

## 1. Deskriptive Statistik

### 1.1 Eindimensionale Häufigkeitsverteilungen (Bezug auf ein einziges Merkmal)

#### a) einfache Häufigkeitsverteilungen

**n** Gesamtzahl der Merkmalsträger  $\sum_{i=1}^v h_i = n$

**$x_i$**  Merkmalswert

**v** Anzahl verschiedener Merkmalswerte

**$h_i$**  absolute einfache Häufigkeit  
Anzahl Merkmalsträger mit Merkmal  $x_i$  (die Summe ergibt n)

**$f_i$**  relative einfache Häufigkeit  
Anteil Merkmalsträger mit Merkmal  $x_i$  (Quotient; die Summe ergibt 1)  
$$f_i = \frac{h_i}{n}$$

**i** Laufindex (Nummerierung)

#### b) kumulierte Häufigkeitsverteilungen

**$H_i$**  absolute kumulierte Häufigkeit,  
Anzahl der Merkmalsträger mit einem Merkmalswert ( $x_i$ ), der kleiner oder gleich  $x_i$  ist.  
$$H_i = \sum_{k=1}^i h_{\text{kleiner gleich}}$$

**$F_i$**  relative kumulierte Häufigkeit,  
Anteil der Merkmalsträger mit einem Merkmalswert ( $x_i$ ), der kleiner oder gleich  $x_i$  ist.  
$$F_i = \frac{H_i}{n}$$

**$HR_i$**  absolute kumulierte Resthäufigkeit  
Anzahl der Merkmalsträger mit einem Merkmalswert ( $x_i$ ), der grösser  $x_i$  ist.  
$$HR_i = n - H_i$$

**$FR_i$**  relative kumulierte Resthäufigkeit  
Anteil der Merkmalsträger mit einem Merkmalswert ( $x_i$ ), der grösser  $x_i$  ist.  
$$FR_i = 1 - F_i$$

### 1.2 Mehrdimensionale Häufigkeitsverteilungen (Bezug auf mehrere Merkmale)

**$h_i$**  Absolute einfache Häufigkeiten  
Vorspalte: Merkmalswerte für Merkmal X ( $x_i$ )      Summenspalte:  $\sum_{k=1}^5 h_{ik}$  Eindimensionale Verteilung für X  
Kopfzeile: Merkmalswerte für Merkmal Y ( $y_k$ )      Summenzeile:  $\sum_{k=1}^4 h_{ik}$  Eindimensionale Verteilung für Y  
Schnittpunkt Summenspalte und -zeile: n

**$h_i + H_i$**  Absolute einfache (einf) und kumulierte (kum) Häufigkeiten  
Vorspalte: Merkmalswerte für Merkmal X ( $x_i$ )      Summenspalte:  $\sum_{k=1}^5 h_{ik}$  Eindimensionale Verteilung für X  
Keine Summe kum  
Kopfzeile: Merkmalswerte für Merkmal Y ( $y_k$ )      Summenzeile:  $\sum_{k=1}^4 h_{ik}$  Eindimensionale Verteilung für Y  
Keine Summe kum  
Spaltenaufteilung: Zwei Zellen einf und kum      Schnittpunkt Summenspalte und -zeile: n  
$$H_{ik} \text{ kum} = \begin{array}{c|ccc|c} & y_1 & y_2 & y_3 & \\ \hline x_1 & h_{11} & h_{12} & h_{13} & \\ \hline x_2 & h_{21} & h_{22} & h_{23} & \\ \hline x_3 & h_{31} & h_{32} & h_{33} & H_{33} \\ \hline x_4 & h_{41} & h_{42} & h_{43} & \end{array} \quad \begin{array}{l} h_{11} + h_{12} + h_{13} \\ + h_{21} + h_{22} + h_{23} \\ + h_{31} + h_{32} + h_{33} = H_{33} \end{array}$$

**$f_i + F_i$**  Relative einfache (einf) und kumulierte (kum) Häufigkeiten (gleiches Verfahren wie  $h_i + H_i$ )  
Variante A:      Variante B:  
$$f_i = \frac{h_i}{n} \quad F_i = \frac{H_i}{n} \quad h_{11} + h_{12} + \dots + H_{ik} \quad \text{Schnittpunkt Summenspalte und -zeile: } 1$$
  
(wie oben  $h_i + H_i$ )

### 1.3 Klassifizierte Häufigkeitsverteilungen

#### a) Darstellung

j	Laufindex für die Klassen (Klassenindex)	
$x_j^u$	Untergrenze der Klasse j (von)	$x_j^o$ Obergrenze der Klasse j (bis unter)
$h_j$	Absolute einfache Klassenhäufigkeit $x_j^u \leq x_i < x_j^o$	
$H_j$	Absolute kumulierte Klassenhäufigkeit $x_i < x_j^o$	
$f_j$	Relative einfache Klassenhäufigkeit $\frac{h_j}{n}$	
$F_j$	Relative kumulierte Klassenhäufigkeit $\frac{H_j}{n}$	

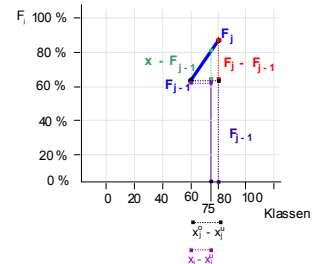
#### b) Interpolation (Näherungsweise Häufigkeitsberechnungen)

Gesucht ist ein Merkmalswert  $x_i$  zwischen  $x_j^u$  und  $x_j^o$

Formel für die **relative kumulierte Häufigkeit** für einen Wert  $x_i$  aus der j-ten Klasse:

Steigungsdreieck: (auch für  $H_j$  möglich)

$$F(x_i) = F_{j-1} + \frac{x_i - x_j^u}{x_j^o - x_j^u} \cdot (F_j - F_{j-1}) \quad \text{oder} \quad \frac{x_i - x_j^u}{x_j^o - x_j^u} = \frac{F_j - F_{j-1}}{F_j - F_{j-1}} \quad (\text{nach } x \text{ auflösen})$$



### 1.4 Grafische Darstellungsarten

#### a) Kreisdiagramm (Flächenproportional, einfache Häufigkeitsverteilungen)

$$\text{Fläche} = r^2 \cdot \pi \quad (r = \text{Radius})$$

$$\text{Flächenproportionalität} = \frac{\text{Fläche Kreis 1}}{n(\text{Kreis 1})} \cdot n(\text{Kreis 2}) = \text{Fläche Kreis 2} \quad \rightarrow \text{Radius berechnen Kreis 2} : \pi \sqrt{\quad}$$

#### b) Histogramm (Flächenproportional, einfache klassifizierte Häufigkeitsverteilungen)

$$d = \frac{\text{Absolute einfache Häufigkeit } h_j}{\text{Klassenbreite}}$$

Breite der Säulen: **Klassenbreite**  
Höhe der Säulen: **Häufigkeitsdichte**  
Anzahl Merkmalsträger je Klasse =  $d \cdot \text{Klassenbreite}$

#### c) Polygonzug (einfache klassifizierte Häufigkeitsverteilungen, Flächenproportional)

Abzisse (x)	Klassen	
Ordinate (y)	Häufigkeitsdichte d	
$x_i$	Klassenmitte im Koordinatensystem	
Anfangsrand	= $\frac{\text{Erste Häufigkeitsdichte}}{2}$	Endrand = $\frac{\text{Letzte Häufigkeitsdichte}}{2}$

#### d) Treppenfunktion (kumulierte nicht-klassifizierte Häufigkeitsverteilungen)

Abzisse (x)	Merkmal
Ordinate (y)	$H_i$

#### e) Summenpolygon (Ogive) (kumulierte klassifizierte Häufigkeitsverteilungen)

Abzisse (x)	Klassen
Ordinate (y)	$H_i$ der Klassenobergrenze und/oder $F_i$ (= Empirische Verteilungsfunktion)
Anfangspunkt der ersten Klasse = 0	

### 1.5 Statistische Messzahlen

#### a) Lageparameter

<b>Mo</b>	<b>Modus</b>
	Merkmalswert $x_i$ , der am meisten absolute Häufigkeiten $h_i$ aufweist
	<b>Klassifizierten Häufigkeitsverteilungen (Feinbestimmung)</b> (Unterschiedliche Klassenbreiten: $h_j = d$ )
	$Mo = x_m^u + (x_m^o - x_m^u) \cdot \frac{h_m - h_{m-1}}{(h_m - h_{m-1}) + (h_m - h_{m+1})}$
$h_m$	Absolute einfache Häufigkeit der Klasse mit der grössten Klassenhäufigkeit

<b>Me</b>	<b>Median (Quantil)</b>
	Mittlere Position aller Merkmalsträger (50%) ausgedrückt in Merkmalswert $x_i$
	$Me = x_{\left(\frac{n+1}{2}\right)}$ bei <b>ungeradem</b> n $x_i$ = Merkmalswert des Merkmalsträgers mit der Positionsnummer i.
	$Me = \frac{1}{2} \cdot \left( x_{\frac{n}{2}} + x_{\frac{n+1}{2}} \right)$ bei <b>geradem</b> n $x_i$ = Merkmalswert des Merkmalsträgers mit der Positionsnummer i.

**Me Median bei Klassifizierten Häufigkeitsverteilungen**  
 1. Bestimmung der Medianklasse = Klasse (x) mit dem Merkmalsträger mit der Positionsnummer  $i = \frac{(n+1)}{2}$  (Hj-Nr)  

$$Me = x_m^u + \frac{(n/2) - H_{m-1}}{H_m - H_{m-1}} \cdot (x_m^o - x_m^u)$$
  
 (sh. Abb. bei Interpolation, für  $H_j$ )

**D<sub>x</sub> Dezile**  
 Zerlegung der Gesamtheit n in zehn Zehntel  
 1. Bestimmung der Dezilklass = Klasse (x) mit dem Merkmalsträger mit der Positionsnummer  $i = \frac{(n+1)}{10}$  (Hj-Nr)  

$$D_9 = x_m^u + \frac{(n \cdot 9/10) - H_{m-1}}{H_m - H_{m-1}} \cdot (x_m^o - x_m^u)$$
 (neuntes Dezil, Merkmalswert  $x_i$  mit 90 % aller Merkmalsträger)  
 (sh. Abb. bei Interpolation, für  $H_j$ )  
**P<sub>x</sub> Perzentile: Hundert Hundertstel**  
**Q<sub>x</sub> Quartile: vier Viertel**

**$\bar{x}$  Arithmetisches Mittel** (häufigster Merkmalswert)  

$$\bar{x} = \frac{\sum_{i=1}^k (x_i \cdot h_i)}{n}$$
  
 Ergebnis ist der häufigste Merkmalswert  
**Arithmetisches Mittel bei klassifizierten Häufigkeitsverteilungen**  

$$\bar{x} = \frac{\sum_{i=1}^k (x_i^l \cdot h_i)}{n}$$
  
 $x_i^l$  = Klassenmitte

**F<sub>GM</sub> Geometrisches Mittel**  
 Durchschnittlicher Vergrößerungs- oder Verkleinerungsfaktor  
**Wachstumsfaktor F =**  $\frac{\text{Wert zum späteren Zeitpunkt}}{\text{Wert zum früheren Zeitpunkt}}$  Wert<sub>Jahr j</sub> · Wachstumsfaktor F = Wert<sub>Jahr j+1</sub>  
 Wachstumsfaktor – 1 · 100 = Jährl. Prozentuale Veränderung  

$$F_{GM} = \sqrt[n]{\frac{\text{Endwert}}{\text{Anfangswert}}}$$
 n = Anzahl Wachstumsfaktoren  

$$F_{GM} = \sqrt[n]{F_1 \cdot F_2 \cdot \dots \cdot F_n}$$
 n = Anzahl Wachstumsfaktoren

**b) Streuungsparameter** 

**R Spannweite (Variabilität, Variationsbreite, Range)** (z.B. bei Börsenkursen)  
 Distanz zwischen dem kleinsten und dem größten Merkmalswert (intervallskaliert, reagiert sehr auf Ausreisser)  
 Für die aufsteigend geordneten Merkmalswerte  $x_1, x_2, \dots, x_n$  gilt:  $R = x_n - x_1$  ( $x_n$  ist der letzte Merkmalswert)  
**Klassifizierte Häufigkeitsverteilungen:  $R = x_v^o - x_1^u$**

**ZQA Interquartilsabstand (Zentraler Quartilsabstand)**  
 Entfernung zwischen den beiden Merkmalswerten x, welche die in der Rangordnung gelegenen 50 % der Merkmalsträger eingrenzen (intervallskaliert, keine Ausreisser)  
 $ZQA = Q_3 - Q_1$   $Q_1 = x_{\left(\frac{n+1}{4}\right)}$   $Q_3 = x_{\left(\frac{3(n+1)}{4}\right)}$   $x_i$  = Merkmalswert des Merkmalsträgers mit der Positionsnummer i.  
**Klassifizierte Häufigkeitsverteilungen:**  
 Zuerst müssen die Klassen von 25 % (25 % von n = ? → in welcher Klasse?) und von 75 % bestimmt werden  
 $ZQA = Q_3 - Q_1$   $Q_1 = x_m^u + \frac{(n \cdot 1/4) - H_{m-1}}{H_m - H_{m-1}} \cdot (x_m^o - x_m^u)$   $Q_3 = x_m^u + \frac{(n \cdot 3/4) - H_{m-1}}{H_m - H_{m-1}} \cdot (x_m^o - x_m^u)$

**$\delta$  Mittlere absolute Abweichung** (sehr geeignet in der beschreibenden Statistik)  
 Durchschnittliche Entfernung aller beobachteten Merkmalswerte vom arith. Mittel (intervallskaliert, Ausreisser)  

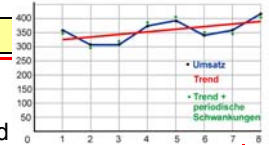
$$\delta = \frac{1}{n} \cdot \sum_{i=1}^k (|x_i - \bar{x}| \cdot h_i)$$
 (|Betrag| bedeutet Rechnen ohne Vorzeichen)  
**Klassifizierte Häufigkeitsverteilung:** 
$$\delta = \frac{1}{n} \cdot \sum_{i=1}^k (|x_i^l - \bar{x}| \cdot h_i)$$
 Klassenmitteln  $x_i^l$

**$\sigma$  Varianz und Standardabweichung** (intervallskaliert, Ausreisser, fehlende Anschaulichkeit, für schliess. Statistik)  
 Varianz: Summe der quadrierten Abweichungen der Merkmalswerte vom arith. Mittel dividiert durch n  

$$\text{Varianz } \sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^k [(x_i - \bar{x})^2 \cdot h_i]$$
 (auch  $f_i$  möglich) **Standardabweichung**  $\sigma = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^k [(x_i - \bar{x})^2 \cdot h_i]}$   
**Klassifizierte Häufigkeitsverteilungen:** Merkmalswerte  $x_i$  durch die Klassenmitteln  $x_i^l$  ersetzen

**VK Variationskoeffizient** (geeignet für den Vergl. von Streuungen von Häufigkeitsverteilungen unterschiedl. Lagen)  
 Die Standardabweichung wird als Prozentsatz des arithmetischen Mittels ausgedrückt (intervallskaliert)  

$$VK \% = \frac{\sigma}{\bar{x}} \cdot 100$$
 bzw. falls  $x < 0$ : 
$$VK \% = \frac{\sigma}{|\bar{x}|} \cdot 100$$



## 1.6 Zeitreihenanalyse (Zeitreihe: Zeitlich geordnete Folge von Merkmalswerten)

### Komponenten der Zeitreihe

**Trend**

Grundrichtung der Entwicklung

**Periodische Schwankungen**

Periodisch wiederkehrende Schwankungen um den Trend

**Restkomponente**

Einmalig wirkende Grössen die wiederholt aber unregelmässig wirken

Quartal  $y_i$  Umsatz im Quartal  $T_i$  Trend Quartal  $i$   $S_i$  periodische Schwankung Quartal  $i$   $R_i$  Restkomponente Quartal  $i$

$y_i = T_i + S_i + R_i$  (Additiver Zusammenhang: Die Komponenten wirken unabhängig voneinander) (Multiplikativ:  $y_i = T_i \cdot S_i \cdot R_i$  abhängig)

**Methode der gleitenden Durchschnitte (Trendermittlung)** Tabelle: (Den Perioden Durchschnitte zuteilen!)

**g D a** gleitender Durchschnitt ungerader (a) Ordnung (mit zunehmenden Wert wird d. gleitende Durchschnitt glatter)

a ist ungerade (3, 5)  $g D 3 = \frac{\text{Wert}_1 + \text{Wert}_2 + \text{Wert}_3}{3}$  (g D für die mittlere Periode<sub>2</sub>)

**g D b** gleitender Durchschnitt gerader (b) Ordnung

b ist gerade (2, 4)  $g D 4 = \frac{(\text{Wert}_1:2) + \text{Wert}_2 + \text{Wert}_3 + \text{Wert}_4 + (\text{Wert}_5:2)}{4}$  (g D für Periode<sub>3</sub>)

Laufen die periodischen Schwankungen über 4 Perioden muss g D mind. von der Ordnung 4 sein, 8, 12 auch möglich.

**Methode der kleinsten Quadrate (Trendermittlung)**

**a) Linearer Verlauf**

Quartale 1, 2, 3, 4, 5, 6, 7 ...  
nicht Quartal 1, 2, 3, 4, 1, 2 etc.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad b = \frac{\left[ \sum_{i=1}^n x_i \cdot y_i \right] + \left[ -n \cdot \bar{x} \cdot \bar{y} \right]}{\left[ \sum_{i=1}^n x_i^2 \right] + \left[ -n \cdot \bar{x}^2 \right]} \quad a = \bar{y} - (b \cdot \bar{x})$$

Trendgerade  $\hat{y} = b \cdot x + a$

**b) Nicht-Linearer Verlauf**

(nur für Exponentialfunktion  
nicht Potenzfunktion)

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \overline{\ln(y)} = \frac{\sum_{i=1}^n \ln(y_i)}{n} \quad \ln(b) = \frac{\left[ \sum_{i=1}^n x_i \cdot \ln(y_i) \right] + \left[ -n \cdot \bar{x} \cdot \overline{\ln(y)} \right]}{\left[ \sum_{i=1}^n x_i^2 \right] + \left[ -n \cdot \bar{x}^2 \right]}$$

$\ln(a) = \overline{\ln(y)} - \ln(b) \cdot \bar{x}$   $\ln(a)$  und  $\ln(b)$  delogarithmieren mit  $e^x$  **Trendgerade  $\hat{y} = a \cdot b^x$**

### Ermittlung der periodischen Schwankungen

Quartal Umsatz Trend Saison-Normale Saisonale Schwankungen  
1. Q 2. Q 3. Q 4. Q = Durchschnitt der Abweichungen = Trend + Saison-Normale  
= Umsatz – Trend des jeweils 1. Q (1. Q + 5 Q):2 etc.

## 1.7 Regressions- und Korrelationsrechnung (Beschreibung des Zusammenhangs zwischen zwei Datensätzen)

**$\hat{y}$**  Regressionsgerade bestimmen mittels Methode der kleinsten Quadrate Streudiagramm erstellen

**$\sigma^2_{xy}$**  Kovarianz (kein Ausmass der Abhängigkeit, sondern nur der Gleich- oder Gegenläufigkeit)

Gemeinsame Varianz der x- und y-Werte

Positive Kovarianz  $\sigma^2_{xy} = \frac{1}{n} \cdot \sum_{i=1}^n [(x_i - \bar{x}) \cdot (y_i - \bar{y})]$  n = Anzahl, nicht Summe der  $x_i$ -Werte

□ x und y zunehmend (gleichläufig) =  $\sigma^2_{xy}$  positiv □ für x zu- und y abnehmend (gegenläufig) =  $\sigma^2_{xy}$  negativ

**r** **Korrelationskoeffizient** (Masskorrelation, Produkt-Moment-Koeffizient, Bravais-Pearson)

Es können nur mehr Werte zwischen -1 und 1 entstehen, daher  $r \in [-1; 1]$ .

$r = \frac{\sigma^2_{xy}}{\sigma_x \cdot \sigma_y}$   $\sigma_x$  = Standardabweichung  $\sigma_x = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$   $\sigma_y$  = Standardabweichung  $\sigma_y = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2}$

□ Vorzeichen gibt gleichläufig oder gegenläufig an

□ Je näher r bei ±1 liegt, desto stärker ist der Zusammenhang zwischen Merkmalswerten und Regressionsgerade

**$R^2$**  **Bestimmtheitsmass** (x und y intervallskaliert)

$R^2$  informiert, welcher Teil der Varianz durch die Regression bestimmt werden kann.  $R^2 \in [-1; 1]$ . Also ob die Regressionsgerade (Methode der kl. Quadrate) den Zusammenhang von x und y gut oder schlecht wiedergibt. Je näher der Wert von  $R^2$  bei ±1 (100 %), desto stärker ist der Zusammenhang. 0 = kein Zusammenhang.

Das Bestimmtheitsmass wird 1, wenn die Fläche der kleinsten Quadrate 0 ist →  $\sum_{i=1}^n (y_i - b \cdot (x_i - \bar{x}) - \bar{y})^2 = 0$

$R^2 = \frac{(\sigma^2_{xy})^2}{\sigma^2_x \cdot \sigma^2_y}$  (% · 100)  $\sigma^2_x$  = Varianz  $\sigma^2_x = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$   $\sigma^2_y$  = Varianz  $\sigma^2_y = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2$

$R^2 = r^2$   $R^2$  drückt aus, dass ... % der Gesamtvarianz  $\sigma^2_{xy}$  durch die Varianz der Regressionswerte  $\sigma^2_{\hat{y}}$  bestimmt werden kann. (Das heisst, dass ... % von  $y_i$  von  $x_i$  abhängig ist.)

**Weitere Berechnungsmöglichkeiten:**

$\sigma^2_{\hat{y}} = \sigma^2_y$  (y auf der Regressionskurve berechnen) +  $\sigma^2_{y-\hat{y}}$

$R^2 = \frac{\sigma^2_{\hat{y}}}{\sigma^2_y}$   $R^2 = b^2 \cdot \frac{\sigma^2_x}{\sigma^2_y}$

**$U^2$**  **Unbestimmtheitsmass:**  $U^2 = 1 - R^2$  (Anteil der Abweichung, der nicht durch die Regression bestimmt wird)

**$\rho$**  **Rangkorrelation** (Merkmale, die nur ordinalskaliert sind)

□ Die Merkmale werden hinsichtlich der beiden Merkmale in eine Rangordnung gebracht

□ Sind zwei oder mehr Merkmale gleich, dann wird beiden Merkmalen das arith. Mittel der Ränge zugeordnet

$\rho = r = \frac{\sigma^2_{xy}}{\sigma_x \cdot \sigma_y}$  Der Zusammenhang zwischen den Rängen, nicht zwischen Merkmalswerten wird untersucht.

□ Sind die Merkmale **vollständig** oder fast vollständig unterschiedlich (keine arith. Mittel bei der Rangzuordnung):

$\rho$  (griech. Rho) =  $r = 1 - \frac{6 \cdot \sum_{i=1}^n D_i^2}{n^3 - n}$  wobei  $D_i = \text{Rang}(x_i) - \text{Rang}(y_i)$

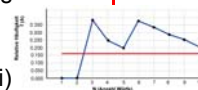
Wein	Bewertung $x_i$	Preis $y_i$	Rang $x_i$	Rang $y_i$
A	ausreichend	13.50	5	5
B	mangelhaft	15.20	6	4
C	gut	16.30	2.5	3
D	sehr gut	18.50	1	1
E	befriedigend	17.40	4	2
F	gut	12.90	2.5	6

**2. Wahrscheinlichkeitsrechnung** (Ausmass der Sicherheit, mit der ein möglicher Ausgang eintritt ausdrücken)**Zufallsvorgang:** Ausgang kann aufgrund von Unkenntnis oder Unwissenheit nicht vorhergesagt werden **$\Omega$  Ereignisraum:** Menge aller möglichen Elementarereignisse**Diskreter Ereignisraum:** endlich, abzählbar viele Elementarereignisse (nicht unendlich)**Stetiger Ereignisraum:** unendlich, überabzählbar viele Elementarereignisse (unendlich)**Elementarereignis:** Einzelne, sich gegenseitig ausschliessende möglichen Ausgänge eines Zufallsvorganges**A B C Ereignis:** Gesuchter Ausgang des Zufallsvorganges  $A = \{...\}$ **Sicheres Ereignis:** Ereignis, dass alle Elementarereignisse eines Ereignisraumes umfasst**Unmögliches Ereignis  $E = \emptyset$ :** Ereignis, dass keine Elementarereignisse eines Ereignisraumes umfasst**2.1 Direkte Ermittlung von Wahrscheinlichkeiten** (Zufallsvorgang wird tatsächlich oder gedanklich durchgeführt)**Klassische Wahrscheinlichkeitsermittlung (a-priori-Wahrscheinlichkeit, Laplace-Wahrscheinlichkeit, von Bernoulli)**Voraussetzung: diskrete Elementarereig. u. diese sind gleich wahrscheinlich; Rein-gedanklich, nicht tatsächlich

$$W(A) = \frac{\text{Anzahl der für A günstigen Elementarereignisse}}{\text{Anzahl der gleich möglichen Elementarereignisse (Anzahl Elemente } \Omega)}$$

**Geometrische Wahrscheinlichkeitsermittlung** für stetige Elementarereignisse: Einteilung in Intervalle**Statistische Wahrscheinlichkeitsermittlung (a-posteriori-Wahrscheinlichkeit, R. von Mises, von Bernoulli)**Voraussetzung: unter identischen Bedingungen tatsächlich wiederholte Durchführung des Zufallsvorganges

$$W(A) = \text{Relative Häufigkeit } f_i = \frac{\text{Anzahl der Zufallsvorgänge mit Ereignis A}}{\text{Anzahl der Zufallsvorgänge insgesamt (n)}}$$

 $W(A)$  entspricht der relativen Häufigkeit (Grenzwert) wenn  $n \rightarrow \infty$ . (Gesetz der grossen Zahl von Bernoulli)**Subjektive Wahrscheinlichkeitsermittlung (a-priori-Wahrscheinlichkeit)**

Sachkundige Personen beurteilen rein gedanklich die Möglichkeit des Eintretens eines Ereignisses zahlenmässig.

**2.2 Indirekte Ermittlung von Wahrscheinlichkeiten** (Ableitung der Wahrscheinlichkeit aus anderen Ereignissen)**Relationen von Ereignissen**

- **Vereinigung von Ereignissen** (mindestens eines von mehreren Ereignissen soll eintreffen)

Alle Elementarereignisse, die entweder in **A oder B oder in A und B** vorkommen gehören zu  **$A \cup B$** .

- **Durchschnitt von Ereignissen** (mehrere Ereignissen sollen zugleich eintreffen)

Alle Elementarereignisse, die **sowohl in A als auch in B** vorkommen gehören zu  **$A \cap B$** .**Disjunkte Mengen:** Kein gemeinsames Elementarereignis; Durchschnitt der Mengen ist leer  $A \cap B = \{ \} = \emptyset$ 

- **Komplementäre Ereignis**

Das zu A komplementäre Ereignis  $\bar{A}$  trifft dann ein, wenn das Ereignis A nicht eintritt.

- **Logische Differenz**

 $A \setminus B$  umfasst alle Elementarereignisse von A, die nicht Elementarereignisse von B sind.

- **Symmetrische Differenz**

Die symmetrische Differenz besteht aus den Elementarereignissen der Vereinigung ohne die Elementarereignisse des Durchschnitts.  $A \oplus B = A \setminus B \cup B \setminus A$ 

- **Vollständiges Ereignissystem**

Zerlegung des Ereignisraumes  $\Omega$  in paarweise disjunkte Ereignisse.

- **Teilergebnis**

Sind die Elemente eines Ereignisses B in A enthalten, so wird Ereignis B als Teilergebnis von A bezeichnet.

**Eigenschaften von Wahrscheinlichkeiten (Axiome von Andrej Kolmogoroff)**1 Nichtnegativität  $W(A) \geq 0$ 2 Normierung  $W(\Omega) = 1 = 100\%$ 3 Additivität  $W(A \cup B) = W(A) + W(B)$ **Additionssatz**Die Wahrscheinlichkeit, dass **mindestens eines von zwei Ereignissen A und B eintritt** beträgt:

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

$$W(A \cup B \cup C) = W(A) + W(B) + W(C) - W(A \cap B) - W(A \cap C) - W(B \cap C) + W(A \cap B \cap C)$$

$$\text{Disjunkte Ereignisse: } \sum_{i=1}^n W(A_i) \quad (\text{Einzelwahrscheinlichkeiten aufaddieren})$$

$\Omega = \{\text{Zweimaliges Werfen eines Würfels}\}$   
 $A = \{\text{Werfen eines Pasches}\} \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$   
 $B = \{\text{Augenzahlsumme} \leq 4\} \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}$   
 $W(A \cap B) = 6/36 + 6/36 - 2/36 = 11/36$

**Bedingte Wahrscheinlichkeit**Die Wahrscheinlichkeit für das **Ereignis A unter der Bedingung des Ereignisses B** (B ist eingetreten).

Die bedingte Wahrscheinlichkeit ist das Verhältnis der verbleibenden günstigen Fälle zu den verbleibenden gleichmöglichen Fälle. Welche Elemente, die sowohl in A als auch in B vorkommen, kommen in B vor? Den B tritt sowieso ein.

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

**Unabhängigkeit von Ereignissen** (Wird durch den Eintritt eines Ereignisses A der Eintritt eines Ereign. B beeinflusst?)Zwei Ereignisse sind voneinander **unabhängig** wenn gilt:

$$W(A) = W(A|B) \quad \text{oder} \quad W(A) = W(A|\bar{B}) \quad \text{oder} \quad W(A|B) = W(A|\bar{B})$$

**Multiplikationssatz**Die Wahrscheinlichkeit, dass **zwei Ereignisse A und B gemeinsam eintreten** beträgt (**abhängige Ereignisse**):

$$W(A \cap B) = W(A) \cdot W(B|A) \quad \text{oder} \quad W(A \cap B) = W(B) \cdot W(A|B)$$

$$W(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n) = W(A_1) \cdot W(A_2|A_1) \cdot W(A_3|A_1 \cap A_2) \cdot \dots \cdot W(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \quad (\text{Baumdiagramm!})$$

**Unabhängige Ereignisse:**  $W(A \cap B \cap C) = W(A) \cdot W(B) \cdot W(C)$ 

- Die Endwahrscheinlichkeiten ergeben sich durch multiplizieren der auf dem Weg des **Baumdiagramms** liegenden  $W(A)$ .

**Wahrscheinlichkeit des Komplementäreignisses**

Das zu A komplementäre Ereignis  $\bar{A}$  trifft dann ein, wenn das Ereignis A nicht eintritt.

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

**Totale Wahrscheinlichkeit**

Bilden die Ereignisse  $A_1, A_2, \dots, A_n$  ein **vollständiges Ereignissystem** (Alle  $A_i$  bilden als paarweise disjunkte Ereignisse  $\Omega$ ) und ist B ein beliebiges Ereignis:

$$W(B) = \sum_{i=1}^n W(A_i) \cdot W(B|A_i)$$

**Satz von Bayes**

Bilden die Ereignisse  $A_1, A_2, \dots, A_n$  ein **vollständiges Ereignissystem** (Alle  $A_i$  bilden als paarweise disjunkte Ereignisse  $\Omega$ ) und ist B ein beliebiges Ereignis:

$$W(A_i|B) = \frac{W(A_i) \cdot W(B|A_i)}{\sum_{i=1}^n W(A_i) \cdot W(B|A_i)}$$

**Logische Differenz**

$A \setminus B$  umfasst alle Elementarereignisse von A, die nicht Elementarereignisse von B sind.

$$W(A \setminus B) = W(A) - W(A \cap B)$$

**Symmetrische Differenz**

Symmetrische Differenz:  $A \circ B = A \setminus B \cup B \setminus A$

$$W(A \circ B) = W(A) - W(A \cap B) + W(B) - W(A \cap B)$$

**2.3 Elementare Kombinatorik**

Die Kombinatorik ermittelt die **Anzahl Möglichkeiten für das Auswählen** oder das **Anordnen der Elemente**.

**Additionsregel**

Die Menge  $\Omega$  sei in die disjunkten Teilmengen  $A_i$   $i = 1$  bis  $k$  zerlegt.

$m$  sei die Anzahl der Elemente in  $\Omega$

$n_i$ ,  $i = 1$  bis  $k$  sei die Anzahl Elemente in  $A_i$

$$m = n_1 + n_2 + \dots + n_k \quad 31 = 4 + 2 + 4 + 3 + 9 + 3 + 6$$

**Multiplikationsregel**

In einer Abfolge von  $k$  Schritten habe man für den ersten Schritt  $m_1$  Möglichkeiten, für den zweiten Schritt  $m_2$  Möglichkeiten. Dann bestehen für die Ausführung aller  $k$  Schritte  $m = m_1 \cdot m_2 \cdot \dots \cdot m_k$  Möglichkeiten.

$$\text{für } k = 3: \quad 2 \cdot 3 \cdot 2 = 12 \text{ Möglichkeiten}$$

**Permutationen**

Keine Auswahlmöglichkeit: Es werden immer alle vorgegebenen Elemente eingebracht.

**Permutation ohne Wiederholung** (1, 2, 3, 4, 5, 6, 7)

Jedes Element ist genau einmal vorgegeben und wird genau einmal in die Anordnung eingebracht.

**Fakultät:**  $n! = n \cdot (n-1) \cdot (n-2) \dots 3 \cdot 2 \cdot 1$

Wie viele verschiedene siebenstellige Zahlen können wir bilden, wenn wir jeweils nacheinander alle sieben Kugeln ziehen (ohne Zurücklegen) um eine siebenstellige Zahl zu erhalten?

$$7 \text{ (Möglichkeiten)} \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040 \text{ Möglichkeiten}$$

**Permutation mit Wiederholung** (1, 1, 2, 2, 2, 3, 4)

Gewisse Elemente sind mehr als einmal vorgegeben, es ist also keine Menge vorgegeben.

Sieben Kugeln sind mit den Zahlen 1, 1, 2, 2, 2, 3 und 4 beschriftet. Wie viele verschiedene siebenstellige Zahlen können wir bilden, wenn wir jeweils nacheinander alle sieben Kugeln ziehen (ohne Zurücklegen) um eine siebenstellige Zahl zu bilden?

$$P_{n_1, n_2, \dots, n_k}^w(n) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} \quad \frac{7!}{2! \cdot 3!} = 420 \text{ Permutationen (Möglichkeiten)}$$

Anzahl Zahlen zur Auswahl

1 kommt 2 mal vor 2 kommt 3 mal vor

**Variation ohne Wiederholung** (1, 2, 3, 4)

• Es werden **nicht alle Zahlen** verwendet für die Anordnung (im Beispiel  $k = 2, n = 4$ ).

• Die **Anordnung** der Zahlen wird berücksichtigt, d.h. es wird unterschieden ob ein Element vorher oder nachher gezogen wird (**2,1 ist nicht 1,2**)

• Kein Zurücklegen, d.h. keine Wiederholungen ( $\rightarrow$  kein 1,1, 2,2 möglich)

$$V_k(n) = \frac{n!}{(n-k)!} \quad V_2(4) = \frac{4!}{(4-2)!} = \frac{4!}{2!} = 12$$

$n = \text{Anzahl Zahlen zur Auswahl}$

$k = \text{Anzahl Stellen}$

Variation ohne Wiederholung

	1,2	1,3	1,4
2,1		2,3	2,4
3,1	3,2		3,4
4,1	4,2	4,3	

**Variationen mit Wiederholung** (1, 2, 3, 4)

• Es werden **nicht alle Zahlen** verwendet für die Anordnung (im Beispiel  $k = 2, n = 4$ ).

• Die **Anordnung** der Zahlen wird berücksichtigt, d.h. es wird unterschieden ob ein Element vorher oder nachher gezogen wird (**2,1 ist nicht 1,2**)

• Zurücklegen, d.h. Wiederholungen sind möglich ( $\rightarrow$  1,1, 2,2 möglich)

$$V_k^w(n) = n^k \quad 4^2 = 16$$

$k = \text{Anzahl Stellen}$

Variation mit Wiederholung

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
4,1	4,2	4,3	4,4

$n = \text{Anzahl Zahlen zur Auswahl}$



**Kombinationen ohne Wiederholung** (1, 2, 3, 4)

- Es werden **nicht alle Zahlen** verwendet für die Anordnung (im Beispiel  $k = 2, n = 4$ ).
- Die **Anordnung** der Zahlen wird **nicht** berücksichtigt, d.h. es wird **nicht** unterschieden ob ein Element vorher oder nachher gezogen wird (**2,1 ist dasselbe wie 1,2**)
- Kein Zurücklegen, d.h. **keine Wiederholungen** ( $\rightarrow$  kein 1,1, 2,2 möglich)

$$K_k(n) = \frac{n!}{(n-k)! \cdot k!} \quad K_k(4) = \frac{4!}{(4-2)! \cdot 2!} = 6$$

$n = \text{Anzahl Zahlen zur Auswahl}$   
 $k = \text{Anzahl Stellen}$

Kombination ohne Wiederholung		
1,2	1,3	1,4
	2,3	2,4
		3,4

**Kombination mit Wiederholung** (1, 2, 3, 4)

- Es werden **nicht alle Zahlen** verwendet für die Anordnung (im Beispiel  $k = 2, n = 4$ ).
- Die **Anordnung** der Zahlen wird **nicht** berücksichtigt, d.h. es wird **nicht** unterschieden ob ein Element vorher oder nachher gezogen wird (**2,1 ist dasselbe wie 1,2**)
- Zurücklegen, d.h. **Wiederholungen sind möglich** ( $\rightarrow$  1,1, 2,2 möglich)

$$K_k^w(n) = \frac{(n+k-1)!}{[(n+k-1)-k]! \cdot k!} \quad K_k^w(4) = \frac{(4+2-1)!}{[(4+2-1)-2]! \cdot 2!} = 10$$

$n = \text{Anzahl Zahlen zur Auswahl}$   
 $k = \text{Anzahl Stellen}$

Kombination mit Wiederholung			
1,1	1,2	1,3	1,4
	2,2	2,3	2,4
		3,3	3,4
			4,4

**2.4 Zufallsvariable****Begriff der Zufallsvariablen**

Ein Unternehmen führe auf dem Markt zwei Produkte A und B ein. Die Wahrscheinlichkeit für eine erfolgreiche Markteinführung wird auf 80 % und 90 % geschätzt.

A = {Erfolg mit A}  $W(A) = 80\%$   $W(\bar{A}) = 20\%$

B = {Erfolg mit B}  $W(B) = 90\%$   $W(\bar{B}) = 10\%$

Zufallsvariable: Anzahl der Erfolge	
Realisation	Wahrscheinlichkeit
0	0.02
1	$0.08 + 0.18 = 0.26$
2	0.72

Die Zufallsvariable X ist eine Funktion, die jedem Elementarereignis aus dem Ereignisraum  $\Omega$  eine reelle Zahl x zuordnet.  
Die Zufallsvariable X ist eine Variable, die bestimmte Realisationen x mit bestimmten Wahrscheinlichkeiten annimmt.

**Diskrete Zufallsvariable**

Die Realisation der Zufallsvariablen wird durch einen Zählvorgang festgestellt (z.B. Anzahl Kunden)

**Stetige Zufallsvariable**

Die Realisation der Zufallsvariablen wird durch einen Messvorgang festgestellt (z.B. Benzinverbrauch)

**Als Synonyme verwendete Begriffe in der Wahrscheinlichkeitsrechnung**

Zufallsvariable X	Merkmal X
Wahrscheinlichkeit	Relative Häufigkeit
Wahrscheinlichkeitsfunktion	Einfache relative Häufigkeitsverteilung
Verteilungsfunktion	Kumulierte relative Häufigkeitsverteilung
Erwartungswert	Arithmetisches Mittel
Varianz	Varianz

**Wahrscheinlichkeitsfunktion**

Die Wahrscheinlichkeitsfunktion ordnet jedem Wert  $x_i$  der Zufallsvariablen X die Wahrscheinlichkeit zu.

Für das Lotto (6 aus 45) ergibt sich folgende Wahrscheinlichkeitsfunktion: **Kombination ohne Wiederholung:**  $\frac{n!}{(n-k)! \cdot k!}$

i	Anzahl Richtige $x_i$	Wahrscheinlichkeit für $x_i$ -Richtige
1	0	0.4005646367
2	1	0.4241272624
3	2	0.1514740223
4	3	0.0224405959
5	4	0.0013646308
6	5	0.0000287291
7	6	0.0000001228

$$f(x) = \begin{cases} 0.4005646367 & \text{für } x = 0 \\ 0.4241272624 & \text{für } x = 1 \\ 0.1514740223 & \text{für } x = 2 \\ 0.0224405959 & \text{für } x = 3 \\ 0.0013646308 & \text{für } x = 4 \\ 0.0000287291 & \text{für } x = 5 \\ 0.0000001228 & \text{für } x = 6 \end{cases} \quad f(x) = \frac{\begin{bmatrix} 6 \\ x \end{bmatrix} \cdot \begin{bmatrix} 39 \\ 6-x \end{bmatrix}}{\begin{bmatrix} 45 \\ 6 \end{bmatrix}}$$

**Verteilungsfunktion**

Die Verteilungsfunktion einer diskreten Zufallsvariablen informiert darüber, wie gross die Wahrscheinlichkeit für eine Realisation ist, deren Wert kleiner oder gleich dem Wert x ist.

Die kumulierten Wahrscheinlichkeiten müssen bestimmt werden:

i	Anzahl Richtige $x_i$	Wahrscheinlichkeit für $x_i$ -Richtige $W(X = x_i) \quad f(x_i)$	$W(X \leq x_i)$ **Wahrscheinlichkeit, dass 6 oder weniger Richtige kommen
1	0	0.4005646367	0.4005646367
2	1	0.4241272624	0.8246918991
3	2	0.1514740223	0.9761659214
4	3	0.0224405959	0.9986065173
5	4	0.0013646308	0.9999711482
6	5	0.0000287291	0.9999998772
7	6	0.0000001228	**1.0000000000

**Parameter**

**Erwartungswert (Arithmetisches Mittel)**

$$E(x) = \sum_{i=1}^n [x_i \cdot f(x_i)]$$

**Varianz**

$$\sigma^2 = \sum_{i=1}^n \{[x_i - E(x)]^2 \cdot f(x_i)\}$$

**Standardabweichung**

$$\sigma = \sqrt{\sum_{i=1}^n \{[x_i - E(x)]^2 \cdot f(x_i)\}}$$

**Die Ungleichung von Tschebyscheff**

Mit welcher Wahrscheinlichkeit realisiert sich eine Zufallsvariable  $X$  in einem Intervall, der symmetrisch um den Erwartungswert liegt.

$$W(E(x) - c \cdot \sigma < X < E(x) + c \cdot \sigma) > 1 - \frac{1}{c^2}$$

Die Ungleichung ergibt eine gute Schätzung.

Bsp.:  $c = 2$   $W(-0.7678 < X < 2.3678) > 1 - \frac{1}{2^2} = 0.75 = 75\%$   
Die Wahrscheinlichkeit 0, 1, oder 2 richtige zu haben ist mindestens 75 %. (In Wirklichkeit: 0.976  $\rightarrow$  97.6 %)

**2.5 Binominalverteilung****Beispiel**

Die Aufrechterhaltung einer Pumpstation ist gewährleistet, wenn 5, 6 oder sieben Motoren einsatzfähig sind.

$A = \{\text{Motor läuft}\}$   $W(A) = 0.95$  Zufallsvariable  $X$ : Anzahl der laufenden Motoren

$\bar{A} = \{\text{Motor fällt aus}\}$   $W(\bar{A}) = 0.05$  Realisationen  $x$ : 0, 1, 2, 3, 4, 5, 6, 7

Gesucht ist die Wahrscheinlichkeit, dass mindestens 5 Motoren laufen:  $F(X \geq 5) = f(5) + f(6) + f(7)$

1. Schritt: Wahrscheinlichkeit, dass genau fünf Motoren laufen?

a) Wie viele Möglichkeiten gibt es, dass genau fünf Motoren laufen?

$$\text{Kombination ohne Wiederholung} = \frac{7!}{(7-5)! \cdot 2!} = 21$$

b) Wie gross ist die Eintretenswahrscheinlichkeit für jede der 21 Möglichkeiten?

Die Wahrscheinlichkeit, dass die Motoren 1 – 5 laufen, während die Motoren 6 und 7 ausfallen beträgt:

$$0.95 \cdot 0.95 \cdot 0.95 \cdot 0.95 \cdot 0.95 \cdot 0.05 \cdot 0.05 = 0.001934$$

Für die 20 anderen Möglichkeiten sind die gleichen Faktoren zu multiplizieren, nur die Reihenfolge ändert.

$$\text{c) Damit beträgt } f(5) = 21 \cdot 0.001934 = 0.0406 = \underline{4.06\%}$$

2. Schritt: Dasselbe ausrechnen für  $f(6)$  und  $f(7)$

3. Schritt: Die Wahrscheinlichkeiten addieren:  $0.0406 + 0.2573 + 0.6983 = 0.9962 = \underline{99.62\%}$

**Eigenschaften damit die Binominalverteilung angewendet werden kann**

1. Der Zufallsvorgang wird  $n$ -mal identisch durchgeführt.

2. Der Zufallsvorgang besitzt genau zwei mögliche Ausgänge.

3. Die Wahrscheinlichkeiten für beide Ausgänge sind bei jedem Zufallsvorgang mit  $W(A) = \Theta$  und  $W(\bar{A}) = 1 - \Theta$

**Formel**  $f_B(x|n;\Theta) = \binom{n}{x} \cdot \Theta^x \cdot (1 - \Theta)^{n-x}$   $\Theta = 0.95$  im Beispiel  $W(A)$

$$f_B(5|7;0.95) = \binom{7}{5} \cdot \frac{n!}{(n-k)! \cdot k!} \cdot 0.95^5 \cdot (1 - 0.95)^{7-5} = \underline{4.06\%} \quad (\text{Wahrscheinlichkeit, dass genau 5 von sieben Motoren laufen})$$

Analog für die anderen  $x$ -Werte

**Berechnung mit der Tabelle**

Die Gegenwahrscheinlichkeit muss angewendet werden:  $f_B(2|7;0.05) = \binom{7}{2} \cdot 0.05^2 \cdot (1 - 0.05)^5 = \underline{4.06\%}$