebige Mengen A und B gilt:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

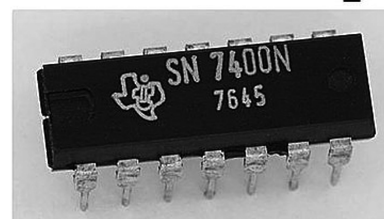
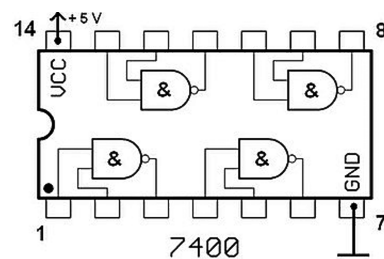
$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

Die De-Morganschen Gesetze haben wichtige Anwendungen in der diskreten Mathematik, der Elektrotechnik, der Physik und der Informatik. Insbesondere werden die De-Morganschen Gesetze beim Entwurf von digitalen Schaltungen genutzt, um die Typen der verwendeten logischen Schaltelemente gegeneinander auszutauschen oder Bauteile einzusparen.



De Morgansches Gesetz mit Logikgattern dargestellt

A	B	$\bar{A}$	$\bar{B}$	$A+B$	$\overline{A+B}$	$\bar{A} \cdot \bar{B}$
0	0	1	1	0	1	1
0	1	1	0	1	0	0
1	0	0	1	1	0	0
1	1	0	0	1	0	0



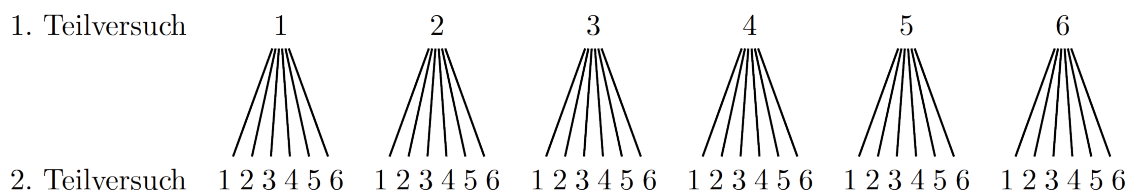
2.3.3 Zusammengesetzte Versuche

Im Folgenden betrachten wir Versuche, die aus Teilversuchen zusammengesetzt sind. Für das Weitere ist es wichtig, eine vollständige Übersicht der **Anzahl an Ausgangsmöglichkeiten** bei solchen zusammengesetzten Versuchen zu erhalten.

Beispiel 1:

Wenn wir zwei Mal hintereinander einen Würfel werfen, handelt es sich dabei um einen zusammengesetzten Versuch. Es treten folgende Möglichkeiten auf:

1. Wurf hat 6 Möglichkeiten mit Stichprobenraum $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$
2. Wurf hat 6 Möglichkeiten mit Stichprobenraum $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$



Dieser **Ereignisbaum** zeigt, dass zu jeder Zahl im ersten Teilversuch jede Zahl im zweiten Teilversuch kombiniert werden kann. Daraus ergeben sich $6 \cdot 6 = 36$ Ausgangsmöglichkeiten beim zusammengesetzten Versuch.

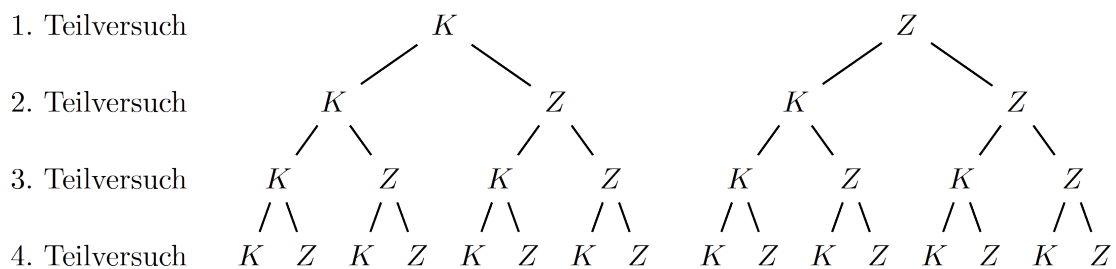
Der Stichprobenraum beim zweimaligen Würfeln ist demzufolge

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (1, 6), (2, 1), (2, 2), (2, 3), \dots, (6, 6)\}.$$

Beispiel 2:

Wenn wir vier Mal hintereinander eine Münze werfen, haben wir bei jedem Wurf die Möglichkeit Kopf K oder Zahl Z zu erhalten.

Eine vollständige Übersicht erhalten wir wiederum mit dem Ereignisbaum.



Aus diesem Ereignisbaum ergeben sich $2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$ Ereignismöglichkeiten (Ausgangsmöglichkeiten) beim zusammengesetzten Versuch. Der Stichprobenraum ist demzufolge

$$\Omega = \{(K, K, K, K), (K, K, K, Z), (K, K, Z, K), \dots, (Z, Z, Z, Z)\}.$$

Dieses Prinzip lässt sich verallgemeinern.

Produktregel

Besteht ein zusammengesetzter Versuch aus m unabhängigen Teilversuchen mit jeweils $n_1, n_2, n_3, \dots, n_m$ Möglichkeiten, so besitzt der zusammengesetzte Versuch

$$n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_m \quad \text{Ausgangsmöglichkeiten.}$$

2.4 Wahrscheinlichkeit

Die Wahrscheinlichkeit ist ein allgemeines Maß der Erwartung für ein unsicheres Ereignis. Auf der einen Seite sollen Vorhersagen (Prognosen) über den Ausgang zukünftiger Ereignisse gemacht werden. Auf der anderen Seite soll aber auch bei bereits eingetretenen Ereignissen beurteilt werden, wie gewöhnlich oder ungewöhnlich sie sind.

Die **Wahrscheinlichkeit** ist eine Abbildung P (*engl. probability*) von den Ereignismengen in das Intervall $[0,1]$.

P ist ein Wahrscheinlichkeitsmaß, das folgende **3 Kolmogorow-Axiome** erfüllen muss:

1. Für jedes Ereignis A bei einem Experiment ist die Wahrscheinlichkeit eine reelle Zahl zwischen 0 und 1: $0 \leq P(A) \leq 1$
2. Das sichere Ereignis Ω bei einem Experiment hat die Wahrscheinlichkeit 1 (Normierung): $P(\Omega) = 1$
3. Die Wahrscheinlichkeit einer Vereinigung abzählbar vieler sich gegenseitig ausschließender (**inkompatibler**) Ereignisse entspricht der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse.
Inkompatible Ereignisse sind **disjunkte** Mengen
$$P(A_1 \cup A_2 \dots) = \sum_{i=1}^{\infty} P(A_i)$$

Folgerungen:

4. $P(\bar{A}) = 1 - P(A)$ (Gegenwahrscheinlichkeit)

Beweis: A und $\bar{A}$ sind **disjunkte** Mengen mit $A \cup \bar{A} = \Omega$

Nach dem 3. Axiom folgt $P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(\Omega) = 1$ und somit:

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

5. $P(\{\}) = 0$ (Unmögliches Ereignis)

Beweis: Da $\{\} = \bar{\Omega}$ folgt $P(\{\}) = 1 - P(\Omega) = 1 - 1 = 0$

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

Beweis: Es ist $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ und diese Mengen sind disjunkt. Somit ist nach dem 3. Axiom: $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B)$

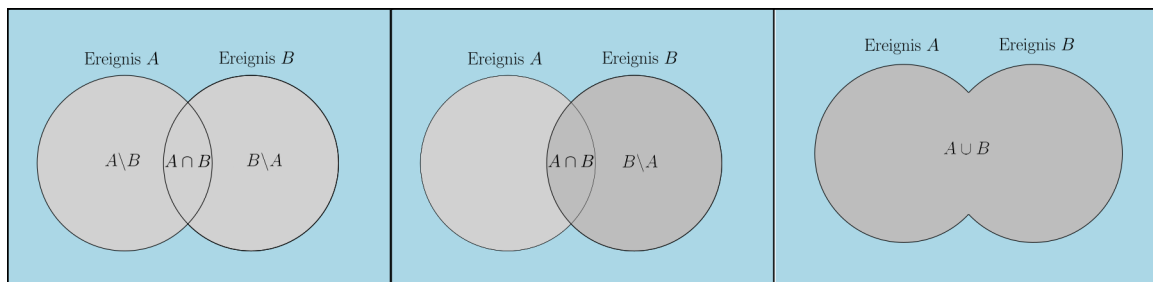
A und B lassen sich jeweils durch die disjunkten Mengen darstellen:

$A = (A \setminus B) \cup (A \cap B)$ sowie $B = (B \setminus A) \cup (A \cap B)$, so dass gilt:

$$P(A) = P(A \setminus B) + P(A \cap B) \quad \text{und} \quad P(B) = P(B \setminus A) + P(A \cap B)$$

Daraus ergibt sich insgesamt:

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) + P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$



2.4.1 Der Additionssatz

Sind die Ereignisse A und B elementfremd, d.h., es gilt $A \cap B = \emptyset$, dann gilt für die Wahrscheinlichkeiten

$$P(A \cup B) = P(A) + P(B).$$

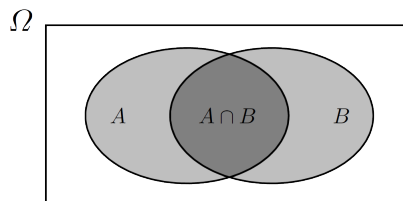
Sind sie **nicht** elementfremd, dann gilt die allgemeinere Version

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

Die folgenden Beispiele illustrieren den Additionssatz.

Beispiele:

1. Wir schießen zufällig auf eine rechteckige Zielscheibe und fragen nach der Wahrscheinlichkeit, entweder die Fläche A oder B zu treffen.



Die Wahrscheinlichkeit $P(A)$ ins Gebiet A zu treffen ist nichts anderes als das Verhältnis des Flächeninhalts von A zum gesamten Flächeninhalt von Ω und analog für die Fläche B . Somit folgt

$$P(A) = \frac{\text{area}(A)}{\text{area}(\Omega)} \quad \text{und} \quad P(B) = \frac{\text{area}(B)}{\text{area}(\Omega)}$$

Mit dem Additionssatz folgt nun

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{\text{area}(A) + \text{area}(B) - \text{area}(A \cap B)}{\text{area}(\Omega)}$$

Wir sehen, dass die Fläche der Schnittmenge abgezogen werden muss, da sie sonst doppelt gezählt würde

2. Wir werfen einen fairen Würfel und betrachten die Ereignisse $A = \{1, 3, 5\}$ und $B = \{1, 2, 3\}$. Es gilt $A \cup B = \{1, 2, 3, 5\}$ und $A \cap B = \{1, 3\}$. Uns interessiert die Wahrscheinlichkeit dafür, dass das Ereignis $A \cup B$ eintritt.

2.4.2 Laplace-Experiment

Zufallsexperimente, bei denen nur endlich viele Elementarereignisse möglich und diese gleichwahrscheinlich sind, (wie z.B. beim Werfen einer idealen Münze) nennt man ein Laplace-Experiment.

Beim Laplace-Experiment mit der Ergebnismenge $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ besitzen alle Elementarereignisse $\{\omega_i\}$ die gleiche Wahrscheinlichkeit

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

Die Wahrscheinlichkeit eines Ereignisses A ist dann gegeben durch

$$P(A) = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega} = \frac{\text{Zahl der günstigen Fälle}}{\text{Zahl der möglichen Fälle}} = \frac{|A|}{|\Omega|}$$

Beispiele:

1. Wie groß ist die Wahrscheinlichkeit für das Ereignis A , mit einem idealen Würfel eine "2" zu würfeln?

Es gibt $n = |\Omega| = 6$ mögliche Fälle und $|A| = 1$ günstige Fälle. Damit ist die Wahrscheinlichkeit:

2. Wie groß ist die Chance, beim Zahlenlotto "6 aus 49" sechs Richtige (A_6) zu tippen?

3. Wie groß ist die Wahrscheinlichkeit dafür, dass unter k Personen mindestens 2 am gleichen Tag Geburtstag haben (die Geburtsjahre müssen nicht übereinstimmen)?

Man betrachtet einfacherweise das komplementäre Ereignis $\bar{A}$: alle k Personen haben an verschiedenen Tagen Geburtstag.

Dann ist

$$P(\bar{A}) = \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - k + 1)}{365^k} \quad \text{und somit:}$$

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - k + 1)}{365^k}$$

Dies kann man für einige Werte von k berechnen:

k	10	20	30	60
$P(A)$	0.117	0.411	0.706	0.994

2.4.3 Mehrstufige Zufallsexperimente

Die Baumdiagramme können sehr geschickt zur Berechnung von Wahrscheinlichkeiten von Ereignissen mehrstufiger Zufallsexperimente benutzt werden.

Dazu werden an den Zweigen die jeweiligen Wahrscheinlichkeiten eingetragen, mit denen das zum Zweig gehörige Ereignis des Zufallsexperimentes eintritt. Diese Wahrscheinlichkeiten nennt man kurz Zweigwahrscheinlichkeiten.

Ein Baumdiagramm, das Zweigwahrscheinlichkeiten enthält, nennt man auch kurz Wahrscheinlichkeitsbaum. Üblicherweise gibt man alle Zweigwahrscheinlichkeiten entweder komplett als Brüche oder Dezimalzahlen an.

Grundlegend ist aus der Aufgabenstellung zu entnehmen, ob es sich bei dem Zufallsexperiment um ein Experiment **mit** oder **ohne** Zurücklegen handelt.

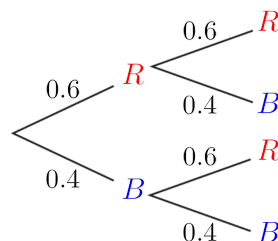
Zufallsexperiment „Mit Zurücklegen“

In einer Urne befinden sich 60 rote Kugeln und 40 blaue Kugeln und wir ziehen zwei Kugeln **mit Zurücklegen**.

Mit der Laplace-Wahrscheinlichkeit erhält man die folgenden Wahrscheinlichkeiten:

$$P(R) = \frac{60}{100} = 0.6$$

$$P(B) = \frac{40}{100} = 0.4$$



Erste Ziehung:

Wie man sehen kann hat man im **ersten** Zug jeweils die Chance eine rote oder eine blaue Kugel zu ziehen. Addiert man die Wahrscheinlichkeiten für beide Ereignisse, so erhält man als Summe eins: $P(\Omega) = 1$.

Zweite Ziehung:

Beim zweiten Zug hat man wieder die gleiche Chance eine rote oder eine blaue Kugel zu ziehen, da man die Kugeln wieder zurücklegt. Dementsprechend ist festzuhalten, dass beim Ziehen mit Zurücklegen bei jedem Zug die **gleichen** Eintrittswahrscheinlichkeiten vorliegen (Laplace-Wahrscheinlichkeit). Auch hier muss die Summe der Wahrscheinlichkeiten an den einzelnen Knoten 1 betragen.

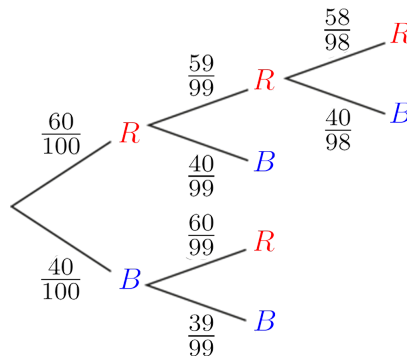
Zufallsexperiment „Ohne Zurücklegen“

In einer Urne befinden sich 60 rote Kugeln und 40 blaue Kugeln und wir ziehen zwei Kugeln **ohne Zurücklegen**.

Mit der Laplace-Wahrscheinlichkeit erhält man wieder die folgenden Wahrscheinlichkeiten:

$$P(R) = \frac{60}{100} = 0.6$$

$$P(B) = \frac{40}{100} = 0.4$$



Erste Ziehung:

Im Baumdiagramm sehen wir die Wahrscheinlichkeiten im **ersten** Zug eine rote oder eine blaue Kugel zu ziehen. Addiert man die Wahrscheinlichkeiten für beide Ereignisse, so erhält man als Summe eins: $P(\Omega)=1$.

Zweite Ziehung:

Im Gegensatz zum Ziehen mit Zurücklegen ändern sich die Wahrscheinlichkeiten beim Ziehen ohne Zurücklegen im zweiten Zug. Zieht man beispielsweise im ersten Zug eine rote Kugel, so hat man im zweiten Zug eine geringere Wahrscheinlichkeit eine rote Kugel zu ziehen.

Warum? Weil sich die Anzahl der günstigen und der möglichen Ereignisse (eine rote Kugel weniger) um 1 verringert.

Es befinden sich also nur noch 59 rote und insgesamt 99 Kugeln in der Urne. Die Wahrscheinlichkeit im zweiten Zug eine rote Kugel zu ziehen, ändert sich von $60/100$ auf $59/99$.

Merke:

Bei Zufallsexperimenten ohne Zurücklegen ist es sinnvoller, Brüche statt Dezimalzahlen für die Wahrscheinlichkeiten zu verwenden.

2.4.4 Wahrscheinlichkeit mit Pfadregeln

Um die Wahrscheinlichkeit eines Ereignisses zu erhalten, multipliziert man die Wahrscheinlichkeit entlang des Pfades, der dieses Ereignis beschreibt.

Wichtig: Die Pfadregel gilt bei jedem mehrstufigen Zufallsexperiment, gleichgültig, ob z.B. mit oder ohne Zurücklegen.

Vorgehen zur Ermittlung einer Wahrscheinlichkeit

- Baumdiagramm zeichnen
- Pfadregel anwenden

1. **Multiplikationsregel:** Die Wahrscheinlichkeiten eines **einzelnen** Ereignisses ist das Produkt der Wahrscheinlichkeiten entlang des Pfades, der zu diesem Ergebnis führt.
2. **Summenregel:** Die Wahrscheinlichkeit eines Ereignisses ist die Summe der Wahrscheinlichkeiten der Pfade, die zu diesem Ereignis gehören.

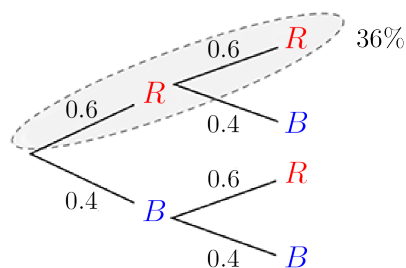
Beispiele:

1. In einer Urne befinden sich wieder 60 rote Kugeln und 40 blaue Kugeln.

- **Wahrscheinlichkeit für das Ziehen einer blauen oder roten Kugel mit Zurücklegen:**

⇒ Laplace-Wahrscheinlichkeit ⇒ $P(R) = 0.6$, $P(B) = 0.4$

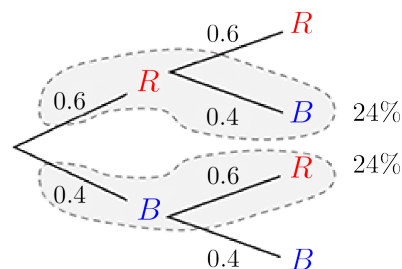
- **Wahrscheinlichkeit für das Ziehen von zwei roten Kugeln:**



⇒ Multiplikationsregel ⇒ $P(RR) =$

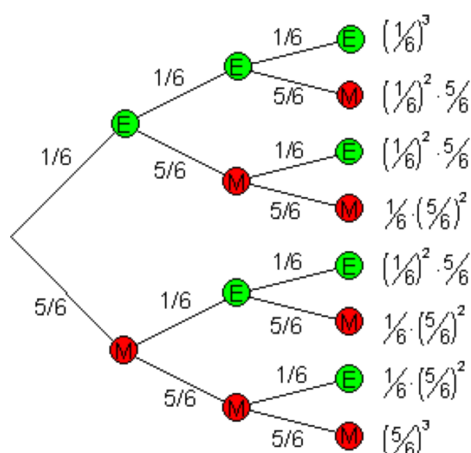
- **Wahrscheinlichkeit für das Ziehen einer roten und einer blauen Kugel:**

(Reihenfolge egal)



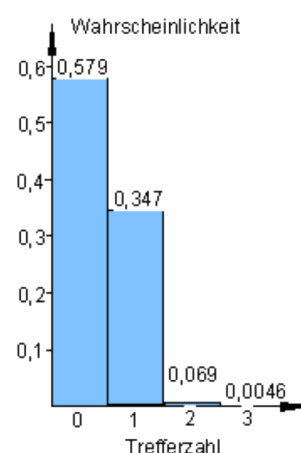
⇒ Summenregel ⇒ $P(RB) + P(BR) =$

2. Wir werfen einen Würfel dreimal nacheinander. Wir interessieren uns dabei für die Wahrscheinlichkeit, mit der die Zahl 6 bei diesem Versuch 0, 1, 2 oder 3 mal auftritt. E bedeutet dabei Erfolg oder Treffer, M bedeutet Misserfolg oder kein Treffer.



Treffer - zahl	Wahrschein- lichkeit
0	$1 \cdot \left(\frac{5}{6}\right)^3 \approx 0,579$
1	$3 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 \approx 0,347$
2	$3 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} \approx 0,069$
3	$1 \cdot \left(\frac{1}{6}\right)^3 \approx 0,0046$

Anzahl
der Pfade

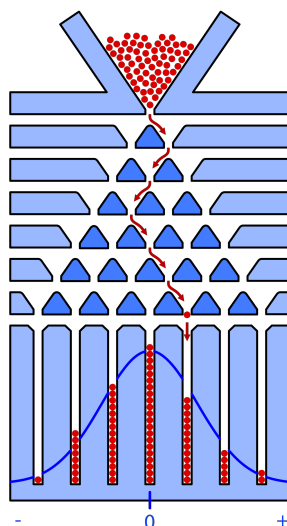


Man nennt eine solche Verteilung auch **hypergeometrische** Verteilung (siehe Kapitel 2.5: Wahrscheinlichkeitsverteilungen).

2.4.5 Galton-Brett

Ein Galton-Brett (nach Francis Galton), auch Zufallsbrett genannt, ist ein mechanisches Modell zur Demonstration und Veranschaulichung der Binominalverteilung.

Das Galton-Brett besteht aus einer regelmäßigen Anordnung von Hindernissen, an denen eine von oben eingeworfene Kugel jeweils nach links oder rechts abprallen kann. Nach dem Passieren der Hindernisse werden die Kugeln in Fächern zum Zählen aufgefangen.

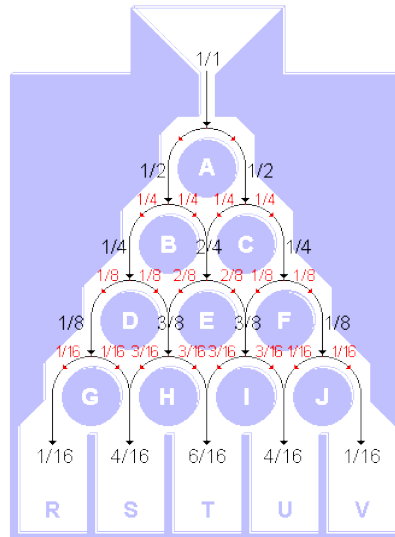


Das Galton-Brett simuliert z.B. ein physikalisches Messgerät, dessen Messwert verrauscht ist. Die horizontale Position der Kugel ist dabei der zu messende Wert, der am oberen Eingang noch exakt vorliegt, während er unten in einem der Fächer durch ein **Rauschsignal** verändert wurde.

Die Hindernisse symbolisieren kleine Störungen, die den Messwert positiv oder negativ beeinflussen. In der Summe können sie zu einer größeren Störung anwachsen, sich aber auch zu Null

addieren. Die Füllhöhen der Fächer geben am Ende Auskunft über die **Häufigkeitsverteilung** der verschiedenen Stärken der aufsummierten Störungen.

Jedes Aufprallen einer Kugel auf eines der Hindernisse ist ein **Bernoulli-Versuch**. Die beiden möglichen Ausgänge sind „Kugel fällt nach rechts“ ($X=1$) oder „Kugel fällt nach links“ ($X=0$).

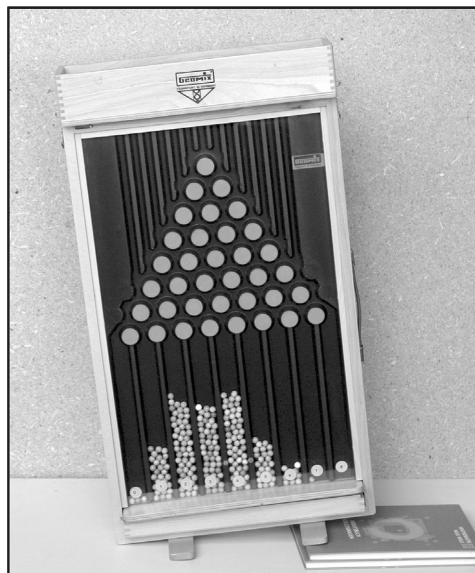


Bei symmetrischem Aufbau ist die Wahrscheinlichkeit, nach rechts zu fallen, $P(X = 1) = p = 1/2$ und die Wahrscheinlichkeit, nach links zu fallen, $P(X = 0) = q = 1 - p = 1/2$.

Indem die Kugel nach Passieren des ersten Hindernisses auf ein neues trifft, bei dem die gleichen Voraussetzungen gelten, wird hier ein weiterer Bernoulli-Versuch durchgeführt; das Durchlaufen des ganzen Gerätes ist also eine **mehrstufige Bernoulli-Kette**, wobei die Zahl der waagerechten Reihen von Hindernissen die Länge dieser Kette ist.

Im oben dargestellten rechten Bild handelt es sich demnach um eine 4-malige Wiederholung eines Bernoulli-Versuchs, d. h. eine Bernoulli-Kette der Länge 4.

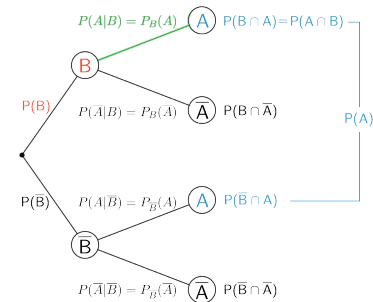
Durch **unsymmetrischen** Aufbau oder durch Schiefstellen des Brettes kann man auch einen anderen Wert als $p = 1/2$ erreichen, wobei aber natürlich weiterhin $q = 1 - p$ ist, denn die Kugeln, die nicht nach rechts fallen, fallen nach links.



2.4.6 Bedingte Wahrscheinlichkeit

Bei der Berechnung von Wahrscheinlichkeiten bei Laplace-Experimenten kann man manchmal eine gegebene **Zusatzinformation** B nutzen. Die mit der Vorinformation B berechnete Wahrscheinlichkeit wird **bedingte Wahrscheinlichkeit** von A unter der Bedingung B genannt und mit $P(A|B)$ (oft auch $P_B(A)$) abgekürzt.

$$P(A|B) = P_B(A) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$



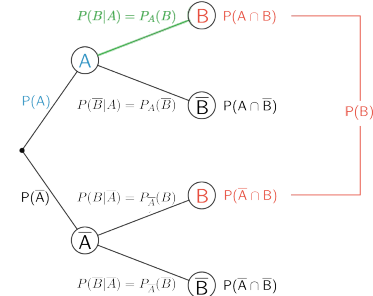
Diese Formel bedeutet:

$$P(\text{Was suchen wir} \mid \text{Was wissen wir bereits bzw. was ist schon eingetreten})$$

In der Aufgabenstellung wird die Bedingung, also das was wir wissen, immer klar durch die Phrasen/Wörter „unter der Bedingung“, „wenn“, „obgleich“.

Analog ist die bedingte Wahrscheinlichkeit $P(B|A)$

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



Beispiel:

Die Wahrscheinlichkeit, aus einem Skatblatt eine Herz-Karte zu ziehen (Ereignis A), beträgt $1/4$, denn es gibt 32 Karten (mögliche Fälle) und darunter 8 Herzkarten (günstige Fälle). Dann ist

$$P(A) = \frac{8}{32} = \frac{1}{4}$$

Wenn nun aber bereits das Ereignis B , die Karte ist „rot“ eingetreten ist, man also nur noch die Auswahl 16 rote Karten (mögliche Fälle) unter den 8 Herz-Karten (günstige Fälle) hat, dann ist

$$P(A|B) = \frac{8}{16} = \frac{1}{2} \quad \text{bzw.} \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2} \quad \text{mit} \quad P(A \cap B) = \frac{4}{16} = \frac{1}{4}$$

Satz von Bayes

Zwischen den bedingten Wahrscheinlichkeiten $P(A|B)$ und $P(B|A)$ besteht die auch als Satz von Bayes bezeichnete Beziehung:

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

Der Satz von Bayes wird verwendet, wenn man das Ergebnis schon kennt und die Wahrscheinlichkeiten für eine mögliche Ursache herausfinden möchte. Wenn man die Gleichung näher betrachtet, erkennt man obige Vorschrift $P(A|B) = P(\text{Was suchen wir} \mid \text{Was wissen wir})$.

Ferner steht im Nenner der Satz der **totalen** Wahrscheinlichkeit, welcher die Summe der möglichen Ausgänge darstellt.

Beispiel:

Bei der Herstellung von Festplatten sind 1% der Speichermedien defekt. Die Wahrscheinlichkeit der Entdeckung einer defekten Festplatte über einen Test durch die QS beträgt 86%.

Problem: Bei 10% der intakten Festplatten wird auch ein Fehler angezeigt.

Gesucht: Wahrscheinlichkeit, dass eine als defekt erkannte Festplatte auch defekt ist.

1. Festlegung der Ereignisse:

A : Festplatte defekt $\bar{A}$: Festplatte intakt

B : Test zeigt Defekt $\bar{B}$: Test zeigt keinen Defekt $\Rightarrow$ Gesucht: $P(A|B)$

2. Auswertung Gegebenheiten aus Aufgabe

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

$$3. \text{ Anwendung Formel: } P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})}$$

$$\Rightarrow P(A|B) = \text{—————} =$$

4. Bewertung:

Eine als defekt getestete Festplatte ist nur mit einer Wahrscheinlichkeit von 8% wirklich defekt! $\rightarrow$ Test verbessern!

2.4.7 Stochastische Unabhängigkeit

Zwei zufällige Ereignisse A und B werden als unabhängig oder auch als stochastisch unabhängig bezeichnet, wenn das Eintreten eines Ereignisses keinen Einfluss auf das andere Ereignis hat.

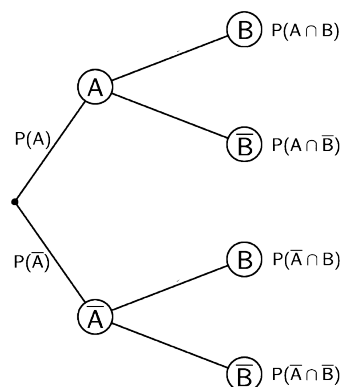
Zwei Ereignisse A, B mit $P(A) > 0$ und $P(B) > 0$ heißen unabhängig, wenn die Bedingung gilt:

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

Vierfeldertafel:

Eine Vierfeldertafel eignet sich zur Bestimmung von Wahrscheinlichkeiten der Verknüpfungen zweier Ereignisse A und B . Im folgenden ist eine allgemeine Vierfeldertafel mit Wahrscheinlichkeitsbaum dargestellt.

	A	$\bar{A}$	Σ
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
$\bar{B}$	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
Σ	$P(A)$	$P(\bar{A})$	1



Beispiel:

In einer Schule werden neben dem Unterricht ein Förderkurs (F) und eine Arbeitsgemeinschaft (AG) angeboten. 70% der Schüler besuchen keinen Förderkurs und 40% der Schüler keine AG . 25% der Schüler nehmen weder an dem Förderkurs noch an der AG teil.

Alle Informationen des Aufgabentextes werden zunächst übersichtlich im ersten Schritt in eine Vierfeldertafel geschrieben.

	F	$\bar{F}$	Σ
AG			
$\bar{AG}$			
Σ			

Wie groß ist nun die Wahrscheinlichkeit, dass...

1. ... ein Schüler den Förderkurs nicht wählt, unter der Bedingung, dass er auch keine AG besucht?

Aus der Aufgabenstellung muss man die Bedingung identifizieren, in dem man auf die Signalwörter („unter der Bedingung“, „wenn“, „obgleich“) achtet. Mit der Kenntnis, welches Ereignis bereits eingetreten ist, folgt:

$$P(\text{Was suchen wir} | \text{Was wissen wir}) = P(\bar{F} | \bar{AG})$$

Im nächsten Schritt berechnet man mit der Formel der bedingten Wahrscheinlichkeit die gesuchte Wahrscheinlichkeit.

$$P(\bar{F} | \bar{AG}) = \frac{P(\bar{F} \cap \bar{AG})}{P(\bar{AG})} =$$

2. ... ein Schüler die AG nicht wählt, unter der Bedingung, dass er auch keinen Förderkurs besucht?

$$P(\bar{AG} | \bar{F}) = \frac{P(\bar{AG} \cap \bar{F})}{P(\bar{F})} =$$

3. ... ein Schüler die AG wählt, unter der Bedingung, dass er schon den Förderkurs gewählt hat?

$$P(AG | F) = \frac{P(AG \cap F)}{P(F)} =$$

4. Sind die Ereignisse unabhängig?

Wir verwenden die Formel $P(A \cap B) = P(A) \cdot P(B)$ und setzen die für uns passenden Ereignisse ein. Es folgt:

$$\begin{aligned} P(F \cap AG) &= P(F) \cdot P(AG) \\ &= \\ &= \end{aligned}$$

In dem Beispiel sind die beiden Ereignisse voneinander abhängig, also nicht stochastisch unabhängig.

Demnach wählt ein Schüler den Förderkurs (Förderkurs nicht) abhängig davon ob er die AG (AG nicht) wählt.

KAPITEL 3

Wahrscheinlichkeitsverteilungen

Bis jetzt haben wir ganz allgemein Zufallsexperimente behandelt. Deren Ausgang waren entweder Zahlen (z.B. Augenzahl beim Würfel) oder “abstraktere” Dinge wie eine Kombination von K und Z (z.B. zweimaliger Wurf mit einer Münze).

In der Praxis sind Messungen, z.B. von einem physikalischen Versuch (Zufallsexperiment), in der Regel **Zahlen** (Werte).

Man führt für diesen Spezialfall den Begriff der **Zufallsvariable** ein. Oft weist man den verschiedenen “abstrakten” Ausgängen eines Zufallsexperiments einfach auch Zahlen bzw. Werte zu. In beiden Fällen haben wir zufällige Zahlen als Ausgänge.

3.1 Zufallsvariable

Als Zufallsvariable bezeichnet man eine Funktion $X : \Omega \rightarrow \mathbb{R}$, die den Elementar-Ereignissen eines Zufallsexperimentes reelle Werte zuordnet. Diese Werte werden als **Realisation** der Zufallsvariablen bezeichnet.

Zufallsvariablen werden üblicherweise mit einem Großbuchstaben bezeichnet, z.B. X . Es hat sich dieser etwas irreführende Begriff durchgesetzt, obwohl Zufallsvariablen eigentlich Funktionen sind.

Wir unterscheiden dabei zwischen diskreten und stetigen Verteilungen (bzw. Zufallsvariablen).

Diskrete Verteilung:

- Eine Zufallsvariable, die abzählbar viele Werte annehmen kann, heißt **diskret** (z.B. die Zahlen beim Würfeln mit einem Würfel).

Stetige Verteilung:

- Eine Zufallsvariable, die überabzählbar viele Werte annehmen kann, heißt **stetig** (z.B. die Lebensdauer eines elektronischen Bauteils, die in einem Intervall der reellen Zahlen jeden Wert annehmen kann).

Beispiel:

Beim Zufallsexperiment Münzwurf kann man eine (diskrete) Zufallsvariable X definieren, indem man dem Ergebnis $\omega = \text{“Kopf”}$ den Wert 0 und dem Ergebnis $\omega = \text{“Zahl”}$ den Wert 1 zuordnet:

$$X(\omega) = \begin{cases} 0 & \text{Kopf} \\ 1 & \text{Zahl} \end{cases}$$

3.2 Verteilungsfunktionen

Von Interesse ist die Frage, mit welchen Wahrscheinlichkeiten eine Zufallsvariable in welchen Bereichen liegt. Man spricht von der sogenannten **Wahrscheinlichkeitsverteilung** bzw. kurz von der **Verteilung** von X .

Was ist z.B. die Wahrscheinlichkeit, dass eine Druckfestigkeit ≤ 30 MPa ist oder im Intervall $[25, 30]$ MPa liegt? Oder was ist die Wahrscheinlichkeit, dass wir in einer Lieferung von 100 Bauteilen weniger als 5 defekte Teile vorfinden?

Wenn wir die **Verteilung einer Zufallsvariablen X** kennen, können wir auf jede beliebige solcher Fragen die entsprechende Antwort geben.

Wie wir später sehen werden, gibt es für die Modellierung von gewissen unsicheren Phänomenen bestimmte Verteilungen, die sich speziell gut dafür eignen.

Wenn man die wichtigsten Verteilungen kennt, kann man diese Sammlung als “Toolbox” verwenden. Man muss für die Modellierung von einem Phänomen dann einfach diejenige heraus suchen, die am besten passt.

Die **Verteilungsfunktion** einer Zufallsvariablen X ist definiert als

$$F_X(t) = P(X \leq t)$$

Die Verteilungsfunktion gibt die Wahrscheinlichkeit an, dass die Zufallsvariable X kleiner gleich als der fixe Wert t ist.

Häufig ist die Wahrscheinlichkeitsverteilung einer Zufallsvariablen nicht bekannt. Man führt dann eine Anzahl von Experimente durch, um diese Verteilung näherungsweise zu bestimmen.

Eigenschaften:

1. F_X ist monoton wachsend, d.h. aus $t_1 < t_2$ folgt $F(t_1) \leq F(t_2)$
2. F_X ist rechtsseitig stetig, d.h. $\lim_{h \rightarrow 0} F_X(t + h) = F_X(t)$
3. $\lim_{t \rightarrow -\infty} F_X(t) = 0$; $\lim_{t \rightarrow +\infty} F_X(t) = 1$

Im folgenden unterscheiden wir zwischen der **Verteilung** einer **stetigen** und einer **diskreten** Zufallsvariablen.

3.3 Diskrete Verteilungen

Nimmt die Zufallsvariable X nur diskrete Werte x_i ($i = 1, 2, \dots$) an, mit den Wahrscheinlichkeiten $P(X = x_i) = p_i$ ($i = 1, 2, \dots$), so ist die Verteilungsfunktion diskret und lautet:

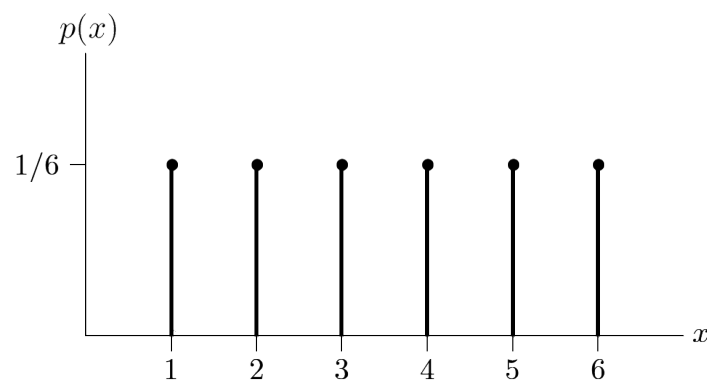
$$F(x) = \sum_{x_i \leq x} p_i$$

Beispiel (Würfelwurf):

Die zufälligen Elementarereignisse beim einmaligen Würfeln sind ω_i : Zahl i gewürfelt ($i = 1, \dots, 6$).

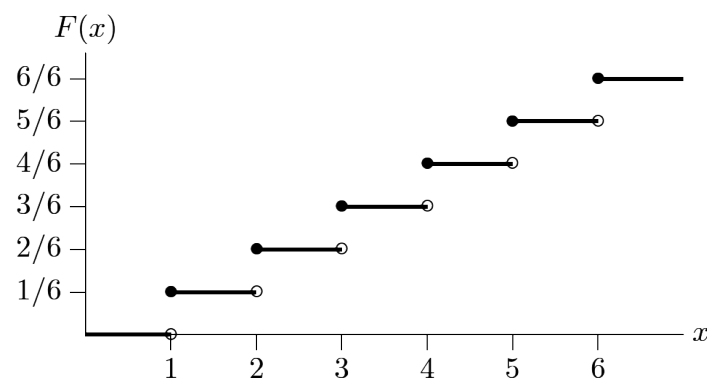
Die Zufallsvariable X kann die Werte $x_1 = 1, x_2 = 2, \dots, x_6 = 6$ annehmen, wobei $P(X = x_i) = 1/6$ gilt.

Zufallsvariable X für alle x_i , **Wahrscheinlichkeitsfunktion** $p(x)$ (bei diskreten Verteilungen):



Dann hat die **Verteilungsfunktion** $F(x)$ die Gestalt:

$$F(x) = \begin{cases} 0 & -\infty < x < 1 \\ 1/6 & 1 \leq x < 2 \\ 2/6 & 2 \leq x < 3 \\ 3/6 & \text{für } 3 \leq x < 4 \\ 4/6 & 4 \leq x < 5 \\ 5/6 & 5 \leq x < 6 \\ 1 & 6 \leq x < \infty \end{cases}$$



Kummulative Verteilungsfunktion beim einmaligen Würfeln:

- charakterisiert einen eingeschlossenen Wert,
- einen ausgeschlossenen Wert.

3.3.1 Kennwerte

Zur Charakterisierung einer Wahrscheinlichkeitsverteilung werden vor allem die Parameter

1. Mittelwert $\bar{x}$ oder Erwartungswert μ
2. Empirische Varianz (Stichprobenvarianz) s^2 oder Varianz σ^2
3. Empirische Standardabweichung s oder Standardabweichung σ

einer Zufallsvariablen verwendet, je nachdem ob es sich um eine Stichprobe mit endlichen Merkmalen, oder eine theoretische Verteilung (mit $n \rightarrow \infty$) handelt.

Erwartungswert, Varianz und Standardabweichung einer diskreten Zufallsvariablen

Die **diskrete** Zufallsvariable X mit der Wahrscheinlichkeitsfunktion $p(x)$ besitzt die folgenden Kennwerte

1. Erwartungswert $\mu = E(X) = \sum_i x_i \cdot p(x_i)$
2. Varianz $\sigma^2 = E((X - \mu)^2) = \sum_i (x_i - \mu)^2 \cdot p(x_i) = \text{Var}(X)$
3. Standardabweichung $\sigma = \sqrt{\text{Var}(X)}$

Beispiele:

1. Berechnen Sie den Erwartungswert der diskreten Zufallsvariablen X mit folgender Verteilung:

x_i	1	2	3	4
$p(x_i)$	1/8	3/8	3/8	1/8

2. Die diskrete Zufallsvariable des "Wurfs eines homogenen Würfels" ist gleichverteilt.

x_i	1	2	3	4	5	6
$p(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

Berechnen Sie den Erwartungswert.

3. Beim Wurf mit zwei homogenen Würfeln ist die diskrete Zufallsvariable " X =erzielte Augensumme" folgendermaßen verteilt.

x_i	2	3	4	5	6	7	8	9	10	11	12
$p(x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Berechnen Sie Erwartungswert, Varianz und Standardabweichung.

3.3.2 Bernoulli-Verteilung [`dbern(x,prob)`]

Die **Bernoulli-Verteilung** mit Parameter $p \in (0,1)$ ist die “einfachste” diskrete Verteilung. Hier kann X nur die Werte 0 oder 1 annehmen, d.h.

$$X = p(x) = \begin{cases} 1 & \text{Wahrscheinlichkeit } p \\ 0 & \text{Wahrscheinlichkeit } 1 - p \end{cases}$$

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

$$\text{Var}(X) = p \cdot (1 - p)$$

Man schreibt: $\mathbf{X} \sim \mathbf{B}(p)$, wobei das Symbol “ $\sim$ ” (Tilde) übersetzt wird als “ist verteilt wie”.

Definition:

Ein Zufallsexperiment mit nur zwei verschiedenen sich gegenseitig ausschließenden Ereignissen A und $\bar{A}$, mit den konstanten Wahrscheinlichkeiten

$P(A) = p$ und $P(\bar{A}) = 1 - p$ heißt **Bernoulli-Experiment**.

Beispiele:

1. Beim Münzwurf sind nur die beiden Ereignisse “Kopf” oder “Zahl” möglich. Bei einer homogenen Münze treten die Ereignisse mit den Wahrscheinlichkeiten

$$P(A) = p = \frac{1}{2} \text{ und } P(\bar{A}) = 1 - p \text{ auf.}$$

2. In einer Urne befinden sich 5 weiße und 3 schwarze Kugeln. Bei der zufälligen Ziehung einer Kugel sind nur die beiden sich gegenseitig ausschließenden Ereignisse:

A : Ziehung einer weißen Kugel

$\bar{A}$: Ziehung einer schwarzen Kugel

möglich. Sie treten mit den Wahrscheinlichkeiten

$$P(A) = \frac{5}{8} \text{ und } P(\bar{A}) = \frac{3}{8} \text{ auf. Es handelt sich also um ein Bernoulli Experiment.}$$

Man schreibt: $\mathbf{X} \sim \mathbf{Ber}\left(\frac{5}{8}\right)$ (Wahrscheinlichkeit für A , Ziehung einer weißen Kugel)

Definition:

Ein Bernoulli-Experiment, welches n -mal nacheinander ausgeführt wird, heißt **Bernoulli-Experiment vom Umfang n** (und führt zur Binomialverteilung).

3.3.3 Binomialverteilung `[dbinom(x,size,prob)]`

Bernoulli-Experimente vom Umfang n führen zur Binomial-Verteilung:

Betrachtet man beispielsweise eine Urne mit weißen und schwarzen Kugeln, wobei die Wahrscheinlichkeit beim Ziehen eine weiße Kugel zu erhalten p sei. Dann ist $q = 1 - p$ die Wahrscheinlichkeit, beim Ziehen eine schwarze Kugel zu erhalten.

Nach jeder Ziehung wird die Kugel **wieder zurückgelegt**, daher bleiben die Wahrscheinlichkeiten p und q für jede Ziehung **konstant**.

Wir ziehen nun nacheinander n Kugeln. Darunter seien dann genau x weiße und $n - x$ schwarze Kugeln. Diese Bernoulli-Kette der Länge n führe zu folgendem Ergebnis:

W W...W S S... S bzw. T T...T N N... N für Treffer bzw. Nichttreffer
 x -mal $(n - x)$ -mal x -mal $(n - x)$ -mal

Die Wahrscheinlichkeit für diese spezielle Realisierung der Zufallsvariable

$X = \text{Anzahl der gezogenen weißen Kugeln bei } n \text{ Ziehungen mit Zurücklegen}$ scheint **zunächst**:

$$P(X = x) = p^x \cdot q^{n-x}$$

Es sind jedoch noch weitere Realisierungen (Treffer) des Ereignisses $X = x$ möglich. Sie entstehen durch Permutation der n gezogenen Kugeln. Die Anzahl der Pfade ist $\binom{n}{x}$,

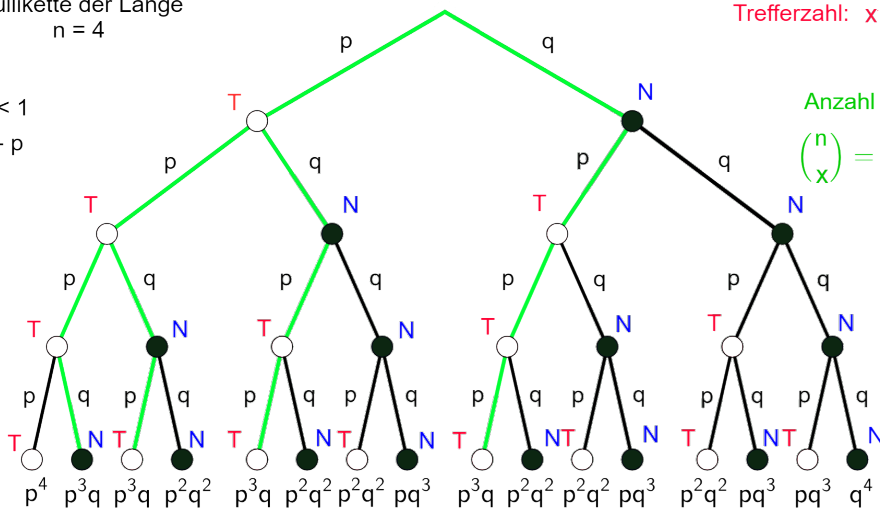
somit ist schließlich
$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Bernoullikette der Länge
 $n = 4$

$0 < p < 1$
 $q = 1 - p$

Trefferzahl: $x = 3$

Anzahl der Pfade:
 $\binom{n}{x} = \binom{4}{3} = 4$



Anmerkung:

Vergleich mit Binomialkoeffizienten (S. 12):

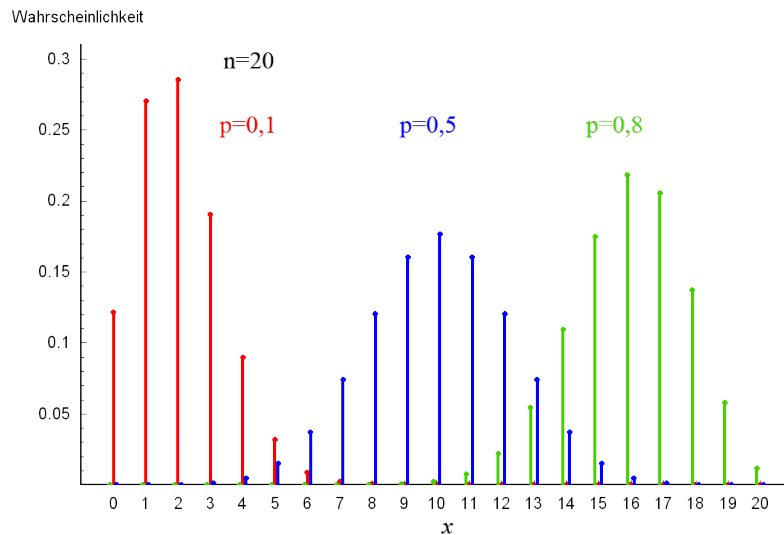
$$\sum_{k=0}^n \binom{n}{k} x^{n-k} \cdot y^k = (x + y)^n$$

$$\sum_{x=0}^n \binom{n}{x} p^x \cdot q^{n-x} = \sum_{x=0}^n \binom{n}{x} q^{n-x} \cdot p^x = (q + p)^n = (p + q)^n$$

Die Wahrscheinlichkeitsfunktion der Binomialverteilung ist:

$$p(x) = P(X = x) = \binom{n}{x} p^x \cdot q^{n-x} \quad \text{mit } x = 0, 1, 2, \dots, n \quad \text{und} \quad q = 1 - p$$

Für großes n hat man schon ein ziemlich “glockenförmiges” Bild. Den Parameter n kennt man in der Regel aus dem Kontext. Die Erfolgswahrscheinlichkeit p nehmen wir bis auf Weiteres als gegeben an. Später sehen wir, wie wir p aus Daten schätzen können.



Die Verteilungstabelle der Binomialverteilung ergibt sich somit zu:

x	0	1	2	...	n
$p(x)$	q^n	$\binom{n}{1} p q^{n-1}$	$\binom{n}{2} p^2 q^{n-2}$	...	p^n

Die Bezeichnung Binomialverteilung erklärt sich aus der Eigenschaft, dass die in der Verteilungstabelle angegebenen Wahrscheinlichkeiten der Reihe nach den Summanden in der binomischen Entwicklung von $(p + q)^n$ entsprechen:

$$(p + q)^n = q^n + \binom{n}{1} p q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \dots + p^n$$

Merkmale der Binomialverteilung:

- Verteilungsfunktion: $F(x) = P(X \leq x) = \sum_{k \leq x} \binom{n}{k} p^k \cdot q^{n-k}$
- Erwartungswert: $E(X) = \mu = np$
- Varianz: $\text{Var}(X) = \sigma^2 = npq$
- Standardabweichung: $\sigma = \sqrt{npq}$

Ist p die Erfolgswahrscheinlichkeit bei einem Versuch und n die Anzahl der Versuche, dann bezeichnet man oft mit $B(k|p,n)$ die Wahrscheinlichkeit, genau k Erfolge zu erzielen.

1. Beispiel:

Wahrscheinlichkeitsfunktion für verschiedene n und p Werte:

$n=4, p=0,1$:

x	0	1	2	3	4
$p(x)$	q^4	$\binom{4}{1} p q^3$	$\binom{4}{2} p^2 q^2$	$\binom{4}{3} p^3 q$	p^4
$p(x)$					

$n=4, p=0,5$:

x	0	1	2	3	4
$p(x)$	q^4	$4 p q^3$	$6 p^2 q^2$	$4 p^3 q$	p^4
$p(x)$					

$n=4, p=0,8$:

x	0	1	2	3	4
$p(x)$	q^4	$4 p q^3$	$6 p^2 q^2$	$4 p^3 q$	p^4
$p(x)$					

2. Beispiel:

Ein Student hat auf eine Prüfung nicht gelernt. Es gibt zehn Fragen mit je vier Antwortmöglichkeiten, genau eine davon ist richtig. Er kreuzt alles komplett zufällig an.

Wie hoch ist die Wahrscheinlichkeit, dass er mindestens 80% der Aufgaben richtig löst?

Lösung:

Der Student muss mindestens $10 \cdot 0.8 = 8$ Fragen von $n = 10$ richtig beantworten.

Pro Frage hat er vier gleichwahrscheinliche Antwortmöglichkeiten.

Die Wahrscheinlichkeit richtig anzukreuzen ist $p = \frac{1}{4}$. Es folgt:

Bemerkungen:

- Die Binomialverteilung findet überall dort Anwendung, wo alternative Entscheidungen zu treffen sind, wie z.B. beim Münzwurf oder der Qualitätskontrolle ("einwandfrei", "Ausschuss")
- Die Wahrscheinlichkeitsfunktion der Binomialverteilung ist i.a. unsymmetrisch und besitzt nur im Sonderfall $p = 0.5$ eine Achsensymmetrie
- Es gilt folgende Rekursionsformel für die Binomialverteilung

$$p(x+1) = \frac{(n-x)p}{(x+1)q} \cdot p(x)$$

3. Beispiel:

Ein homogener Würfel wird 100-mal geworfen. Wie oft kann man dabei eine gerade Augenzahl erwarten?

Lösung:

4. Beispiel:

Bei einer Fertigung werden 5% ($p = 0.05$) der Produkte fehlerhaft gefertigt. Zur Qualitätsprüfung werden 5 Produkte ($n = 5$) entnommen. Berechnen Sie die Wahrscheinlichkeiten P für das Vorfinden von genau 1 ($x = 1$) oder 2 ($x = 2$) defekten Produkten rechnerisch sowie mit R.

Lösung:

1 Produkt defekt:

2 Produkte defekt:

3.3.4 Hypergeometrische Verteilung [dhyper(x,m,n,k)]

Wenn eine Stichprobe **ohne Zurücklegen** und **”kleiner” Grundgesamtheit** entnommen wird, liefert die Binomialverteilung nur schlechte Ergebnisse, da die Versuche nicht stochastisch unabhängig voneinander sind. Je kleiner die Menge der Grundgesamtheit ist, desto ungenauer wird die Binomialverteilung das Ergebnis repräsentieren.

Deshalb wird bei diesen Stichproben die **hypergeometrische Verteilung** verwendet (z.B. bei Qualitäts- und Endkontrollen eines Herstellers oder bei Abnahmekontrollen eines Kunden).

Die hypergeometrische Verteilung ist festgelegt durch die Parameter:

N : Anzahl der Elemente in der Grundgesamtheit

M : Anzahl der Elemente in der Grundgesamtheit mit der Eigenschaft A

n : Anzahl der entnommenen Elemente

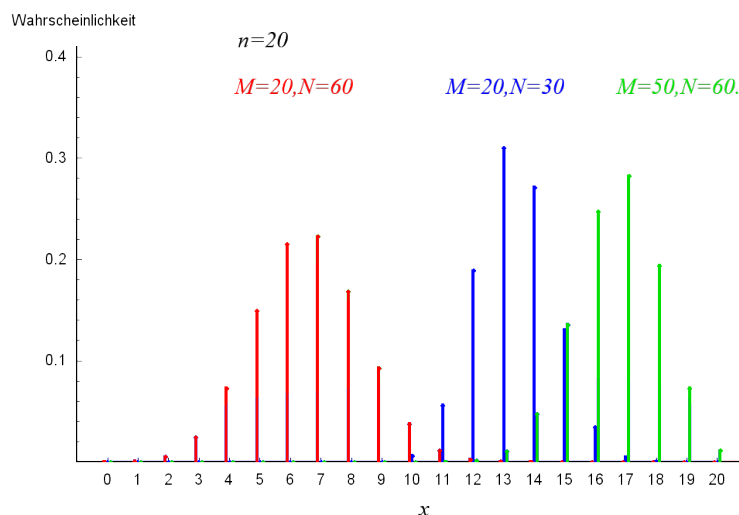
x : Anzahl der Elemente aus M , die in n enthalten sind.

Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung:

$$p(x) = P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Merkmale der hypergeometrischen Verteilung:

- Verteilungsfunktion: $F(x) = P(X \leq x) = \sum_{k \leq x} \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$
- Erwartungswert: $E(X) = \mu = n \frac{M}{N}$
- Varianz: $\text{Var}(X) = \sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}$
- Standardabweichung: $\sigma = \sqrt{\frac{nM(N-M)(N-n)}{N^2(N-1)}}$

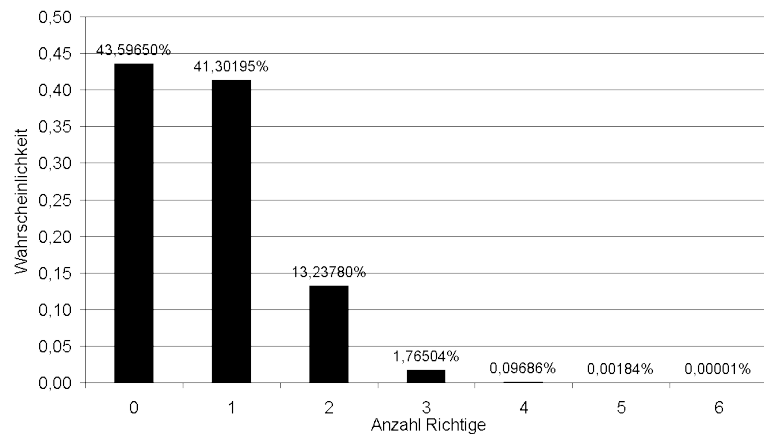


Beispiel:

Beim Zahlenlotto gibt es 49 nummerierte Kugeln; davon werden bei der Auslosung 6 gezogen. Auf dem Lottoschein werden 6 Zahlen angekreuzt.

Die Notation $h(x|49; 6; 6)$ gibt die Wahrscheinlichkeit dafür an, genau $x = 0, 1, 2, 3, \dots, 6$ "Treffer" zu erzielen.

6 aus 49 (N=49)		
H _{49,6,6}		
r	Anzahl	Wahrscheinlichkeit (P) in %
0	6.096.454	43,596
1	5.775.588	41,302
2	1.851.150	13,238
3	246.820	1,7650
4	13.545	0,096862
5	258	0,0018450
6	1	0,0000071511
Σ	13.983.816	100,00
Erwartungswert		0,735
Streuung		0,578

**Bemerkung:**

Bei sehr großer Grundgesamtheit N und wenn die Anzahl n der entnommenen Elemente nicht so groß ist, ändert sich N kaum. In diesem Fall kann man die Hypergeometrische mit den Parametern n , M und N durch die Binomialverteilung mit n und $p = M/N$ annähern, ohne einen „zu großen Fehler“ zu machen.

Faustregel: Die hypergeometrische Verteilung kann näherungsweise durch die rechnerisch bequemere Binomialverteilung ersetzt werden, wenn die Bedingung $n < N/20$ erfüllt ist.

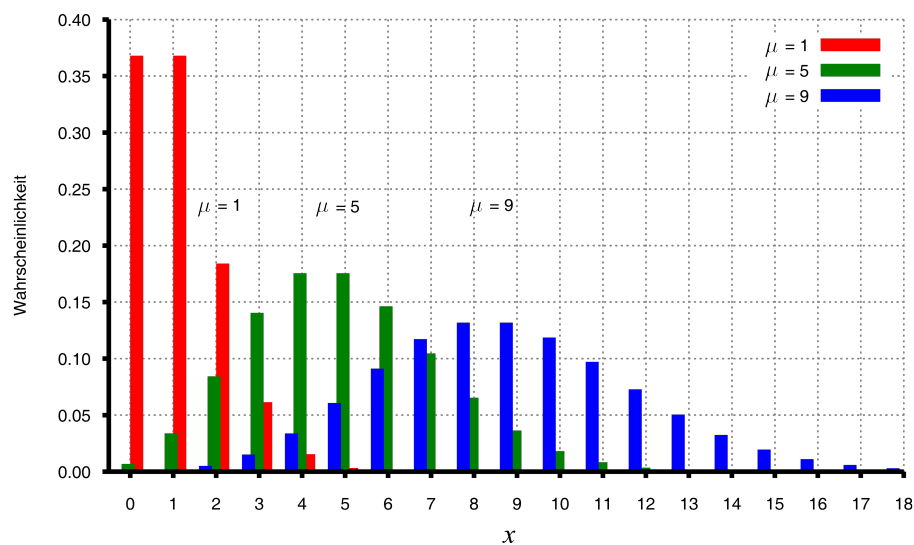
3.3.5 Poisson-Verteilung [`dpois(x,lambda)`]

Die Poisson-Verteilung findet man in Naturwissenschaft und Technik im Zusammenhang mit Bernoulli-Experimenten, bei denen Ereignisse nur mit **geringer Wahrscheinlichkeit** auftreten.

Ein Beispiel hierfür ist der radioaktive Zerfall. Die Anzahl der pro Sekunde zerfallenden Atomkerne ist im Vergleich zur Anzahl der insgesamt vorhandenen Kerne äußerst gering.

Wahrscheinlichkeitsfunktion der Poisson-Verteilung

$$p(x) = P(X = x) = \frac{\mu^x}{x!} \cdot e^{-\mu} \quad \text{mit } x = 0, 1, 2, \dots, \infty, \quad \mu > 0$$



Merkmale der Poisson-Verteilung:

- Verteilungsfunktion $F(x) = P(X \leq x) = e^{-\mu} \cdot \sum_{k \leq x} \frac{\mu^k}{k!}$
- Erwartungswert: $E(X) = \mu$
- Varianz: $\text{Var}(X) = \sigma^2 = \mu$
- Standardabweichung: $\sigma = \sqrt{\mu}$

Bemerkungen:

- Die Wahrscheinlichkeitsfunktion der Poisson-Verteilung ist i.a. unsymmetrisch. Für große Werte von μ wird die Poisson-Verteilung nahezu symmetrisch. Die Symmetrieachse liegt in der Nähe des Erwartungswertes.
- Die Poisson-Verteilung ist vollständig durch den Erwartungswert μ bestimmt (oft auch mit λ bezeichnet).
- Die Poisson-Verteilung lässt sich aus der Binomial-Verteilung für den Grenzübergang $n \rightarrow \infty$ und $p \rightarrow 0$ herleiten, wobei der Mittelwert $\mu = np$ konstant bleiben soll.

- **Faustregel:** Die Binomialverteilung darf näherungsweise durch die Poisson-Verteilung ersetzt werden, wenn die beiden Bedingungen
 $np < 10$ und etwa $n > 1500p$
 erfüllt sind. Da dies eine Faustregel ist, sollte man beide Werte zusammen betrachten und ggf. abwägen.

Beispiele:

1. Beim radioaktiven Zerfall ist die Zufallsvariable X =Anzahl der Atomkerne, die in einer Sekunde zerfallen poisson-verteilt mit dem Parameter μ . Dieser gibt an, wieviele Atomkerne durchschnittlich in einer Sekunde zerfallen.

Bei einem speziellen Präparat zerfallen im Mittel pro Minute 120 Atomkerne. Wie groß ist die Wahrscheinlichkeit, dafür mit einem Zählgerät mehr als zwei Zerfälle pro Sekunde zu registrieren?

Lösung:

Im Mittel zerfallen 2 Atomkerne pro Sekunde.

Somit ist $\mu = 2$ und die Wahrscheinlichkeitsfunktion lautet:

$$p(x) = P(X = x) = \frac{2^x}{x!} \cdot e^{-2}$$

Das interessierende Ereignis $X > 2$ besitzt die folgende Wahrscheinlichkeit:

2. Die Serienproduktion von Glühbirnen erfolgt mit einem Ausschussanteil von 1%. Aus der laufenden Produktion wird eine Stichprobe vom Umfang $n=100$ entnommen. Mit welcher Wahrscheinlichkeit enthält diese Stichprobe A drei oder mehr defekte Glühbirnen.

Lösung:

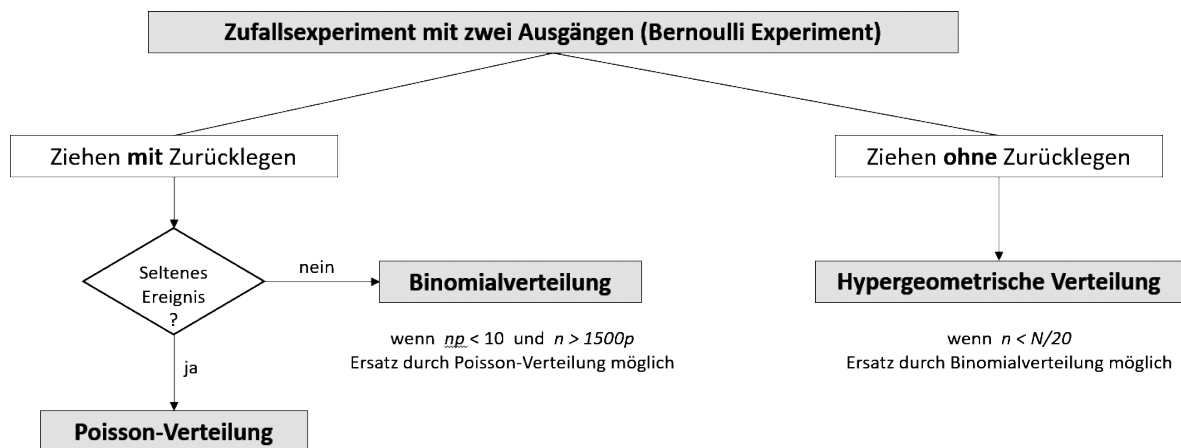
Die Zufallsvariable X =Anzahl der defekten Glühbirnen in der entnommenen Stichprobe vom Umfang $n = 100$ ist binomialverteilt mit den Parametern $n = 100$ und $p = 0.01$.

Wegen $np = 1 < 10$ genügt die Zufallsvariable X nach der Faustregel näherungsweise einer Poisson-Verteilung mit dem Parameter $\mu = np = 1$ und somit der diskreten Wahrscheinlichkeitsfunktion

$$p(x) = P(X = x) = \frac{1}{x!} e^{-1}.$$

Um den Rechenaufwand zu reduzieren berechnet man die Wahrscheinlichkeit für das komplementäre Ereignis: $\bar{A}$ = „In einer Stichprobe vom Umfang $n = 100$ befinden sich höchstens 2 defekte Glühbirnen“.

3.3.6 Überblick: Zufallsexperimente mit zwei Ausgängen



n Umfang der Stichprobe/Anzahl der entnommenen Elemente

N Anzahl der Elemente in der Grundgesamtheit

p Wahrscheinlichkeit (Dezimalzahl)

3.4 Stetige Verteilungen

Ist die Zufallsvariable kontinuierlich (stetig), so ist die Wahrscheinlichkeit dafür, dass sie einen bestimmten Wert x_i annimmt gleich 0. Man betrachtet daher die Wahrscheinlichkeit dafür, dass X in einem endlichen Intervall $[a, b]$ liegt. Lässt sich diese mit Hilfe einer Funktion $f(t)$, der **Wahrscheinlichkeitsdichte**, in der Form

$$P(a \leq X \leq b) = \int_a^b f(t) \, dt$$

darstellen, dann spricht man von einer **stetigen** Verteilungsfunktion:

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

Bei einer stetigen Verteilungsfunktion tritt die Integration über die Dichtefunktion $f(t)$ an die Stelle der Summation über die p_i im Fall einer diskreten Zufallsvariablen.

Es gilt für die Dichtefunktion:

$$\begin{aligned} f(x) &\geq 0 \quad (\text{Nichtnegativität}) \\ \int_{-\infty}^{\infty} f(x) \, dx &= 1 \quad (\text{Normiertheit}) \end{aligned}$$

Es gilt außerdem: Die Wahrscheinlichkeit dafür, dass eine **stetige** Zufallsvariable einen beliebigen Wert x_0 annimmt, ist gleich Null:

$$P(X = x_0) = 0$$

Daraus erklärt sich die Tatsache, dass man in der praktischen Anwendung von stetigen Zufallsvariablen nur an Ereignissen der Gestalt „ X nimmt Werte zwischen a und b an“ und nicht an sogenannten Punktereignissen $X = x_i$ interessiert ist.

Beispiel:

Bei der stetigen Zufallsvariablen X : „Dauer des Schlangestehens vor der Kasse im Supermarkt“ wäre das Ereignis „der Kunde wartet exakt 8 Minuten und 12 Sekunden bis er bezahlen kann“ ohne Interesse.

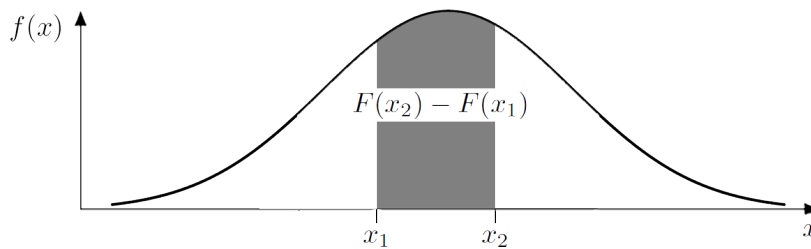
Vielmehr fragt man danach, wie groß die Wahrscheinlichkeit dafür ist, dass die Wartezeit z.B. zwischen 7 und 8 Minuten liegt.

Die Wahrscheinlichkeit dafür, dass eine stetige Zufallsvariable X z.B. in einem Intervall $[x_1, x_2]$ liegt, lässt sich bestimmen zu

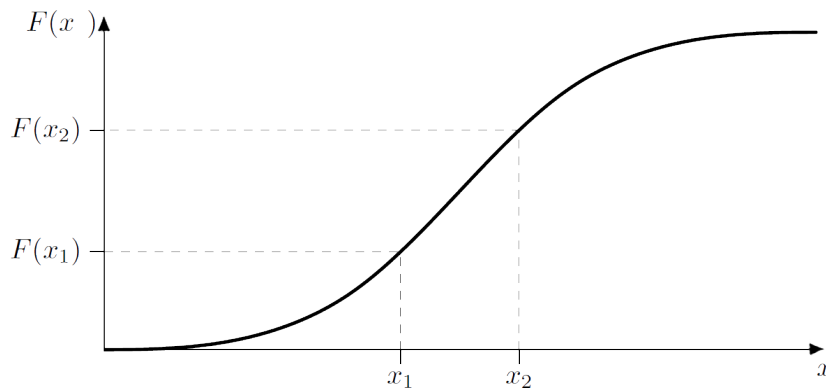
$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$$

Da die Wahrscheinlichkeiten für die Endpunkte des Intervalls Null sind, ist es unerheblich, ob sie zu dem Intervall gehören: $[x_1, x_2] \hat{=} (x_1, x_2] \hat{=} [x_1, x_2) \hat{=} (x_1, x_2)$.

Dichtefunktion $f(x)$:



Verteilungsfunktion $F(x)$:



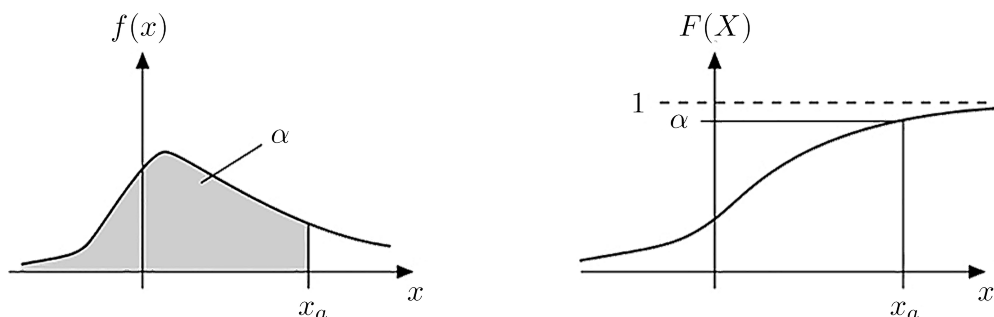
Grafische Darstellung der Wahrscheinlichkeit $P(x_1 \leq X \leq x_2)$

Flächeninterpretation, Quantil (Lagemaß)

Unter dem α -Quantil x_α einer Verteilungsfunktion $F(x)$ einer stetigen Zufallsgröße X versteht man den Wert, der folgende Gleichung erfüllt:

$$F(x_\alpha) = \alpha.$$

Das bedeutet, x_α teilt den Wertebereich von X in zwei Teile so, dass gerade $\alpha \cdot 100\%$ aller Beobachtungen von X links von x_α liegen, also kleiner oder gleich x_α sind, und $(1 - \alpha) \cdot 100\%$ rechts von x_α liegen, also größer oder gleich x_α sind.



Links ein Quantil mit Verteilungsdichte, rechts Quantil und Verteilungsfunktion.

Erwartungswert, Varianz und Standardabweichung einer stetigen Zufallsvariablen

Die stetige Zufallsvariable X mit der Wahrscheinlichkeitsdichte $f(x)$ besitzt die folgenden Kennwerte

1. Erwartungswert: $\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$
2. Varianz: $\text{Var}(X) = \sigma^2 = E((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \, dx = \text{Var}(X)$
3. Standardabweichung: $\sigma = \sqrt{\text{Var}(X)}$

3.4.1 Gaußsche Normalverteilung [`dnorm(x, mean, sd)`]

Eine stetige Verteilung mit der Dichtefunktion

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

heißt Gaußsche Normalverteilung, oder kurz Normalverteilung $\mathcal{N}(\mu, \sigma^2)$, mit

Erwartungswert $\mu \in \mathbb{R}$

Standardabweichung $\sigma > 0$

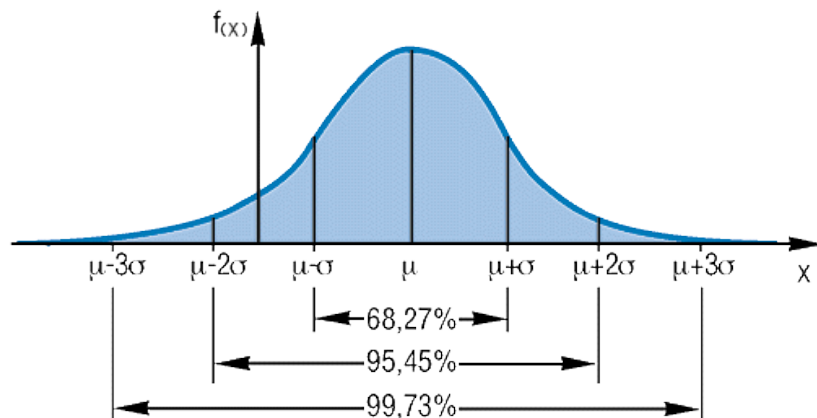
Varianz $\sigma^2 > 0$

Die Verteilungsfunktion der Normalverteilung ist:

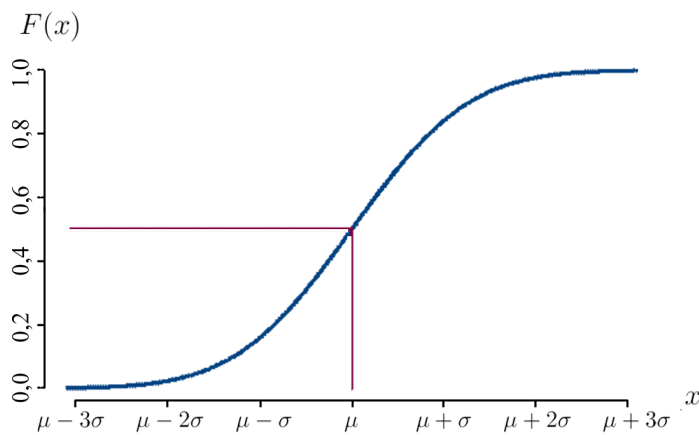
$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^x e^{-\frac{1}{2} \left(\frac{t - \mu}{\sigma} \right)^2} \cdot dt$$

Zahlreiche Zufallsvariable, wie z.B. physikalische, technische Messgrößen genügen dieser Verteilung.

Dichtefunktion $f(x)$ der Normalverteilung, bzw. $\mathcal{N}(\mu, \sigma^2)$:



Verteilungsfunktion $F(x)$ der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$:



Bemerkungen:

- Laplace konnte zeigen: $\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} \cdot dt = \sqrt{2\pi}$
- Die Dichtefunktion der Normalverteilung ist achsensymmetrisch bezüglich $x = \mu$
- Das einzige Maximum liegt bei $x_1 = \mu$
- Die beiden Wendepunkte liegen symmetrisch zum Maximum an den Stellen $x_{2,3} = \mu \pm \sigma$
- $f(x)$ ist normiert, d.h. $\int_{-\infty}^{\infty} f(x) = 1$
- Je kleiner die Standardabweichung ist, umso höher liegt das Maximum und umso steiler fällt die Dichtefunktion nach beiden Seiten hin ab

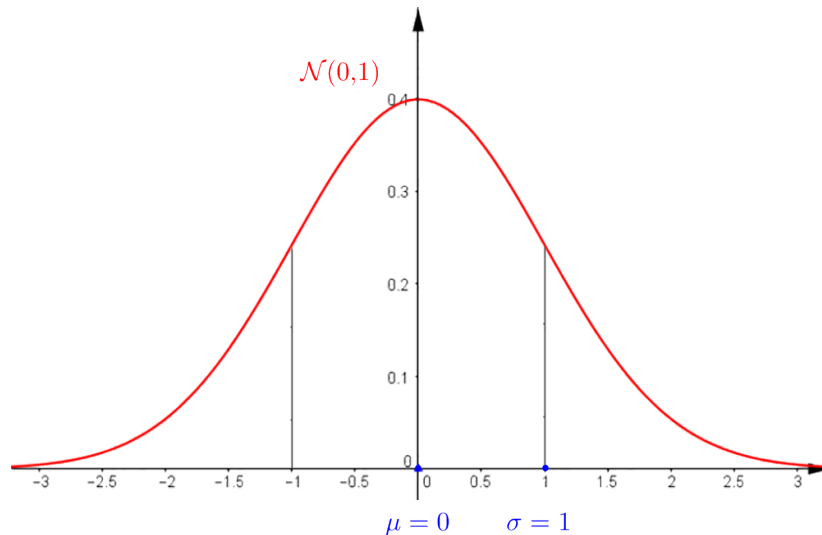
Eine Normalverteilung mit den Parametern $\mu = 0$ und $\sigma = 1$ heißt **Standardnormalverteilung**.

Eine normalverteilte Zufallsvariable X mit den Parametern μ und σ lässt sich immer mit Hilfe der Variablentransformation $U = \frac{X - \mu}{\sigma}$ in die standardnormalverteilte Zufallsvariable U überführen.

Die Standardnormalverteilung wird oft durch $\mathcal{N}(0,1)$ gekennzeichnet.
Ihre Verteilungsfunktion ist

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{1}{2}t^2} \cdot dt$$

Standardnormalverteilung $\mathcal{N}(0,1)$:



3.4.2 Wahrscheinlichkeiten bei normalverteilten Zufallsvariablen

Die Berechnung von Wahrscheinlichkeiten normalverteilter Zufallsvariablen erfolgt in der Anwendung stets mit Hilfe der Verteilungsfunktion der Standardnormalverteilung

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{1}{2}t^2} dt.$$

Dieses uneigentliche Integral ist nicht elementar lösbar, sondern muss numerische angenähert werden. Es gibt z.B. Interpolationstabellen (vgl. Anhang) mit den Funktionswerten $\Phi(u)$ für $u \geq 0$.

Damit kann man

für $u \geq 0$: die Werte von $\Phi(u)$ direkt ablesen

für $u < 0$: $\Phi(u) = 1 - \Phi(-u)$ bestimmen, mit $-u > 0 \rightarrow$ Entnahme aus Tabelle

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

Beispiele:

$$1. \Phi(-1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056$$

`pnorm(x, mean, sd)`

`pnorm(-1.25, 0, 1)`

`[1] 0.1056498`

$$2. \Phi(-2.423) =$$

3.4.3 Transformation zur Standardnormalverteilung

1. $P(X \leq x)$

$$\text{Es ist } P(X \leq x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^x e^{-\frac{1}{2} \left(\frac{t-\mu}{\sigma} \right)^2} \cdot dt$$

Mit der Substitution $u = \frac{t-\mu}{\sigma}$ und somit $\frac{du}{dt} = \frac{1}{\sigma}$ folgt:

$$\begin{aligned} P(X \leq x) &= \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^x e^{-\frac{1}{2} \left(\frac{t-\mu}{\sigma} \right)^2} \cdot dt = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2} u^2} \cdot \sigma \, du \\ &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2} u^2} \cdot du = \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

2. $P(X \geq x)$

$$P(X \geq x) = 1 - P(X \leq x) = 1 - \Phi\left(\frac{x-\mu}{\sigma}\right)$$

3. $P(a \leq X \leq b)$

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

4. $P(|X - \mu| \leq k\sigma)$

Sehr oft interessieren Intervalle, die symmetrisch um den Erwartungswert μ liegen.

$$\text{Es ist: } P(|X - \mu| \leq k\sigma) \iff -k\sigma \leq X - \mu \leq k\sigma \iff -k\sigma + \mu \leq X \leq k\sigma + \mu$$

und somit ist

$$P(|X - \mu| \leq k\sigma) = P(-k\sigma + \mu \leq X \leq k\sigma + \mu) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

$k = 1$	$P(X - \mu \leq k\sigma) = 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826$	$\approx 68\%$ liegen in diesem Intervall
$k = 2$	$P(X - \mu \leq k\sigma) = 2 \cdot \Phi(2) - 1 = 2 \cdot 0.9772 - 1 = 0.9544$	$\approx 95\%$ liegen im Intervall
$k = 3$	$P(X - \mu \leq k\sigma) = 2 \cdot \Phi(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9974$	$\approx 99\%$ liegen im Intervall

Beispiele:

- a) In einem Werk werden Gewindeschrauben erstellt, deren Durchmesser eine normalverteilte Zufallsvariable X darstellt mit Mittelwert $\mu = 10$ mm und Standardabweichung $\sigma = 0.2$ mm. Toleriert werden dabei noch zufallsbedingte Abweichungen vom Soll Durchmesser bis maximal 0.3 mm.

Welcher Anteil an Ausschussware ist zu erwarten?

Lösung:

Man berechnet zunächst $P(9.7 \leq X \leq 10.3)$, die Wahrscheinlichkeit, dass die Schraubendurchmesser im Toleranzbereich sind. Um obige Gleichung anwenden zu können, ist 0.3 mm durch $k\sigma$ zu ersetzen, d.h. $0.3 = k \cdot 0.2 \implies k = 1.5$.

Somit ist...

- b) Die Zufallsvariable X sei normalverteilt mit dem Mittelwert $\mu = 10$ und der Varianz $\sigma^2 = 4$.
Wieviel Prozent aller Werte liegen dann im Intervall $5 \leq X \leq 12$?

Lösung:

- c) Die normalverteilte Zufallsvariable X mit dem Mittelwert μ und der Standardabweichung σ soll 50% ihrer Werte in dem symmetrischen Intervall $|X - \mu| \leq k \cdot \sigma$ annehmen. Wie lautet der Faktor k ?

Lösung:

- d) Die standardnormalverteilte Zufallsvariable U soll mit der Wahrscheinlichkeit $p = 0.9$ einen Wert aus dem Intervall $U \leq c$ annehmen.
Wie lautet c ?

Lösung:

- e) Die Werte der standardnormalverteilten Zufallsvariablen U sollen mit der Wahrscheinlichkeit $p = 0.9$ oberhalb der Schranke c liegen.
Wie lautet c ?

Lösung:

- f) Bei einer zweiseitigen Abgrenzung sollen die Werte der standardnormalverteilten Zufallsvariablen U mit der Wahrscheinlichkeit $p = 0.95$ im symmetrischen Intervall $-c \leq U \leq c$ liegen.
Wie lautet c ?

Lösung:

3.4.4 Exponentialverteilung [`dexp(x,rate)`]

Die Exponentialverteilung ist eine stetige Verteilung. Mit Hilfe der Exponentialverteilung können v.a. Lebensdauer- oder Wartezeitenprobleme modelliert werden.

Die Exponentialverteilung ist eng mit der Poisson-Verteilung verwandt. Während letztere eine diskrete Verteilung ist und die Zufallsvariable die Anzahl des Eintretens eines bestimmten Ereignisses (eine diskrete Größe) widerspiegelt, ist die Exponentialverteilung eine stetige Verteilung mit einer Zufallsvariablen, die z.B. den zeitlichen Abstand (eine stetige Größe) zwischen 2 Ereignissen darstellt.

Der Durchschnitts- bzw. Erwartungswert der Exponentialverteilung ist entsprechend gleich dem Kehrwert des Erwartungswerts der Poissonverteilung.

Die Verteilung einer stetigen Zufallsvariablen X mit der Dichtefunktion

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \\ \lambda \cdot e^{-\lambda x} & \text{für } x \geq 0 \end{cases}$$

heißt Exponentialverteilung mit dem Parameter $\lambda > 0$ (Ausfallrate).

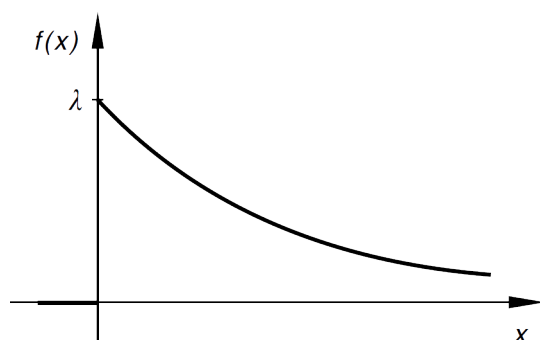
Für die Verteilungsfunktion ergibt sich dann:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } x \geq 0 \end{cases} \quad (\text{Ausfallwahrscheinlichkeit})$$

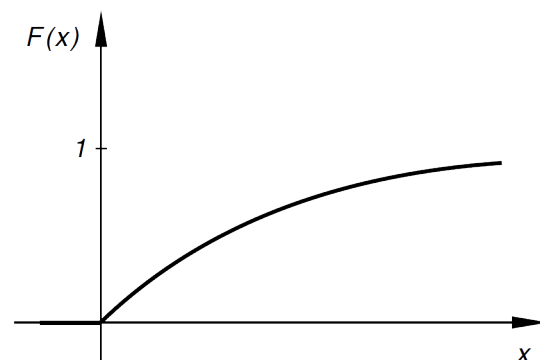
Die mittlere Lebensdauer ist durch den Erwartungswert μ gegeben.

- Erwartungswert: $E(X) = \mu = \frac{1}{\lambda}$
- Varianz: $\text{Var}(X) = \sigma^2 = \frac{1}{\lambda^2}$
- Standardabweichung: $\sigma = \frac{1}{\lambda}$

Bemerkung: Bei einer exponentialverteilten Zufallsvariablen stimmen Mittelwert und Standardabweichung stets überein: $\mu = \sigma = \frac{1}{\lambda}$.



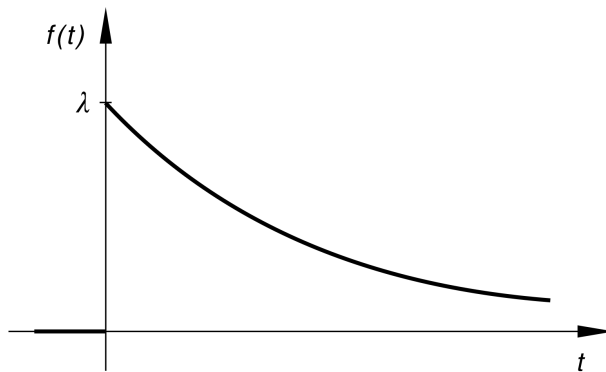
Dichtefunktion



Verteilungsfunktion

Berechnung des Erwartungswertes, die Varianz und die Standardabweichung:

Die mittlere Lebensdauer ist durch den Erwartungswert $E(T)$ gegeben:



Dichtefunktion der exponentialverteilten Zufallsvariablen „ T = Lebensdauer eines elektronischen Bauelements“

$$\begin{aligned}
 E(T) &= \int_{-\infty}^{\infty} t \cdot f(t) dt = \int_0^{\infty} t \cdot \lambda \cdot e^{-\lambda t} dt = \lambda \cdot \underbrace{\int_0^{\infty} t \cdot e^{-\lambda t} dt}_{\text{Formelsammlung mit } a = -\lambda} = \\
 &= \lambda \left[\frac{-\lambda t - 1}{\lambda^2} \cdot e^{-\lambda t} \right]_0^{\infty} = \frac{1}{\lambda} \left[(-\lambda t - 1) \cdot e^{-\lambda t} \right]_0^{\infty} = \\
 &= \frac{1}{\lambda} (0 + 1) = \frac{1}{\lambda}
 \end{aligned}$$

Um die Varianz σ^2 berechnen zu können, benötigen wir noch den Erwartungswert $E(T^2)$:

$$\begin{aligned}
 E(T^2) &= \int_{-\infty}^{\infty} t^2 \cdot f(t) dt = \int_0^{\infty} t^2 \cdot \lambda \cdot e^{-\lambda t} dt = \lambda \cdot \underbrace{\int_0^{\infty} t^2 \cdot e^{-\lambda t} dt}_{\text{Formelsammlung mit } a = -\lambda} = \\
 &= \lambda \left[\frac{\lambda^2 t^2 + 2\lambda t + 2}{-\lambda^3} \cdot e^{-\lambda t} \right]_0^{\infty} = \\
 &= -\frac{1}{\lambda^2} \left[(\lambda^2 t^2 + 2\lambda t + 2) \cdot e^{-\lambda t} \right]_0^{\infty} = -\frac{1}{\lambda^2} (0 - 2) = \frac{2}{\lambda^2}
 \end{aligned}$$

$$\sigma^2 = E(T^2) - \mu^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Die *Standardabweichung* der exponentialverteilten Zufallsvariablen T beträgt somit $\sigma = 1/\lambda$ und stimmt mit dem Mittelwert μ überein.

Beispiel:

Für ein empfindliches mikroelektronisches Bauteil gilt eine Ausfallrate (auch Zerfallskonstante) $\lambda = 0.1$ pro Jahr (a). In einem Jahr fallen also 10% der gesamten Bauelemente (Anzahl: C) aus. Wieviele Bauelemente sind in den ersten 3 Jahren ausgefallen ($t = 3a$)?

Die Anzahl der Teile pro Jahr die ausfallen ist

$$f(t) = e^{-\lambda \cdot t}$$

$$P(X \leq t) = 1 - e^{-\lambda \cdot t}$$

$$F(3) = 1 - e^{-\frac{0.1}{a} \cdot 3a} = 1 - e^{-0.3} \approx 0.26 \Rightarrow \text{in den ersten 3 Jahren sind etwa 26\% der } C \text{ Bauteile ausgefallen.}$$

Für $t \rightarrow \infty$ geht die Anzahl der Ausfälle dann gegen 100% von C .

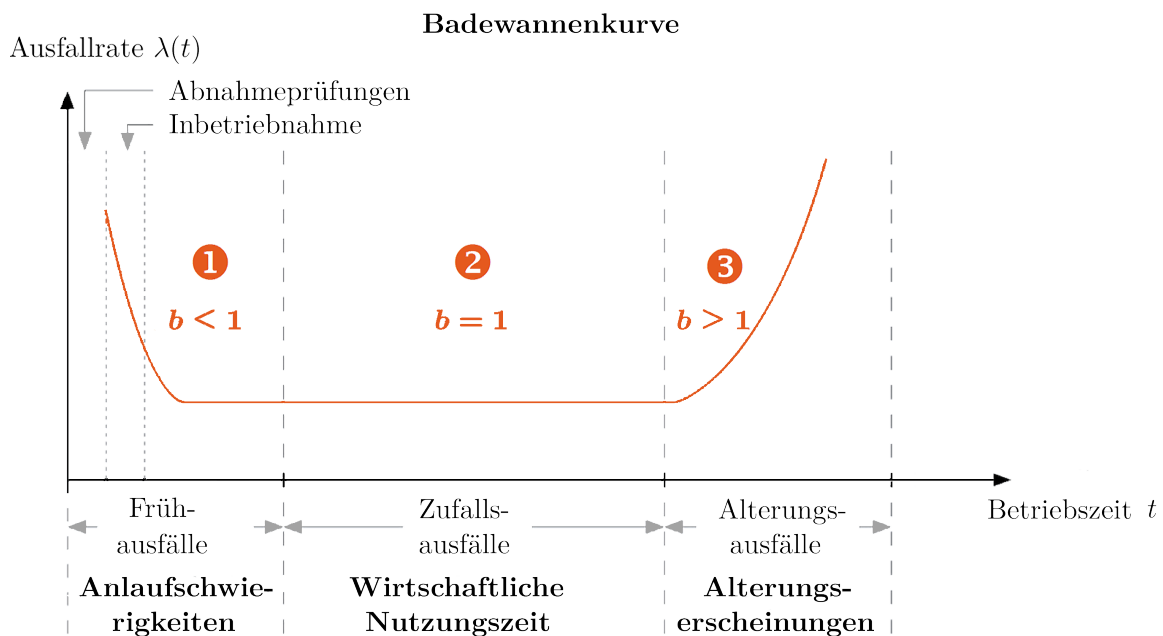
3.4.5 Weibullverteilung [`dweibull(x,shape,scale)`]

Die Weibullverteilung ist eine stetige zweiparametrische Verteilung, die für positive reelle Zahlen definiert ist und die flexibel für die Modellierung vieler Prozesse verwendet werden kann.

Die Exponentialverteilung ist ein Spezialfall der Weibullverteilung, die ebenfalls in der Zuverlässigkeitstheorie und Lebensdaueranalyse etabliert ist. Benannt ist die Verteilung nach dem schwedischen Ingenieur und Mathematiker Waloddi Weibull.

Aus empirischen Untersuchungen für die Lebensdauer gewinnt man folgende Erkenntnis:

- erste Phase → zunächst hohe, dann sinkende Ausfallwahrscheinlichkeit (etwa bis die optimale Einrichtung, Einstellung, ... erfolgt ist)
- zweite Phase → gleichbleibend niedrige Ausfallrate (Verlauf hat sich eingependelt)
- dritte Phase → altersbedingte ansteigende Ausfallrate



Die Ausfallrate $\lambda(t)$ wird durch

$$\lambda(t) = \lambda b \cdot (\lambda t)^{b-1} = \frac{b}{T} \cdot \left(\frac{t}{T}\right)^{b-1}$$

beschrieben. Wobei gilt:

b = Formparameter oder Ausfallsteilheit

λ = die mittlere Ausfallwahrscheinlichkeit pro Intervall.

Sie berechnet sich insbesondere bei Zeitabhängigkeiten für den Bereich $b = 1$ als Kehrwert der mittleren/charakteristischen Lebensdauer T : $\lambda = 1/T$.

T = Lageparameter und kann verwendet werden, um die durchschnittliche Lebensdauer zu verändern. Für $b = 1$ wird T als charakteristische Lebensdauer bezeichnet ($\hat{=}$ Ausfallwahrscheinlichkeit von 63.2% der betrachteten Elemente).

Er gibt allerdings im Allgemeinen nicht die durchschnittliche Lebensdauer an.

Den Formparameter b verwendet man, um Veränderungen der Ausfallrate im Zeitablauf modellieren zu können.

Entsprechend der hauptsächlichen Verwendung der Weibullverteilung als Lebensdauerverteilung wird im Folgenden mit Ausfällen u.Ä. argumentiert. Der Formparameter b kann dazu genutzt werden, um zu modellieren, ob Früh- oder Spätausfälle häufiger sind.

Wird $b < 1$ gewählt, treten verstärkt Frühausfälle auf, bei $b > 1$ verstärkt Spätausfälle. Wird der Formparameter $b = 1$ gewählt, ergibt sich exakt die Exponentialverteilung mit der Voraussetzung einer konstanten Ausfallwahrscheinlichkeit wie z.B. bei der Lebensdauer von elektronischen Bauteilen.

Für die **Ausfallraten** gilt für die Bereiche:

$b < 1$: abnehmend

$b = 1$: konstant

$b > 1$: zunehmend

In der Praxis liegen typische Werte bei $0.25 < b < 5$.

Für die Weibullverteilung gilt allgemein:

Dichtefunktion

$$f(t) = \lambda b \cdot (\lambda t)^{b-1} \cdot e^{-(\lambda t)^b} = \frac{b}{T} \cdot \left(\frac{t}{T}\right)^{b-1} \cdot e^{-\left(\frac{t}{T}\right)^b}$$

Verteilungsfunktion

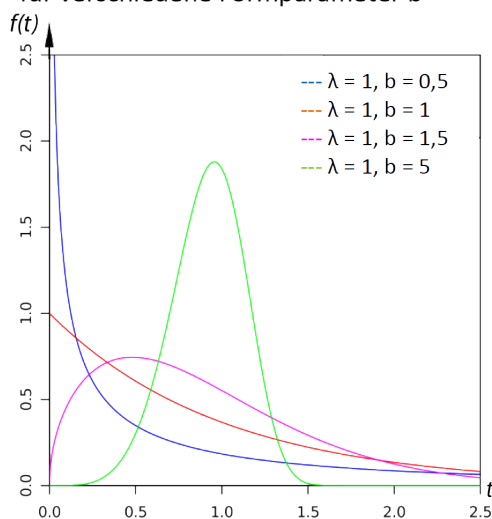
$$F(t) = 1 - e^{-(\lambda t)^b} = 1 - e^{-\left(\frac{t}{T}\right)^b}$$

Überlebensfunktion (d.h. kein Ausfall)

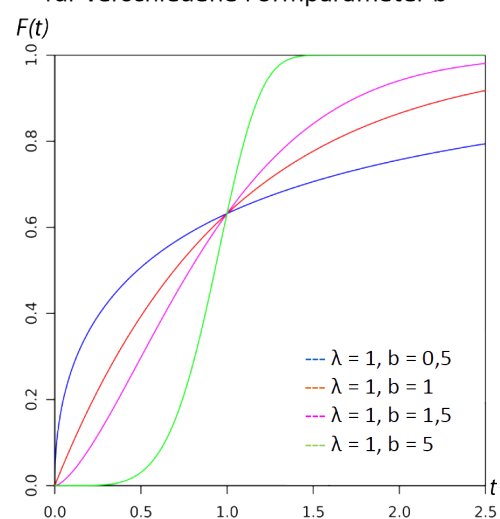
$$R(t) = e^{-(\lambda t)^b} = e^{-\left(\frac{t}{T}\right)^b}$$

Wird kein Skalenparameter λ angegeben, so ist implizit $\lambda = 1$ gemeint (wird meist zur normierten Darstellung der Funktionen verwendet).

Dichtefunktion
für verschiedene Formparameter b



Verteilungsfunktion
für verschiedene Formparameter b



Für die Bestimmung von Erwartungswert und Varianz der Weibullverteilung benötigt man zuerst die Gammafunktion $\Gamma(\alpha)$:

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} \cdot e^{-t} dt$$

Mit den Eigenschaften

$\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha)$ und $\Gamma(1) = 1$ und mit $L = \text{Lebensdauer}$, erhält man

$$\text{Erwartungswert: } E(L) = \mu = T \cdot \Gamma\left(1 + \frac{1}{b}\right)$$

$$\text{Varianz: } \text{Var}(L) = \sigma^2 = T^2 \left[\Gamma\left(1 + \frac{2}{b}\right) - \left(\Gamma\left(1 + \frac{1}{b}\right) \right)^2 \right]$$

Neben der Analyse von Ausfall- und Zuverlässigkeitswahrscheinlichkeiten von Anlagen, Systemen und Komponenten findet die Weibull-Verteilung zum Beispiel auch Anwendung in der Wettervorhersage für die Windkraftindustrie und in allgemeinen Versicherungsmodellen.

Beispiel:



Die Lebensdauer eines mechanischen Ventils kann durch die Weibullverteilung beschrieben werden. Aus Untersuchungen ist bekannt:

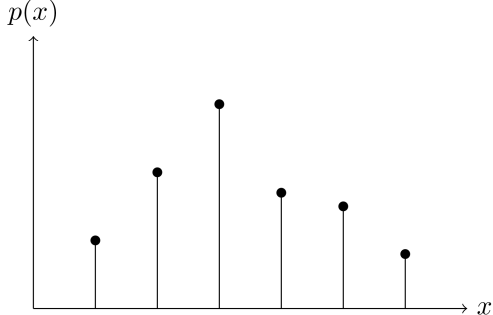
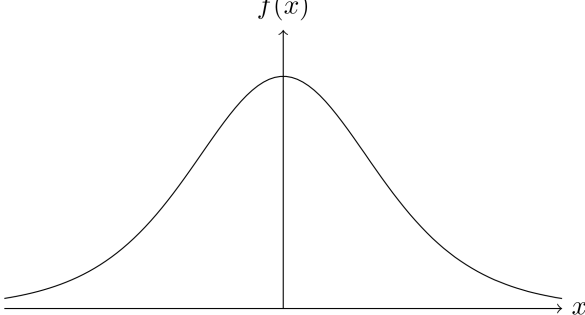
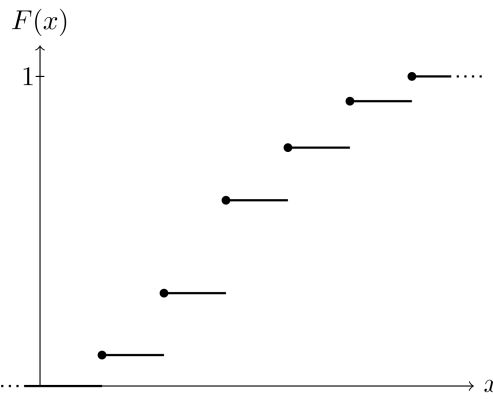
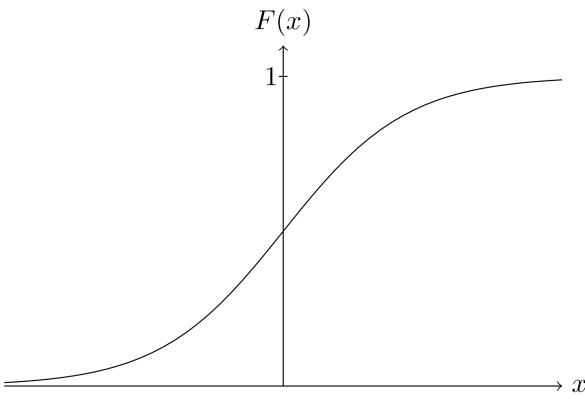
$b = 2$ und die charakteristische Lebensdauer $T = 10^3$ h.

Bestimmt werden soll:

- Überlebenswahrscheinlichkeit des Ventils für eine Betriebszeit von $t = 200$ h.
- Mittlere Zeit bis zum Ausfall
- Ausfallrate des Ventils unter Zugrundelegung der Betriebszeit (200 h).

Lösung:

3.5 Vergleich der Konzepte: Diskrete vs. stetige Verteilungen

diskret	stetig
Wahrscheinlichkeitsfunktion  $P(X = x_k) = p(x_k) \in [0, 1], x_k \in W$	Dichte  $P(X = x) = 0, x \in W$
Kumulative Verteilungsfunktion  $F(x) = \sum_{k: x_k \leq x} p(x_k)$	Kumulative Verteilungsfunktion  $F(x) = \int_{-\infty}^x f(u) du$
Erwartungswert $E(X) = \sum_{k \geq 1} x_k p(x_k)$	Erwartungswert $E(X) = \int_{-\infty}^{\infty} x f(x) dx$

KAPITEL 4

Statistik

4.1 Deskriptive Statistik

Eine grundlegende Aufgabe der Statistik besteht darin, Kenntnisse und Informationen über die Eigenschaften einer bestimmten Menge von Objekten zu gewinnen, ohne dass man dabei alle Objekte in die Untersuchung mit einbeziehen muss. Letzteres ist meist auch gar nicht möglich, z.B. aus folgenden Gründen:

- Zu hoher Zeit- und Kostenaufwand
- Die Anzahl der Objekte ist zu groß
- Die Objekte müssen oder könnten bei der Untersuchung zerstört werden

Unter einer **Grundgesamtheit N** versteht man die Gesamtheit gleichartiger Elemente, die hinsichtlich eines bestimmten Merkmals untersucht werden sollen. Das interessierende Merkmal beschreibt man dabei durch eine Zufallsvariable X .

Eine aus der Grundgesamtheit zufällig herausgegriffene Teilmenge mit n Elementen wird als **Zufallsstichprobe vom Umfang n** bezeichnet.

Die beobachteten Merkmalswerte der n Elemente sind **Realisierungen** der Zufallsvariablen X und heißen **Stichprobenwerte n** .

Die Aufgabe der Statistik besteht dann darin, aus einer Zufallsstichprobe mit Hilfe der Wahrscheinlichkeitsrechnung Rückschlüsse auf die Grundgesamtheit zu ermöglichen.

Beispiele:

1. Die Kapazität der Kondensatoren aus einer Serienproduktion ist ein quantitatives Merkmal. Zu Kontrollzwecken wurde der Grundgesamtheit eine Stichprobe vom Umfang $n = 10$ entnommen:

Nr. der Stichprobe	1	2	3	4	5	6	7	8	9	10
Kapazität in μF	99	101	102	102	99	98	100	101	101	102

2. Bei der Funktionsprüfung seriengleicher Bauelemente interessiert das qualitative Merkmal "das Bauelement ist funktionstüchtig". Mögliche Ergebnisse einer solchen Prüfung sind dann "ja" oder "nein".

4.1.1 Grundbegriffe

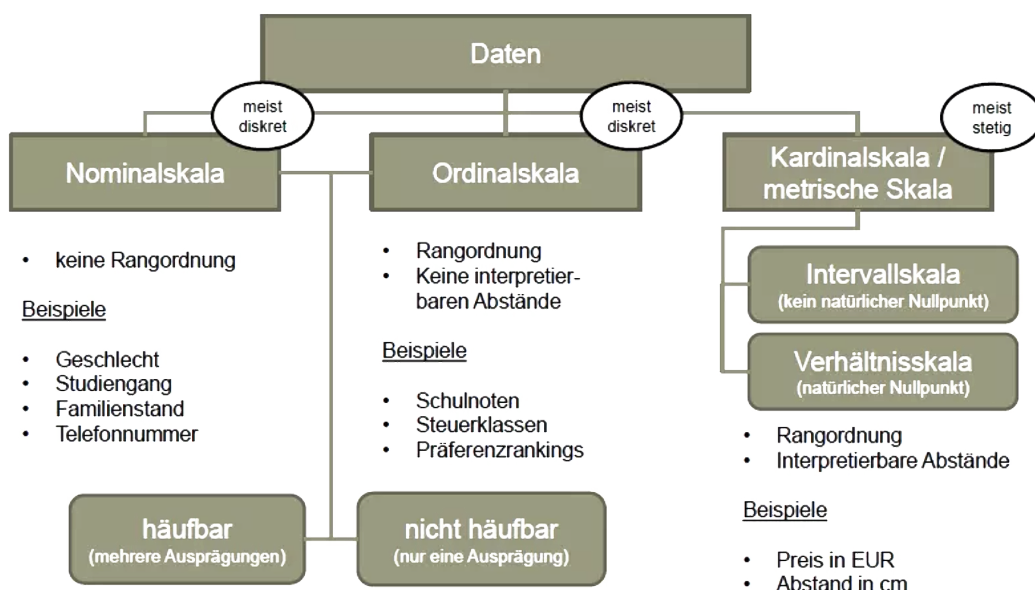
Merkmale

- **Merkmalsträger:** Untersuchte statistische Einheit (Person, Objekt...)
- **Merkmal:** Interessierende Eigenschaft des Merkmalsträgers (Alter, Größe...)
- **(Merkmals-)Ausprägung:** Konkret beobachteter Wert des Merkmals
- **Grundgesamtheit:** Menge aller relevanten Merkmalsträger
- **Typen von Merkmalen:**
 1. qualitativ - quantitativ
 - qualitativ: es können keine mathematische Werte angenommen werden (z.B. Geschlecht)
 - quantitativ: z.B. Schuhgröße
 - qualitative Merkmale sind quantifizierbar (z.B.: weiblich 1, männlich 0)
 2. diskret - stetig
 - diskret: abzählbar viele unterschiedliche Ausprägungen (z.B. Noten)
 - stetig: alle Zwischenwerte realisierbar (z.B. Temperatur)
 3. häufbar - nicht häufbar
 - häufbar: mehre Ausprägungen annehmbar (z.B. Hobbies)
 - nicht häufbar: nur eine Ausprägung pro Merkmal (z.B. Geschlecht)

Skalenniveaus

Die Statistik unterteilt die möglichen Merkmalsausprägungen in drei Kategorien, die als Skalenniveaus bezeichnet werden. Diese Datentypen bestimmen die mit den gefundenen Merkmalsausprägungen möglichen Rechenverfahren.

Der Versuch, nicht anwendbare Methoden zu nutzen, führt zu unbrauchbaren Ergebnissen. Man muss also wissen, welche Art Daten vorliegen, bevor man mit der numerischen Auswertung beginnt.



4.1.2 Kennwerte einer Stichprobe

Häufigkeitsverteilung einer Stichprobe

Gegeben sei eine Stichprobe vom Umfang n aus einer (endlichen oder unendlichen) Grundgesamtheit und die Stichprobenwerte des interessierenden Merkmals. Der Abstand zwischen den kleinsten und dem größten Wert heißt Spannweite der Stichprobe.

In der Stichprobe vom Umfang n treten k verschiedene Werte auf. Man bestimmt, wie oft jeder Stichprobenwert x_i in der Stichprobe enthalten ist.

Diese Zahl $n_i \in \mathbb{N}$ heißt absolute Häufigkeit n_i des Stichprobenwertes x_i . Dabei gilt:

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = n.$$

Die relative Häufigkeit h_i erhält man, indem man die absolute Häufigkeit n_i durch die Anzahl n der Stichprobe dividiert:

$$h_i = \frac{n_i}{n}$$

Dabei gelten folgende Beziehungen:

$$0 \leq h_i \leq 1 \quad \text{und} \quad \sum_{i=1}^k h_i = h_1 + h_2 + \dots + h_k = 1$$

Die Stichprobe kann vollständig durch die Verteilungstabelle beschrieben werden:

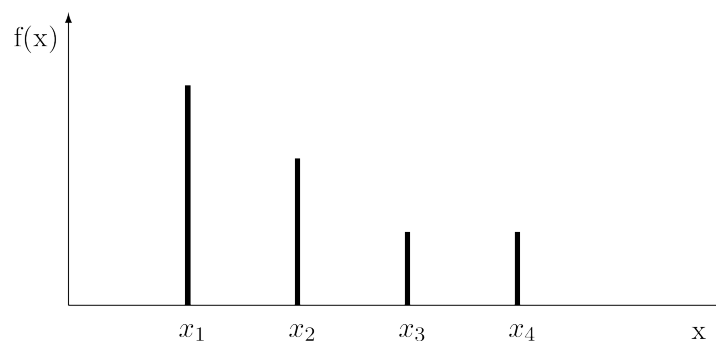
Stichprobenwert	x_1	x_2	x_3	...	x_k
absolute Häufigkeit	n_1	n_2	n_3	...	n_k
relative Häufigkeit	h_1	h_2	h_3	...	h_k

Die Verteilung der einzelnen Stichprobenwerte in der Stichprobe lässt sich durch die wie folgt definierte Häufigkeitsfunktion $f(x)$ darstellen:

$$f(x) = \begin{cases} h_i & \text{für } x = x_i \\ 0 & \text{für alle weiteren } x \end{cases}$$

Die Häufigkeitsfunktion besitzt also die Eigenschaften einer Wahrscheinlichkeitsfunktion. Sie ist die empirische Häufigkeitsfunktion für die Stichprobe.

Die Häufigkeitsfunktion wird oft durch ein Stabdiagramm dargestellt



Beispiel

Aus der laufenden Tagesproduktion von Gewindeschrauben mit einem Solldurchmesser von 5.0 mm ist eine Stichprobe vom Umfang $n = 25$ entnommen worden. Dabei ergaben sich folgende Stichprobenwerte:

4.9 4.8 5.0 5.2 5.2
 5.1 4.7 5.0 5.0 4.9
 4.8 4.9 5.1 5.0 5.0
 5.1 5.0 4.9 4.8 4.9
 4.9 5.0 5.0 5.1 5.0

Es treten nur 6 verschiedene Werte auf:

Stichprobenwert	4.7	4.8	4.9	5.0	5.1	5.2
absolute Häufigkeit	1	3	6	9	4	2
relative Häufigkeit, Häufigkeits-fktn. $f(x)$	0.04	0.12	0.24	0.36	0.16	0.08

Anmerkung:

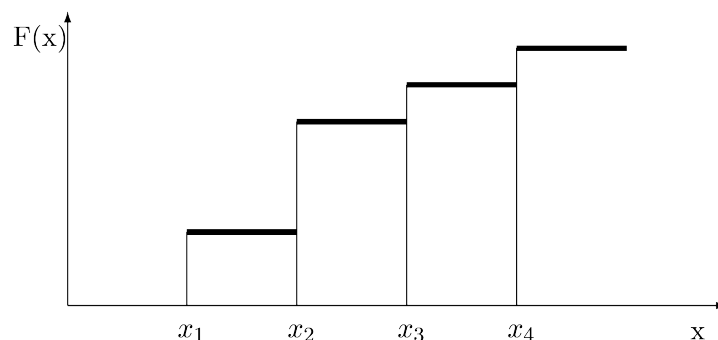
Bei vielen verschiedenen Einzelwerten ist es sinnvoll diese in Klassen einzuteilen und die Häufigkeit des Auftretens in dieser Klasse auszuwerten. Um die Klassenbreite N_k zu bestimmen, bedient man sich oft der **Faustformel** $N_k = \sqrt{n}$ mit $n = \text{Anzahl der Werte}$.

Verteilungsfunktion einer Stichprobe

Die Summe der relativen Häufigkeiten aller Stichprobenwerte, die $\leq x$ sind, heißt Verteilungsfunktion $F(x)$ der Stichprobe:

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

In der graphischen Darstellung erhält man eine Treppenfunktion:

**Beispiel:**

Für das vorige Beispiel ergibt sich folgende Verteilungsfunktion:

Stichprobenwert	4.7	4.8	4.9	5.0	5.1	5.2
Häufigkeitsfunktion $f(x)$	0.04	0.12	0.24	0.36	0.16	0.08
Verteilungsfunktion $F(x)$	0.04	0.16	0.40	0.76	0.92	1

Mittelwert, Varianz und Standardabweichung einer Stichprobe

Neben der Beschreibung einer Stichprobe über die Häufigkeitsfunktion oder die Verteilungsfunktion gibt es weitere Kennwerte um sie zu charakterisieren:

Der **Mittelwert $\bar{x}$** einer Stichprobe $x_1, x_2, \dots, x_n$ vom Umfang n ist das arithmetische Mittel der Stichprobenwerte:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Der **Median $x_{0.5}$** ist der Stichprobenwert, der genau in der **Mitte** einer geordneten Stichprobe liegt, d.h. darüber und darunter befinden sich gleich viele Stichprobenwerte.

Der **Modalwert x_{mod}** , ist derjenige Stichprobenwert, der in der Stichprobe am **häufigsten** vorkommt.

Ein **Streumaß** um den Mittelwert wird aus den Abweichungsquadraten gebildet. Die Größe

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

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

oder auch

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

heißt **Varianz s^2** der Stichprobe. Die Quadratwurzel aus der Varianz heißt **Standardabweichung s** der Stichprobe.

Beispiel:

Die Untersuchung des Wirkungsgrades X von 5 seriengleichen Kesseln einer Ölheizungsanlage eines bestimmten Fabrikats führte zu folgenden Messergebnissen:

92.4% 91.9% 92.0% 91.8% 91.9%

Lösung:

Der Mittelwert ergibt sich zu $\bar{x} =$

Die Varianz ergibt sich zu $s^2 =$

Die Standardabweichung ist $s \approx$

$\Rightarrow$ Der mittlere Wirkungsgrad der Heizkessel beträgt **92%** mit einer Varianz von ca. **0.06%** und einer Standardabweichung von **$\approx 0.23\%$** .

Berechnung von Mittelwert und Varianz mit Hilfe der Häufigkeitsfunktion:

Bei einer Stichprobe vom Umfang n und k Stichprobenwerten $x_1, x_2, \dots, x_k$, sowie der Häufigkeitsfunktion $f(x)$, berechnet sich

$$\text{Mittelwert: } \bar{x} = \sum_{i=1}^k x_i \cdot f(x_i)$$

$$\begin{aligned} \text{Varianz: } s^2 &= \frac{n}{n-1} \cdot \sum_{i=1}^k (x_i - \bar{x})^2 \cdot f(x_i) \quad \text{oder auch} \\ &= \frac{n}{n-1} \cdot \left(\sum_{i=1}^k x_i^2 \cdot f(x_i) - \bar{x}^2 \right) \end{aligned}$$

Beispiel:

Aus einer Serienfabrikation von Gewindeschrauben wurden wahllos 100 Schrauben entnommen und der jeweilige Durchmesser bestimmt. Dies führt zu folgender Häufigkeitsfunktion. Bestimmen Sie Mittelwert, Varianz und Standardabweichung.

x_i	3.50	3.51	3.52	3.53	3.54	3.55	3.56	3.57
$f(x_i)$	0.03	0.08	0.22	0.30	0.18	0.10	0.06	0.03

Lösung:

Mittelwert $\bar{x} =$

Varianz $s^2 =$

Die Standardabweichung ist somit:

Durchschnitt und Mittelwert

Der Unterschied zwischen Durchschnitt und Mittelwert ist, dass beim Durchschnitt selten erwähnt wird wie dieser berechnet wird, während zum Mittelwert meist die Berechnungsgrundlage genannt wird.

Umgangssprachlich wird oft der Durchschnitt mit dem arithmetischen Mittel gleichgesetzt.

Der Durchschnitt bzw. das arithmetische Mittel darf genau genommen nur für normalverteilte Werte verwendet werden, welche eine niedrige Anzahl an Ausreißern aufweist (siehe folgende Kapitel). Außerdem darf man den arithmetischen Mittelwert eigentlich nur für Merkmale verwenden bei denen mathematische Operationen Sinn machen.

Beispiel:

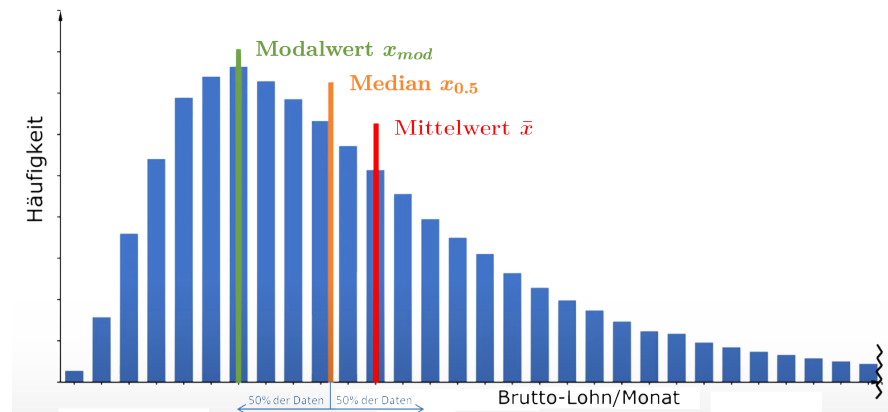
Die Noten einer Klasse sind meist nicht normalverteilt, da die Schüleranzahl oft nicht ausreicht und der Leistungsstand sowie die Aufgaben das nicht ermöglicht. Einzelne Ausreißer beeinflussen das Ergebnis somit unverhältnismäßig. Außerdem stimmen die Abstände der Noten meist nicht mit den Punkten zusammen (Note 5 ab 50% der Punkte). Besser wäre hier die Bestimmung des Medians (siehe nächstes Kap.), um eine anschauliche Bewertung der Klasse zu bekommen.

4.1.3 Quantile und Quartile

Quantile gehören zu den Lagemaßen der Statistik. Sie teilen eine bestimmte Menge an Daten so ein, dass ein Teil p kleiner oder gleich und der andere Teil $1 - p$ größer oder gleich dem Quantil ist.

Das 20%-Quantil oder auch 0.2-Quantil zum Beispiel sagt aus, dass genau 20 Prozent der Werte einer Verteilung unter dem Quantil liegen. Der Rest der Werte liegt darüber. Die Quantile sind für die wichtigsten praktischen Verteilungen tabelliert worden (siehe Anhang).

Manche Quantile sind in der Statistik so wichtig, dass sie einen eigenen Namen bekommen haben. So ist das 50%-Quantil nichts anderes als der **Median** bzw. $x_{0.5}$. Er teilt die Verteilung genau in der Mitte.



Quantile

Quantile sind in der Statistik die am häufigsten verwendete Form der Quantile. Sie unterteilen die Verteilung in vier gleich große Teile. Es gibt also das 0.25-Quantil ($x_{0.25}$, unteres Quartil), das 0.5-Quantil ($x_{0.5}$, Median) und das 0.75-Quantil ($x_{0.75}$, oberes Quartil).

Man muss unterscheiden ob $n \cdot p$ ganzzahlig oder nicht ganzzahlig ist:

$$x_p = \begin{cases} \frac{1}{2}(x_{(np)} + x_{(np+1)}) & \text{falls } n \cdot p \text{ ganzzahlig} \\ x_{(\lfloor np+1 \rfloor)} & \text{falls } n \cdot p \text{ nicht ganzzahlig} \end{cases}$$

Die Klammer $\lfloor \dots \rfloor$ im Index bedeutet, dass der Wert zwischen der Klammer immer abgerundet wird, egal wie nah er am nächsthöherem Wert liegt.

Beispiel:

Von zwei Stichproben (1,2,3,3,4,5,6,7) und (1,1,2,3,3,7,8) soll das 25% Quartil berechnet werden.

Lösung:

Normalerweise werden die Verteilungen der Größe nach aufsteigend sortiert. Hier liegen die Daten bereits sortiert vor.

Erste Stichprobe: $n \cdot p = 8 \cdot 0.25 = 2 \Rightarrow$ ganzzahlig:

$$x_{0.25} = \frac{1}{2}(x_2 + x_{2+1}) = \frac{1}{2}(2 + 3) = 2.5 \Rightarrow \text{Das erste Quartil der Verteilung ist 2.5.}$$

Das bedeutet, dass 25% der Werte kleiner als 2.5 sind.

Zweite Stichprobe: $n \cdot p = 7 \cdot 0.25 = 1.75 \Rightarrow$ nicht ganzzahlig:

$$x_{0.25} = (x_{(\lfloor 1.75 \rfloor + 1)}) = (x_{(1+1)}) = x_2 = 1$$

4.1.4 Boxplot

Der Box-Plot (auch Box-Whisker-Plot oder deutsch Kastengrafik) ist ein Diagramm, das zur grafischen Darstellung der Verteilung eines mindestens ordinalskalierten Merkmals verwendet wird.

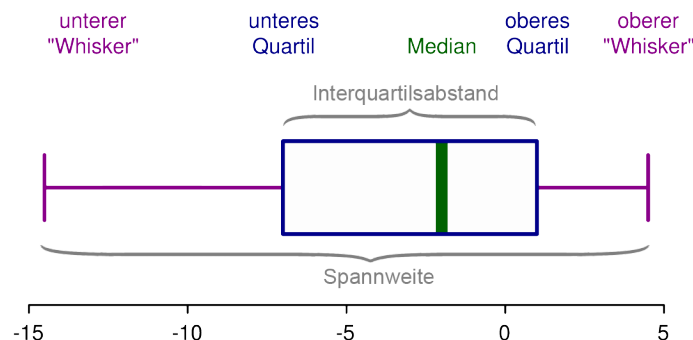
Es fasst dabei verschiedene **robuste** Streuungs- und Lagemaße in einer Darstellung zusammen.

Ein Box-Plot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich über diesen Bereich verteilen.

Deshalb werden alle Werte der sogenannten **Fünf-Punkte-Zusammenfassung**, also der Median, die zwei Quartile und die beiden Extremwerte, dargestellt.

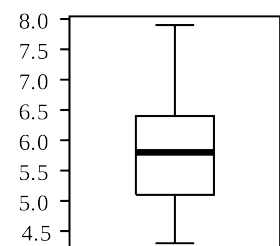
1. Den **Medianwert** als den mittleren der aufsteigend geordneten Beobachtungswerte: unterhalb und oberhalb dieses Wertes liegen dann je 50 % der Beobachtungen.
2. Das **untere Quartil** als den Wert, unterhalb dem 25 % der Werte liegen
3. Das **obere Quartil** als den Wert, unterhalb dem 75 % der Werte liegen.
4. Der **unterster Extremwert** (unterer "Whisker")
5. Der **oberste Extremwert** (oberer "Whisker")

Die Differenz zwischen oberem und unterem Quartil, den Wertebereich in dem die mittleren 50 % der Daten liegen, nennt man **Interquartilsabstand (IQR)**.



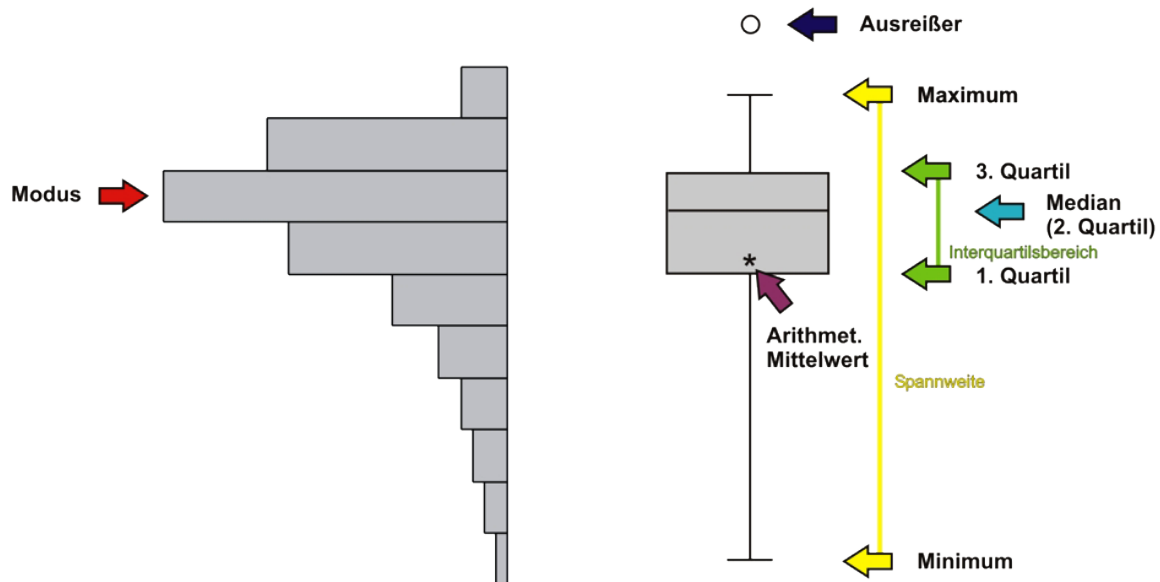
Beispiel:

```
⇒ par(las = 1)           # alle Achsenbeschriftungen
                           # horizontale Ausrichtung
⇒ boxplot(iris$Sepal.Length) # Boxplot Variable Sepal.Length
                           # des Irisdatensatzes aus R
```



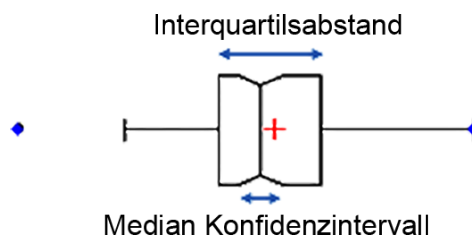
Boxplot Abwandlungen:

Eine Abwandlung besteht darin, das **arithmetische Mittel** und die **Ausreißer** in einen Box-Plot mit einzutragen. Es wird $\bar{x}$ dabei meist als Stern eingetragen. Da der Box-Plot ansonsten nur robuste Streuungs- und Lagemaße enthält, sollte das arithmetische Mittel als nicht-robustes Lagemaß eigentlich nicht in einen Box-Plot aufgenommen werden.



Im gekerbten (engl. notched) Box-Plot werden auch **Konfidenzintervalle** (Vertrauensbereich) für den Median aufgenommen.

Das Konfidenzintervall gibt den Bereich an, der bei unendlicher Wiederholung eines Zufallsexperiments mit einer gewissen Wahrscheinlichkeit (dem Konfidenzniveau) die wahre Lage des Parameters einschließt.

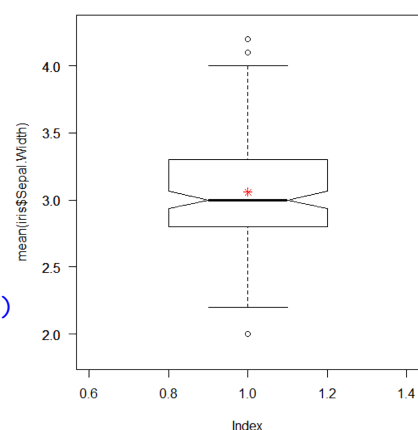


Beispiel:

⇒ `par(las = 1)`

⇒ `boxplot(iris$Sepal.Width, notch=TRUE)`

⇒ `points(mean(iris$Sepal.Width), col="red", pch=8)`



4.2 Schließende Statistik

4.2.1 Parameterschätzungen

Eine Grundgesamtheit mit einem Merkmal X ist durch die Verteilungsfunktion $F(x)$ vollständig charakterisiert. In der Anwendung stellt sich jedoch häufig das Problem, dass die Verteilungsfunktion vom Typ her bekannt ist, aber noch einige Parameter unbekannt sind. Bekannt sei, dass X normalverteilt ist, aber die Verteilung gemäß der Normalverteilung noch unbekannt, also die Parameter μ und σ .

Zwei wichtige Fragen die sich dabei ergeben, sind

1. Wie erhält man aus einer Stichprobe Schätzwerte für die Parameter?
2. Wie genau und sicher sind solche Schätzwerte?

Aufgaben der Parameterschätzung: Die Verteilungsfunktion einer Zufallsvariablen X sei vom Typ her bekannt, enthalte aber noch unbekannte Parameter. Zu den Aufgaben der Parameterschätzung gehören:

1. Bestimmung von Schätzwerten für die unbekannten Parameter unter Verwendung einer Stichprobe. Da der Schätzwert einem Punkt auf der reellen Achse entspricht, nennt man solche Schätzungen auch Punktschätzung
2. Konstruktion von sog. Konfidenzintervallen, in denen die unbekannten Parameter mit einer vorgegebenen Wahrscheinlichkeit vermutet werden. Man spricht von Intervallschätzung

Für die Schätzung der unbekannten Parameter einer Wahrscheinlichkeitsverteilung werden spezielle Funktionen benötigt. Sie ermöglichen die näherungsweise Berechnung dieser Parameter unter Verwendung einer konkreten Stichprobe, die man der Grundgesamtheit entnimmt. Schätzwerte für die Parameter spezieller Wahrscheinlichkeitsverteilung sind z.B.: (Anmerkung: In der Statistik werden Schätzwerte traditionell mit einem Dach gekennzeichnet).

Verteilung	Schätzwert für	Bemerkungen
Binomialverteilung $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$	Parameter p $\hat{p} = \frac{k}{n}$	k = Anzahl der Erfolge bei n -facher Ausführung eines Bernoulli-Experiments
Poisson-Verteilung $f(x) = \frac{\mu^x}{x!} \cdot e^{-\mu}$	Mittelwert μ $\hat{\mu} = \bar{x}$	$\bar{x}$ = Mittelwert der Stichprobe
Exponentialverteilung $f(x) = \lambda \cdot e^{-\lambda x}$	Parameter λ $\hat{\lambda} = \frac{1}{\bar{x}}$	$\bar{x}$ = Mittelwert der Stichprobe
Normalverteilung $f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}$	Mittelwert μ $\hat{\mu} = \bar{x}$ Varianz σ^2 $\hat{\sigma}^2 = s^2$	$\bar{x}$ = Mittelwert der Stichprobe s^2 = Varianz der Stichprobe

Die Herleitung der angegebenen Formeln zur Berechnung der Schätzwerte erfolgt z.B. mit Hilfe der Maximum-Likelihood-Methode oder mit der Methode der kleinsten Quadrate (MKQ).

Beispiele:

1. Die Lebensdauer T eines bestimmten elektronischen Bauelements genüge einer Exponentialverteilung mit dem unbekannten Parameter λ . Wir ermitteln einen Schätzwert $\hat{\lambda}$ für diesen Parameter anhand der folgenden Stichprobe:

i	1	2	3	4	5	6	7	8
t_i	950	980	1150	770	1230	1210	990	1120

Aus der Stichprobe erhält man den Mittelwert

2. Es soll der Ausschussanteil p einer Serienproduktion von Glühbirnen mittels einer Stichprobenuntersuchung geschätzt werden. Hierzu wurde eine Stichprobe von $n = 300$ Stück entnommen, wobei sich $k = 6$ Glühbirnen als defekt erwiesen.

4.2.2 Lineare Regressionsanalyse

Oft möchte man eine Größe y durch eine andere Größe x erklären oder vorhersagen.

Können wir zum Beispiel vorhersagen, wie sich der Absatz von Produkten (y) in Abhängigkeit des Verkaufspreises (x) (in gewissen Grenzen) entwickelt?

Im Grunde ist jede Darstellung einer mathematischen Funktion eine solche Vorhersage: Welches $y = f(x)$ ergibt sich für welches x ?

Hier geht es speziell nur um **lineare Funktionen**.

Anhand einiger Beispiele soll die (lineare) Regressionsanalyse vorgestellt werden.

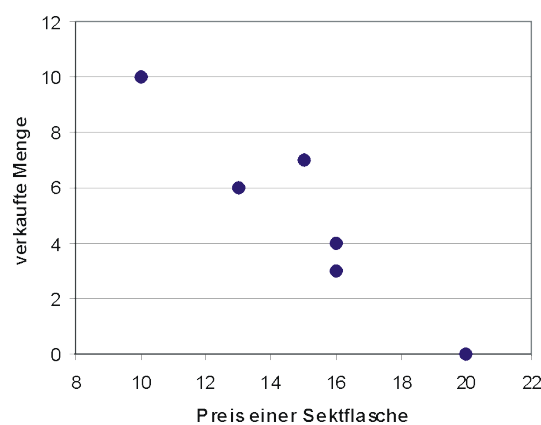
Beispiel:

Eine Sektkellerei möchte einen hochwertigen Sekt auf den Markt bringen. Für die Festlegung des Abgabepreises soll zunächst eine Preis-Absatz-Funktion ermittelt werden.

Dazu wurde in $n = 6$ Geschäften ein Testverkauf durchgeführt. Man erhielt sechs Wertepaare mit dem Ladenpreis x (in Euro) einer Flasche und die verkaufte Menge y an Flaschen:

Laden	i	1	2	3	4	5	6
Preis einer Flasche	x_i	20	16	15	16	13	10
verkaufte Menge	y_i	0	3	7	4	6	10

Streudiagramm Preis - Absatz



Man geht von folgendem stochastischen Modell aus:

Man betrachtet zwei Variablen, die vermutlich (Annahme) ungefähr in einem linearen Zusammenhang aufweisen:

$$y \approx \alpha + \beta x$$

stehen. Dabei sind y als abhängige und x als unabhängige Variable definiert.

Man nennt x auch erklärende oder **exogene** Variable und y Zielvariable oder **endogene** Variable. Es existieren von x und y je n Beobachtungen x_i und y_i ($i = 1, \dots, n$).

Der funktionale Zusammenhang $y = f(x)$ zwischen x und y kann nicht exakt festgestellt werden, da $\alpha + \beta x$ von einer Störgröße u überlagert wird, die nichterfassbare Einflüsse (menschliches Verhalten, Messungenauigkeiten usw.) mit einschließt. Es ergibt sich also das Modell:

$$y = \alpha + \beta x + u$$

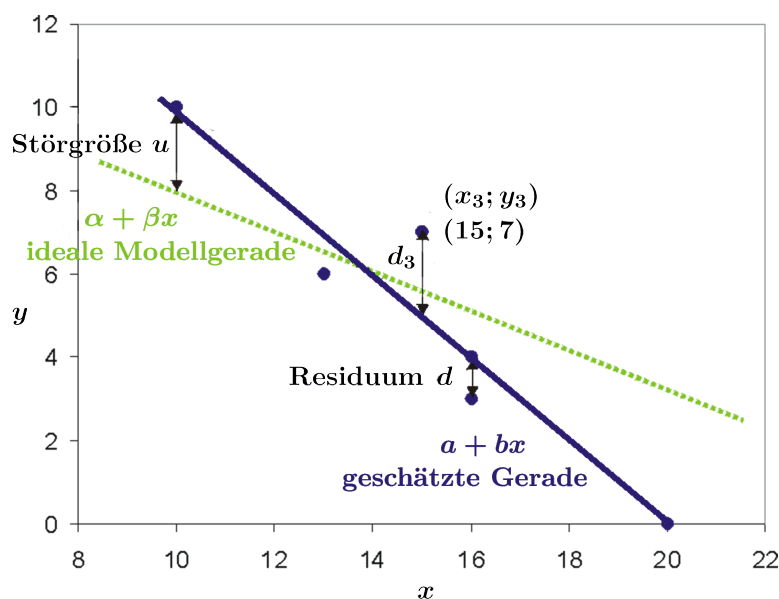
Da α und βx nicht bekannt sind, kann y auch nicht in die Komponenten $\alpha + \beta x$ und u zerlegt werden.

Es soll eine mathematische Schätzung für die Parameter α und β durch zwei Konstanten a und b gefunden werden, und zwar so, dass sich ergibt

$$y_i = a + bx_i + d_i$$

wobei d_i das **Residuum** bezeichnet, die Abweichung des beobachteten y - Wertes vom Geschätzten. Das Residuum d ist also der (lokale) Schätzwert für u . Es gibt verschiedene Möglichkeiten, die Regressionsgerade zu schätzen.

Man könnte z. B. eine Gerade so durch den Punkteschwarm legen, dass die Quadratsumme der Residuen, also der senkrechten Abweichungen d_i der Punkte von dieser Ausgleichsgeraden minimiert wird (**Methode der kleinsten Quadrate**).



Bei d_i handelt es sich um einen zufälligen Fehlerterm (error). Man kann sich z.B. **Messfehler** oder nicht-systematische Effekte darunter vorstellen. Dies gilt allerdings nur, wenn das funktionale Modell $y = a + bx$ stimmt. Das kann man entweder statistisch prüfen oder einfach annehmen.

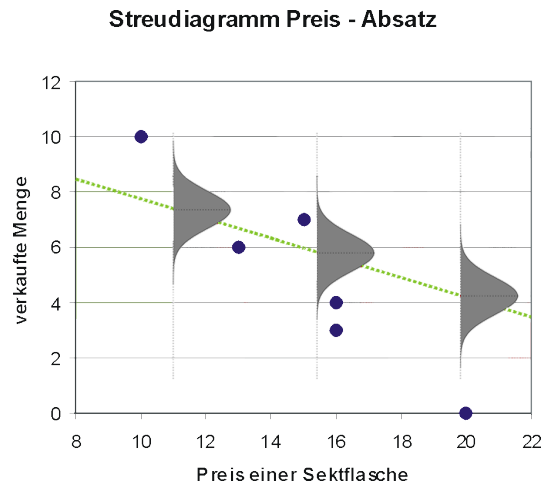
Typischerweise nehmen wir für die Fehler an, dass sie **normalverteilt** sind, d.h.

$$u_i \sim \mathcal{N}(0, \sigma^2)$$

Die Streuung um die Gerade herum wird durch den Fehlerterm verursacht und ist durch die Dichte der Normalverteilung illustriert.

Da $E(u_i) = 0$, gibt es keine systematischen Abweichungen von der Geraden.

Zusätzlich ist $\text{Var}(u_i) = \sigma^2$, d.h. die Streuung um die Gerade ist überall gleich groß. Diese Eigenschaft nennt man **Homoskedastizität**.



Axiome des linearen Regressionsmodells:

Damit dieses Verfahren also sinnvolle Ergebnisse liefert, wurden für das Lineare Regressionsmodell verteilungstheoretische Annahmen getroffen. Es gilt

$$y_i = \alpha + \beta x_i + u_i$$

und wir definieren die Störgröße u_i als Zufallsvariable. Die Annahmen des linearen Regressionsmodells sind dann:

1. Alle u_i haben den Erwartungswert Null:

$$E(u_i) = 0, \quad (i = 1, \dots, n)$$
2. Alle u_i haben die gleiche Varianz:

$$\text{Var } u_i = \text{Var } u_j, \quad (i, j = 1, \dots, n, i \neq j)$$
3. Die u_i sind sämtlich stochastisch unabhängig voneinander.

Minimierung: Methode der kleinsten Quadrate

Die herkömmliche Methode, die sich auf der Basis der Axiome ergibt, ist die Minimum-Quadrat-Methode oder Methode der kleinsten Quadrate. Man minimiert also die summierten Quadrate der Residuen bzgl. a und b ,

$$RSS = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2 \rightarrow \min \quad (\text{engl. } RSS = \text{Residual Sum of Squares})$$

Ausmultiplizieren der Klammer:

$$\begin{aligned} S &= \sum_{i=1}^n (y_i - a - bx_i) \cdot (y_i - a - bx_i) \\ S &= \sum_{i=1}^n (y_i^2 - y_i a - y_i b x_i - a y_i + a^2 + a b x_i - y_i b x_i + a b x_i + b^2 x_i^2) \\ S &= \sum_{i=1}^n (y_i^2 - 2y_i a - 2y_i b x_i + a^2 + 2a b x_i + b^2 x_i^2) \\ S &= \sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n y_i - 2b \sum_{i=1}^n y_i x_i + n a^2 + 2ab \sum_{i=1}^n x_i + b^2 \sum_{i=1}^n x_i^2 \end{aligned}$$

Minimieren durch (partiell) Ableiten und anschließendes Nullsetzen:

$$\begin{aligned} \frac{\partial S}{\partial a} &= -2 \sum_{i=1}^n y_i + 2na + 2b \sum_{i=1}^n x_i \\ \frac{\partial S}{\partial b} &= -2 \sum_{i=1}^n x_i y_i + 2a \sum_{i=1}^n x_i + 2b \sum_{i=1}^n x_i^2 \end{aligned}$$

ergibt:

$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

Wir erhalten die gesuchten **Regressionskoeffizienten** als die Lösungen

$$b = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad \text{und}$$

$$a = \bar{y} - b \bar{x}$$

mit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ dem Mittelwert der x -Daten (y entsprechend).

Mit dem Verschiebungssatz kann man b auch darstellen als:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} = \frac{\text{Kovarianz}}{\text{Varianz}}$$

Korrelationskoeffizient

Der Korrelationskoeffizient r ist ein Maß für den Grad des linearen Zusammenhangs zwischen zwei metrisch skalierten Merkmalen, das nicht von den Maßeinheiten der Messung abhängt und somit dimensionslos ist.

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

dabei gilt:

r = Korrelationskoeffizient

x_i = Werte der Variablen x in einer Stichprobe

$\bar{x}$ = Mittel der Werte der Variablen x

y_i = Werte der Variablen y in einer Stichprobe

$\bar{y}$ = Mittel der Werte der Variablen y

r kann Werte zwischen -1 und +1 annehmen. Bei einem Wert von +1 (bzw. -1) besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen.

Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab. Allerdings können diese ungeachtet dessen in nichtlinearer Weise voneinander abhängen, z.B. exponentiell.

Damit ist der Korrelationskoeffizient kein geeignetes Maß für die (reine) stochastische Abhängigkeit von Merkmalen.

Der Korrelationskoeffizient wurde erstmals vom britischen Naturforscher Sir Francis Galton (1822–1911) in den 1870er Jahren verwendet. Karl Pearson lieferte schließlich eine formal-mathematische Begründung für den Korrelationskoeffizienten. Da er von Auguste Bravais und Pearson populär gemacht wurde, wird der Korrelationskoeffizient auch Pearson-Korrelation oder **Bravais-Pearson-Korrelation** genannt.

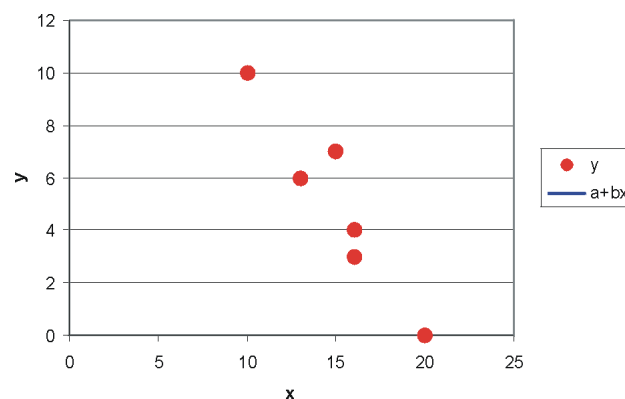
Das Quadrat des Korrelationskoeffizienten r stellt das **Bestimmtheitsmaß R^2** dar. Das Bestimmtheitsmaß gibt allerdings nur darüber Auskunft, wie gut die Anpassung ist - ein Wert nahe 1 heißt also wirklich nur, dass unser Model recht gut an die Werte angepasst ist.

Aufgabe:

Berechnen Sie aus dem Sekt-Beispiel die Regressionsgerade $\hat{y} = a + bx$ und bestimmen den Korrelationskoeffizient nach Bravais-Pearson.

Lösung:

	Preis einer Flasche	verkaufte Menge	$x_i - \bar{x}$	$y_i - \bar{y}$				
#	x_i	y_i	x^*	y^*	x^*y^*	x^*x^*	y^*y^*	$\hat{y}$
1								
2								
3								
4								
5								
6								
Σ								



Interpretation: Da der Korrelationskoeffizient mit -0,949 sehr nahe an -1 liegt, besteht ein fast vollständiger negativer (linearer) Zusammenhang zwischen dem Preis eine Flasche Sekt und der verkauften Menge.

4.2.3 Gewinnung von Schätzfunktionen

Schätzung des Mittelwertes

Man kann zeigen, dass der Mittelwert $\bar{x}$ einer Zufallsstichprobe $x_1, x_2, \dots, x_n$ als geeigneter Schätzwert für den unbekannten Mittelwert μ der Wahrscheinlichkeitsverteilung der Zufallsvariablen X genommen werden kann:

$$\mu \approx \hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Diesen Schätzwert kann man als spezielle Realisierung der Funktion

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

ansetzen. Dabei sind $X_1, X_2, \dots, X_n$ unabhängige Zufallsvariable, die alle die gleiche Verteilung besitzen, und $x_1, x_2, \dots, x_n$ sind die Werte dieser Zufallsvariablen.

Die Funktion $\bar{X}$ ist ebenfalls eine Zufallsvariable. Sie wird als Schätzfunktion für den unbekannten Mittelwert μ der Zufallsvariablen X bezeichnet.

Definition:

Eine **Stichprobenfunktion** ist eine Zufallsvariable, die von n unabhängigen Zufallsvariablen abhängt, die alle der gleichen Verteilungsfunktion genügen.

Die oben beschriebene Schätzfunktion für den unbekannten Mittelwert ist eine Stichprobenfunktion.

Eigenschaften der Schätzfunktion $\bar{X}$

1. Die Schätzfunktion $\bar{X}$ besitzt den Erwartungswert μ :

$$E(\bar{X}) = \mu$$

Eine Schätzfunktion mit der Eigenschaft $E(\bar{X}) = \mu$ heißt erwartungstreu

2. Die Schätzfunktion $\bar{X}$ besitzt die Varianz

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Das bedeutet, dass die Varianz der Schätzfunktion $\bar{X}$ mit zunehmendem Stichprobenumfang n abnimmt und gegen Null konvergiert für $n \rightarrow \infty$. Dies bedeutet, dass die Werte der Zufallsvariablen $\bar{X}$ mit zunehmendem n immer weniger um den Mittelwert μ streuen. Die Schätzfunktion $\bar{X}$ heißt dann konsistent.

3. Es gibt weitere erwartungstreue Schätzfunktionen für den Mittelwert. Man kann jedoch zeigen, dass $\bar{X}$ die erwartungstreue Schätzfunktion mit der kleinsten Varianz ist.

Kriterien für eine optimale Schätzfunktion

Schätzfunktionen für einen unbekannten Parameter ϑ sind spezielle Stichprobenfunktionen vom Typ

$$\Theta = g(X_1, X_2, \dots, X_n)$$

die für jede konkrete Stichprobe $x_1, x_2, \dots, x_n$ einen Schätzwert

$$\hat{\vartheta} = g(x_1, x_2, \dots, x_n)$$

für den Parameter ϑ liefern. Dabei sind $X_1, X_2, \dots, X_n$ unabhängige Zufallsvariable, die alle die gleiche Verteilungsfunktion $F(x)$ besitzen.

Eine Schätzfunktion ist optimal, wenn sie die folgenden Eigenschaften besitzt:

1. Die Schätzfunktion Θ ist erwartungstreu, d.h. ihr Erwartungswert ist gleich dem zu schätzenden Parameter

$$E(\Theta) = \vartheta$$

2. Die Schätzfunktion Θ ist konsistent, d.h. Θ konvergiert mit zunehmendem Stichprobenumfang n gegen den Parameter ϑ
3. Die Schätzfunktion Θ ist effizient, d.h. es gibt bei gleichem Stichprobenumfang n keine andere erwartungstreue Schätzfunktion mit einer kleineren Varianz

Schätzungen für die Varianz

Die Varianz s^2 einer Zufallsstichprobe $x_1, x_2, \dots, x_n$ liefert einen geeigneten Schätzwert $\hat{\sigma}^2$ für die unbekannte Varianz σ^2 der Wahrscheinlichkeitsverteilung der zugehörigen Zufallsvariablen X .

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

Die zugehörige Schätzfunktion ist die Stichprobenfunktion

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

Sie ist erwartungstreu. Ein Freiheitsgrad fehlt $(n-1)$, da $\bar{X}$ bereits enthalten ist.

Vereinzelte wird auch die Stichprobenfunktion

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

als Schätzfunktion genommen. Sie ist jedoch nicht erwartungstreu (verzerrt mit systematischen Fehler).

Als Schätzfunktion für die Standardabweichung σ verwendet man die Stichprobenfunktion $S = \sqrt{S^2}$. Sie ist ebenfalls nicht erwartungstreu.

Grund: Wurzelfunktion und Erwartungswertbildung sind keine linearen Funktionen und lassen sich daher nicht vertauschen.

Maximum-Likelihood-Methode

Es gibt mehrere Verfahren eine Schätzfunktion zu gewinnen, die Maximum-Likelihood-Methode ist das wohl wichtigste Verfahren zur Gewinnung einer Schätzfunktion.

Die Idee des Verfahrens ist es, als Schätzwerte für die wahren Parameter der Grundgesamtheit diejenigen auszuwählen, bei denen die beobachteten Stichprobenrealisationen am wahrscheinlichsten sind.

X sei eine Zufallsvariable, deren Wahrscheinlichkeitsdichte f von einem Parameter ϑ abhängt. Liegt eine Stichprobe mit n Realisierungen $x_1, x_2, \dots, x_n$ von n unabhängigen und identisch verteilten Zufallsvariablen $X_1, X_2, \dots, X_n$ vor, so lässt sich die Wahrscheinlichkeitsfunktion wie folgt faktorisieren:

$$f(x_1, x_2, \dots, x_n, \vartheta) = f_{X_1}(x_1, \vartheta) \cdot f_{X_2}(x_2, \vartheta) \cdot \dots \cdot f_{X_n}(x_n, \vartheta)$$

Hier kann nun für feste Realisierungen $x_1, x_2, \dots, x_n$ die Dichte als Funktion von ϑ betrachtet werden. Dies führt zur Likelihood-Funktion

$$L(\vartheta) = \prod_{i=1}^n f_{X_i}(x_i, \vartheta)$$

Da der Logarithmus streng monoton wachsend ist, liegt das Maximum von $L(\vartheta)$ an derselben Stelle wie das der **Log-Likelihood-Funktion** $\ell(\vartheta) = \ln L(\vartheta) = \sum_{i=1}^n \ln f_{X_i}(x_i, \vartheta)$, die aus dem Produkt eine Summe macht und sich dadurch meist deutlich einfacher (ableiten und) maximieren lässt.

Wird diese Funktion in Abhängigkeit von ϑ maximiert, so erhält man die Maximum-Likelihood-Schätzung für ϑ . Es wird also der Wert von ϑ gesucht, bei dem die Stichprobenwerte $x_1, x_2, \dots, x_n$ das höchste Dichtemaximum haben.

Beispiele:

1. Die Zahl der Anrufe bei zwei Telefonisten in einer Stunde kann mit einer Poisson-Verteilung

$$X_1 \sim P_0(\lambda) \text{ und } X_2 \sim P_0(\lambda)$$

modelliert werden. Beim ersten Telefonisten gehen drei und beim zweiten fünf Anrufe pro Stunde unabhängig voneinander ein. Die Likelihood-Funktion für den unbekannten Parameter λ ergibt sich als

$$L(\lambda) = P(\{X_1 = 3\} \cap \{X_2 = 5\}) = P(X_1 = 3) \cdot P(X_2 = 5)$$

Setzt man die Werte in die Wahrscheinlichkeitsfunktion

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

ein, so folgt

$$L(\lambda) = \frac{\lambda^3}{3!} e^{-\lambda} \cdot \frac{\lambda^5}{5!} e^{-\lambda} = \frac{\lambda^8}{3! \cdot 5!} e^{-2\lambda}$$

Zur Bestimmung des Maximums, wird die Funktion abgeleitet:

$$L'(\lambda) = \frac{1}{3! \cdot 5!} (8\lambda^7 e^{-2\lambda} - 2\lambda^8 e^{-2\lambda}) = \frac{2e^{-2\lambda}}{3! \cdot 5!} \cdot \lambda^7 (4 - \lambda)$$

und die Nullstellen bestimmt: $\lambda_1 = 0$ und $\lambda_2 = 4$.

An der Stelle $\lambda = 4$ hat die Funktion ein Maximum und die ist der Maximum- Likelihood- Schätzwert.

Im allgemeinen Fall, mit n Telefonisten, die jeweils x_i Anrufe pro Stunde erhalten, ergibt sich die Likelihoodfunktion als

$$L(\lambda) = \frac{\lambda^{x_1+x_2+\dots+x_n}}{x_1!x_2!\dots x_n!} \cdot e^{-n\lambda}$$

Die Ableitung nach λ ergibt

$$\begin{aligned} L'(\lambda) &= \frac{(x_1 + x_2 + \dots + x_n)\lambda^{x_1+x_2+\dots+x_n-1}}{x_1!x_2!\dots x_n!} \cdot e^{-n\lambda} - n \cdot \frac{\lambda^{x_1+x_2+\dots+x_n}}{x_1!x_2!\dots x_n!} \cdot e^{-n\lambda} \\ &= \frac{e^{-n\lambda}}{x_1!x_2!\dots x_n!} \cdot \lambda^{x_1+x_2+\dots+x_n-1}(x_1 + x_2 + \dots + x_n - n\lambda) \end{aligned}$$

Die Nullstellen hiervon sind

$$\lambda_1 = 0 \text{ und } x_1 + x_2 + \dots + x_n - n\lambda = 0 \implies \lambda_2 = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Als Schätzung ergibt sich also

$$\hat{\lambda} = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}$$

und die zugehörige Schätzfunktion als

$$\hat{\lambda} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

2. Eine Urne enthält $N = 8$ Kugeln, die entweder rot oder schwarz sind. Die genaue Anzahl M der roten Kugeln ist nicht bekannt. Nacheinander werden $n = 4$ Kugeln gezogen und jeweils wieder zurück in die Urne gelegt. Beobachtet werden

x_1	x_2	x_3	x_4
1	1	0	1
rot	rot	schwarz	rot

Gesucht ist die nach dem Maximum-Likelihood-Prinzip plausibelste Zusammensetzung der Kugeln in der Urne.

In jedem Zug ist die Wahrscheinlichkeit, eine rote Kugel zu ziehen, gleich M/N . Wegen der Unabhängigkeit der Ziehungen ist die Wahrscheinlichkeit des beobachteten Ergebnisses und damit die zugehörige Likelihood-Funktion in Abhängigkeit vom unbekannten Parameter M gegeben durch

$$L(M) =$$

Es ergeben sich hierfür folgende Funktionswerte:

Die Likelihoodfunktion ist also maximal für **M=6** und ist somit der plausibelste Wert für die Realisierung dreier roter Kugeln bei vier Ziehungen und somit ein Schätzwert nach der Maximum-Likelihood-Methode.

3. Um den Ausschussanteil p in der Tagesproduktion von Glühbirnen zu schätzen wird eine Stichprobe vom Umfang $n = 120$ entnommen. dabei sind $k = 6$ Glühbirnen defekt. Die Maximum-Likelihood-Methode liefert für den Ausschussanteil p den Schätzwert

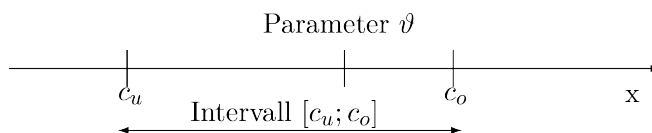
$$\hat{p} =$$

Man kann also davon ausgehen, dass im Mittel jede zwanzigste Glühbirne in der Tagesproduktion defekt ist.

Intervallschätzungen

Im vorigen Kapitel haben wir uns damit beschäftigt, näherungsweise einen unbekannten Parameter der Wahrscheinlichkeitsverteilung zu schätzen. Dies erhielten wir durch eine Schätzfunktion aus einer konkreten Stichprobe. Diese sogenannte Punktschätzung ermöglicht jedoch keine Aussage über die Genauigkeit der Schätzung. Der aus einer Zufallsstichprobe gewonnene Schätzwert kann noch erheblich vom tatsächlichen Wert abweichen.

Es liegt daher die Idee nahe, anhand einer Stichprobe ein Intervall $[c_u; c_o]$ zu bestimmen, das den Parameter mit Sicherheit enthält.



Ein solches Intervall kann es aber nicht geben, da absolut sichere Rückschlüsse von einer Stichprobe auf die Gesamtheit grundsätzlich nicht möglich sind. Auch der Versuch ein Intervall anzugeben, das den Parameter ϑ mit einer festen großen Wahrscheinlichkeit enthält, ist nicht möglich. Daher:

Bestimmung eines Konfidenz- (oder Vertrauens-) Intervalls für den unbekannten Parameter ϑ einer vom Typ her bekannten Wahrscheinlichkeitsverteilung

X sei eine Zufallsvariable, deren Wahrscheinlichkeitsverteilung noch einen unbekannten Parameter ϑ enthalte. Für diesen Parameter kann man aus einer Stichprobe ein Konfidenzintervall wie folgt bestimmen:

1. Zunächst wählt man ein Konfidenzniveau $\gamma = 1 - \alpha$. (mit α = Irrtumswahrsch.)
2. Es werden für den Parameter ϑ zwei Schätzfunktionen

$$\Theta_u = g_u(X_1, X_2, \dots, X_n)$$

$$\Theta_o = g_o(X_1, X_2, \dots, X_n)$$
 bestimmt, die mit der gewählten Wahrscheinlichkeit γ den wahren Wert des Parameters ϑ einschließen:

$$P(\Theta_u \leq \vartheta \leq \Theta_o) = \gamma$$
3. Aus der konkreten Stichprobe werden die Werte der beiden Stichprobenfunktionen Θ_u und Θ_o berechnet:

$$c_u = g_u(x_1, x_2, \dots, x_n)$$

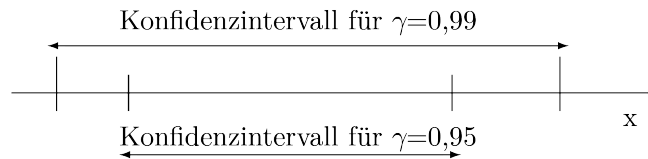
$$c_o = g_o(x_1, x_2, \dots, x_n)$$
 Sie liefern die Grenzen des gesuchten Konfidenzintervalls

$$c_u \leq \vartheta \leq c_o$$

4. Das Konfidenzintervall für den unbekannten Parameter ϑ lautet:

$$c_u \leq \vartheta \leq c_o$$

Der wahre Wert des Parameters ϑ liegt dann mit einem Vertrauen von $\gamma \cdot 100\%$ in diesem Intervall



Beispiel:

In einem konkreten Anwendungsfall wählen wir ein Vertrauensniveau von $\gamma = 95\%$. Entnimmt man der Grundgesamtheit 100 Stichproben, so kann man darauf vertrauen, dass in etwa 95 Fällen der unbekannte Parameter ϑ in das zugehörige Konfidenzintervall fällt.

Die Irrtumswahrscheinlichkeit ist somit 5%. Man trifft also in ca. 95 Fällen eine richtige und in etwa 5 Fällen eine falsche Entscheidung.

Konfidenzintervall für den unbekannten Mittelwert μ einer Normalverteilung bei bekannter Varianz σ^2

X sei eine normalverteilte Zufallsvariable mit dem unbekannten Mittelwert μ und der als bekannt vorausgesetzten Varianz σ^2 . Für den Mittelwert μ lässt sich dann unter Verwendung einer Stichprobe schrittweise ein Konfidenzintervall bestimmen:

1. Man wählt zunächst ein bestimmtes Vertrauensniveau γ (meist $\gamma = 0.95$ oder $\gamma = 0.99$).

2. Man berechnet die Konstante c aus der Bedingung

$$P(-c \leq U \leq c) = \gamma$$

für die standardnormalverteilte Zufallsvariable

$$U = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

unter Verwendung einer Tabelle oder eines SW-Programmes (z.B. R)

Dabei bedeuten:

$\bar{X}$: Schätzfunktion für den unbekannten Mittelwert μ der normalverteilten Grundgesamtheit

σ : Standardabweichung der normalverteilten Grundgesamtheit

n : Umfang der verwendeten Stichprobe

3. Berechnung des Mittelwertes $\bar{x}$ der konkreten Stichprobe
4. Das Vertrauensintervall für den unbekannten Mittelwert μ der normalverteilten Grundgesamtheit lautet dann

$$\bar{x} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}}$$

Der wahre Wert des Mittelwertes μ liegt mit einem Vertrauen von $\gamma \cdot 100\%$ in diesem Intervall

Bemerkungen:

1. Bei vielen Messinstrumenten wird die Varianz σ^2 bereits vom Hersteller als eine Gerätekonstante mitgegeben.
2. Häufig wird auch die Irrtumswahrscheinlichkeit α vorgegeben; oft $\alpha = 0.05$ oder $\alpha = 0.01$
3. Das Konfidenzintervall besitzt die Länge $l = \frac{2c\sigma}{\sqrt{n}}$ und hängt damit von der Standardabweichung σ , dem Stichprobenumfang n und dem Vertrauensniveau γ ab.
Für feste Werte von σ und γ gilt daher

$$l \sim \frac{1}{\sqrt{n}}$$

Das bedeutet, dass eine lineare Kürzung des Konfidenzintervalls stets durch eine quadratische Vergrößerung des Stichprobenumfangs erreicht werden kann.

Beispiele:

1. Die Längenmessung (in mm) von 10 Schrauben, die zufällig aus einem Sortiment ausgewählt wurden, führt zu folgendem Ergebnis:

i	1	2	3	4	5	6	7	8	9	10
x_i	10	8	9	10	11	11	9	12	8	12

Wir setzen voraus, dass die Stichprobe aus einer normalverteilten Grundgesamtheit mit der Varianz $\sigma^2 = 4 \text{ mm}^2$ stammt und wollen für den unbekannten Mittelwert μ ein Konfidenzintervall zum Vertrauensniveau $\gamma = 0.95$ bestimmen:

- a) Das Vertrauensniveau ist mit $\gamma = 0.95$ vorgegeben
- b) Die Bestimmungsgleichung für die Konstante c lautet

Daraus folgt dann:

Damit ergibt sich

- c) Für den Mittelwert der Stichprobe erhält man:

$$\bar{x} =$$

- d) Mit $n = 10$, $\bar{x} = 10 \text{ mm}$ und $\sigma = 2 \text{ mm}$ ist das Konfidenzintervall für den Mittelwert μ der normalverteilten Grundgesamtheit:

Man kann daher mit einem Vertrauen von 95% davon ausgehen, dass der wahre Wert von μ in diesem Intervall der Länge $l = 2.48 \text{ mm}$ liegt

2. Einer normalverteilten Grundgesamtheit mit der Varianz $\sigma^2 = 100$ soll eine Stichprobe vom Umfang n entnommen werden. Wie groß muss man den Stichprobenumfang wählen, damit das Konfidenzintervall für den unbekannten Mittelwert μ bei einem Vertrauensniveau $\gamma = 99\%$ die Länge $l = 2$ hat?
Zunächst bestimmt man n

Die unbekannte Konstante c kann man aus der Bedingung

wie folgt bestimmen:

Der gesuchte Stichprobenumfang beträgt folglich:

$$n =$$

Konfidenzintervall für den unbekannten Mittelwert μ einer Normalverteilung bei unbekannter Varianz σ^2

X sei eine normalverteilte Zufallsvariable mit dem unbekannten Mittelwert μ und der ebenfalls unbekannten Varianz σ^2 . Für den Mittelwert μ lässt sich dann unter Verwendung einer **Stichprobe** schrittweise ein Konfidenzintervall bestimmen:

1. Man wählt zunächst ein bestimmtes Vertrauensniveau γ (meist $\gamma = 0.95$ oder $\gamma = 0.99$)
2. Man berechnet die Konstante c aus der Bedingung

$$P(-c \leq T \leq c) = \gamma$$

für die Zufallsvariable. Achtung: Diese ist NICHT mehr normalverteilt!

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

unter Verwendung einer Tabelle oder eines SW-Programmes.

Dabei bedeuten:

$\bar{X}$: Schätzfunktion für den unbekannten Mittelwert μ der normalverteilten Grundgesamtheit

S : Schätzfunktion für die unbekannten Standardabweichung σ der normalverteilten Grundgesamtheit

n : Umfang der verwendeten Stichprobe

3. Berechnung des Mittelwertes $\bar{x}$ und der Varianz s^2 , bzw. der Standardabweichung s der konkreten Stichprobe

4. Das Intervall für den unbekannten Mittelwert μ der normalverteilten Grundgesamtheit lautet dann

$$\bar{x} - c \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{s}{\sqrt{n}}$$

Der wahre Wert des Mittelwertes μ liegt mit einem Vertrauen von $\gamma \cdot 100\%$ in diesem Intervall

Bemerkungen:

1. Bei unbekannter Varianz σ^2 sind die Intervalle für den Mittelwert μ stets breiter als bei bekannter Varianz (bei gleichem Vertrauensniveau und gleichem Stichprobenumfang)
2. Ist die Varianz σ^2 nicht bekannt, sondern muss geschätzt werden, so werden die Werte der Normalverteilung durch die t-Verteilung ersetzt.
3. Bei umfangreichen Stichproben ($n > 30$) kann die Standardabweichung σ der Grundgesamtheit durch die Standardabweichung s der Stichprobe geschätzt werden: $\sigma \approx s$.

In diesem Sonderfall darf man daher von einer normalverteilten Grundgesamtheit mit der bekannten Varianz $\sigma^2 \approx s^2$ ausgehen und das vorhergehenden Verfahren anwenden

Beispiel:

Die Messung von 8 zufällig aus der Serienproduktion entnommenen Widerständen (in Ω) führt zu folgendem Messprotokoll

i	1	2	3	4	5	6	7	8
x_i	100	104	98	96	101	104	98	99

Wir setzen voraus, dass die Stichprobe aus einer normalverteilten Grundgesamtheit mit der unbekannten Varianz und unbekanntem Mittelwert stammt. Es soll ein Konfidenzintervall für den Mittelwert bestimmt werden.

1. Als Vertrauensniveau wählen wir $\gamma = 0.95$.
2. Die Bestimmungsgleichung für die Konstante c lautet

Daraus folgt dann:

Mit Hilfe von Tabellen oder SW-Programmen erhalten wir daraus für $f = n - 1 = 7$ Freiheitsgrade den folgenden Wert für c :

Damit ergibt sich die gesuchte Konstante c zu

$$F(c) =$$

3. Für den Mittelwert der Stichprobe erhält man:

$$\bar{x} =$$

Für die Varianz erhält man:

$$s^2 =$$

Für die Standardabweichung erhält man daraus

$$s =$$

4. Mit $n = 8$, $c = 2.365$ mm und $s = 2.878$ erhält man das Konfidenzintervall für den unbekannten Mittelwert μ der normalverteilten Grundgesamtheit durch

Man kann daher mit einem Vertrauen von 95% davon ausgehen, dass der wahre Wert von μ in diesem Intervall der Länge $l = 4.812$ Ω liegt.

4.3 Hypothesentests

Unter einer **Hypothese** versteht man Annahmen, Vermutungen oder Behauptungen über die Wahrscheinlichkeitsverteilung einer Zufallsvariablen oder Grundgesamtheit und deren Parameter. Für die Überprüfung von Hypothesen werden sogenannte **Parametertests** durchgeführt.

Dabei handelt es sich um ein statistisches Prüfverfahren für einen unbekannten **Parameter ϑ** in der Wahrscheinlichkeitsverteilung einer Zufallsvariablen oder Grundgesamtheit. Die Art der Verteilung (z.B. Binomialverteilung, Normalverteilung etc.) wird als bekannt vorausgesetzt.

Solche Hypothesentests werden oftmals durchgeführt, wenn man:

- *vermutet*, dass eine aufgestellte Behauptung nicht stimmt
- *eine Beobachtung macht*, die nicht zu der bisher geltenden Lehrmeinung passt
- *hofft* oder *befürchtet*, dass eine Maßnahme zu einer Veränderung des bislang Geltenden führt

Bei einem Hypothesentest nimmt man also an, die dem Zufallsexperiment zugrunde liegende Erfolgswahrscheinlichkeit zu kennen oder man hat zumindest eine Vermutung. Das heißt, man hat eine **Hypothese** über die Wahrscheinlichkeit. Ob diese Hypothese haltbar ist, kann mit Hilfe einer oder mehreren Stichproben aus der Grundgesamtheit getestet werden. Das Ergebnis dieser Stichprobe(n) kann daraufhin untersucht werden, ob es verträglich mit der Hypothese ist oder ob es signifikant von dem zu erwartenden Ergebnis abweicht.

Die ursprüngliche Annahme nennt man die **Nullhypothese H_0** ("Status Quo") das Gegenteil dieser Nullhypothese nennt man die **Gegenhypothese H_1** (was man nachweisen will). Der Parametertest ermöglicht also eine Entscheidung darüber, ob die Nullhypothese H_0 beibehalten werden kann, weil die Analyse des Stichprobenmaterials keine Widersprüche zu ihr aufzeigt, oder ob H_0 zugunsten der Alternativhypothese H_1 abgelehnt werden muss.

Für den Hypothesentest muss zunächst das **Signifikanzniveau α** festgelegt werden: Möchte man, dass ein Stichprobenergebnis mit einer definierten Wahrscheinlichkeit (z. B. 95% oder 99%) innerhalb des definierten Intervalls um den Erwartungswert liegt, so nennt man die Wahrscheinlichkeit mit der das Stichprobenergebnis außerhalb dieses Intervalls liegt, das **Signifikanzniveau** dieses Hypothesentests (in diesen Fällen 5% oder 1%). Die Festlegungen auf ein Signifikanzniveau von 5% und 1% sind willkürlich. Für einzelne Tests sind auch andere Werte denkbar. In der Praxis nutzt man aber häufig einen dieser beiden Werte.

Das Intervall um den Erwartungswert, in dem die Ergebnisse liegen, die mit der Hypothese – also mit der angenommenen Erfolgswahrscheinlichkeit – verträglich sind, nennt man den **Annahmebereich** der Hypothese. Den Bereich, der außerhalb dieses Intervalls liegt, nennt man den **Ablehnungsbereich** der Hypothese.

Wichtig:

Liegt ein Stichprobenergebnis im Annahmebereich einer Hypothese, so bedeutet dies nicht, dass damit die Hypothese bestätigt wird. Man kann in solch einem Fall nur sagen, dass es aufgrund des Stichprobenergebnisses keinen Anlass gibt, die Hypothese zu verwerfen. Man sagt, die Hypothese wird beibehalten. Eine hundertprozentige Sicherheit, dass die angenommene Hypothese auch tatsächlich wahr ist, gibt es also nicht, da immer von einer Stichprobe auf die unbekannte Grundgesamtheit geschlossen wird.

4.3.1 Einseitiger Hypothesentest

Beim Testen von Hypothesen wird zwischen einseitigen und zweiseitigen Hypothesentests unterschieden. Bei einem einseitigen Hypothesentest wird in der Alternativhypothese eine Abweichung nur in eine Richtung vermutet.

Sie besagt entweder, dass die Wahrscheinlichkeit größer ist (rechtsseitiger Hypothesentest) oder dass sie kleiner ist (linksseitiger Hypothesentest). Der Annahme- und Ablehnungsbereich einer Nullhypothese wird durch eine kritische Grenze c definiert. Diese Grenze wird durch das Signifikanzniveau α bestimmt.

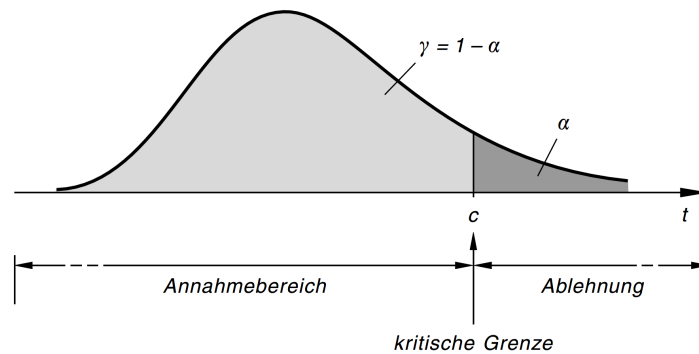


Bild 4.1: Annahme- und Ablehnungsbereich bei einem einseitigen (rechtsseitigen) Hypothesentest.

Beispiel:

Die Inhaberin einer Losbude beschäftigt einen Angestellten, der Besucher eines Volkfestes anspricht, um diese zum Kauf von Losen zu animieren. Sie ist mit der Erfolgsquote des Angestellten unzufrieden.

Dem Angestellten soll das Gehalt gekürzt werden, wenn weniger als 15% der angesprochenen Besucher Lose kaufen. Die Entscheidung über die Gehaltskürzung soll mithilfe eines Hypothesentests auf der Grundlage von 100 angesprochenen Besuchern getroffen werden.

Geben Sie die Nullhypothese H_0 und die Gegenhypothese H_1 an, sowie den Annahme- und Ablehnungsbereich für H_0 .

Lösung:

Gegeben:

$$H_0 : p \geq 0.15$$

$$H_1 : p < 0.15$$

Annahmebereich von $H_0 : a = [c + 1, 100]$

Ablehnungsbereich von $H_0 : \bar{A} = [0, c]$

Es handelt sich um einen einseitigen Hypothesentest, da eine Abweichung nur in eine Richtung angenommen wird (weniger als 15% der angesprochenen Besucher kaufen ein Los).

4.3.2 Zweiseitiger Hypothesentest

Im Unterschied zum einseitigen Hypothesentest untersucht ein zweiseitiger Test die Abweichungen von der Nullhypothese in beide Richtungen.

Der Annahmebereich (nichtkritischer Bereich) wird dabei durch zwei kritische Grenzen c_u (unterer Wert) und c_o (oberer Wert) definiert. Die Ablehnungsbereiche (kritische Bereiche) sind jeweils kleiner c_u oder größer c_o .

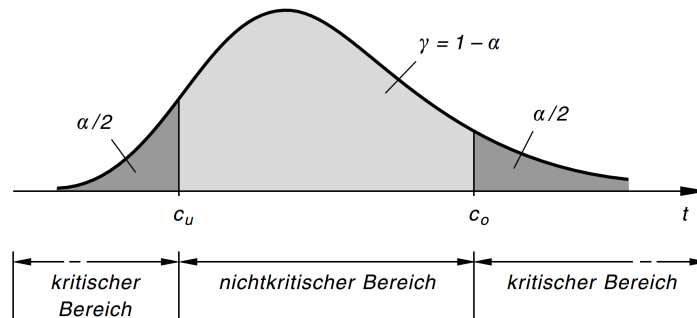


Bild 4.2: Zweiseitiger Hypothesentest

Beispiel:

Ein Hersteller produziert in großen Mengen Zylinderscheiben mit einem Solldurchmesser von $\mu_0 = 20.2$ mm. Um die Einhaltung dieses Sollwertes zu überprüfen, wird eine Stichprobe entnommen und ein Hypothesentest durchgeführt.

Es handelt sich hierbei um einen zweiseitigen Hypothesentest, da die Zylinderscheiben im Durchmesser sowohl größer als auch kleiner als $\mu_0 = 20.2$ mm sein können. Es kann also Abweichungen in beide Richtungen geben.

Geben Sie die Nullhypothese H_0 und die Gegenhypothese H_1 sowie den Annahmebereich für H_0 an.

Lösung:

$$H_0 : \mu = \mu_0 = 20.2 \text{ mm} \quad H_1 : \mu \neq \mu_0 = 20.2 \text{ mm}$$

mit μ = unbekannter Mittelwert der Grundgesamtheit aller produzierten Zylinderschrauben

$$\text{Annahmebereich von } H_0 : A = [c_u, c_o]$$

Der Ablehnungsbereich von H_0 enthält entsprechend Werte unterhalb c_u und oberhalb c_o

In diesem Beispiel sind c_u und c_o Reelle Zahlen.

4.3.3 Fehler 1. und 2. Art

Bei der Durchführung von Hypothesentests gibt es zwei grundlegende Fehlerarten, die mit der Entscheidung der Gültigkeit der Nullhypothese verbunden sind:

- Der **Fehler 1. Art** tritt auf, wenn die Nullhypothese H_0 fälschlicherweise abgelehnt wird, obwohl sie in Wirklichkeit wahr ist.
Dieser Fehler wird durch das Signifikanzniveau α quantifiziert. α wird auch Irrtumswahrscheinlichkeit genannt. Sie gibt also an, mit welcher maximalen Wahrscheinlichkeit eine wahre Nullhypothese fälschlicherweise abgelehnt wird.
- Der **Fehler 2. Art** tritt auf, wenn die Nullhypothese H_0 nicht abgelehnt wird, obwohl die Alternativhypothese H_1 wahr ist. Dieser Fehler hat die Wahrscheinlichkeit β .

Beispiel:

Ein Großhändler bezieht vom Hersteller einen größeren Posten eines elektronischen Bauelements. Bei der Annahme der Ware wird er eine Abnahmekontrolle durchführen, um zu prüfen, ob die vereinbarten Lieferbedingungen (max. 1% Ausschuss) eingehalten wurde. Hierzu wird der Lieferung eine Stichprobe entnommen.

Fällt der berechnete Wert in den nicht-kritischen Bereich, wird die Lieferung angenommen.

Fällt der Wert in den kritischen Bereich, wird die Annahme verweigert.

Ein **Fehler 1. Art** liegt vor, wenn die Lieferung eigentlich in Ordnung ist, aber wegen einer mangelbehafteten Stichprobe (also zufällig eine zu hohe Anzahl schlechte Bauteile) abgelehnt wird.

Die zwischen dem Produzenten und Großhändler vorab vereinbarte Signifikanzzahl α ist dabei die Wahrscheinlichkeit dafür, eine an sich einwandfreie Lieferung aufgrund einer Zufallsstichprobe zurückzuweisen, weil der Wert der Testgröße zufälligerweise in den kritischen Bereich fällt.

Ein **Fehler 2. Art** liegt vor, wenn die Lieferung aufgrund einer guten Stichprobe (also zufällig keine oder sehr wenige schlechte Bauteile) als in Ordnung bezeichnet wird, obwohl die Lieferung als Gesamtheit nicht in Ordnung war.

β ist dabei die Wahrscheinlichkeit dafür, dass der Großhändler die Lieferung annimmt, obwohl sie nicht den vereinbarten Bedingungen (maximaler Ausschuss von 1%) entspricht. Diese Fehlentscheidung tritt genau dann ein, wenn der aus der Stichprobe berechnete Wert der Testvariablen zufällig in den nicht-kritischen Bereich fällt. Daher wird der Fehler 2. Art auch als **Konsumentenrisiko** bezeichnet.

In der Praxis will man Fehler 1. und 2. Art möglichst klein halten. Aus der grafischen Darstellung der Verteilung kann man unmittelbar entnehmen, dass eine Verkleinerung von α automatisch eine Vergrößerung von β nach sich zieht.

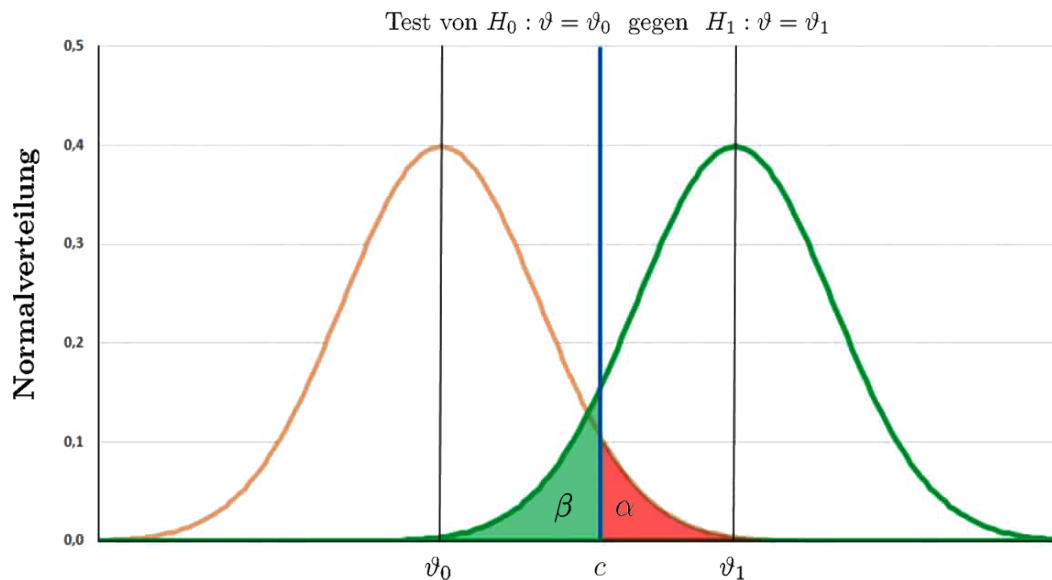


Bild 4.3: Fehler 1. Art (α -Fehler) und Fehler 2. Art (β -Fehler).

Eine Verkleinerung von α bedeutet eine Verschiebung der kritischen Grenze c nach rechts. Dabei nimmt β automatisch zu.

Umgekehrt gilt: Wird β verkleinert, so vergrößert sich dabei α . Dies entspricht einer Verschiebung der kritischen Grenze c nach links.

Entscheidet man sich für ein kleines α , d.h. für ein geringes Risiko, eine richtige Nullhypothese ablehnen zu müssen, so nimmt man ein deutlich erhöhtes Risiko für einen Fehler 2. Art in Kauf. Man muss also im Einzelnen sorgfältig abwägen.

Wichtig:

Bei sicherheitskritischen Anwendungen (Flugzeuge, Atomkraftwerke, etc.) besitzen Fehler 2. Art eine viel höhere Kritikalität als Fehler 1. Art.

Praxistipp:

Wähle zunächst die kleine Signifikanzzahl α für das Risiko des Fehlers 1. Art.

Bestimme anschließend die kritische Grenze c und daraus die Wahrscheinlichkeit β für einen Fehler 2. Art in Abhängigkeit des Parameters ϑ_1 .

Der Fehler 2. Art lässt sich dann nur durch eine Erhöhung des Stichprobenumfangs n verringern.

	H_0 wird angenommen	H_0 wird abgelehnt
H_0 trifft zu	✓	Fehler 1. Art (α -Fehler)
H_0 trifft nicht zu	Fehler 2. Art (β -Fehler)	✓

4.3.4 Durchführung Hypothesentest

Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X ist bekannt, enthält jedoch unbekannte Parameter. In der Praxis treten häufig Normalverteilungen auf, bei denen die Parameter μ und σ^2 unbekannt sind. Verschiedene Arten der Parametertests werden in Kapitel 4.4 beschrieben.

Ein Parametertest für einen unbekannten Parameter ϑ lässt sich folgendermaßen planen und durchführen:

1. Zunächst werden die Nullhypothese H_0 und die Alternativhypothese H_1 formuliert. Für einen zweiseitigen Hypothesentest würde das wie folgt aussehen:

$$\text{Nullhypothese } H_0 : \vartheta = \vartheta_0$$

$$\text{Alternativhypothese } H_1 : \vartheta \neq \vartheta_0$$

2. Im zweiten Schritt wird ein **Signifikanzniveau** α gewählt. Dies ist die Wahrscheinlichkeit dafür, dass die Nullhypothese abgelehnt wird, obwohl sie richtig ist (Fehler 1. Art).

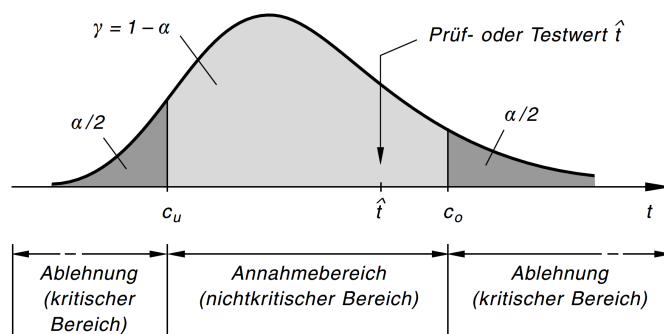
In der Praxis wird üblicherweise $\alpha = 0.05$ oder $\alpha = 0.01$ gewählt.

3. Die **Testvariable** T ist eine Funktion der Stichprobendaten $X_1, X_2, \dots, X_n$, die verwendet wird, um die Hypothese zu überprüfen. Die Wahrscheinlichkeitsverteilung von T ist bekannt. Diese Verteilung wird genutzt, um zu entscheiden, ob die Nullhypothese abgelehnt wird.
4. Mit dem gewählten Signifikanzniveau α werden anschließend zwei kritische Grenzen c_u und c_o festgelegt, sodass die Testvariable T mit der Wahrscheinlichkeit $\gamma = 1 - \alpha$ Werte im Intervall $c_u \leq T \leq c_o$ annimmt. Die Bestimmung der kritischen Grenzen erfolgt dabei durch die Gleichung:

$$P(c_u \leq T \leq c_o \mid H_0) = \gamma = 1 - \alpha$$

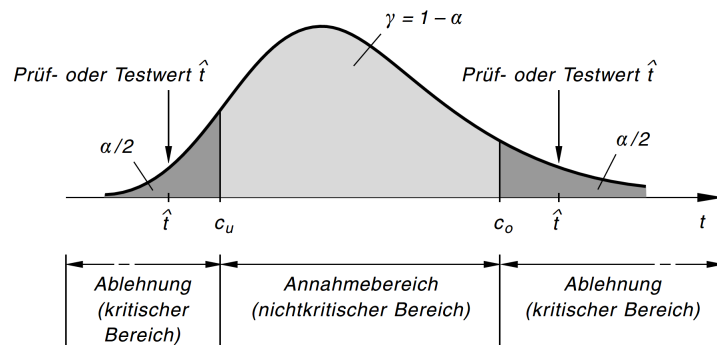
Anhand dieser Gleichung werden die kritischen Grenzen unter Verwendung der Verteilungsfunktion der Testvariable T ermittelt.

5. In diesem Schritt wird der Prüfwert $\hat{t} = g(x_1, x_2, \dots, x_n)$ von T anhand der Stichprobendaten berechnet. Dieser Wert dient dazu, die Hypothese zu überprüfen und ist das Ergebnis der Berechnung aus den gesammelten Daten.
6. Im sechsten Schritt wird eine Entscheidung über die Nullhypothese H_0 getroffen:
 - a) Wenn der Prüfwert $\hat{t}$ im Intervall $c_u \leq \hat{t} \leq c_o$ liegt, wird die Nullhypothese H_0 beibehalten. Dies bedeutet, dass der Schätzwert $\hat{\vartheta}$ mit dem angenommenen Wert ϑ_0 übereinstimmt und die Abweichung auf Zufall zurückzuführen ist. Die Nullhypothese H_0 wird beibehalten.



- b) Wenn der Prüfwert $\hat{t}$ außerhalb des Intervalls $c_u \leq \hat{t} \leq c_o$ liegt, wird die Nullhypothese H_0 abgelehnt. Die Abweichung des Schätzwerts $\hat{\vartheta}$ von ϑ_0 ist

signifikant und deutet darauf hin, dass ϑ einen anderen Wert hat als ϑ_0 . Man kann also davon ausgehen, dass H_0 falsch ist.



4.4 Arten von Parametertests

4.4.1 Zweiseitiger Hypothesentest mit unbekanntem Mittelwert μ einer Normalverteilung bei bekannter Varianz σ^2

Bei dieser Art des Tests ist X eine normalverteilte Zufallsvariable. Der Mittelwert μ ist in diesem Fall unbekannt, die Varianz σ^2 ist bekannt. Es soll geprüft werden, ob der unbekannte Mittelwert μ , wie vermutet, den speziellen Wert μ_0 besitzt. Daher werden folgende Hypothesen aufgestellt:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Die Testvariable U ist eine Zufallsvariable, da der tatsächliche Mittelwert unbekannt ist. Es ergibt sich folgende Formel:

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- $\bar{X}$: Schätzfunktion für den unbekannten Mittelwert μ
- μ_0 : Vermuteter Mittelwert
- σ : Standardabweichung der normalverteilten Grundgesamtheit
- n : Umfang der Stichprobe

Beispiel:

Ein Hersteller produziert in großen Mengen Nägel mit einer Solllänge von $\mu_0 = 22$ mm. Um die Einhaltung dieses Sollwertes zu überprüfen, wird eine Stichprobe von $n = 10$ entnommen. Die Auswertung ergibt:

Die Hypothesen lauten:

$$H_0 : \mu = 22 \text{ mm,}$$

$$H_1 : \mu \neq 22 \text{ mm.}$$

Das Signifikanzniveau beträgt $\alpha = 0.01$ und der Test wird zweiseitig durchgeführt.

Bestimmung der kritischen Grenzen

Die kritischen Grenzen werden durch folgende Gleichung bestimmt:

$$P(-c \leq U \leq c)_{H_0} = 1 - \alpha$$

Da eine Normalverteilung angenommen wird und der Test zweiseitig ist, folgt:

$$\begin{aligned} P(-c \leq U \leq c) &= \Phi(c) - \Phi(-c) \\ &= \Phi(c) - (1 - \Phi(c)) \\ &= 2\Phi(c) - 1 \\ 2\Phi(c) - 1 &= 1 - \alpha = 1 - 0.01 = 0.99 \end{aligned}$$

$$2\Phi(c) - 1 = 0.99 \quad \Rightarrow \quad \Phi(c) = 0.995$$

Aus der Tabelle der Standardnormalverteilung ergibt sich:

$$c = 2.576$$

Der Annahmebereich wird durch das symmetrische Intervall wie folgt beschrieben:

$$-2.576 \leq u \leq 2.576$$

Berechnung des Testwerts

Der Testwert $\hat{u}$ berechnet sich mit $\bar{x} = 20$ mm und der Standardabweichung $\sigma = 3$ mm wie folgt:

$$\begin{aligned} \hat{u} &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ \hat{u} &= \frac{(20 - 22) \text{ mm}}{3 \text{ mm}/\sqrt{10}} \approx -2.108 \end{aligned}$$

Da $\hat{u} = -2.108$, liegt der Testwert im Annahmebereich. Somit wird die Nullhypothese $H_0 : \mu = \mu_0 = 22$ mm, aufgrund der verwendeten Stichprobe angenommen.

Daher kann man davon ausgehen, dass die normalverteilte Grundgesamtheit den Mittelwert $\mu_0 = 22$ mm besitzt.

4.4.2 Test für den unbekannten Mittelwert μ einer Normalverteilung bei unbekannter Varianz σ^2

Bei dieser Art des Tests ist X eine normalverteilte Zufallsvariable. Sowohl Mittelwert μ als auch Varianz σ^2 sind unbekannt. Es soll geprüft werden, ob der ebenfalls unbekannte Mittelwert μ , wie vermutet, den speziellen Wert μ_0 besitzt. Daher werden folgende Hypothesen aufgestellt:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Die Testvariable T ist eine Zufallsvariable, da sowohl der tatsächliche Mittelwert als auch die tatsächliche Varianz unbekannt sind. Es ergibt sich folgende Formel:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

- $\bar{X}$: Schätzfunktion für den unbekannten Mittelwert μ
- μ_0 : Vermuteter Mittelwert
- S : Schätzfunktion für die unbekannte Standardabweichung σ
- n : Umfang der Stichprobe

Beispiel:

Ein Hersteller produziert in großen Mengen Zylinderscheiben mit einem Solldurchmesser von $\mu_0 = 20.2$ mm. Um die Einhaltung dieses Sollwertes zu überprüfen, wird eine Stichprobe von $n = 16$ entnommen.

Die Auswertung der Stichprobe ergibt einen mittlere Durchmesser $\bar{x} = 20.6$ mm und eine Standardabweichung von $s = 0.5$.

Die Hypothesen lauten:

$$H_0 : \mu = 20.2 \text{ mm} \quad (\mu = \mu_0),$$

$$H_1 : \mu \neq 20.2 \text{ mm} \quad (\mu \neq \mu_0).$$

Das Signifikanzniveau soll $\alpha = 0.05$ betragen und der Test wird zweiseitig durchgeführt.

Bestimmung der kritischen Grenzen

Die kritischen Grenzen werden durch folgende Gleichung bestimmt:

$$P(-c \leq T \leq c)_{H_0} = 1 - \alpha$$

Da der Test zweiseitig ist, folgt:

$$2F(c) - 1 = 0.95 \quad \Rightarrow \quad F(c) = 0.975$$

Da σ unbekannt ist, genügt die Testvariable T der t-Verteilung mit:

$$f = n - 1 = 16 - 1 = 15 \text{ Freiheitsgraden}$$

Aus der Tabelle der t-Verteilung ergibt sich:

$$c = t_{\frac{\alpha}{2}; 15} = 2.131$$

Der Annahmebereich wird durch das symmetrische Intervall wie folgt beschrieben:

$$-2.131 \leq t \leq 2.131$$

Berechnung des Testwerts

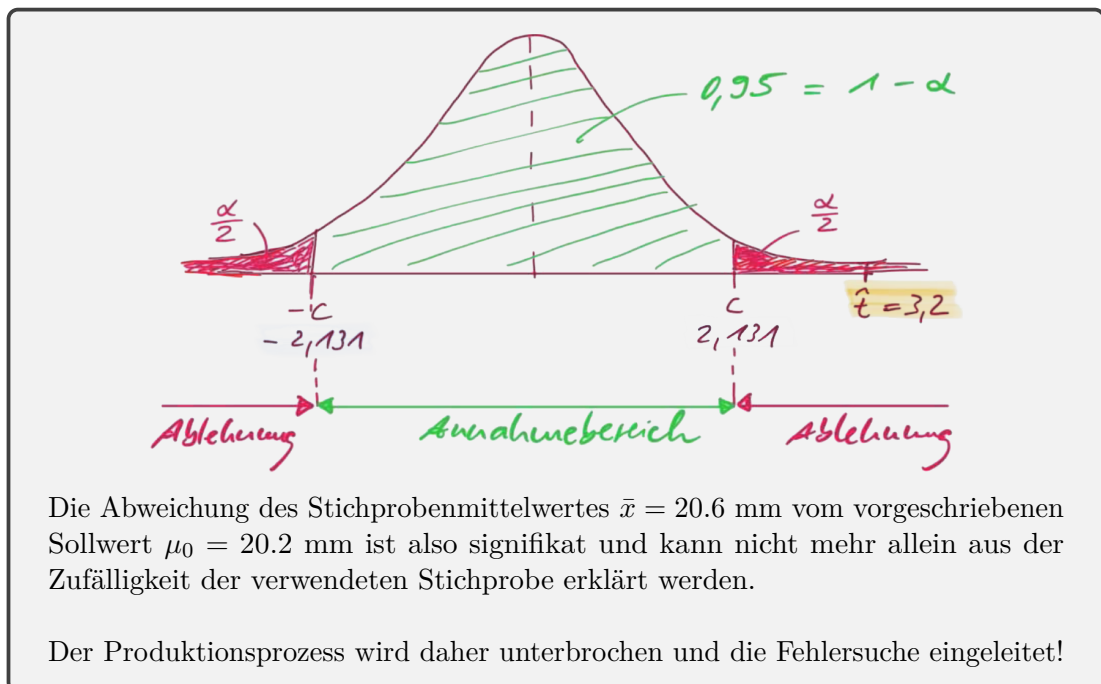
Der Testwert $\hat{t}$ berechnet sich mit $\bar{x} = 20.6$ mm und der empirischen Standardabweichung $s = 0.5$ mm wie folgt:

$$\hat{t} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$\hat{t} = \frac{(20.6 - 20.2) \text{ mm}}{0.5 \text{ mm}/\sqrt{16}} = 3.2$$

Da $|\hat{t}| > 2.131$, liegt der Testwert im kritischen Bereich. Somit wird die Nullhypothese H_0 abgelehnt.

Skizze der Verteilung:



Hinweis:

Bei einer umfangreichen Stichprobe (Faustregel $n > 30$) ist die definierte Testvariable T näherungsweise standardnormalverteilt und man darf das Testverfahren wie bei bekannter Varianz verwenden ($\sigma^2 \approx s^2$).

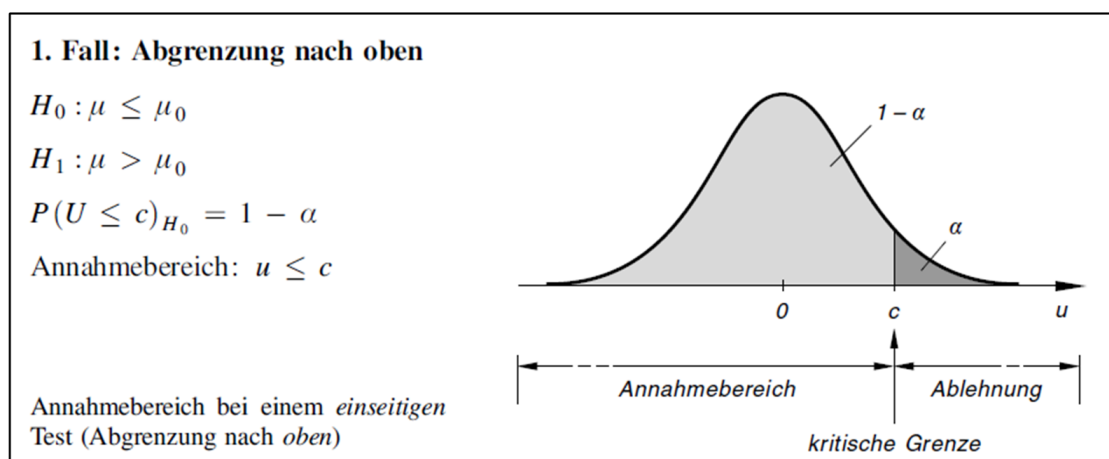
Bemerkungen:

1. Es gibt viele Arten von Parametertests bzw. Hypothesentests. In diesem Skript wenden wir die Hypothesentests auf die in der Praxis besonders wichtige Normalverteilung (Tabelle Standardnormalverteilung und t -Verteilung) an. Die Vorgehensweise ist weitgehend die gleiche wie bei der Herleitung von Vertrauens- oder Konfidenzintervallen.
2. Die Idee hinter Konfidenzintervallen und Hypothesentests ist ähnlich. Tatsächlich ist es so, dass jedes Konfidenzintervall gleichzeitig auch ein Hypothesentest ist. Das Konfidenzintervall enthält all diejenigen Werte (z. B. für μ_0), für die ein Hypothesentest nicht signifikant wäre. D. h. H_0 würde beibehalten. In anderen Worten: Ein Konfidenzintervall ist die Menge aller Punkte, für die die entsprechende Nullhypothese H_0 nicht abgelehnt werden kann. Bei einem Hypothesentest hingegen, wird anhand eines definierten Punktes (errechneter Testwert) geprüft, ob sie stimmt.

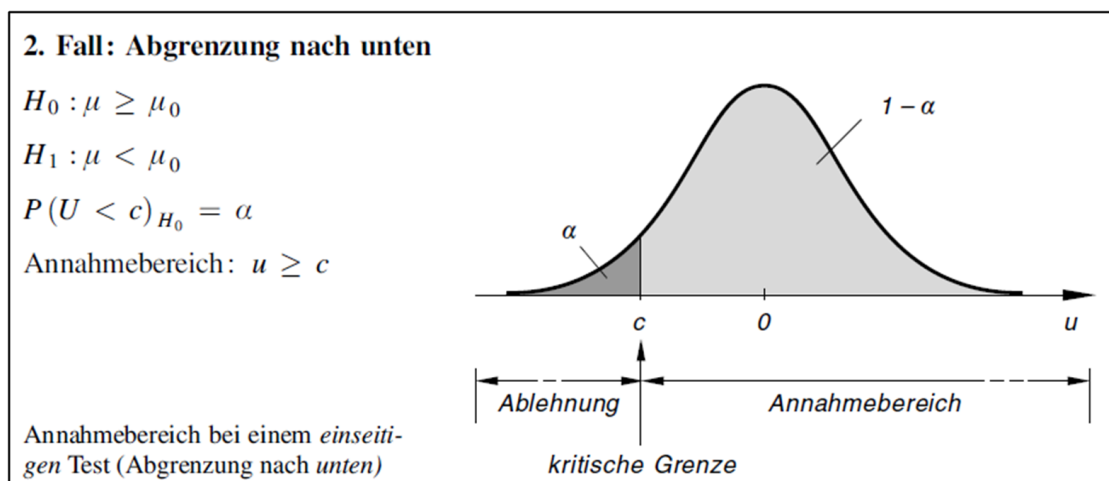
4.4.3 Einseitiger Hypothesentest mit unbekanntem Mittelwert μ einer Normalverteilung bei bekannter bzw. unbekannter Varianz σ^2

Analog zu den in den vorigen Kapiteln beschriebenen zweiseitigen Hypothesentests verlaufen die einseitigen Hypothesentests. In diesen Fällen gibt es jedoch nur eine kritische Grenze c .

Lautet $H_1 : \mu > \mu_0$ prüfen wir, ob μ_0 überschritten wird (einseitiger Test mit Abgrenzung nach oben, rechtsseitiger Hypothesentest)



Lautet $H_1 : \mu < \mu_0$ prüfen wir, ob μ_0 unterschritten wird (einseitiger Test mit Abgrenzung nach unten, linksseitiger Hypothesentest)



Hinweise:

1. In beiden Fällen kann H_0 auch lauten: $\mu = \mu_0$
2. Für t -verteilte Variablen werden häufig U und u durch T und t ersetzt, ansonsten analoge Vorgehensweise
3. In der Alternativhypothese H_1 steht niemals ein $=$ oder $\leq$ oder $\geq$

4.4.4 Tests für einen unbekannten Anteilswert p

Bei einer statistischen Qualitätskontrolle hat man es häufig mit einer binomialverteilten Grundgesamtheit zu tun, deren Parameter p jedoch unbekannt ist. Ein Musterbeispiel dafür ist der Ausschussanteil p bei der Serienfabrikation von speziellen elektronischen Bauelementen, d. h. also der Anteil p an fehlerhaften Teilen in der Gesamtproduktion.

Wird in einem konkreten Fall vermutet, dass dieser Parameter einen bestimmten Wert p_0 besitzt ($p = p_0$), so lässt sich diese Hypothese durch einen geeigneten Parametertest überprüfen. Man testet dann mittels einer Stichprobenuntersuchung

die Nullhypothese $H_0 : p = p_0$
 gegen die Alternativhypothese $H_1 : p \neq p_0$

Beispiel:

Ein Hersteller eines bestimmten elektronischen Bauelements behauptet, dass seine Produktion höchstens $p_0 = 4\%$ Ausschuss enthalte. Bei der Anlieferung eines größeren Postens dieser Elemente wurde von Seiten des Abnehmers eine Gütekontrolle durchgeführt. Sie bestand in einer Stichprobenuntersuchung von $n = 300$ Bauelementen. Unter ihnen befanden sich dabei $k = 15$ defekte Bauelemente. Geben Sie die Nullhypothese H_0 und die Gegenhypothese H_1 an.

Lösung:

$$H_0 : p \leq p_0 = 0.04$$

$$H_1 : p > p_0 = 0.04$$

mit p = unbekannter Ausschusswert der Grundgesamtheit aller elektronischen Bauelemente

Die weitere Vorgehensweise bei diesem Test unterscheidet sich unwesentlich von den bisher beschriebenen Hypothesentests. Man benötigt eine Schätzfunktion $\hat{P}$ (Zufallsvariable) für den unbekannten Parameter p . Diese Schätzfunktion gewinnt man nach der Maximum-Likelihood-Methode aus der Stichprobe, die man für den Test zieht. Gemäß obigem Beispiel ist der Schätzwert (vgl. Kapitel 4.2.1 Parameterschätzungen)

$$\hat{p} = \frac{k}{n} = \frac{15}{300} = 0.05$$

Weiterhin benötigt man eine Testvariable, auf deren Herleitung in diesem Skript verzichtet wird. Sie lautet:

$$U = \sqrt{\frac{n}{p_0(1-p_0)}} \cdot (\hat{P} - p_0)$$

Wie schon erwähnt, werden die beschriebenen Hypothesentests auf die in der Praxis besonders wichtige Normalverteilung angewendet. Es handelt sich bei solchen Qualitätskontrollen zwar um eine binomialverteilte Grundgesamtheit (Bauelement funktioniert oder funktioniert nicht), jedoch verhält sich die Schätzfunktion $\hat{P}$ annähernd normalverteilt, wenn der Stichprobenumfang n folgender Bedingung genügt:

$$n \cdot p_0(1 - p_0) > 9$$

Zweiseitiger Test für einen unbekannten Anteilswert p (Parameter (p) einer Binomialverteilung) unter Verwendung einer umfangreichen Stichprobe

Es soll geprüft werden, ob ein unbekannter Anteilswert p (Parameter p einer Binomialverteilung) einen bestimmten Wert p_0 besitzt ($p = p_0$). Wir entnehmen zu diesem Zweck der binomialverteilten Grundgesamtheit eine umfangreiche Stichprobe, d. h. eine Stichprobe, deren Umfang n der folgenden Bedingung genügt:

$$np_0(1 - p_0) > 9$$

Die Stichprobe selbst besteht dann darin, dass das Bernoulli-Experiment n -mal nacheinander ausgeführt wird und dabei die Anzahl k der „Erfolge“ festgestellt wird. Als „Erfolg“ wird das Eintreten des Ereignisses A gewertet, „Misserfolg“ bedeutet demnach, dass das komplementäre Ereignis $\bar{A}$ eintritt. Die beobachtete relative Häufigkeit für das Ereignis A („Erfolg“) beträgt somit $h(A) = \frac{k}{n}$. Unter Verwendung dieser Stichprobe wird die

$$\text{Nullhypothese: } H_0 : p = p_0$$

gegen die

$$\text{Alternativhypothese: } H_1 : p \neq p_0$$

getestet.

Die Durchführung dieses zweiseitigen Parameter-tests erfolgt schrittweise wie folgt:

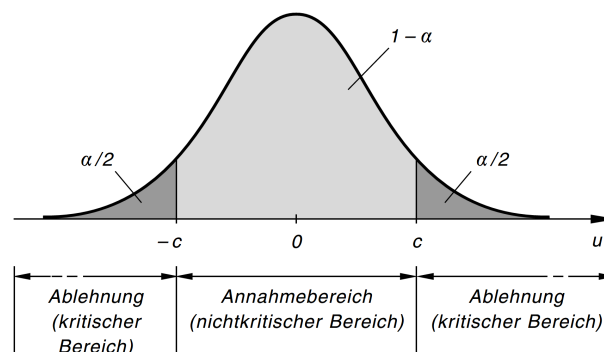
1. Wahl einer bestimmten Signifikanzzahl (Irrtumswahrscheinlichkeit) α (in der Praxis meist $\alpha = 5\%$ oder $\alpha = 1\%$).
2. Test- oder Prüfvariable ist die näherungsweise standardnormalverteilte Zufallsvariable

$$U = \sqrt{\frac{n}{p_0(1 - p_0)}} \cdot (\hat{P} - p_0)$$

Die Berechnung des kritischen Wertes c und damit der kritischen Grenzen $\mp c$ erfolgt dabei aus der Bedingung

$$P(-c \leq U \leq c)_{H_0} = 1 - \alpha$$

Der nichtkritische Bereich (Annahmebereich) lautet dann: $-c \leq u \leq c$



3. Berechnung des Schätzwertes $\hat{p} = h(A) = \frac{k}{n}$ für den Parameter p aus der vorgegebenen konkreten Stichprobe (n -fache Ausführung des Bernoulli-Experimentes, dabei k -mal „Erfolg“) sowie des Test- oder Prüfwertes der Testvariablen U .

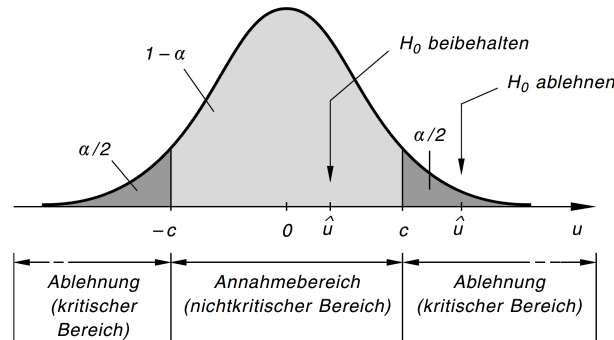
$$\hat{u} = \sqrt{\frac{n}{p_0(1 - p_0)}} \cdot (\hat{p} - p_0)$$

4. **Testentscheidung:** Fällt der Test- und Prüfwert $\hat{u}$ in den nicht-kritischen

Bereich (Annahmebereich), d.h. gilt

$$-c \leq \hat{u} \leq c$$

so wird die Nullhypothese $H_0 : p = p_0$ beibehalten, anderenfalls zugunsten der Alternativhypothese $H_1 : p \neq p_0$ verworfen.



Beispiel:

Der Hersteller eines bestimmten elektronischen Bauelements behauptet, dass seine Produktion höchstens 4% Ausschuss enthalte. Bei der Anlieferung eines größeren Postens dieser Elemente wurde von Seiten des Abnehmers eine Gütekontrolle durchgeführt. Sie bestand in einer Stichprobenuntersuchung von $n = 300$ Bauelementen. Unter ihnen befanden sich dabei $k = 15$ funktionsuntüchtige (d.h. defekte) Teile. Man überprüfe die Angaben des Herstellers auf einem Signifikanzniveau von $\alpha = 0.01$.

Da in diesem Fall nur die Abweichungen nach oben interessieren, wird ein einseitiges Prüfverfahren angewendet.

Testen der Nullhypothese

$$H_0 : p \leq p_0 = 0.04$$

gegen die Alternativhypothese

$$H_1 : p > p_0 = 0.04$$

Die Voraussetzung einer umfangreichen Stichprobe ist gegeben:

$$np_0(1 - p_0) > 9$$

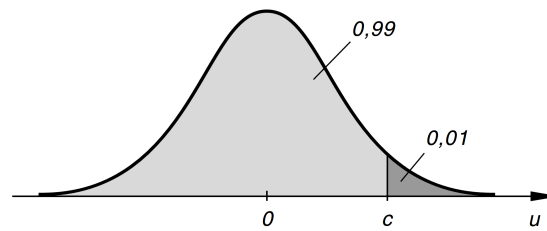
$$np_0(1 - p_0) = 300 \cdot 0.04(1 - 0.04) = 11.52 > 9$$

Schritt 1: Die Signifikanzzahl ist vorgegeben mit $\alpha = 0,01$.

Schritt 2: Berechnung der kritischen Grenze c der standardnormalverteilten Testvariablen.

$$U = \sqrt{\frac{n}{p_0(1 - p_0)}} \cdot (\hat{P} - p_0) = \sqrt{\frac{300}{0.04(1 - 0.04)}} \cdot (\hat{P} - 0.04) = 88.388(\hat{P} - 0.04)$$

$$P(U \leq c)_{H_0} = 1 - \alpha = 1 - 0.01 = 0.99$$



Unter Verwendung des Tabellenwerks ergibt sich folgender Wert für die unbekannte Schranke c :

$$P(U \leq c)_{H_0} = \phi(c) = 0.99 \rightarrow c = u_{0.99} = 2.326$$

Der nicht kritische Bereich lautet damit:

$$u \leq 2.326$$

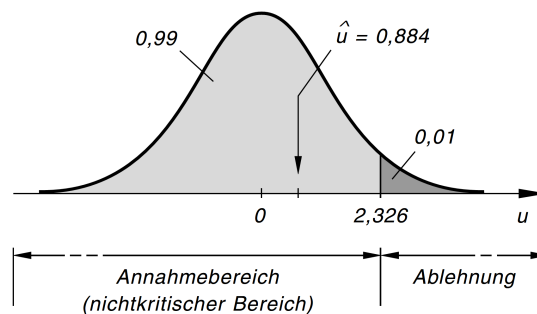
Schritt 3: Der Schätzwert für den unbekannten Parameter p beträgt

$$\hat{p} = \frac{k}{n} = \frac{15}{300} = 0.05$$

Somit besitzt die Testvariable U den Test- oder Prüfwert

$$\hat{u} = 88.388(\hat{p} - 0.04) = 88.388(0.05 - 0.04) = 0.884$$

Schritt 4: Der Test- oder Prüfwert $\hat{u} = 0.884$ fällt in den Annahmebereich $u \leq 2.326$. Die Nullhypothese $H_0 : p \leq p_0 = 0.04$ wird daher beibehalten, d.h. es gibt aufgrund der verwendeten Stichprobe keinen Anlass, an den Angaben des Herstellers bezüglich des maximalen Ausschussanteils von 4% zu zweifeln. Die Abweichung des aus der Stichprobe ermittelten Anteils $\hat{p} = 0.05$ von dem vom Hersteller angegebenen Anteilswert $p_0 = 0.04$ ist bei dem gewählten Signifikanzniveau von $\alpha = 0.01$ nicht signifikant und somit zufallsbedingt.



4.5 Statistische Signifikanz und fachliche Relevanz

Der Begriff der statistischen Signifikanz wird oft missbraucht, um gleichzeitig auch die entsprechende fachliche Relevanz zu untermauern. Diese beiden Begriffe müssen aber nicht unbedingt miteinander einhergehen.

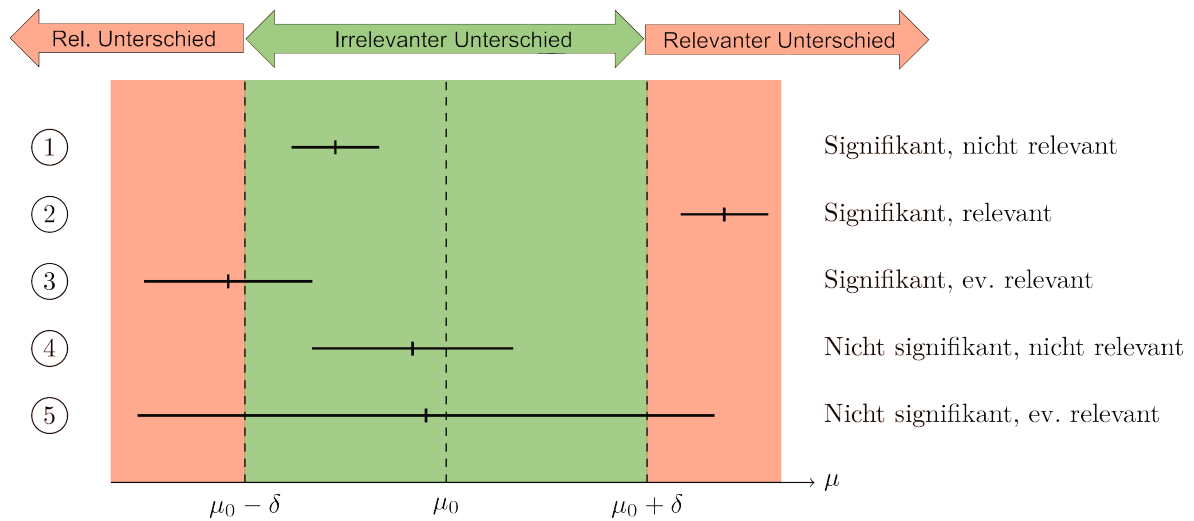
Wenn man genügend viele Beobachtungen sammelt, dann wird man jede Nullhypothese verwerfen können (denn diese stimmt in der Praxis nie exakt).

Hierzu müssen wir das beste aus 'beiden Welten' miteinander kombinieren: Entsprechendes Fachwissen und der statistische Output. Wir müssen zuerst basierend auf Fachwissen definieren, was ein relevanter Effekt oder Unterschied ist (die Statistik kann uns hier nicht helfen).

Wenn wir dies gemacht haben, können wir die Statistik ins Spiel bringen.

Wichtig ist also die Reihenfolge. ERST das Ziel formulieren/festlegen und DANACH die statistische Analyse durchführen, OHNE das Ziel nochmal zu verändern. Oft beobachtet man, dass beides iterativ behandelt wird, um eine bestimmte Argumentation zu befördern.

⇒ Das ist nichts anderes als Manipulation!



Verschiedene Fälle (1 bis 5) von statistischer Signifikanz und fachlicher Relevanz.

Die Vertrauensintervalle für μ sind durch Striche dargestellt. Der 'irrelevante Bereich' geht von $\mu_0 - \delta$ bis zu $\mu_0 + \delta$ (grün), wobei das δ durch entsprechendes Fachwissen definiert wurde.

4.6 Statistische Qualitätskontrolle unter Verwendung von Kontrollkarten

Für Interessierte bietet Papula 2016 (S. 599–604) eine vertiefte Behandlung der statistischen Qualitätskontrolle unter Verwendung von Kontrollkarten.

KAPITEL 5

Anhang

5.1 Wahrscheinlichkeitsbegriff

$P(E)$	Eintreffen des Ereignisses
1	gewiss
>0.5	wahrscheinlich
$=0.5$	zweifelhaft
<0.5	unwahrscheinlich
0	unmöglich

5.2 Binomialverteilung

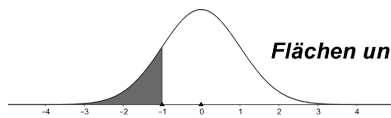
Pascalsches Dreieck

```

      1
     1 1
    1 2 1
   1 3 3 1
  1 4 6 4 1
 1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1
1 11 55 165 330 462 462 330 165 55 11 1
1 12 66 220 495 792 924 792 495 220 66 12 1
1 13 78 286 715 1287 1716 1716 1287 715 286 78 13 1
1 14 91 364 1001 2002 3003 3432 3003 2002 1001 364 91 14 1
1 15 105 455 1365 3003 5005 6435 6435 5005 3003 1365 455 105 15 1
1 16 120 560 1820 4368 8008 11440 12870 11440 8008 4368 1820 560 120 16 1

```

5.3 Gaußsche Normalverteilung



Flächen unter der Normalkurve

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2}} dx \quad \text{für } 0 \leq u \leq 3,9$$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	.500000	.503989	.507978	.511966	.515953	.519938	.523922	.527903	.531881	.535856
0,1	.539828	.543795	.547758	.551717	.555670	.559618	.563560	.567495	.571424	.575345
0,2	.579260	.583166	.587064	.590954	.594835	.598706	.602568	.606420	.610261	.614092
0,3	.617911	.621720	.625516	.629300	.633072	.636831	.640576	.644309	.648027	.651732
0,4	.655422	.659097	.662757	.666402	.670031	.673645	.677242	.680822	.684386	.687933
0,5	.691462	.694974	.698468	.701944	.705402	.708840	.712260	.715661	.719043	.722405
0,6	.725747	.729069	.732371	.735653	.738914	.742154	.745373	.748571	.751748	.754903
0,7	.758036	.761148	.764238	.767305	.770350	.773373	.776373	.779350	.782305	.785236
0,8	.788145	.791030	.793892	.796731	.799546	.802338	.805106	.807850	.810570	.813267
0,9	.815940	.818589	.821214	.823814	.826391	.828944	.831472	.833977	.836457	.838913
1,0	.841345	.843752	.846136	.848495	.850830	.853141	.855428	.857690	.859929	.862143
1,1	.864334	.866500	.868643	.870762	.872857	.874928	.876976	.879000	.881000	.882977
1,2	.884930	.886861	.888768	.890651	.892512	.894350	.896165	.897958	.899727	.901475
1,3	.903200	.904902	.906582	.908241	.909877	.911492	.913085	.914656	.916207	.917736
1,4	.919243	.920730	.922196	.923642	.925066	.926471	.927855	.929219	.930563	.931889
1,5	.933193	.934478	.935744	.936992	.938220	.939429	.940620	.941792	.942947	.944083
1,6	.945201	.946301	.947384	.948449	.949497	.950528	.951543	.952540	.953521	.954486
1,7	.955434	.956367	.957284	.958185	.959070	.959941	.960796	.961636	.962462	.963273
1,8	.964070	.964852	.965620	.966375	.967116	.967843	.968557	.969258	.969946	.970621
1,9	.971283	.971933	.972571	.973197	.973810	.974412	.975002	.975581	.976148	.976704
2,0	.977250	.977784	.978308	.978822	.979325	.979818	.980301	.980774	.981237	.981691
2,1	.982136	.982571	.982997	.983414	.983823	.984222	.984614	.984997	.985371	.985738
2,2	.986097	.986447	.986791	.987126	.987454	.987776	.988089	.988396	.988696	.988989
2,3	.989276	.989556	.989830	.990097	.990358	.990613	.990862	.991106	.991344	.991576
2,4	.991802	.992024	.992240	.992451	.992656	.992857	.993053	.993244	.993431	.993613
2,5	.993790	.993963	.994132	.994300	.994457	.994614	.994766	.994915	.995060	.995201
2,6	.995339	.995473	.995604	.995731	.995855	.995975	.996093	.996207	.996319	.996427
2,7	.996533	.996636	.996736	.996833	.996928	.997020	.997110	.997197	.997282	.997365
2,8	.997445	.997523	.997599	.997673	.997744	.997814	.997882	.997948	.998012	.998074
2,9	.998134	.998193	.998250	.998305	.998359	.998411	.998462	.998511	.998559	.998605
	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3,0	.998650	.999032	.999313	.999517	.999663	.999767	.999841	.999892	.999928	.999952

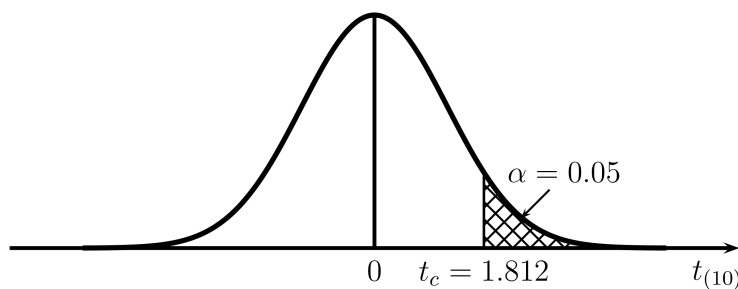
5.4 t-Test

$t_{(q)}$ -Verteilung (Student Verteilung)

q	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.327	2.576

Interpretation: α bezeichnet das Signifikanzniveau und q die Freiheitsgrade einer $t_{(q)}$ -verteilten Zufallsvariablen.

Beispiel: Bei einem *einseitigen* Test mit einem Signifikanzniveau von 5% ($\alpha = 0.05$) und 10 Freiheitsgraden ($q = 10$) ist der kritische Wert $t_c = 1.812$. Das heißt, $\Pr(t_{(10)} > 1.812) = 0.05$. Für einen *zweiseitigen* Test mit $\alpha = 0.05$ und $q = 10$ ist der kritische Wert $t_c = 2.228$.



5.5 Statistiksoftware R

5.5.1 Daten einlesen mit R

Um Datensätze mit R zu analysieren, müssen diese zuerst eingelesen werden. Je nach Struktur des Datenfiles kommen verschiedene Einlese-Befehle zur Anwendung. Es können u.a. Textfiles (.txt, .dat), Datenfiles (.csv) und Excel-Files eingelesen werden.

Excel-Dateien können im Prinzip direkt eingelesen werden. Wir empfehlen jedoch, das Excelfile abzuspeichern als .csv-File (Menu File - Save as - File Type: CSV) und dieses mit read.csv() einzulesen. Warum?

- Grafiken, Rahmen, Formatierungen verunmöglichen das Einlesen.
- Sie müssen entscheiden, welche Daten aus dem ganzen Excelfile tatsächlich ins Datenfile gehören. Sämtliche Daten müssen in einem File stehen, nicht auf mehrere Tabs verteilt.
- Das .csv-File enthält die nackten Daten, die mit jedem Editor (Notepad) angezeigt werden können. So sehen sie genau, was nachher ins R eingelesen wird und können allfällige (Excel-) Fehler noch erkennen.
- Ein .csv-File kann problemlos verschickt werden und bereitet auf keinem Betriebssystem Schwierigkeiten.

R-Befehl	geeignet für	wichtige Argumente
read.table()	.txt .dat	header=TRUE/FALSE, sep= , dec= , fill=TRUE
read.csv()	.csv	header=TRUE/FALSE, sep= , dec=
scan()	.txt .dat ...	file= , what= , header=T/F, sep= , skip= , fill=T/F

Beispiel

```
-----
| Name Alter Grösse Geschlecht
| Anna 19 167 w
| Paul 22 185 m
| Udo 20 179 m

> d.bsp <- read.table("Pfad/bsp_space.txt", header=TRUE)

-----
| Name   Alter   Grösse   Geschlecht
| Anna   19      167      w
| Paul   22      185      m
| Udo    20      179      m

> d.bsp <- read.table("Pfad/bsp_tab.txt", header=TRUE, sep="\t")

## Datenfile in R anzeigen
> d.bsp
  Name Alter Grösse Geschlecht
1 Anna   19   167          w
2 Paul   22   185          m
3 Udo    20   179          m
```

5.5.2 Grundlegende Syntax und Befehle

Befehl	Beschreibung
<code><-</code>	Zuweisung
<code>?function</code>	Hilfe zur Funktion <code>function</code>
<code>#</code>	Beginn Kommentar
<code>+, -, *, /, ^</code>	Addition, Subtraktion, Multiplikation, Division, Potenz
<code>>, <, ==, != >=, <=</code>	größer, kleiner, gleich, ungleich
<code>%*%</code>	Matrix- und Vektormultiplikation
<code>c(1, 2, 3)</code>	verbindet die Werte 1, 2, 3 zu einem Vektor
<code>1:n</code>	erzeugt den Vektor 1, 2, ...,n
<code>seq(from, to, by=)</code>	erzeugt einen Vektor mit Elementen von <code>from</code> bis <code>to</code> in fester Schrittweite <code>by</code>
<code>rep(x, n)</code>	wiederholt den Vektor <code>x</code> <code>n</code> -mal
<code>matrix(data, ncol=n)</code>	erzeugt eine Matrix mit <code>n</code> Spalten aus den Werten in <code>data</code>
<code>max(v), min(v)</code>	Maximum, Minimum aus einem Vektor <code>v</code>
<code>mean(v), median(v)</code>	arith. Mittel, Median aus einem Vektor <code>v</code>
<code>sum(v), prod(v)</code>	Summe, Produkt der Elemente eines Vektors <code>v</code>
<code>sd(v), var(v)</code>	Standardabweichung, Varianz aus einem Vektor <code>v</code>
<code>read.table("path/file")</code>	Laden einer Datei
<code>solve(M)</code>	Inverse der Matrix <code>M</code>
<code>length(v)</code>	Anzahl der Elemente im Vektor <code>v</code>
<code>v[i]</code>	Zugriff auf das <code>i</code> -te Element im Vektor <code>v</code>
<code>dim(M)</code>	Dimension der Matrix <code>M</code> (Reihen und Spalten)
<code>M[i, j]</code>	Zugriff auf das Element der <code>i</code> -ten Reihe und <code>j</code> -ten Spalte der Matrix <code>M</code>
<code>t()</code>	transponieren
<code>sqrt()</code>	Quadratwurzel
<code>abs()</code>	Absolutbetrag

5.5.3 Datenaufbereitung und grafische Darstellung

Befehl	Beschreibung
<code>plot(x, y)</code>	Scatterplot x gegen y
<code>hist()</code>	Histogramm
<code>boxplot()</code>	Boxplot
<code>barplot</code>	Balkendiagramm
<code>pie()</code>	Tortendiagramm
<code>curve(..., add=TRUE)</code>	Zeichnen einer Kurve ins Diagramm
<code>abline()</code>	Zeichnen einer Geraden ins Diagramm
<code>points()</code>	Zeichnen von Punkten ins Diagramm
<code>text()</code>	Einfügen eines Textes ins Diagramm
<code>par(mfrow=c(n, m))</code>	n x m Plots in einem Fenster
<code>?par</code>	Hilfe zu graphischen Parametern
<code>main="..."</code>	Überschrift
<code>xlab=, ylab=</code>	Achsenbeschriftung
<code>col=</code>	Farbwahl
<code>cex=, pch=</code>	Größe und Symbol eines Punktes
<code>lty=, lwd=</code>	Linientyp, Liniendicke
<code>type=</code>	Typ (1=Linie, p=Punkte, b=beides)
<code>xlim=c(a,b), ylim=c(d,e)</code>	Achsenbereich von a bis b bzw. d bis e
<code>legend()</code>	Hinzufügung einer Legende
<code>pdf(), bmp(), jpeg()</code>	Start Speichern einer Graphik
<code>width=, height=</code>	Größe der zu speichernden Graphik
<code>dev.off()</code>	Device schließen → Speichern beenden
<code>read.table()</code>	Einlesen von Daten
<code>setwd()</code>	Arbeitsverzeichnis festlegen
<code>complete.cases()</code>	Überprüfung auf vollständige Zeilen
<code>summary()</code>	Zusammenfassung
<code>table()</code>	Häufigkeits- und Kontingenztabellen
<code>attach()</code>	direkter Zugriff auf Elemente möglich, sonst mit <code>\$</code> oder <code>[, ...]</code>
<code>round(x, n)</code>	Runden von x auf n Nachkommastellen
<code>cor()</code>	Korrelation und Korrelationsmatrix

5.5.4 Befehle zum Erzeugen von Zufallszahlen

Befehl	Beschreibung
<code>rnorm()</code>	Zufallszahl aus der Normalverteilung
<code>runif()</code>	Zufallszahl aus der Gleichverteilung
<code>sample()</code>	Zufälliges Ziehen aus einer Menge
<code>dnorm()</code>	Dichtefunktion der Normalverteilung
<code>pnorm()</code>	Verteilungsfunktion der Normalverteilung
<code>qnorm()</code>	Quantilsfunktion der Normalverteilung
<code>rmvnorm()</code>	Zufallszahl aus der multivariaten Normalverteilung (Funktion aus dem Paket <code>mvtnorm</code>)
<code>set.seed()</code>	Festlegen eines Seeds für Zufallszahlen
<code>library()</code>	Laden von Paketen
<code>library(help="mvtnorm")</code>	Hilfe zu Paket <code>mvtnorm</code>

5.5.5 Wahrscheinlichkeitsverteilungen

Nicht nur für die Binomialverteilung sondern für viele bekannte Wahrscheinlichkeitsverteilungen gibt es R-Funktionen, und zwar jeweils vier:

für die Wahrscheinlichkeits(dichte)funktion (die Zähldichte bei diskreten Zufallsvariablen (**d**) bzw. die Wahrscheinlichkeitsdichte bei stetigen Zufallsvariablen (**p**), für die Verteilungs- und die Quantilfunktion (**q**), und schließlich für die Simulation von Zufallsexperimenten mit Zufallswerten (**r** für random, also zufällig).

Diese R-Funktionen beginnen also jeweils mit den Buchstaben **d**, **p**, **q** bzw. **r**, und nehmen zum Teil dieselben Argumente, zum Teil unterschiedliche, je nach den Parametern der entsprechenden Wahrscheinlichkeitsverteilungen.

Hier sind zusammengefasst die R-Funktionen für die vier vorhin vorgestellten Wahrscheinlichkeitsverteilungen:

Binomialverteilung:

```
dbinom(x, size, prob)
pbinom(q, size, prob, lower.tail = TRUE)
qbinom(p, size, prob, lower.tail = TRUE)
rbinom(n, size, prob)
```

Die Argumente `size` und `prob` entsprechen den Parametern n (Anzahl der Stufen des Zufallsexperiments) bzw. p (Wahrscheinlichkeit des Erfolgs) der Binomialverteilung.

Normalverteilung:

```
dnorm(x, mean = 0, sd = 1)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)
rnorm(n, mean = 0, sd = 1)
```

Die optionalen Argumente `mean` und `sd` nehmen den Erwartungswert bzw. die Standardabweichung der Verteilung; falls diese Argumente weggelassen werden, werden die Werte 0 bzw. 1 übergeben, d.h. die Verteilung ist dann die Standardnormalverteilung.

χ^2 - und t-Verteilungen:

```

dchisq(x, df) bzw. dt(x, df)
pchisq(q, df, lower.tail = TRUE) bzw. pt(q, df, lower.tail = TRUE)
qchisq(p, df, lower.tail = TRUE) bzw. qt(p, df, lower.tail = TRUE)
rchisq(n, df) bzw. rt(n, df)

```

Das Argument `df` nimmt die Anzahl der Freiheitsgrade der Verteilung (das Kürzel steht für die englische Bezeichnung 'degrees of freedom').

Übersicht:

Diskrete Verteilungen		
... (-)Verteilung	R-Name	Verteilungsparameter
Binomial	<code>binom</code>	<code>size, prob</code>
Geometrische	<code>geom</code>	<code>prob</code>
Hypergeometrische	<code>hyper</code>	<code>m, n, k</code>
Multinomial	<code>multinom</code>	<code>size, prob</code> (nur <code>r....</code> und <code>d....</code>)
Negative Binomial	<code>nbinom</code>	<code>size, prob</code>
Poisson	<code>pois</code>	<code>lambda</code>
Wilcoxons Vorzeichen-Rangsummen	<code>signrank</code>	<code>n</code>
Wilcoxons Rangsummen	<code>wilcox</code>	<code>m, n</code>

Stetige Verteilungen		
... (-)Verteilung	R-Name	Verteilungsparameter
Beta	<code>beta</code>	<code>shape1, shape2, ncp = 0</code>
Cauchy	<code>cauchy</code>	<code>location = 0, scale = 1</code>
χ^2	<code>chisq</code>	<code>df, ncp = 0</code>
Exponential	<code>exp</code>	<code>rate = 1</code>
F	<code>f</code>	<code>df1, df2 (ncp = 0)</code>
Gamma	<code>gamma</code>	<code>shape, rate = 1</code>
Log-Normal	<code>lnorm</code>	<code>meanlog = 0, sdlog = 1</code>
Logistische	<code>logis</code>	<code>location = 0, scale = 1</code>
Multivariate Normal (im package <code>mvtnorm</code>)	<code>mvnorm</code>	<code>mean = rep(0, d), sigma = diag(d)</code> (mit <code>d</code> = Dimension)
Multivariate <i>t</i> (im package <code>mvtnorm</code>)	<code>mvt</code>	(etwas komplizierter; siehe seine On-line-Hilfe)
Normal	<code>norm</code>	<code>mean = 0, sd = 1</code>
Students <i>t</i>	<code>t</code>	<code>df, ncp = 0</code>
Uniforme	<code>unif</code>	<code>min = 0, max = 1</code>
Weibull	<code>weibull</code>	<code>shape, scale = 1</code>

5.5.6 Interner Datensatz

Um mit R sofort starten zu können ohne eigene Daten einzugeben, ist in der Standardinstallation bereits ein umfangreicher Datensatz implementiert (Iris Datensatz).

Der Iris Datensatz beinhaltet 150 Beobachtungen von 4 Attributen von Schwertlilien.

Gemessen wurden dabei jeweils die Breite und die Länge des Kelchblatts (Sepalum) sowie des Kronblatts (Petalum) in Zentimeter.

Des weiteren ist für jeden Datensatz die Art der Schwertlilie (Iris setosa, Iris virginica oder Iris versicolor) angegeben. Für jede der 3 Schwertlilienarten liegen 50 Datensätze vor.

150 Beobachtungen, 5 Merkmale

- Länge und Breite des Kelch- bzw. Blütenblattes
(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) in cm.
- sowie Schwertlilienart
(Species)

Man kann sehr schnell einen Überblick über einen Datensatz erhalten:

⇒ `summary(iris)`

```

Sepal.Length      Sepal.Width      Petal.Length      Petal.Width      Species
Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa   :50
1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500

```

Durch Einbindung von Libraries, (z.B. hier für die Bibliothek 'mosaic'), sind viele weitere Befehle und Funktionen erweiterbar:

⇒ `require(mosaic) ⇒ glimpse(iris)`

```

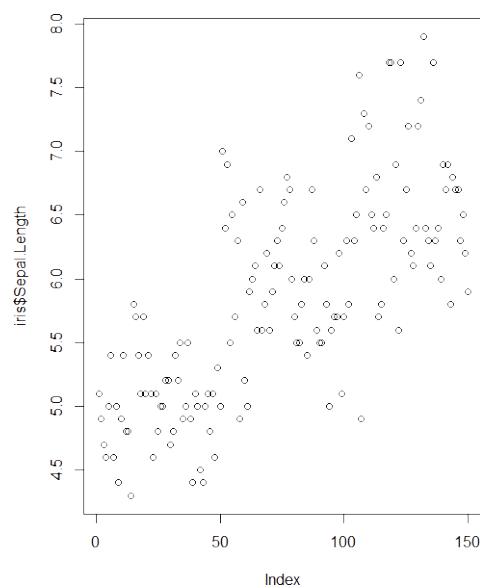
observations: 150
variables: 5
$ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.4, 4.9, 5.4, 4.8, 4.8, 4.3, ...
$ Sepal.Width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 2.9, 3.1, 3.7, 3.4, 3.0, 3.0, ...
$ Petal.Length  <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5, 1.5, 1.6, 1.4, 1.1, ...
$ Petal.Width   <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.2, 0.2, 0.1, 0.1, ...
$ Species       <fct> setosa, setosa, setosa, setosa, setosa, setosa, setosa, setosa, setosa, setosa...

```

5.5.7 Graphische Darstellungen

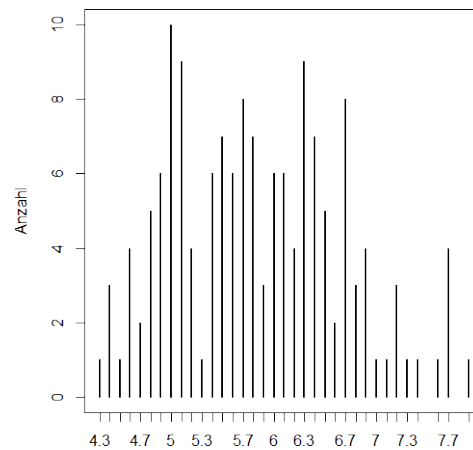
Einfacher Plot

⇒ `plot(iris$Sepal.Length)`



Häufigkeiten

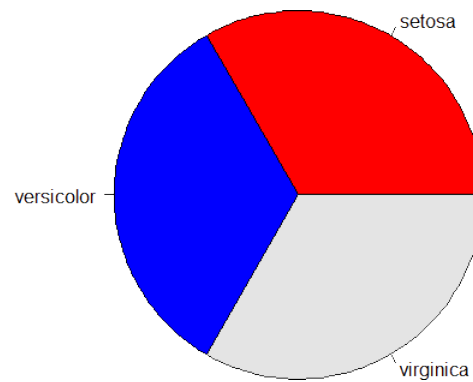
```
⇒ Anzahl<-table(iris$Sepal.Length)
⇒ plot(Anzahl)
```



Kuchendiagramme

Abs. Häufigkeiten der Variable X , im data.frame z

```
⇒ pie((table(iris$Species)),
    col=c("red","blue","grey"))
```



Histogramm

```
⇒ hist.default(iris$Sepal.Length)
⇒ help(hist)
```

