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 = $\frac{h}{N}$
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

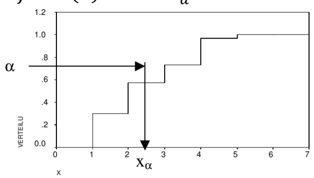
Deskriptive Statistik

Messniveau	Beschreibung	Beispiel	Diagramm
Nominal	Nur Kategorisierung, Kennzeichnung	Wohnort, Studiengang, 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

$$\alpha Quantil = kleinste\ Zahl\ für\ F(x) = \alpha = x_\alpha$$

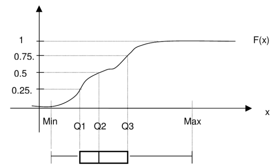
Q1 = $x_{0.25}$
Q2 = $x_{0.5}$ = **Median**
Q3 = $x_{0.75}$
q = $x_{0.75}$ - $x_{0.25}$ = **Quartilsabstand**



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:

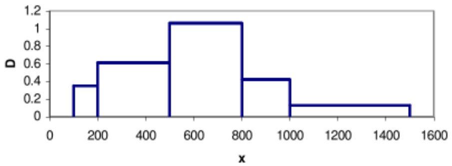


Boxplot mit Rechner erstellen
1) In Tabelle alle Werte eintragen
2) Spalte bei A, B, C, … beschriften
3) Mit **CTRL+I** in **Data & Statistics** importieren
4) Mit TAB die Seiten (link, unten) beschriften
5) **Menu** > **Plot-Typ** > **Boxplot**

Klassierung

Histogramm

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



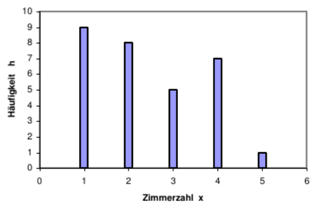
Δx = Klassenbreite
D = Absolute Dichte (Höhe)
d = Relative Dichte (Höhe)

$$D = \frac{h}{\Delta x}$$

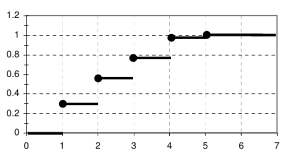
$$d = \frac{f}{\Delta x}$$

Histogramm mit Rechner erstellen
6) In Tabelle alle Werte eintragen
7) Spalte bei A, B, C, … beschriften
8) Mit **CTRL+I** in **Data & Statistics** importieren
9) Mit TAB die Seiten (link, unten) beschriften
10) **Menu** > **Plot-Typ** > **Histogramm**

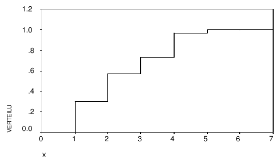
Stab-/Balkendiagramm:



Summenkurve/Verteilungsfkt:



oder auch



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

α-Quantil bei klassierten Daten durch lineare Interpolation berechnen:
Q_α = α-Quantil
α = Gesuchtes Quantil der Klasse, z. B. 0.5 für den Median, 0.0..1.0
$Fi_{vorhergehend}$ = Kumulierte relative Häufigkeit Fi von vorhergehender Kl.
Δx = Klassenbreite der Klasse, die das Quantil enthält
fi = relative Häufigkeit der Klasse, die das Quantil enthält
UKG = Untere Klassengrenze der Klasse, die das Quantil enthält

$$Q_\alpha = \frac{(\alpha - Fi_{vorhergehend}) * \Delta x}{fi} + UKG$$

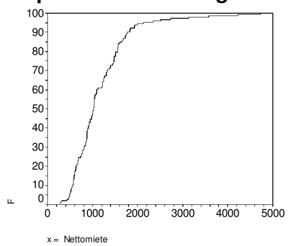
Quantil-Wert eines Merkmals bei klassierten Daten durch lineare Interpolation berechnen:
x = Merkmal (Zahlenwert)
α = Gesuchtes Quantil-Wert des Merkmals, 0.0..1.0
UKG = Untere Klassengrenze der Klasse, die das Merkmal enthält
fi = relative Häufigkeit der Klasse, die das Merkmal enthält
Δx = Klassenbreite der Klasse, die das Merkmal enthält
$Fi_{vorhergehend}$ = Kumulierte relative Häufigkeit Fi von vorhergehender Kl.

$$\alpha = \frac{(x - UKG) * fi}{\Delta x} + Fi_{vorhergehend}$$

Mittelwert mit gewichteten Klassenmittelwerten berechnen:
$\bar{x}$ = Mittelwert aller klassierten Werte
$\bar{x}_k$ = Mittelwert der Klasse k
hi_k = Absolute Häufigkeit der Klasse k
Hi_j = Kumulierte absolute Häufigkeit der letzten Klasse

$$\bar{x} = \frac{1}{Hi_j}(\bar{x}_0 * hi_0 + \dots + \bar{x}_j * hi_j), \qquad k = 0..j$$

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 gestellt 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) = $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_n$
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 = s^2 = Varianz
s = Standardabweichung
n = Anzahl verschiedene Werte

Die Varianz ist die mittlere quadrierte Abweichung vom Durchschnitt:

$$s^2 = \sum_{j=1}^{n}(x_j - \bar{x})^2 = \sum_{j=1}^{n} x_j^2 - n\bar{x}^2$$

$$s = \sqrt{s^2}$$

Arithmetisches Mittel $\bar{x}$ mit zwei Variablen berechnen:
Dies wird im Rechner nicht direkt angezeigt. Die Spalten sind x bzw. y. $\sum xy$ und $\sum y$ können abgelesen werden.

$$\bar{x} = \frac{\sum xy}{\sum y}$$

$\bar{x}$ 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

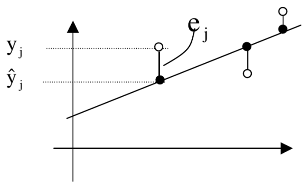
Lineare Regression

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

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

x = Ausgangsgrösse
y = Zielvariable, abhängig von x

y_j = beobachtete y-Werte
$\hat{y}_j$ = erklärte y-Werte (berechnet durch Regressionsgleichung: $\hat{y}_j = m * x_j + b$)
e_j = 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

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.

S = Summe der Residuenquadrate
S_{yy} = Quadrierte Summe aller y-Werte
$S_{\hat{y}\hat{y}}$ = Quadrierte Summe aller $\hat{y}$-Werte
S_{ee} = Quadrierte Summe aller e-Werte

Es gilt die **Quadratesummen-Zerlegung**:

$$S_{yy} = S_{\hat{y}\hat{y}} + S_{ee}$$

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

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

Totale Varianz = Erklärte Varianz + Fehler-Varianz

$$s_{\hat{y}}^2 = s_y^2 * R^2$$

Die Totale Varianz ist die Varianz der y-Werte!

Totale Varianz mit Rechner berechnen
5) In Tabelle alle Werte eintragen
6) Beide Spalten markieren
7) **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit zwei Variable…**
8) Den Resultatwert $sx := s_{n-1}x$ kopieren und quadrieren, dann ist dies die Totale Varianz s_y^2

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 $sx := s_{n-1}x$ kopieren und quadrieren, dann ist dies die Totale Varianz s_y^2
5) Den Resultatwert r kopieren und quadrieren, dann ist dies die das Bestimmtheitsmass R^2
6) Nun mit $s_{\hat{y}}^2 = s_y^2 * R^2$ die Erklärte Varianz $s_{\hat{y}}^2$ berechnen

Stichprobenvarianz s_x^2

s_x^2 = Stichprobenvarianz (wird bei nur einer Variable, wie hier, auch als s^2 bezeichnet)
S_{xx} = Quadrierte Summe aller x-Werte

$$s_x^2 = \frac{S_{xx}}{n-1}$$

$$S_{xx} = \sum_{j=1}^{n}(x_j - \bar{x})^2 = \sum_{j=1}^{n} x_j - n * \bar{x}^2 = s_x^2 * (n-1)$$

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

Stichprobenvarianz mit Rechner berechnen
7) In Tabelle alle Werte eintragen
8) **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit einer Variable**
9) Den Resultatwert sx kopieren und im Rechner als s_x^2 verwenden

Kovarianz s_{xy}

Die Kovarianz ist die gemeinsame Varianz von x und y.

S_{xy} = Summe aller Abweichungsrechtecke von dem Mittelwert

$$s_{xy} = \frac{S_{xy}}{n-1}$$

$$S_{xy} = \sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y}) = \sum_{j=1}^{n} x_j y_j - n * \overline{xy} = s_{xy} * (n-1)$$

Kovarianz mit Rechner berechnen
1) In Tabelle alle Werte in Spalten A bzw. B eintragen
2) **Menu** > **Statistik** > **Statistische Berechnung…** > **Statistik mit zwei Variablen**, Resultat in Spalte C bzw. D
3) Wie in Tabellenkalkulation die mittlere Formel ausrechen:
   a. Spalte E mit $x * y$ berechnen (`=a1*b1`) und herunterziehen mit **Menu** > **Daten** > **Nach unten ausfüllen**
   b. Davon die Summe bilden und den Rest in einer Zeile ausrechen: s_{xy} := `"=(sum(e1:e10)-10*d2*d8)/9"`
   $n = 10$, $x * y$-Resultat in Spalte e,
   $\bar{x}$ in Feld D2, $\bar{y}$ in Feld D8.

Bestimmtheitsmass R^2
= Anteil der erklärten Varianz $s_{\hat{y}}^2$ an der totalen Varianz s_y^2.

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{s_{xy}^2}{s_x^2 * s_y^2} \quad 0 \le R^2 \le 1$$

Wenn $R^2 = 1$, dann ist der Anteil der erklärten Varianz 100 %, es liegen alle Punkte auf der Regressionsgeraden.
Wenn $R^2 = 0$, dann korrelieren die x- und y-Werte nicht (haben nichts miteinander zu tun. Die Steigung m der Regressionsgerade ist 0.

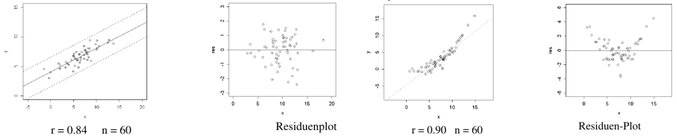
Korrelation r
r hat dasselbe Vorzeichen wie s_{xy}

$$r = \pm\sqrt{R^2} = \frac{s_{xy}}{s_x \; * s_y} \quad -1 \le r \le 1$$

Interpretation von R^2 und r
Aus dem R^2 alleine lässt sich aber **nicht** folgern, dass bei grossem R^2 ein linearer Trend besteht und bei kleinem nicht, denn es gibt Gegenbeispiele!

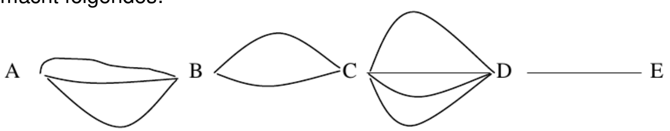
Man muss immer das Streudiagramm betrachten: Die Residuen sollten zufällig und überall etwa gleich und die x-Achte streuen, kleine Residuen sind häufiger als grosse.

Idealfall, Regressionsgerade gut: Kein Trend, Regressionsgerade unpassend:



Kombinatorik
z = Anzahl verschiedene Kombinationen (Permutationen)

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



$$z = \quad 3 \quad * \quad 2 \quad * \quad 4 \quad * \quad 1 \quad = 24$$

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?

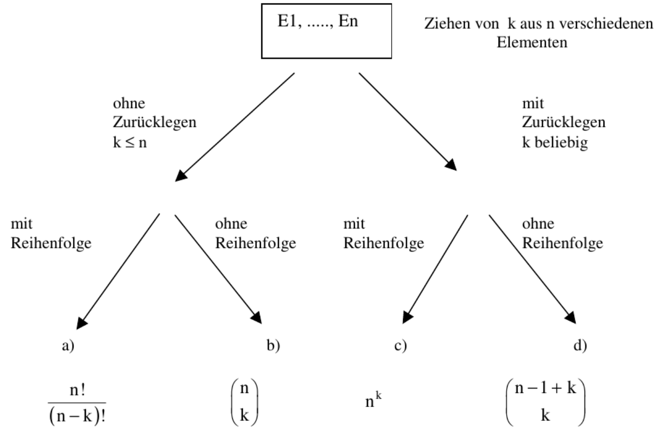
$$z = 9 * 9 * 8 * 7$$

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

$$z = \frac{12 * 11}{2}$$

Zuerst hat man $12 * 11$ verschiedene Möglichkeiten, die Birnen rauszunehmen. Da es bei diesem Paar aber nicht auf die Reihenfolge an kommt, muss die Menge noch halbiert werden.

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



Wsk

Wsk $P(A)$ für das Eintreffen von A
A = Ereignis, das eintreffen soll
g = Anzahl für das Ereignis A günstige Fälle
m = Maximale Anzahl Ereignisse (gegenseitig ausschliessend)

$$P(A) = \frac{g}{m}, \quad 0 \le P(A) \le 1$$

Absolute und Relative Häufigkeiten H_n und h_n
H_n = Absolute Häufigkeit, z. B. aus laufender Produktion
h_n = Relative Häufigkeit
n = Anzahl Proben/Messungen

$$h_n = \frac{H_n}{n}, \quad 0 \le h_n \le 1, \;\; H_n \ge 0$$

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

$$h_n \xrightarrow{n \to \infty} P(A)$$

Abstraktes Rechnen
Ω = Ergebnisraum, also Menge aller möglichen Ausgänge (Münze 2x werfen: $\Omega = \{kk, kz, zk, zz\}$)
A = Ergebnis, $A \subset \Omega$
$P(A)$ = **Wsk**, dass das Ergebnis A eintrifft
$P(\overline{A})$ = Gegen**wsk** des Ergebnisses A

Eigenschaften:

$$0 \le P(A) \le 1$$

$$P(\Omega) = 1$$

$$A \cap B = \{\,\} \longrightarrow P(A \cup B) = P(A) + P(B)$$

Wsksfunktion heisst eine Zuordnung wie $A \to P(A)$.

Folgerungen aus den Eigenschaften:

$$\Omega = A \cup \overline{A} \;\wedge A \cap \overline{A} \longrightarrow P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$$

$$P(\{\,\}) = 1 - P(\Omega) = 0$$

$$A \subseteq B \Rightarrow P(A) \le P(B)$$

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

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

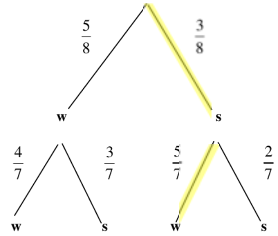
Bsp: In einer Urne liegen 5 weisse und 3 schwarze Kugeln. Man soll zwei ohne Zurücklegen ziehen. Gemäss $P(A) = \frac{g}{m}$ ergeben sich folgende **Wsk**:

$$P(w, w) = \frac{5 * 4}{8 * 7}$$

$$P(w, s) = P(s, w) = \frac{5 * 3}{8 * 7}$$

$$P(s, s) = \frac{3 * 2}{8 * 7}$$

Als **Ereignisbaum** dargestellt, der **Pfad** von $P(s, w) = \frac{3*5}{8*7}$ ist markiert:



Pfadregeln
- Entlang eines Pfades werden **Wsk** multipliziert
- Pfade sind disjunkt, ihre **Wsk** werden addiert ($P(alle\;Pfade) \;=\; 1$)

Bsp:

$$P(Min.\,eine\ schwarze\ K.) =$$
$$P(ws) + P(sw) + P(ss) = \frac{5*3}{8*7} + \frac{3*5}{8*7} + \frac{3*2}{8*7} = \frac{36}{56}$$

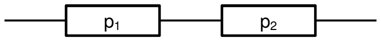
Oder mit dem Gegenereignis:

$$P(Ohne\ schwarze\ K.) = P(ww) = \frac{5*4}{8*7} = \frac{20}{56}$$
$$P(Min.\,eine\ schwarze\ K.) = 1 - P(Ohne\ schwarze\ K.) = 1 - \frac{20}{56} = \frac{36}{56}$$

Systeme

P = Ausfallwahrscheinlichkeit des Systems
p_i = Ausfallwahrscheinlichkeit einer Komponente
Z = Zuverlässigkeit des Systems
z_i = Zuverlässigkeit einer Komponente

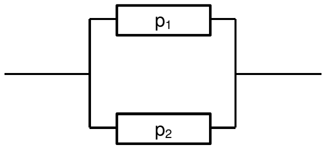
Serieschaltungen



$$P = 1 - z_1 * z_2 = p_1 + p_2 - p_1 * p_2$$

$$Z = z_1 * z_2 = (1 - p_1)(1 - p_2)$$

Parallelschaltungen



$$P = (1 - z_1)(1 - z_2) = p_1 * p_2$$

$$Z = z_1 + z_2 - z_1 * z_2 = 1 - p_1 * p_2$$

Bedingte Wsk $P(A|B)$

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

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:

$$P(M_B) = \frac{P(B \cap M)}{P(M)}$$

$$P(F_N) = \frac{P(N \cap F)}{P(F)}$$

$$P(B_F) = \frac{P(F \cap B)}{P(B)}$$

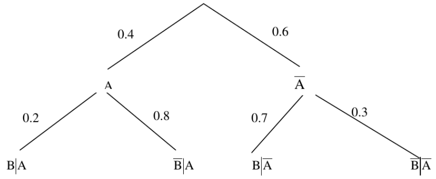
Allgemein:
$P(A|B)$ = **Wsk**, dass A unter der Bedingung von B eintrifft
$P(A \cap B)$ = Geschnittene **Wsk**

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

$\boldsymbol{P(A \cap B)}$ 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!

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

Ereignisbäume angepasst:



Die Knoten werden jeweils mit dem Zustand $(A, \overline{A}, B, \overline{B})$ beschriftet. Die Abhängigkeit (bei $P(B|A)$ ist dies A) wird zuerst/oben eingetragen.

$$P(B) = P(B|A) * P(A) + P(B|\overline{A}) * P(\overline{A}) = 0.2 * 0.4 + 0.7 * 0.6$$

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.

α = false negative, Kranker wird als gesund erkannt
β = false positive, Gesunder wird als krank erkannt

$$Sensitivität = 1 - \alpha$$

$$Spezivität = 1 - \beta$$

Unabhängige Ereignisse

A ist von B stochastisch unabhängig, wenn gilt:

$$P(A|B) = P(A)$$

Dann ist auch

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

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

Überprüfen kann man die Unabhängigkeit, indem man beide Seiten der obigen Formel $P(A \cap B)$ und $P(A) * P(B)$ berechnet und die Resultate vergleicht.

Verteilungen

$f(x)$ = Punktuelle Verteilungsfunktion, Wert einer Ausprägung
$F(x)$ = Bereitchsfunktion, Wert nach x %, 0..1

Punktueller Wert:

$$P(x) = f(x) \quad \rightarrow \quad ...Pdf(..)$$

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

$$P(X \le x) = F(x) \quad \rightarrow \quad ...Cdf(..)$$

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

$$P(X \ge x) = 1 - F(x) \quad \rightarrow \quad ...Cdf(..)$$

Quantile

Die Flächen und $F(x)$ sind die Quantile!

$$Q_\alpha = solve(\alpha = F(x), x)$$

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
μ = Erwartungswert E[X]
σ^2 = Varianz V[X]

$$X \sim HG(N, K, n)$$

$$\mu = n * \frac{K}{N}$$

$$\sigma^2 = n * \frac{K}{N} * \frac{N-K}{N} * \frac{N-n}{N-1}$$

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

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
μ = Erwartungswert E[X]
σ^2 = Varianz V[X]

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

$$\mu = n * p$$

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

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

Geometrische Verteilung
Warten auf den ersten Erfolg.

X = Anz. Misserfolge
p = Misserfogls**wsk** (ohne den ersten Erfolg)
μ = Erwartungswert E[X]
σ^2 = Varianz V[X]

$$X \sim GM(p)$$

$$\mu = \frac{p}{1-p}$$

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

$$P(X = k) = p^k (1-p)$$

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
μ = Erwartungswert E[X]
σ^2 = Varianz V[X]

$$T \sim Exp(\mu = ..)$$

$$f(t) = c * e^{-c*t}$$

$$F(t) = 1 - e^{-c*t}$$

$$\mu = \frac{1}{c}$$

$$\sigma^2 = \frac{1}{c^2}$$

Alter bzw. Überlebende:

$$P(T > t) = 1 - F(t) = e^{-c*t}$$

Uniforme Verteilung
Wenn etwas stetig ist.

a = Anfangswert
b = Endwert
μ = Erwartungswert E[X]
σ^2 = Varianz V[X]

$$X \sim U(a, b)$$

$$\mu = \frac{a+b}{2}$$

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

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

φ = Normalverteilungs**fkt** f(x)
ϕ = Kumulierte Normalverteilungs**fkt** F(x)
μ = Erwartungswert E[X]
σ^2 = Varianz V[X]

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{1}{2}*x^2}$$

Rechnen mit Erwartungswert E[X] und Varianz V[X]
Das fette $\boldsymbol{f(x)}$ muss jeweils durch die effektive Funktion der Verteilung ersetzt werden.

Bei stetigen Zufallsvariablen:
x = Effektiv das x als Buchstabe, es wird einfach verrechnet

$$E[X] = \int_{-\infty}^{\infty} x * \boldsymbol{f(x)} dx$$

$$V[X] = \int_{-\infty}^{\infty} (x-\mu)^2 * \boldsymbol{f(x)} dx = \int_{-\infty}^{\infty} x^2 * \boldsymbol{f(x)} dx - \mu^2$$

Varianz aus Erwartungswert:

$$V[X] = E[(X-\mu)^2] = E[X^2] - \mu^2$$

Rechenregeln:

$$E[aX] = a * E[X]$$

$$E[X+Y] = E[X] + E[Y]$$

$$V[aX] = a^2 * V[X]$$

$$V[X+Y] = V[X] + V[Y] \ (X, Y \ seien \ unabh.)$$

Summen von unabhängigen Variablen
μ_S = Summe von den Erwartungswerten
σ_S = Summe von den Standardabweichungen
n = Anz. Merkmale

$$\mu_S = n * \mu$$

$$\sigma_S = \sqrt{n} * \sigma$$

Mittelwerte von Kennwerten
$\mu_{\bar{x}}$ = Mittelwert von den Erwartungswerten
$\sigma_{\bar{x}}$ = Mittelwert von den Standardabweichungen
X = Zufallsvariable

$$\mu_{\bar{x}} = \mu_X$$

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{n}} * \sigma_X$$