**Zusammenfassung MSTO**

René Bernhardsgrütter – 05.08.2013

**Wsk** = Wahrscheinlichkeit

# Begriffe

N oder X= Anzahl verschiedener Merkmale (Messobjekte)

n oder x = Merkmalswert = Anzahl gezählter/gemessener Objekte

h = Absolute Häufigkeit

H = Kumulierte absolute Häufigkeit

f = Relative Häufigkeit =

F = Kumulierte relative Häufigkeit

hi = Absolute Klassenhäufigkeit (0..∞)

Hi = Kumulierte Absolute Klassenhäufigkeit (0..∞)

fi = Relative Klassenhäufigkeit (0.0..1.0)

Fi = Kumulierte relative Klassenhäufigkeit (0.0..1.0)

M = Median

q = Quartilsabstand

# Deskriptive Statistik

|  |  |  |  |
| --- | --- | --- | --- |
| **Messniveau** | **Beschreibung** | **Beispiel** | **Diagramm** |
| **Nominal** | Nur Kategorisierung, Kennzeichnung | Wohnort, Studien-gang, Telefonnummer | Balken |
| **Ordinal** | Grössenordnung vorhanden, Unterschied aber nicht messbar | Kleidergrösse (s, m, l) | Balken |
| **Metrisch-stetig** | Grössen beliebig genau messbar | Masse | Graph,  Klassenbildung, Histogramm |
| **Metrisch-diskret** | Nur bestimmte Werte vorgesehen | Kinderzahl | Klassenbildung, Histogramm |

## Quartil/Quantile

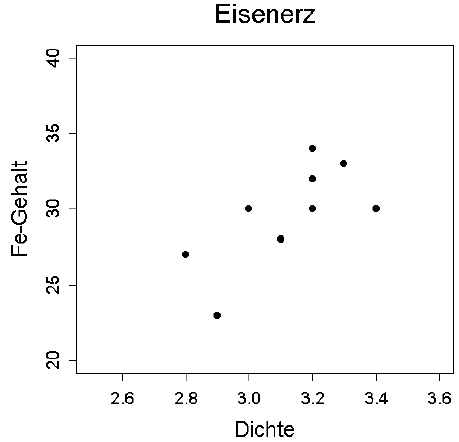
|  |  |
| --- | --- |
| Q1 = x0.25  Q2 = x0.5 = **Median**  Q3 = x0.75  q = x0.75 - x0.25 = **Quartilsabstand** |  |

|  |
| --- |
| **Median vs Mittelwert**  **Der Median teilt eine Stichprobe in zwei gleiche Hälften.** Er wird von extremen Werten (Ausreissern) praktisch kaum beeinflusst. Deshalb kann der Median zum Beispiel bei schiefen, unsymmetrischen Verteilungen (Laborwerte) oder bei der Betrachtung von Überlebenszeiten besser interpretiert werden.  Wenn es heiss: Ist jemand **„über/unter Form“**, dann ist die „Form“ der **Median**, also ist die Frage, ob jemand mit einem gewissen Wert in der oberen oder unteren Hälfte der verfügbaren Daten ist. |

**Boxplot:**

|  |  |
| --- | --- |
|  | **Boxplot mit Rechner erstellen**   1. In Tabelle alle Werte eintragen 2. Spalte bei A, B, C, … beschriften 3. Mit **CTRL+I** in **Data & Statistics** importieren 4. Mit TAB die Seiten (link, unten) beschriften 5. **Menu** > **Plot-Typ** > **Boxplot** |

**Streudiagramm:**



## Klassierung

**Histogramm**

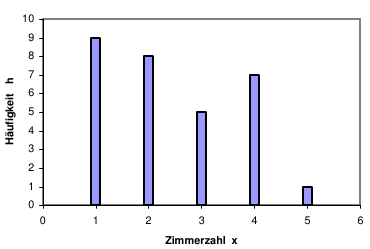
|  |  |  |
| --- | --- | --- |
| X  von..bis unter | hi |  |
| 100-200 | 182 |
| 200-500 | 35 |
| 500-800 | 317 |
| 800-1000 | 75 |

∆x = Klassenbreite

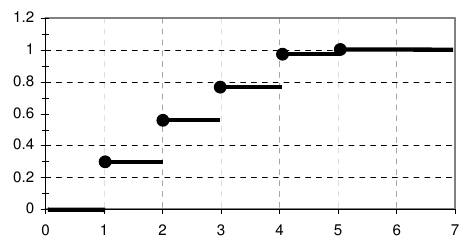
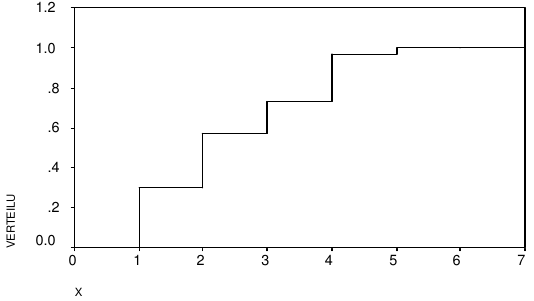
D = Absolute Dichte (Höhe)

d = Relative Dichte (Höhe)

|  |
| --- |
| **Histogramm mit Rechner erstellen**   1. In Tabelle alle Werte eintragen 2. Spalte bei A, B, C, … beschriften 3. Mit **CTRL+I** in **Data & Statistics** importieren 4. Mit TAB die Seiten (link, unten) beschriften 5. **Menu** > **Plot-Typ** > **Histogramm** |

**Stab-/Balkendiagramm**:  


**Summenkurve/Verteilungsfkt**:

 oder auch 

**Medianklasse bestimmen:**Die Medianklasse ist diejenige Klasse, bei der **Fi erstmals ≥ 0.5** ist.

**Merkmalswert eines α-Quantils bei klassierten Daten durch lineare Interpolation berechnen:**

Qα = Merkmalswert des gegebenen α-Quantils

α = Gegebenes Quantil der Klasse, z. B. 0.5 für den Median, 0.0..1.0

Fivorhergehend = Kumulierte relative Häufigkeit Fi von vorhergehender Kl.

∆x = Klassenbreite der Klasse, die das α-Quantil enthält

fi = relative Häufigkeit der Klasse, die das α-Quantil enthält

UKG = Untere Klassengrenze der Klasse, die das α-Quantil enthält

**Quantil-Wert eines Merkmals bei klassierten Daten durch lineare Interpolation berechnen:**

x = Merkmal (Zahlenwert)

α = Gesuchter Quantil-Wert des Merkmals, 0.0..1.0

UKG = Untere Klassengrenze der Klasse, die das Merkmal enthält

fi = relative Häufigkeit der Klasse, die das Merkmal enthält

∆x = Klassenbreite der Klasse, die das Merkmal enthält

Fivorhergehend = Kumulierte relative Häufigkeit Fi von vorhergehender Kl.

**Mittelwert mit gewichteten Klassenmittelwerten berechnen:**

= Mittelwert aller klassierten Werte

= Mittelwert der Klasse k

= Absolute Häufigkeit der Klasse k

= Kumulierte absolute Häufigkeit der letzten Klasse

**Empirische Verteilungsfunktion F(x) = Fi:**

|  |  |
| --- | --- |
|  | Gut, wenn man schnell einen Überblick über die Verteilung haben möchte.  Hier sieht man schnell:  ca. 70 % sind ≤ 1400 Fr.  ca. 50 % sind < 1020 Fr. |

**Häufigkeitsverteilung:**

Zeigt auf, welches Merkmal **wie oft** vorgekommen ist. Dies kann relativ oder absolut gargestellt werden. Oft werden Balkendiagramme verwendet.

Bsp: Tote/Jahr: 4, 4, 3, 2, 5, 1, 4,1,0 für 1990 – 2000. Dies gibt ein Balkendiagramm mit X-Achse 0-5, Y-Achse 0-3, die 4 kommt am häufigsten vor.

## Kennwerte

Kennwerte sind Zahlen, die die Verteilung von Merkmalswerten charakterisieren.

**Lagemasse** zeigen die mittlere Lage an. Verschiedene Möglichkeiten:

1. **Median** = M = Q2
2. **Mittelwert** (Durchschnitt) =
3. **Modus** (bezeichnet den häufigsten Wert)

Der Median ist im Gegensatz zum Mittelwert **robuster**, denn das Hinzufügen von einem extrem grossen Wert (Ausreisser) beeinflusst den Median nicht, den Mittelwert hingegen schon.

**Streumasse** drücken die mittlere Abweichung von *einer* Mitte aus.

V = s2 = Varianz

s = Standardabweichung

n = Anzahl verschiedene Werte

Die Varianz ist die mittlere quadrierte Abweichung vom Durchschnitt:

**Arithmetisches Mittel mit zwei Variablen berechnen:**

Dies wird im Rechner nicht direkt angezeigt. Die Spalten sind x bzw. y.

und können abgelesen werden.

|  |
| --- |
| **mit zwei Variablen mit Rechner berechnen**   1. In Tabelle alle Werte eintragen 2. Beide Spalten markieren 3. **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit zwei Variable…** 4. Die oben beschriebene Gleichung lösen |

## Lineare Regression

= Untersuchung, ob zwei Variablen etwas mit einander zu tun haben.

Regressionsgerade = Gerade, die man durch eine Punktewolke legt (auch: Ausgleichsgerade)

|  |  |
| --- | --- |
| = Ausgangsgrösse  = Zielvariable, abhängig von x  = beobachtete y-Werte  = erklärte y-Werte (berechnet durch Regressionsgleichung: )  = Residuen oder Fehler |  |

Mit der Methode der **kleinsten Quadrate** erreicht man eine optimale Regressionsgerade.

|  |
| --- |
| **Lineare Regression mit Rechner grafisch erstellen**   1. Alle Werte in Tabelle in zwei Spalten eintragen, erste Spalte die x-Werte, zweite Spalte die y-Werte 2. Spalte bei A, B beschriften 3. Mit **CTRL+I** in **Data & Statistics** importieren 4. Mit TAB die Seiten (link, unten) beschriften 5. **Menu** > **Analysieren** > **Regression** > **Lineare Regression (mx+b) anzeigen**   **Lineare Regression mit Rechner numerisch berechnen**   1. Alle Werte in Tabelle in zwei Spalten eintragen, erste Spalte die x-Werte, zweite Spalte die y-Werte 2. Beide Spalten markieren 3. **Menu** > **Statistik** > **Statistische Berechnung…** > **Lineare Regression (mx+b)…** 4. **m** und **b** können nun direkt weiterverwendet werden |

**Regressionsgerade aus Drittwerten berechnen:**

m = Steigung der Geraden

= Kovarianz

= x-Varianz

= Summe aller Abweichungsrechtecke von dem Mittelwert

= Quadrierte Summe aller x-Werte

= Mittel aller y-Werte

= Mittel aller x-Werte

= Verschiebung der Regressionsgeraden auf der Y-Achse

Folglich ist:

Die Gerade ist auch:

### Varianz

Die Varianz ist eine Eigenschaft der Verteilung einer Zufallsvariable und hängt nicht vom Zufall ab. Sie misst die Streuung der Werte relativ zum Erwartungswert, dabei werden die Quadrate der Abweichungen entsprechend ihrer **Wsk** gewichtet.

= Summe der Residuenquadrate

= Quadrierte Summe aller -Werte

= Quadrierte Summe aller -Werte

= Quadrierte Summe aller e-Werte

Es gilt die **Quadratesummen-Zerlegung**:

Teilt man diese durch n-1, so erhält man die **Varianz-Zerlegung**:

Totale Varianz = Erklärte Varianz + Fehler-Varianz

**Die Totale Varianz ist die *Varianz der y-Werte*!**

|  |
| --- |
| **Totale Varianz mit Rechner berechnen**   1. In Tabelle alle Werte eintragen 2. Beide Spalten markieren 3. **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit zwei Variable…** 4. Den Resultatwert kopieren und quadrieren, dann ist dies die Totale Varianz |

|  |
| --- |
| **Erklärte Varianz mit Rechner berechnen**   1. In Tabelle alle Werte eintragen 2. Beide Spalten markieren 3. **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit zwei Variable…** 4. Den Resultatwert kopieren und quadrieren, dann ist dies die Totale Varianz 5. Den Resultatwert kopieren und quadrieren, dann ist dies die das Bestimmtheitsmass 6. Nun mit die Erklärte Varianz berechnen |

### Stichprobenvarianz

= Stichprobenvarianz (wird bei nur einer Variable, wie hier, auch als bezeichnet)

= Quadrierte Summe aller x-Werte

**Die Stichprobenvarianz ist die Varianz der x-Werte (Stichproben)!**

|  |
| --- |
| **Stichprobenvarianz mit Rechner berechnen**   1. In Tabelle alle Werte eintragen 2. **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit einer Variable** 3. Den Resultatwert kopieren und im Rechner als verwenden |

### Kovarianz

Die Kovarianz ist die gemeinsame Varianz von und und der Korrelation (alles im TR enthalten, muss zusammenkopiert werden).

= Summe aller Abweichungsrechtecke von dem Mittelwert

Diese Formeln dürften am besten sein:

### Bestimmtheitsmass R2

= Anteil der erklärten Varianz an der totalen Varianz .

Wenn R2 = 1, dann ist der Anteil der erklärten Varianz 100 %, es liegen alle Punkte auf der Regressionsgeraden.

Wenn R2 = 0, dann korrelieren die x- und y-Werte nicht (haben nichts miteinander zu tun. Die Steigung m der Regressionsgerade ist 0.

### Korrelation r

r hat dasselbe Vorzeichen wie

### Interpretation von R2 und r

Aus dem R2 alleine lässt sich aber **nicht** folgern, dass bei grossem R2 ein linearer Trend besteht und bei kleinem nicht, denn es gibt Gegenbeispiele!

Man muss immer das Streudiagramm betrachten: Die Residuen sollten zufällig und überall etwa gleich und die x-Achte streuen, kleine Residuen sind häufiger als grosse.

|  |  |
| --- | --- |
| Idealfall, Regressionsgerade gut: | Kein Trend, Regressionsgerade unpassend: |

# Schubfachprinzip

|  |
| --- |
| **Schwache Form**  Falls man Objekte auf Mengen verteilt und grösser als ist, dann gibt es mindestens eine Menge, in der mehr als Objekt landet. |

|  |
| --- |
| **Starke Form**  Verteilt man Objekte auf Mengen, so gibt es mindestens eine Menge, in der sich zumindest Objekte befinden.  **Formel**  n = Anzahl Objekte  m = Anzahl Schubfächer  q = Gewünschte Anzahl („Wie viele Personen braucht es, damit mindestens q = 5 dieselbe Anzahl Haare auf dem Kopf haben?“) |

# Binomialkoeffizienten

1. Binomische Formel:

2. Binomische Formel:

**Allgemeine binomische Formel:**

**Multinomial-Formel mit 3 Summanden:**

Diese Form muss immer beibehalten werden. Wenn beispielsweise **nur ein Faktor** vorhanden ist, muss dieser als eine Addition/Subtraktion dargestellt werden.

**Bsp: 0.99910 ausrechen:**

### Regeln der Binomialkoeffizienten

|  |
| --- |
| **Binomialkoeffizienten mit Rechner berechnen**   1. Folgenden Befehl verwenden: nCr(n, k) |

**Zahlenfaktor einer Form berechnen:**

= Faktor der Form

= Exponent von

= Exponent von

Bsp: Faktor in aus berechnen:

**Grösster Zahlenfaktor einer Form bestimmen:**

= Faktor bei (kann auch 1 sein)

= Faktor bei (kann auch 1 sein)

= Faktor der Form in der Summenformel

= Laufvariable von 0..n

Durchtesten und schauen, wo das grösstes raus kommt:

Das beim grössten ist das Ergebnis.

Bsp: Welcher Summand von hat den grösst

en Koeff.?

k=0 => 531441, k=1 => 3897234, **k=2 => 4330260**, k=3 => 2347695

### Pascal’sches Dreieck

|  |  |
| --- | --- |
| **Berechnen der Zahlen**  Oben mit 1 beginnen. In den folgenden Reihen immer die oberen zwei Zahlen zusammenzählen. | n=0  n=1  n=2  n=3  n=4 |

**Anwendung**

Die Zahl im Dreieck ist der Faktor jeder Form.

Der Exponent von a beginnt bei n und nimmt mit jeder Form eins ab.

Der Exponent von b beginnt mit 0 und nimmt mit jeder Form 1 zu.

Das Vorzeichen wechselt regelmässig zwischen und (dort nimmt man das, welches zu der gegebenen Aufgabe passt).

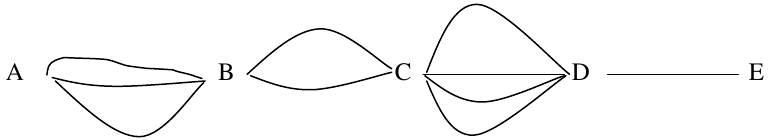
Bsp:

# Kombinatorik

z = Anzahl verschiedene Kombinationen

## Produktregel

Man fragt sich, wie viele Wege es von A nach E gibt. Die Produktregel macht folgendes:



Man nimmt immer die jeweils zur Verfügung stehende Anzahl Möglichkeiten und multipliziert diese mit den vorherigen Möglichkeiten bis zum Schluss.

Bsp: Wie viele vierstellige Passwortkombinationen mit den Ziffern 0-9 gibt es, wenn die erste Ziffer keine 0 sein soll?

Bsp: Aus einer Kiste mit 12 Glühbirnen nimmt man zwei zur Kontrolle raus. Wie viele verschiedene Paare gibt es?

Zuerst hat man verschiedene Möglichkeiten, die Birnen rauszunehmen. Da es bei diesem Paar aber nicht auf die Reihenfolge an kommt, muss die Menge noch halbiert werden.

**Permutationen mit Einschränkung** müssen halbmanuell ausgerechnet werden. Dabei rechnet man zuerst die Permutationen ohne Einschränkung aus, dann die anderen und zählt sie ab, dividiert sie raus oder addiert sie (je nach dem).

Bsp: Um einen Tisch soll man Personen (A..F) auf verschiedene Arten anordnen. Zwei Anordnungen werden aber als gleich angesehen, wenn die eine durch Rotieren aus der anderen erreicht werden kann.

|  |  |
| --- | --- |
|  | Diese beiden Anordnungen sind gleich, da die linke durch Rotieren aus der rechten erreicht werden kann. |

Ohne Einschränkung hätte man

Personen.

Jede Sitzordnung kommt total 6-mal vor (dieselbe einfach rotiert), daher kann man von den Anordnungen nur einen Sechstel nehmen:

## Mit Wdh => nicht unterscheidbare Elemente

N = Anzahl Permutationen

n = Anzahl Elemente

p = Anzahl nicht unterscheidbarer Elemente

Man startet mit allen Permutationen und dividiert nachher von jeden nicht unterscheidbaren Permutationen alle möglichen Anordnungen ab, die sich ergäben, wenn die Elemente doch unterscheidbar wären.

Bsp: Das Wort BALLADE soll auf möglichst viele verschiedene Möglichkeiten angeordnet werden.

Bsp: Auf wie viele Arten können sich 9 Leute auf 12 Plätze verteilen?

Die drei leeren Plätze können als Wiederholungen aufgefasst werden.

### Torfolgen-Beispiel

A, B = Zwei Fussballteams

**Beispiel**: A:B spielen 7:4. Wie viele ***Torfolgen*** gibt es? Eine Torfolge ist zum Beispiel: . Sofort klar:

Wenn so etwas kommt, immer ein Bespiel aufschreiben! Hier war das Schwierigste, herauszufinden, was genau eine *Torfolge* ist!

## Ohne Wdh => unterscheidbare Elemente

### Lotto-Formel

N = Anz. Möglichkeiten, wo k Zahlen richtig sind

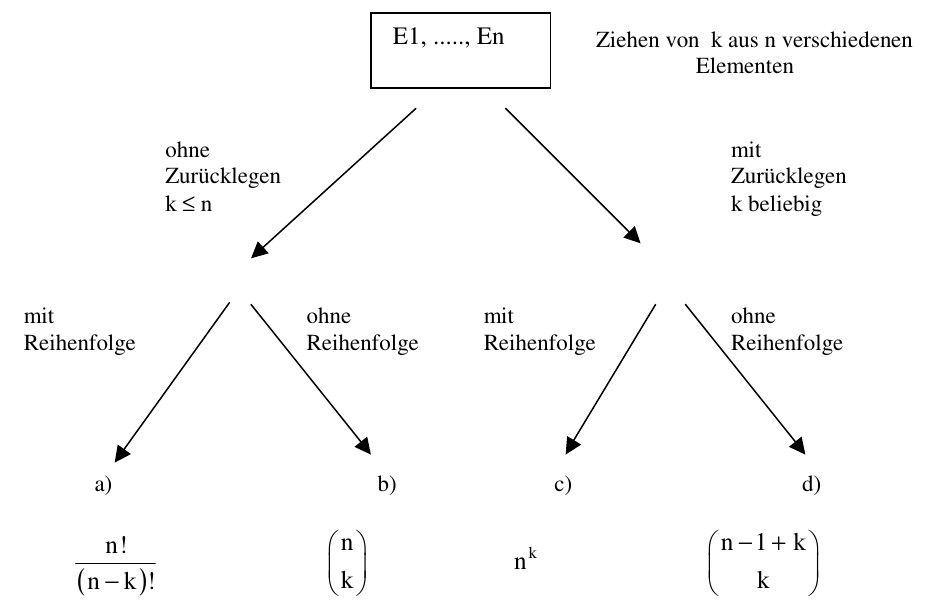
6 = Anz. Züge

45 = Anz. Lose (Elemente)

k = Richtige Zahlen

## Verschiedene Auswahlarten

Es wird zwischen mit/ohne Zurücklegen und mit/ohne Beachtung der Reihenfolge unterschieden.



# Wsk

### Wsk für das Eintreffen von

= Ereignis, das eintreffen soll

= Anzahl für das Ereignis A günstige Fälle

= Maximale Anzahl Ereignisse (gegenseitig ausschliessend)

### Absolute und Relative Häufigkeiten und

= Absolute Häufigkeit, z. B. aus laufender Produktion

= Relative Häufigkeit

= Anzahl Proben/Messungen

Die relative Häufigkeit nähert sich dem **Wsk**wert des Ereignisses. Diese Zahl heisst **statistische Wsk**:

## Abstraktes Rechnen

= Ergebnisraum, also Menge aller möglichen Ausgänge (Münze 2x werfen: )

A = Ergebnis,

= **Wsk**, dass das Ergebnis eintrifft

= Gegen**wsk** des Ergebnisses

Eigenschaften:

**Wsksfunktion** heisst eine Zuordnung wie .

**Folgerungen aus den Eigenschaften:**

### Ereignisbaum zur Darstellung mehrstufiger Versuche

Bei mehrstufigen Versuchen kann man mit einem Ereignisbaum gut darstellen, was für **Wsk** existieren.

Bsp: In einer Urne liegen 5 weisse und 3 schwarze Kugeln. Man soll zwei ohne Zurücklegen ziehen. Gemäss ergeben sich folgende **Wsk**:

Als **Ereignisbaum** dargestellt, der **Pfad** von ist markiert:

|  |  |
| --- | --- |
|  | **Pfadregeln**   * Entlang eines Pfades werden **Wsk** multipliziert * Pfade sind disjunkt, ihre **Wsk** werden addiert () |

Bsp:

Oder mit dem Gegenereignis:

### Geburtstagsparadoxon

P = Wahrscheinlichkeit, dass 2 Personen am selben Tag Geb. haben

n = Anzahl Personen, die man nimmt

### Jasskarten-n-Blatt

P = Wsk, dass dies eintritt, wenn man k Karten aus 36 zieht

k = Anzahl Züge von Karten

n = Anzahl Karten in Folge von aus einer Farbe (Bsp: Rose 6,7,8,9,10)

4 = Anzahl Farben (müssen hier einfliessen, da Farben relevant sind)

## Systeme

P = Ausfallwahrscheinlichkeit des Systems

pi = Ausfallwahrscheinlichkeit einer Komponente

Z = Zuverlässigkeit des Systems

zi = Zuverlässigkeit einer Komponente

### Serieschaltungen

p2

p1

### Parallelschaltungen

p2

p1

## Bedingte Wsk

= **Wsk** von Ereignissen beziehen sich auf andere Ereignisse, die bereits eingetreten sind.

|  |
| --- |
| **Immer zuerst Bedingte Wsk-Schreibweise!!!** |

Es werden verschiedene **Wsk**, die bool’sch errechenbar sind, miteinander kombiniert.

Bsp: Man kennt die Zahlen von farbenblind B ja/N nein und ob jeweils Mann M/Frau F. Es lassen sich unter verschiedenen Bedingungen die **Wsk** ausrechnen:

**Allgemein:**

= **Wsk**, dass A unter der Bedingung von B eintrifft

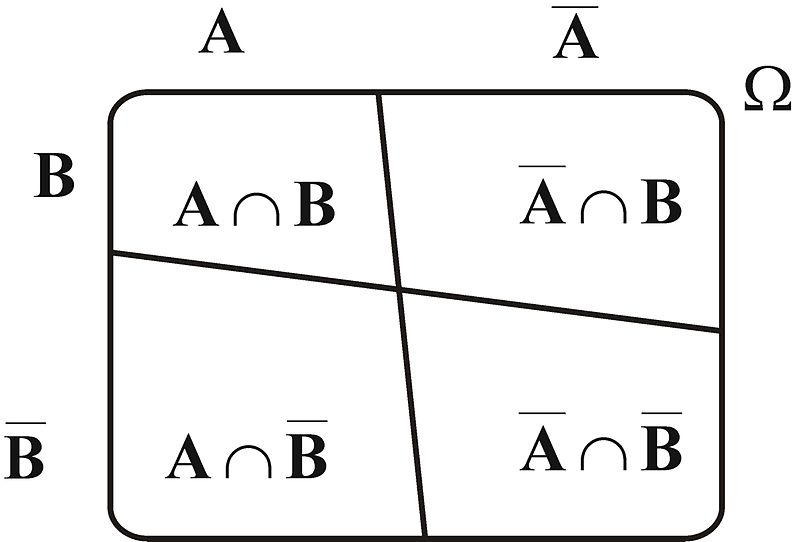
= Geschnittene **Wsk**

kann bei abhängigen Problemen nicht mit einer Formel errechnet werden sondern nur mit dem Baum! Es ist das, wo P(A) und P(B) gemeinsam erfüllt sind!

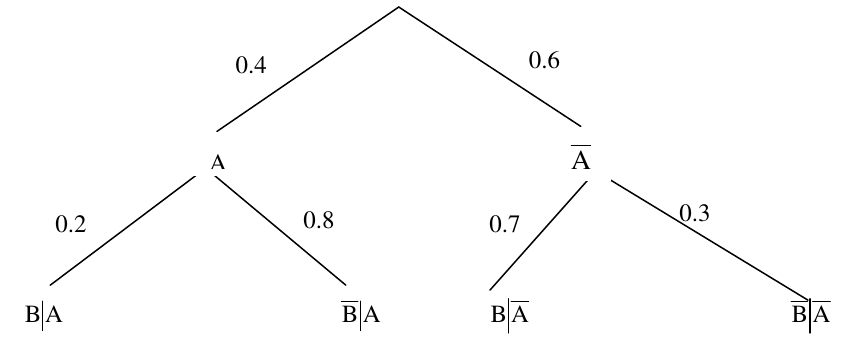
**Additionssatz:**

**De Morgansche Regeln:**

Daraus folgt:



**Ereignisbäume angepasst:**



Die Knoten werden jeweils mit dem Zustand () beschriftet. Die Abhängigkeit (bei ist dies ) wird zuerst/oben eingetragen.

|  |
| --- |
| **Wenn Unbekannte im Baum sind, Variablen einsetzen  und mit solver danach auflösen!** |

## Fehlalarme/Medikamentenwirkung

**Prävelenz** Empirische Daten, wie viele Brände passieren bzw. wie viele Krank sind.

**Sensitivität** Wie gut Brände erkannt werden bzw. wie gut Kranke erkannt werden.

**Spezivität** Wie gut erkannt wird, dass es nicht brennt bzw. wie gut Gesunde erkannt werden.

= false negative, Kranker wird als gesund erkannt

= false positive, Gesunder wird als krank erkannt

### Beispiel Feuermelder

0.95 = Sensitivität, also korrekt erkannte Feuer

0.99 = Spezivität, also korrekt erkannt, dass es nicht brennt

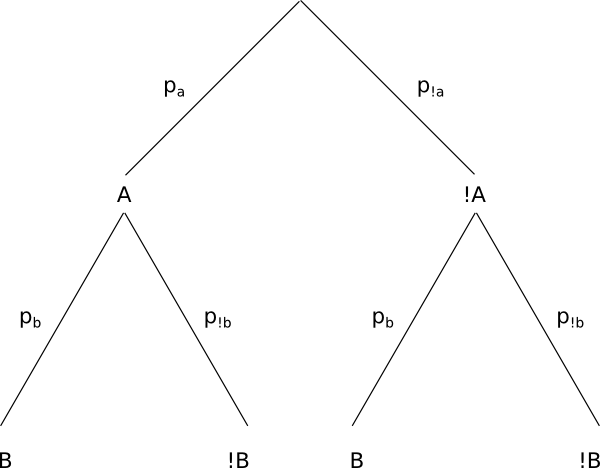
1/365 = Brennhäufigkeit (1x pro Jahr)

A = für Feuer, es brennt oder nicht (!A)

B = für Meldung, es meldet oder nicht (!B)

Frage: Es wird Feuer gemeldet. Mit welcher Wsk brennt es wirklich?

1. In Frageform schreiben:
2. Baum zeichnen:



Meldung

Feuer

A = Feuer vorhanden, !A = kein Feuer

pa = , p!a

B = Meldung, !B = keine Meldung

Wenn A (es brennt): pb = 0.95, p!b = 0.05

Wenn !A (es brennt nicht): pb = 0.01, p!b = 0.99

1. Nun normal ausrechnen:

## Unabhängige Ereignisse

ist von stochastisch unabhängig, wenn gilt:

Dann ist auch

Denn egal ob mit oder ohne , kommt immer gleich häufig vor.

**Überprüfen** kann man die Unabhängigkeit, indem man beide Seiten der obigen Formel und berechnet und die Resultate vergleicht.

## Verteilungen

= Punktuelle Verteilungsfunktion, Wert einer Ausprägung

= Verteilungsfunktion, Wert nach x %, 0..1

Punktueller Wert:

%-Wert von 0 bis zu bestimmtem **stetigen** x:

%-Wert ab **stetigem** x bis zu Ende:

### Rechnen mit Erwartungswert E[X] und Varianz V[X]

Das fette blaue muss jeweils durch die effektive Funktion der Verteilung ersetzt werden.

**Bei stetigen Zufallsvariablen:**

E[X] = Erwartungswert (= )

V[X] = Varianz (= )

= Standardabweichung

x = Effektiv das x als Buchstabe, es wird einfach verrechnet

**Varianz aus Erwartungswert:**

**Rechenregeln:**

### Summen von unabhängigen Variablen

= Summe von den Erwartungswerten

= Summe von den Standardabweichungen

n = Anz. Merkmale

### Mittelwerte von Kennwerten

= Mittelwert von den Erwartungswerten

= Mittelwert von den Standardabweichungen

X = Zufallsvariable

### Quantile

Die Flächen und sind die Quantile!

### Zeichnungen

|  |  |
| --- | --- |
| **Wahrscheinlichkeitsverteilung:**  Summe aller Stäbe \* p = 1 | http://wikis.zum.de/rsg/images/5/5f/Verteilung3.jpg |
| **Verteilungsfunktion:**  Geht immer von 0 bis 1.0. |
| http://www.bb-sbl.de/assets/images/tutorial/Vf_chi2.jpg |

### Hypergeometrische Verteilung

Man zählt Erfolge, z. B. beim Ziehen von Stichproben ohne Zurücklegen.

N = Anzahl Kugeln

K= Schwarze Kugeln

N-K = Weisse Kugeln

X = Anzahl Schwarzer Kugeln in der Stichprobe

n = Ausprägung in Verteilung

= Erwartungswert E[X]

= Varianz V[X]

Das wird auch bei **Annahmekontrolle** verwendet, z. B. max. 3 % dürfen fehlerhaft sein. Es ist ein Ziehen ohne Zurücklegen.

### Binomialverteilung

Anzahl Erfolge/Misserfolge oder funktionstüchtig/defekt.

n = Anz. Versuche

k = Anz. Misserfolge bzw. Defekte, 0, 1, 2, .. n

p = Wsk, dass Erfolg bzw. funktionstüchtig

= Erwartungswert E[X]

= Varianz V[X]

### Geometrische Verteilung

Warten auf den ersten Erfolg.

X = Anz. Misserfolge

p = Misserfogls**wsk** (ohne den ersten Erfolg)

= Erwartungswert E[X]

= Varianz V[X]

|  |
| --- |
| **q = Erfolgswsk = 1-p**  Das k+1 ist, da beim TR die **obere Schranke der erste Erfolg ist**! |

### Exponentialverteilung

Ist das Grundmodell für Lebenszeiten von Systemen mit Geburten und Toten.

T = Zufallsvariable der Zeit

t = Anz. Zeiteinheiten eines Merkmals, z. B. Jahre

c = Konstante des Systems

= Erwartungswert E[X]

= Varianz V[X]

**Alter bzw. Überlebende:**

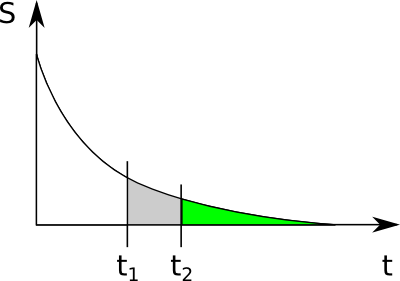
**Überlebende, wenn bereits bestimmtes Alter erreicht:**

c = 0.03

t1 = 20 Jahre. Alter, welches Bedingung ist

t2 = 25 Jahre. Alter, das erreicht werden soll

Wie gross ist die Wsk, dass ein Gerät, dass t1 Jahre alt ist, t2 Jahre alt wird?

1. Frage umschreiben:
2. Es geht darum, dass ein Gerät, das im grauen Bereich ist, in den Grünen kommt. Wie gross ist die Wsk davon?  
   
3. => Kleinerer Teil über grösserer Teil rechnen:

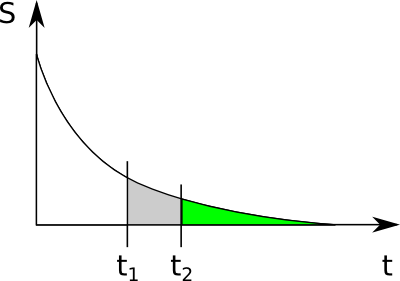
**Überlebende, wenn bereits bestimmtes Alter erreicht, die höchstens ein weiteres Alter erreichen:**

c = 0.03

t1 = 20 Jahre. Alter, welches Bedingung ist

t2 = 30 Jahre. Alter, das als obere Grenze gilt

Wie gross ist die Wsk, dass ein Gerät, dass t1 Jahre alt ist, höchstens t2 Jahre alt wird?

1. Frage umschreiben:
2. Es geht darum, dass ein Gerät, das im grauen Bereich ist, höchstens an die Grenze zum Grünen kommen darf. Wie gross ist die Wsk, dass das Gerät vorher kaputt geht?  
   
3. => Anteil von t1..t2 durch 0..t1 teilen:

### Uniforme Verteilung

Wenn etwas stetig ist.

a = Anfangswert

b = Endwert

= Erwartungswert E[X]

= Varianz V[X]

### Normalverteilung

Standard für alle Dinge, irgendwie… Es steht, wenn es NV ist.

= Normalverteilungs**fkt** f(x)

= Kumulierte Normalverteilungs**fkt** F(x)

= Erwartungswert E[X], berechnen aus gegebenen Daten

= Varianz V[X], Berechnen aus gegebenen Daten

**Normalverteilungstransformation:**

Wenn man in einer Normalverteilung eine „Subverteilung“ braucht, kann man das transformieren, also verschieben. Dazu gibt es dann eine neue Normalverteilung. Z hat , .

X = ZV der alten Normalverteilung

Z = Standardnormalverteilung

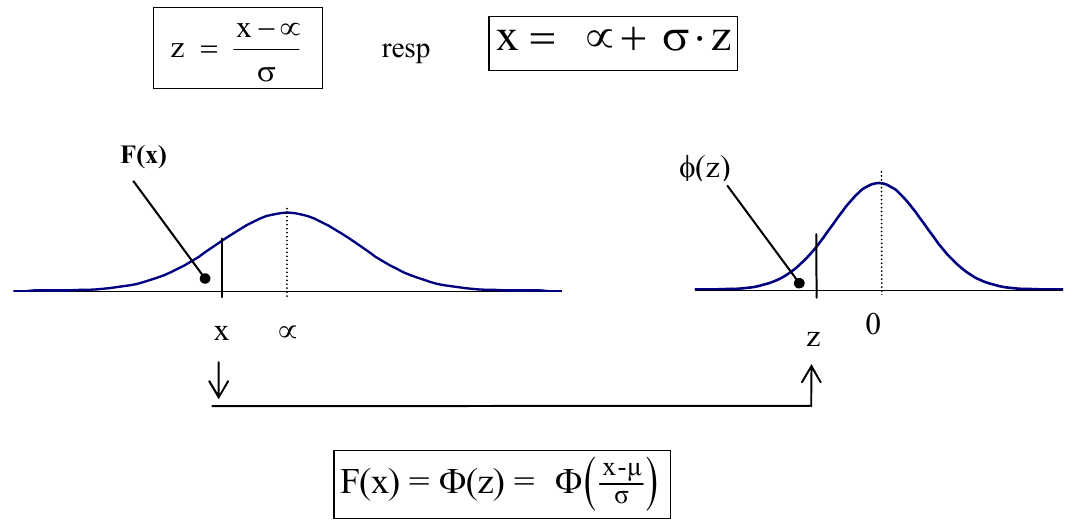
x = Merkmal von X, z. B. 2065g einer Zuckerpackung

= Erwartungswert von X

= Standardabweichung von X, z. B. 30g bei Zuckerpackung

= Merkmal von Z, beim Zuckerbsp (Ü11, A4a) ist es 75%, also wo das neue steht.

Wenn nun gesucht ist, kann dies hergeleitet werden:



**Standardnormalverteilung:**

= Erwartungswert E[X]

= Standardabweichung

**Summe von unabhängigen Normalverteilungen:**

Sind X und Y unabhängige, normalverteilte Zufallsvariablen so ist die Summe ebenfalls normalverteilt mit den Parametern:

Ist die Summe von normalverteilten, gleichverteilten ZV , dann gilt:

Bsp: Standardkuh ist

40 Kühe:

Das sagt, dass die Milchmenge pro Tag 18.97 Liter variieren kann.

### Poission-Verteilung

Gesetzmässigkeit für **seltene Ereignisse/Treffer, z. B. Brände, Ausfälle**, ...

X = Anzahl Treffer pro Zeiteinheit

= Mittle Anzahl von Treffer (Brand, neuer Kunde) pro Zeitintervall (bei Warteschlange treffen pro Minute 30 Kunden ein )

n = Anzahl Zeitintervalle

N = Mittlere Ereignisanzahl in einem Zeitintervall =

**Von Binomial zu Poisson:**

Bei und von den Binos kann man zu Poission wechseln:

**Wartezeit bis zum nächsten Eintreffen / Poisson-Prozess:**

Ist exponentiell verteilt mit dem Parameter .

T = Zeit bis zum nächsten Treffer, ist eine stetige ZV

f(t) = Dichtefunktion von T

= Erwartungswert der Wartezeit zwischen 2 Ereignissen

Wsk, dass nächstes Ereignis zwischen und eintritt:

## Näherung an Normalverteilung

Man nähert die Binomialverteilung an die Normalvert. Je grösser das , desto eher passt es.

**De Moivre-Laplace**: Näherung bei grossen von :

Dabei muss man einsetzen:

Faustregel, ab wann anwendbar:

(wenn geht auch schon weniger als 9)

## Zentraler Grenzwertsatz

X = ZV mit unabhängigen Elementen

= Mittelwert über alle Elemente von X ( ist also auch eine ZV)

Wichtig: Der Verteilungstyp von X überträgt sich i.d.R. nicht auf .

Zentraler Grenzwertsatz:

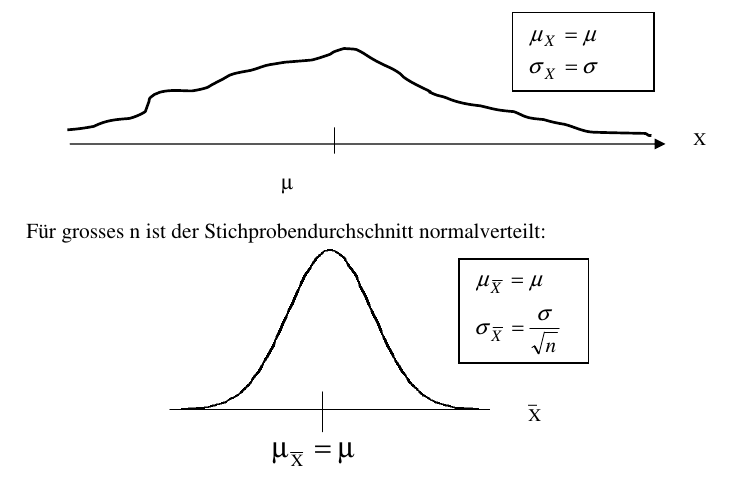
= voneinander unabhängige ZVs, i = 1..n

ist gegeben

ist gegeben

Die Standardisierung von strebt gegen die NV:

**Praktisch relevant bei :**



## Schätzen

Man soll Schätzfunktion erstellen und ihre Genauigkeit bestimmen.

Werte, die ein Tächli haben, sind geschätzt. Bsp: .

### Punktschätzung

= gesch. Erwartungswert

= Punktschätung

A = Anzahl Erfolge

n = Totale Anzahl

**Genauigkeit:**

= Stichprobenvarianz

= Standardfehler der Stichprobe